UNIVERSIDADE FEDERAL DE SANTA CATARINA
CENTRO TECNOLÓGICO
DEPARTAMENTO DE INFORMÁTICA E ESTATÍSTICA
CURSO DE SISTEMAS DE INFORMAÇÃO

Artur Antonio Dal Prá

# Avaliação imobiliária através de inteligência computacional

Florianópolis (SC)
2023

Artur Antonio Dal Prá

# Avaliação imobiliária através de inteligência computacional

Trabalho de Conclusão de Curso submetido ao curso de Sistemas de Informação do Centro Tecnológico da Universidade Federal de Santa Catarina como requisito parcial para a obtenção do título de Bacharel em Sistemas de Informação.

Orientador: Prof. Élder Rizzon Santos, Dr.

Florianópolis (SC)

2023

Ficha de identificação da obra elaborada pelo autor,
através do Programa de Geração Automática da Biblioteca Universitária da UFSC.

Dal Prá, Artur Antonio
Avaliação imobiliária através de inteligência computacional / Artur Antonio Dal Prá ; orientador, Élder Rizzon Santos, 2023.
136 p.

Trabalho de Conclusão de Curso (graduação) - Universidade Federal de Santa Catarina, Centro Tecnológico, Graduação em Sistemas de Informação, Florianópolis, 2023.

Inclui referências.

1. Sistemas de Informação. 2. avaliação imobiliária. 3. inteligência computacional. 4. lógica difusa. I. Santos, Élder Rizzon. II. Universidade Federal de Santa Catarina. Graduação em Sistemas de Informação. III. Título.

Artur Antonio Dal Prá

**Avaliação imobiliária através de inteligência computacional**

Este Trabalho de Conclusão de Curso foi julgado adequado para obtenção do título de Bacharel em Sistemas de Informação e aprovado em sua forma final pelo Curso de Sistemas de Informação.

Florianópolis (SC), 19 de junho de 2023.

Coordenação do Curso

**Banca examinadora**

Prof. Élder Rizzon Santos, Dr.
Orientador

Profª. Andréa Cristina Konrath, Dra.
Universidade Federal de Santa Catarina

Prof. Mauro Roisenberg, Dr.
Universidade Federal de Santa Catarina

Florianópolis (SC), 2023.

Aos meus queridos pais, cujo apoio
foi essencial durante a minha jornada
acadêmica, sou imensamente grato
por tudo aquilo que fizeram por mim.

## AGRADECIMENTOS

Agradeço ao corpo docente do curso de graduação em Sistemas de Informação da UFSC, por toda base que permitiu atingir o conhecimento necessário na elaboração deste trabalho. Agradeço aos professores e colegas do Programa de Pós Graduação em Estatística da UFRGS que muito contribuíram durante as disciplinas de Modelagem Estatística e Inferência enquanto aluno especial, aos professores e colegas do Programa de Pós Graduação em Engenharia de Transportes e Gestão Territorial da UFSC que contribuíram nas disciplinas de Engenharia de Avaliações e de Tópicos Avançados em Engenharia de Avaliações. Agradeço à professora Silvia Modesto Nassar pelos conhecimentos adquiridos durante a disciplina Sistemas Especialistas Difusos enquanto aluno especial no Programa de Pós Graduação em Ciências da Computação da UFSC. Em especial agradeço ao orientador pela disponibilidade e pelas contribuições. Este trabalho encerra o ciclo da graduação, onde aproveito para agradecer aos colegas que foram importantes ao longo desta jornada. Agradeço à Daiana Resnel Vieira pela parceria. Por fim, agradeço a todos aqueles que participaram direta ou indiretamente à elaboração deste trabalho.

## RESUMO

Este trabalho apresenta uma visão geral sobre o estado da arte da inteligência computacional aplicada à avaliação imobiliária e sua implementação através de *scripts* usando pacotes em linguagem R para a predição do preço de venda de apartamentos situados no Bairro Trindade na data base março de 2023. Um estudo comparativo é realizado confrontando a técnica consagrada do método comparativo direto de dados de mercado por inferência estatística usando Regressão Linear Múltipla (MLR), comparada com outras técnicas, como o sistema de inferência difuso de Mamdani (FIS), o sistema de inferência difuso de Takagi-Sugeno-Kang (TSK), o sistema de inferência difuso neuroadaptativo (ANFIS). Além disso, o trabalho aborda brevemente o balanceamento de classes por subamostragem da classe majoritária (*undersampling*), e destaca considerações importantes para o êxito na aplicação de modelos baseados em lógica difusa na avaliação imobiliária. O estudo conclui que a MLR apresenta os melhores resultados, seguido da técnica TSK. A técnica FIS apresenta um desempenho razoável, enquanto que a técnica ANFIS mostra potencial, mas requer ser aplicada sob certos cuidados. Foram implementadas as métricas Erro Absoluto Médio (MAE), Erro Quadrático Médio (MSE), Raiz do Erro Quadrático Médio (RMSE) e Erro Médio Percentual Absoluto (MAPE), analisadas conjuntamente sob a aplicação de conjuntos treino x teste com diversas proporções. Sob a ótica da métrica MAPE, o melhor modelo MLR resultou 7,39%, o TSK 8,51%, seguido pelo FIS 10,40% e ANFIS com 13,05%.



**Palavras-chave**: *Inteligência computacional; avaliação imobiliária; regressão linear múltipla; lógica difusa.*

## ABSTRACT

This study provides an overview of the state-of-the-art in computational intelligence applied to real estate valuation and its implementation through scripts using R libraries for predicting the selling price of apartments located in the Trindade neighborhood as of the reference date of march 2023. A comparative study is conducted, comparing the well-established technique of direct comparison method using statistical inference through Multiple Linear Regression (MLR) with other techniques such as Mamdani's fuzzy inference system (FIS), Takagi-Sugeno-Kang's fuzzy inference system (TSK), and adaptive neuro-fuzzy inference system (ANFIS). Additionally, the work briefly addresses class balancing through undersampling of the majority class and highlights important considerations for the successful application of fuzzy logic-based models in real estate valuation. The study concludes that MLR yields the best results, followed by the TSK technique. The FIS technique shows reasonable performance, while the ANFIS technique demonstrates potential but requires a careful application. Metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE) were implemented and jointly analyzed using various train-test set ratios. From the perspective of the MAPE metric, the best MLR model resulted in 7.39%, TSK achieved 8.51%, followed by FIS with 10.40%, and ANFIS with 13.05%.



**Keywords**: *Computational intelligence; real estate valuation; multiple linear regression; fuzzy logic*.

## LISTA DE FIGURAS

## LISTA DE QUADROS

## LISTA DE TABELAS

## LISTA DE ABREVIATURAS E SIGLAS

| | |
|---|---|
| ABNT | Associação Brasileira de Normas Técnicas |
| ANFIS | *Adaptative Neuro Fuzzy Inference System* |
| ANN | *Artificial Neural Network* |
| CSV | *Comma Separated Values* |
| CUB | Custo unitário básico |
| FIS | *Fuzzy Inference System* |
| FLSR | *Fuzzy Least-Squares Regression* |
| GA | *Genetic Algorithm* |
| IA | Inteligência Artificial |
| IC | Inteligência Computacional |
| MLR | *Multiple Linear Regression* |
| MSE | *Mean Squared Error* |
| NBR | Norma Brasileira Regulamentadora |
| RLM | Regressão Linear Múltipla |
| RNA | Redes Neurais Artificiais |
| SMOTE | *Synthetic Minority Over-sampling Technique* |
| TBM | *Tree Based Models* |
| TCC | Trabalho de Conclusão de Curso |
| TFN | *Triangular Fuzzy Number* |
| TSK | Takagi-Sugeno-Kang |
| UFSC | Universidade Federal de Santa Catarina |

**SUMÁRIO**

# 1 INTRODUÇÃO

O mercado imobiliário traz uma realidade que está correlacionada com diversos outros setores da economia. Uma variável impactante à estabilidade do mercado imobiliário é a precificação dos imóveis. No julgamento dos possíveis preços de mercado que podem ser atribuídos a um imóvel urbano, muitas interpretações e técnicas coexistem.

Segundo Thofehrn (2008), o ato de avaliar um imóvel urbano consiste em determinar qual é o seu preço justo, que será pago por um determinado comprador (onde este deseja por livre e espontânea vontade comprar o imóvel) para um vendedor (em igual condição de livre arbítrio), seguindo ao requisito que ambos conheçam qual é o aproveitamento do terreno, e este preço justo se estabelecerá numa determinada data, observando-se a tendência mercadológica que existe nas circunvizinhanças. Deste modo, estima-se um cenário de mercado com média liquidez, ou seja, a comercialização do imóvel (absorção do produto imobiliário pelo mercado) ocorrerá em menos de três meses.

Segundo Santello (2004), a avaliação de imóveis demanda conhecimentos específicos da área e também uma base de dados confiáveis. Ainda, os métodos convencionais demandam uso da estatística inferencial, onde o decisor arbitra o valor do imóvel conforme uma faixa de valores (intervalo de confiança) em função da qualidade dos dados que originou o modelo matemático. Segundo Bussab e Morettin (2017), a inferência estatística é um ramo da estatística que produz afirmações sobre determinada característica (variável) da população de interesse, a partir de informações colhidas de uma parte dessa população.

Usualmente os avaliadores aplicam a técnica de inferência estatística com regressão linear múltipla quando o mercado disponibilizar uma quantidade de elementos suficiente para compor uma amostra representativa e enquadramento nos graus de fundamentação e de precisão citados na NBR 14.653-2 (2011), que é a norma brasileira de avaliação de bens, parte de imóveis urbanos. Não havendo uma quantidade de elementos para compor uma base amostral suficiente, os avaliadores aplicam a técnica da homogeneização por fatores ou outras alternativas, como o método da capitalização da renda ou os métodos involutivo vertical ou horizontal, estas abordagens não são aplicáveis ao domínio do problema.

Em outras regiões, principalmente no estado de São Paulo, os avaliadores costumam ignorar a inferência estatística e fazer uso somente da avaliação por homogeneização de fatores, mesmo para cenários com uma base amostral composta por muitos elementos.

Ambos os métodos são aceitos pela NBR 14.653-2. Em 2011 foi incorporado ao texto da NBR 14.653-2 a alternativa de avaliação por Redes Neurais Artificiais, que veio acompanhada por premissas e recomendações para seu uso.

Segundo Pelli (2004), grandes avanços na área de engenharia são notados, no que tange às modelagens de sistemas reais por meio de mecanismos de inferência. Nesta ótica, este autor cita a restrição em uso de Estimadores dos Mínimos Quadrados e elenca a possibilidade do uso de sistemas híbridos, como por exemplo, as redes neuro-difusas e redes neurais artificiais, que são aplicáveis às avaliações de imóveis urbanos.

O uso de muitas alternativas conduz a resultados diferentes, e então surge a possibilidade de usar um método para validar os resultados do outro.

Assim, como a técnica de Redes Neurais Artificiais faz necessário uma base amostral bastante superior, muitas vezes seu uso resta prejudicado e então a lógica difusa surge como alternativa.

A importância em realizar uma avaliação imobiliária com boa técnica reside no fato de que uma avaliação inadequada pode conduzir a consequências negativas, como transações desvantajosas, pode ocasionar prejuízos financeiros e passivos judiciais. A norma de avaliação de imóveis traz recomendações para que os erros usuais (que podem causar sub ajustamento nos modelos) sejam evitados, tais como base amostral que não representa o imóvel avaliando, seleção de elementos em mercado distinto do avaliando, imperícia do avaliador.

O emprego de uma técnica científica objetiva traz precisão aos resultados do modelo, e garante a transparência, imparcialidade e consistência no trabalho do avaliador.

## 1.1 OBJETIVOS

### 1.1.1 Objetivo geral

Analisar modelos para avaliação de imóveis do tipo apartamento utilizando técnicas de inteligência computacional.

### 1.1.2 Objetivos específicos

Os objetivos específicos são estes numerados a seguir:

1 - verificar as técnicas de inteligência computacional disponíveis, teorias e modelos que podem ser empregados na avaliação de apartamentos;

2 - definir os requisitos da avaliação para estabelecer uma representação de conhecimento do domínio;

3 - especificar e executar um estudo de caso através de *scripts* em linguagem de programação R para a avaliação imobiliária;

4 - propor modelos de IA para a realização da avaliação imobiliária de apartamentos e compará-los com as alternativas convencionais;

5 - identificar quais variáveis são relevantes na aplicabilidade do modelo para avaliar valor de mercado de apartamentos;

6 - demonstrar a validade dos modelos de inteligência computacional (IC) aplicados à avaliação de apartamentos, elencar as suas vantagens e desafios para efetivação;

7 - identificar os requisitos mínimos para haver a aceitação de uma avaliação imobiliária usando os modelos propostos;

8 – comparar e analisar o desempenho dos modelos por meio de métricas.

## 1.2 MÉTODO DE PESQUISA

Para fins de organização, os objetivos específicos foram numerados e os respectivos métodos de pesquisa conforme a seguir:

### 1.2.1 Objetivo 1

a) Listar algumas das técnicas de inteligência artificial que podem ser usadas para a avaliação imobiliária de apartamentos e destacar quais se mostram mais adequadas para esta finalidade;
b) levantamento do estado da arte; e
c) elaboração da fundamentação teórica.

### 1.2.2 Objetivo 2

a) Listar as principais técnicas de avaliação de imóveis que podem ser aplicadas à apartamentos e destacar quais se mostram mais apropriadas para esta finalidade;
b) revisar os normativos que abrangem o tema;
c) levantamento do estado da arte; e
d) elaboração da fundamentação teórica.

### 1.2.3 Objetivo 3

a) Estabelecer quais os requisitos para viabilizar um *script* para avaliação automatizada;
b) revisar a teoria para abranger todos os requisitos;
c) especificar o estudo de caso; e
d) executar o estudo de caso.

### 1.2.4 Objetivo 4, 5 e 6

a) Estudar as técnicas de avaliação de imóveis atualmente disponíveis que podem ser aplicadas à avaliação de apartamentos;
b) estudar a aplicabilidade da lógica difusa nas avaliações imobiliárias de apartamentos;
c) definir as variáveis mais significativas;
d) estabelecer um modelo difuso para a realização de avaliação imobiliária de apartamentos;
e) a partir dos resultados da execução do estudo de caso, comparar com os métodos clássicos de avaliação imobiliária, identificando as convergências de resultados;
f) identificar se há alguma especificidade para a tipologia de apartamento;
g) estabelecer critérios e recomendações para a aplicabilidade da lógica difusa nas avaliações;
h) propor recomendações para a utilização de modelos difusos em avaliações de apartamentos.

### 1.2.5 Objetivo 7

a) Estudar pesquisas realizadas na área de avaliação imobiliária que fazem uso de técnicas de inteligência artificial;
b) avaliar as bibliotecas (*libraries*) disponíveis com funcionalidades relacionadas com a lógica difusa; e
c) estabelecer parâmetros para haver a aceitação da avaliação imobiliária usando lógica difusa.

## 1.3 ESCOPO DO TRABALHO E DELIMITAÇÕES

Em um procedimento avaliatório cotidiano, não havendo uma quantidade de elementos para compor uma base amostral suficiente, os avaliadores aplicam outras técnicas previstas na NBR 14.653-2 (2011), como o método da capitalização da renda ou método involutivo que não serão abordados neste trabalho em razão do enfoque ser um empreendimento (muitas vezes fictício) e não o conjunto de dados em si.

A limitação espacial para a obtenção da base amostral é o Bairro Trindade e regiões aproximadas, atendendo ao critério disposto no item 8.2.1.1 da NBR 14.653-2 (2011). A necessidade da contemporaneidade dos dados é citada no item 8.2.1.3.2 da NBR 14.653-2 (2011), então o conjunto de dados é composto por elementos obtidos em prazo de no máximo um mês para atender ao requisito da contemporaneidade e a variável tempo não é usada.

Este trabalho foi dimensionado para ocorrer ao longo de três semestres e não se propõe a desenvolver pacotes estatísticos, mas somente fazer uso de pacotes existentes, onde os resultados são coletados para compor o estudo comparativo.

Obviamente, este trabalho compara métricas para avaliar modelos gerados por meio de determinadas técnicas, essas métricas são calculadas usando informações dos dados de teste aplicadas no respectivo modelo, não sendo necessário fazer a avaliação de determinado imóvel em específico, então não haverá um imóvel avaliando.

Finalmente, toda e qualquer atividade não citada neste espaço excetua-se do escopo.

## 1.4 JUSTIFICATIVA

Este trabalho traz algumas das diversas técnicas disponíveis para predição dos valores de mercado para um imóvel urbano. Para tanto, estima-se um cenário de mercado com média liquidez, ou seja, a comercialização do imóvel ocorrerá em menos de três meses, o que é compatível com as características do objeto de estudo.

Usualmente os avaliadores fazem uso da técnica de inferência estatística com regressão linear múltipla (MLR), seguindo as recomendações do Anexo A da NBR 14.653-2 (2011) quando o mercado disponibilizar uma quantidade elementos para compor uma base amostral (conjunto de dados) representativa e enquadramento nos graus de fundamentação e de precisão citados na NBR 14.653-2 (2011), que é a norma brasileira de avaliação de bens, parte de imóveis urbanos. Esta mesma norma traz recomendações para o emprego da técnica de redes neurais artificiais (ANN) em seu Anexo E.

A literatura científica traz outras técnicas aplicáveis na avaliação imobiliária que também já mostraram resultados satisfatórios em estudos anteriores, tais como a o sistema de inferência difuso de Mamdani (FIS) e de Takagi-Sugeno-Kang (TSK), o método dos mínimos quadrados difusos (FLSR), o sistema difuso neuroadaptativo (ANFIS) e os métodos baseados em árvore (TBM).

A motivação é elencar e explorar possíveis técnicas capazes de gerar modelos tão bons quanto aqueles gerados pelas técnicas convencionais. O uso de muitas alternativas conduz a resultados diferentes, e então surge a possibilidade de usar um método para validar os resultados do outro, em especial quando há técnicas já consagradas em outros meios que ainda não estão previstas no texto normativo.

A técnica ANN faz necessário uma base amostral mais numerosa que aquela que torna possível a aplicação da regressão linear múltipla, então muitas vezes o uso de ANN resta prejudicado, razão pela qual outros modelos, como aqueles baseados em lógica difusa surgem como alternativa, em especial pelas vantagens no tratamento de dados qualitativos.

Segundo Kamire *et al* (2021), quando relações não lineares entre os parâmetros impactam na determinação do preço dos imóveis (usualmente variável dependente), então se torna difícil fazer as predições empregando os métodos tradicionais (como MLR).

# 2 REVISÃO BIBLIOGRÁFICA

## 2.1 CONSIDERAÇÕES INICIAIS SOBRE ECONOMIA URBANA E AVALIAÇÃO IMOBILIÁRIA

De acordo com Sarip *et al* (2016), um imóvel é uma personificação da terra física e de todas as suas melhorias, juntamente com todos os direitos, interesses, benefícios e responsabilidades decorrentes da propriedade do referido elemento. Isto significa que a avaliação de bens imóveis é uma atividade que se presta a fornecer uma medida quantitativa dos ativos (benefícios) e passivos resultantes da propriedade.

Estes autores explicam que o valor de mercado do imóvel é considerado como sendo a quantia estimada pela qual um imóvel pode ser comercializado na data da avaliação entre um comprador interessado e um vendedor disposto em uma transação em condições normais de mercado e em que as partes tenham agido com conhecimento, prudência e sem compulsão. De fato, este é o mesmo conceito difundido no Conselho Internacional de Padrões para Avaliação (*International Valuation Standards Council* / IVS).

No Brasil esta atividade é realizada por engenheiros, arquitetos e, questionavelmente, por corretores de imóveis e oficiais de justiça.

Segundo Sarip *et al* (2016), ao chegar a uma estimativa do valor do imóvel avaliando, os profissionais precisam se relacionar com importantes princípios econômicos que fundamentam o funcionamento do mercado imobiliário, como por exemplo, os princípios de oferta e demanda, concorrência, substituição, antecipação e mudança, onde comum a todos esses princípios é o seu efeito direto e indireto sobre o grau de utilidade e produtividade de uma propriedade. Os autores esclarecem que, consequentemente, pode-se afirmar que a utilidade do imóvel reflete as influências combinadas de todas as forças de mercado que vêm a suportar o valor de uma parcela de imóveis.

## 2.2 EVOLUÇÃO E TENDÊNCIAS PARA A NORMA BRASILEIRA

A engenharia de avaliações no Brasil começou de maneira empírica, seguido de alguns estudos isolados que culminaram na publicação de um livro chamado Engenharia de Avaliações, em 1941. Depois de 1950 notou-se no Brasil um fenômeno intenso de urbanização por causa do êxodo rural, então rodovias e empreendimentos imobiliários fizeram necessário estimar o valor justo de imóveis para fins de desapropriação.

A primeira norma brasileira de avaliações (NB-502) foi publicada em 1970 pela ABNT. Algumas décadas depois (em 2001) a norma de avaliações mudou de nome, passando a se chamar NBR 14.653-2, trazendo o tratamento científico por Regressão Linear Múltipla. Em 2011 foi incorporado ao texto da NBR 14.653-2 a alternativa de avaliação por Redes Neurais Artificiais, que veio acompanhada por premissas e recomendações para seu uso, bem como regressão espacial e envoltória de dados sob dupla ótica.

A tendência é progredir para outros campos da IA como lógica difusa (abordagens de Mamdani, Takagi-Sugeno-Kang), sistemas híbridos (como o ANFIS), e aprofundar o tratamento com redes neurais artificiais.

## 2.3 TÉCNICAS APLICÁVEIS À AVALIAÇÃO IMOBILIÁRIA

Inicialmente, cabe destacar a existência prevista na NBR 14.653-2 (2011) de outros métodos alternativos, tais como:

a) homogeneização por fatores, onde o tratamento dos preços observados se dá mediante a aplicação de transformações matemáticas para expressar as diferenças entre os atributos do mercado e do avaliando, em termos relativos;
b) método involutivo, que identifica o valor de mercado de um imóvel a partir de seu aproveitamento eficiente, partindo de um estudo de viabilidade técnico-econômica, com um empreendimento hipotético compatível com as características e condições do local onde se insere, considerando cenários para execução e comercialização do produto hipotético. Este método pode ser horizontal (quando o

empreendimento hipotético é uma gleba urbanizável) ou vertical, quando o empreendimento hipotético é uma edificação;

c) método da capitalização da renda, onde o valor do imóvel é identificado com base na capitalização presente da sua renda líquida prevista, fazendo uso de cenários viáveis.

Segundo Sarip *et al* (2016), é válido afirmar que mais recentemente tem havido um crescente interesse em descobrir novas abordagens para estimar os valores dos bens imóveis que a principal motivação é a expectativa crescente do mercado de avaliações mais rápidas, para impulsionar decisões que sejam ágeis e seguras em termos de não comprometer a qualidade dos julgamentos.

Não obstante, Sarip *et al* (2016) ressaltam que a exigência de eficiência pode melhorar os meios pelos quais as estimativas de valor são geradas, e que esta característica pode conduzir à experimentação de novas técnicas, algumas das quais foram consideradas promissoras para a avaliação imobiliária. Ainda, pode-se afirmar, segundo os autores, que as novas abordagens incluem regressões múltiplas, redes neurais e aplicações mais complexas, como sistemas especialistas e lógica difusa.

O tradicional uso de inferência estatística por regressão linear múltipla possui, segundo Pelli (2004), a limitação de restringir e dificultar o conhecimento e análise dos processos. Este autor explica que, dado este panorama, a necessidade de novas técnicas para que ocorra a representação dos processos de avaliação de imóveis são evidentes e algumas possibilidades são o uso de sistemas híbridos, como redes neuro-fuzzy (ANFIS) e redes neurais artificiais (ANN).

Pelli (2004) adverte que os estimadores dos mínimos quadrados usados nas regressões lineares dependem muito do mapeamento das aproximações lineares dos dados do mercado e que, eventualmente, esta pode não refletir o valor de mercado do imóvel avaliando, especialmente quando os dados apresentarem alta dispersão e características variadas, então ressalta o uso da inteligência computacional, enfatizando a possibilidade de uso das redes neuro-fuzzy e das redes neurais artificiais, devido à sua capacidade de representar problemas não lineares por meio do aprendizado e de sua capacidade de generalização.

Já Sarip *et al* (2016) observaram que estudos recentes têm sido conduzidos para investigar técnicas mais complexas para modelar as relações não-lineares

subjacentes entre os fatores de precificação e o preço da propriedade, tais como Redes Neurais Artificiais (ANN) e Sistemas de Inferência Difusos (FIS).

Por outro lado, Santello (2004) alerta que as estimativas com uso de estatística inferencial envolvem condicionantes que são difusas e que oscilam conforme diversos interesses, sendo impactadas por variáveis ambientais e temporais, políticas e negócios e até mesmo alocação de recursos.

Pelli (2004) notou grandes avanços na área de engenharia, no que tange às modelagens de sistemas reais por meio de mecanismos de inferência. Nesta ótica, este autor cita que existem restrições em uso de Estimadores dos Mínimos Quadrados e elenca a possibilidade do uso de sistemas híbridos, como por exemplo, as redes neuro-fuzzy e redes neurais artificiais, que são aplicáveis às avaliações de imóveis urbanos.

Deste modo, Pelli (2004) continua explicando que a regressão linear múltipla é a metodologia mais aplicada pelos profissionais da área de avaliações imobiliárias, onde usualmente modelos linearizáveis por meio de transformações matemáticas são usados no intuito de representar o comportamento do mercado imobiliário, onde o uso de tais transformações encontra embasamento por serem variáveis não lineares. O autor esclarece que o uso da transformação de variáveis (principalmente quando aplicado na variável dependente) pode restringir e dificultar o conhecimento e análise dos processos.

Sarip *et al* (2016) explicam que diversos estudos foram realizados para modelar a relação não-linear subjacente entre as variáveis de preço e preço da propriedade para prever os preços de venda da habitação e que nos últimos anos, técnicas de modelagem não-linear mais avançadas, como Redes Neurais Artificiais (ANN) e Sistemas de Inferência Difusa (FIS) surgiram como técnicas eficazes para prever os preços de imóveis.

Conforme Sarip *et al* (2016), a ANN tem sido amplamente usada para prever o preço dos imóveis por muitos anos, enquanto que a técnica de sistemas difusos neuroadaptativos (ANFIS) foi introduzida recentemente. Segundo os autores, por outro lado, o modelo baseado em regressão de mínimos quadrados difusos (FLSR) é comparativamente novo. Os autores chegam a citar que, até onde possuem conhecimento, nenhuma previsão de preços de propriedade usando FLSR foi desenvolvida até a época de sua publicação, que deste modo seria a pioneira no assunto.

De fato, além do que fora posto por Sarip *et al* (2016), por Pelli (2004) e Santello (2004), outras técnicas também são aplicáveis, como os modelos de regressão baseados em árvores. Segundo Lasota *et al* (2019), os modelos baseados em árvore (TBM) têm especial desempenho quando os dados possuem muitos *outliers*.

Para a abordagem há também técnicas já amplamente conhecidas e técnicas menos conhecidas, como os modelos híbridos e os modelos de *ensemble learning*. Os modelos híbridos são modificações dos modelos já conhecidos. Os modelos de *ensemble learning* são modelos de aprendizado onde técnicas distintas são aplicadas em conjunto, então modelos de *machine learning* podem operar como sendo diversos modelos (cada modelo é um estimador), estes estimadores trabalham em conjunto e usualmente nota-se um incremento na precisão do modelo.

### 2.3.1 Regressão Linear Múltipla (*Multiple Linear Regression* / MLR)

A Regressão Linear Múltipla (Multiple Linear Regression / MLR) trata-se de uma técnica estatística para modelar a relação entre a variável dependente e as variáveis independentes. Eventualmente esta técnica é citada na bibliografia internacional por Análise de Regressão Múltipla.

Segundo ÇETINKAYA-RUNDEL (2019), a MLR é uma extensão da regressão simples entre duas variáveis para o caso em que permanece havendo uma variável dependente, mas são diversos os preditores (variáveis independentes).

Dessa forma, um modelo é geralmente escrito como:

$$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + erro \quad (1)$$

onde

$\hat{y}$ é o valor predito

$\beta_0$ é o termo independente

k é a quantidade de preditores

$\beta_i$ são os parâmetros usualmente estimados por *software* estatístico.
*erro* é a parcela que se refere ao erro, que é a própria variação não explicada na variável dependente.

Para calcular os parâmetros faz-se uso de funções estatísticas denominadas por estimadores. Quanto aos estimadores em amostras pequenas e com comportamento linear é esperado que sejam:

a) eficientes: deve apresentar uma variância mínima (dispersão muito pequena em torno do verdadeiro valor do parâmetro) quando comparado com outros possíveis estimadores. Bolfarine (2000) define como sendo aquele estimador que atinge o limite inferior da variância dos estimadores não viciados;
b) suficientes: deve usar todos os dados da amostra, Bolfarine (2000) define por ter a capacidade de condensar os dados sem perder nenhuma informação que esteja contida nos dados;
c) consistentes: segundo Bolfarine (2000), à medida que a o tamanho da base amostral, os estimadores ficam tão próximos do parâmetro que está sendo estimado quanto desejado;
d) invariantes: deve se manter caso seja aplicada alguma transformação linear ou escala nos dados;
e) não tendenciosos (não viesado): na média o valor deve coincidir com o valor médio do parâmetro da população estudada, Bolfarine (2000) explica que o vício deve diminuir à medida que o tamanho da base amostral aumenta.

Os modelos de MLR servem para tomar decisões de uma população a partir de uma amostra e então é necessário haver um teste de hipóteses, para averiguar se as informações sobre a população são consistentes com aquelas obtidas a partir da amostra. No objeto de estudo deste trabalho, coletam-se diversos elementos para compor uma base amostral que seja representativa do tipo apartamento. Aplica-se neste caso o teste de hipóteses que traz por hipótese nula que os parâmetros estimados são iguais a zero, ou seja, que não há modelo estatístico válido. A hipótese alternativa é o contrário da hipótese nula. O objetivo do teste é rejeitar a hipótese nula em favor da hipótese básica. A NBR 14.653-2 (2011) permite que a

hipótese nula seja rejeitada quando o valor p (*p-value*) associado ao teste F for menor que o nível de significância de 0,05.

Ao ajustar um modelo, deve-se levar em conta a significância alcançada, então havendo rejeição da hipótese nula, considera-se que uma ou mais variáveis independentes podem explicar a variação na variável dependente. Neste trabalho, nenhum ajuste de MLR acima de 0,05 será empregado.

Em seguida, testa-se a significância estatística de cada variável independente por meio do valor p, sob o mesmo critério do nível de significância de 0,05.

Obviamente, o teste de cada variável independente deve vir acompanhado de testes complementares, como a observação da coerência no valor do sinal e do coeficiente de determinação $r^2$. O coeficiente $r^2$ demonstra a proporção da variação na variável dependente que é explicada pelas variáveis independentes, sendo seu complemento devido a outros fatores, tais como variáveis não incluídas no modelo e à parcela de erro.

Segundo ÇETINKAYA-RUNDEL (2019), as correlações estarão entre -1 e +1, sendo as correlações fortes e negativas se aproximando de -1 e as correlações fortes e positivas se aproximando de +1, enquanto que as correlações nulas se aproximam de zero.

Conforme ÇETINKAYA-RUNDEL (2019), há condições que precisam ser atendidas para a linha de mínimos quadrados, a saber:

a) linearidade: os dados precisam apresentar tendência linear;
b) quase normalidade dos resíduos: os resíduos precisam apresentar aspecto semelhante à curva normal;
c) variabilidade constante: a variabilidade dos pontos em volta da linha de mínimos quadrados deve permanecer aproximadamente constante;
d) observações independentes: os resíduos das observações devem ser independentes, então não deve haver um padrão, se frustrado esta premissa, pode haver viés e então a estimação dos parâmetros poderá estar comprometida.

Moreira (2010) ressalta que a análise de regressão linear é a principal técnica usada na avaliação de imóveis. Certamente esta impressão se dá pela flexibilidade

em testar a inclusão de variáveis, pela interpretabilidade dos coeficientes de regressão e por ser amplamente reconhecida neste meio.

Assim como é feito no Brasil, Sarip *et al* (2016) citam que, na prática, o método de comparação de vendas ou abordagem de valor de mercado tem sido o tradicional e, de longe, o método mais comum adotado para a avaliação de imóveis, particularmente para imóveis residenciais. Explicam ainda que usando este método, o valor de um imóvel imobiliário é estimado com base nas vendas de outras propriedades comparáveis, que seria um meio para estabelecer uma estimativa do valor de mercado do imóvel.

Sarip *et al* (2016) avançam citando que há meios para valoração imobiliária, então a MLR consiste na aplicação de um método estatístico que estima o valor de uma variável dependente (por exemplo, preço da habitação) com base em determinados valores das variáveis independentes (por exemplo, os atributos físicos da casa).

Segundo Sarip *et al* (2016) o método de comparação de vendas baseia-se na noção de que existe uma relação direta entre o valor de mercado de um imóvel imobiliário e os preços de venda das propriedades comparáveis em que estes últimos representam alternativas de investimento competitivas no mercado.

Sarip *et al* (2016) justificam que influências de valor, como as características físicas e as qualidades de localização, são consideradas na análise da comparabilidade, num processo que embute a consideração de oferta e demanda, levando à opinião final sobre a estimativa de valor e que, na prática, os avaliadores utilizam dados históricos de transações como referência, realizando ajustes nos valores para explicar as diferenças que existem entre os bens imóveis e as propriedades comparáveis. Esta prática é prevista na NBR 14.653-2 (2011), quando os avaliadores têm a liberdade para precificar dentro do campo de arbítrio.

Estes autores citam ainda que esta abordagem tem sido amplamente utilizada em aplicações de predição apesar de suas limitações quando se trata de lidar com a relação entre variáveis independentes e dependentes na presença de *outliers*, não-linearidade e não-normalidade.

Sarip *et al* (2016) verificaram que as literaturas costumam indicar que a melhor técnica é a MLR, então a aplicabilidade dos outros métodos seria complementar para fins de melhorar esta técnica, no entanto, a MLR é inaplicável quando existem poucos elementos na amostra.

A avaliação imobiliária é uma ciência que carrega imprecisões. Matematicamente esta imprecisão consta no modelo como um termo de erro aleatório. Muitas vezes este erro aleatório pode ser parcialmente explicado através da introdução de variáveis qualitativas no modelo. Por meio destas variáveis é possível mensurar a vagueza (imprecisão) das informações. Algumas vezes as variáveis qualitativas admitem certo subjetivismo, dada a dificuldade (inerente ao processo de amostragem) de diferenciar o limite entre uma classe e outra. Exemplificando, a determinação do ponto exato em que um imóvel passa da qualidade área grande para área média resta prejudicada e carrega certo subjetivismo. Mesmo diante de variáveis qualitativas (como padrão construtivo e estado de conservação, por exemplo), a NBR 14.653-2 pede critérios objetivos, usualmente são modeladas como variáveis *dummy* ou *proxy* mas o pressuposto de linearidade nem sempre pode ser bem atendido. Este é um tratamento frágil que pode causar imprecisões nos resultados do modelo. A NBR 14.653-2 (2011) em seu item 8.2.1.2.2 recomenda que sempre que possível sejam usados modelos com variáveis quantitativas, justamente por reconhecer que modelos de regressão linear múltipla não são adequados para o tratamento das variáveis qualitativas, ocasião em que estabelece uma ordem de prioridades para variáveis qualitativas, sendo variáveis *dummy*, variáveis *proxy* (CUB para expressar padrão construtivo, IDH para a localização, coeficientes de depreciação para o estado de conservação), códigos alocados e por último códigos ajustados.

A NBR 14.653-2 (2011) recomenda a escolha da variável dependente como sendo aquela que melhor expresse os preços, como preço total ou unitário. Já quanto às variáveis independentes, a NBR 14.653-2 (2011) recomenda a escolha conforme o senso comum e conhecimentos adquiridos, citando exemplos (distância ao polo, área e localização).

Os próximos tópicos trazem uma alternativa para que as variáveis qualitativas sejam modeladas de maneira mais adequada.

### 2.3.2 Sistemas de Inferência Difusos (*Fuzzy Inference Systems* / FIS)

O sistema de inferência difuso de Mamdani foi uma primeira abordagem para resolver um sistema de inferência baseado em lógica difusa.

Segundo Caixa (2018), aspectos qualitativos causam subjetividades, imprecisões, incertezas e ambiguidades e o tratamento costuma se dar pelo discernimento e bom senso do avaliador. Muitas das variáveis tais como pequena (área), alto (padrão), perto (distância), bastante (vagas), velho (idade), e outros podem ser tratadas por sistemas difusos. Caixa (2018) destaca que os advérbios podem ser diminuidores ou aumentadores conforme a área de pertinência dos conjuntos difusos é ampliada ou reduzida.

Desta maneira, os Sistemas de Inferência Difusos (*Fuzzy Inference Systems* / FIS) partem da análise da dispersão das variáveis independentes (chamadas de variáveis linguísticas) para criar os conjuntos difusos que são relacionados com as funções de pertinência, onde por meio da relação entre as possibilidades de valores de cada conjunto difuso é traçado o grau de pertinência, sendo este o processo de fuzzificação.

Em seguida, com auxílio de um especialista na área de estudo, uma base de regras é estabelecida com o objetivo de relacionar as variáveis independentes com a variável dependente. Nesse contexto, os conjuntos difusos são utilizados para obter os consequentes, estabelecendo assim uma relação entre as variáveis de entrada (independentes) e a variável de saída (dependente).

No processo de inferência as regras são combinadas e é obtida a saída difusa do sistema num processo que se chama de agregação.

Posteriormente, como se deseja uma saída *crisp* (um número real), então ocorre o processo de defuzzificação.

Quanto aos processos de defuzzificação, tem-se os métodos:

a) método da média ponderada (WAM), comumente tratado na literatura especializada por *means weighted average method*, onde calcula-se a saída (*crisp*) como sendo uma média ponderada dos valores difusos de saída, então cada valor é ponderado por sua magnitude no cálculo da média ponderada, notando-se que o resultado final é um valor nítido (*crisp*) que representa a saída do sistema de lógica difusa;

b) método do primeiro máximo (FIRST.MAX), comumente tratado por *means first maxima*, onde calcula-se a saída (*crisp*) como o valor *crisp* que corresponde ao primeiro pico de máximo que se observa no conjunto difuso de saída, ou seja, é definido como o valor de entrada correspondente ao primeiro conjunto difuso que atinge seu valor máximo;
c) método do último máximo (LAST.MAX), comumente tratado por *means last maxima*, onde calcula-se a saída (*crisp*) como o valor *crisp* que corresponde ao último pico de máximo do conjunto difuso de saída, ou seja, é definido como o valor de entrada correspondente ao último conjunto difuso que atinge seu valor máximo;
d) método da média dos máximos (MEAN.MAX), comumente tratado por *means mean maxima*, onde calcula-se a saída (*crisp*) como o valor *crisp* que corresponde à média dos valores *crisp* correspondentes aos picos de máximos do conjunto difuso de saída;
e) método do centro de gravidade modificado (COG), comumente tratado por *means modified center of gravity*, onde calcula-se a saída (*crisp*) como a posição central do conjunto difuso de saída ponderado conforme suas magnitudes, então calculado o centro de gravidade, deve-se incluir um fator de ajuste que varia conforme a assimetria do conjunto difuso. Usualmente a escolha por esta alternativa pode melhorar a precisão da saída *crisp*.

A escolha do método de defuzzificação mais adequado dependerá de cada contexto, mas sob o escopo da avaliação imobiliária, convém testar os resultados (observando o desempenho e as métricas) pelos métodos WAM (se destaca quando houverem diversos picos de máximo) e COG (melhor aplicável caso de assimetrias significativas). Obviamente, a escolha equivocada poderá comprometer a precisão dos resultados.

*Sarip et al* (2016) explicam que a aplicação da lógica difusa para prever preços de imóveis geralmente mostra resultados encorajadores na modelagem do comportamento complexo e não linear entre variáveis de entrada e saída. Ainda, os autores destacam que o Sistema de Inferência *Fuzzy* (FIS) foi adotado recentemente

em vários estudos como um modelo de previsão e que essa técnica foi considerada útil especialmente quando a amostra de dados inclui variáveis linguísticas (por exemplo, aquelas relacionadas a atributos ambientais) ou quando os dados são extraídos principalmente de fontes não numéricas, como questionários. Os autores continuam, explicando que os sistemas de inferência *fuzzy* consistem num conjunto de regras IF-THEN que são disparadas simultaneamente quando são fornecidas com uma observação e que a força de disparo de cada regra é relativa ao grau de correspondência da parte antecedente. Isto significa que há desafios em termos de complexidade para a criação de um modelo FIS e estes incluem a determinação de conjuntos e regras *fuzzy*. Os autores advertem que a determinação requer conhecimento de domínio de especialistas humanos e destacam um relevante ponto negativo do método: a tarefa de ajustar as regras e conjuntos difusos pode consumir muito tempo.

Sarip *et al* (2016) verificaram que a técnica de lógica difusa aplicada à avaliação de imóveis apresenta bons resultados quando verifica-se haver comportamento não linear entre as variáveis. Ademais, observaram que os sistemas de inferência difusos são relevantes em variáveis linguísticas e não numéricas, ao passo que o desafio se torna determinar os conjuntos e as regras difusas a serem aplicados à determinada observação.

Desta maneira, Santello (2004) ressalta a necessidade que as atuais técnicas e métodos sejam melhorados, então propõe a lógica difusa como ferramenta de melhor visualização e entendimento das características dos imóveis em avaliação e operações lógicas do processo avaliatório.

Santello (2004) ressalta que com o uso da lógica difusa, é possível visualizar o dendograma de decisão com todos os critérios de análise, podendo inclusive serem observados os pesos usados nos critérios. O autor destaca que a possibilidade de visualização gráfica conduz à transparência, facilitando o entendimento das decisões por parte do leigo.

A grande propaganda de Santello (2004) é que numa atuação conforme papel de vendedor, se houvesse a comportamento atribuindo pesos maiores aos decisores que atuam no polo vendedor, como qualidades positivas e pesos menores para qualidades negativas, haveria a evidência deste fato, então, por este motivo a técnica da lógica difusa apresentaria transparência.

Desta maneira, Santello (2004) explica que é possível avaliar as variações que ocorrem nos preços dos imóveis a partir da modelagem das variáveis, tendo por critérios indicadores físicos, sociais e econômicos que são agrupados em conjuntos chamados de *fuzzy sets*, ou seja, operações lógicas em sistemas difusos. O autor explica que os dendogramas são grafos que são estruturados e são do tipo *top down*, onde são organizados os critérios e nestes grafos, cada nó, saída ou resultado é definido conforme determinado bloco de regras, onde as variáveis de entrada e saída podem ser discretas ou difusas, sendo operadas por operadores “e” ou “ou”.

Naturalmente, como se observa no mercado, dois eventos opostos podem ocorrer simultaneamente, e o uso da lógica difusa como apoio à decisão contempla esta possibilidade.

Santello (2004) destaca que uma possível aplicação para a lógica difusa é avaliar as variações nos valores de imóveis tendo o conhecimento prévio de que variáveis influenciam na formação do valor. Ele então adverte que a modelagem difusa não deve substituir os métodos existentes para avaliação de imóveis, mas que se presta para uma contribuição complementar a tais métodos, uma vez que as intervenções ocorrem no campo de arbítrio e que têm sua aplicabilidade quando houver ambiente com informações incompletas ou difusas.

### 2.3.3 Sistema de Inferência Difuso de Takagi-Sugeno-Kang (Takagi-Sugeno-Kang Fuzzy Inference System / TSK)

O Sistema de Inferência Difuso de Takagi-Sugeno-Kang (Takagi-Sugeno-Kang *Fuzzy Inference System* / TSK) é uma técnica que partiu do sistema Mamdani, mantendo o antecedente e substituindo o consequente por uma equação polinomial que resulta na própria saída (*output*) do sistema.

Segundo Valle (2021), um sistema TSK possui as componentes:

a) dicionário para definir os conjuntos difusos;
b) base de regras, para definir a relação entre as variáveis; e
c) método de inferência, que é como será determinada a saída para determinada entrada.

A técnica TSK se constitui num misto entre uma regressão linear convencional e um sistema difuso. Na regressão linear convencional, o método de mínimos quadrados é usado para encontrar os parâmetros βi que minimizam a soma dos erros quadrados. Em geral, a diferença entre os valores estimados e observados surge da incerteza da estrutura do modelo ou de observações imprecisas.

A imprecisão incorporada (que é bastante comum em situações de previsões de preços de venda) torna importante a aplicação da técnica TSK. A técnica TSK consiste em usar as regras para constituir filtros e a partir deles clusterizar os dados para criar cada função polinomial. Note que a cada regra corresponderá uma única função polinomial.

Como no TSK as equações são ajustadas diretamente aos dados, então a modelagem é mais precisa que aquela obtida através da abordagem por Mamdani (FIS) por capturar mais precisamente a relação funcional entre a variável dependente e as variáveis independentes do setor imobiliário.

Ross (2017) explica que sob a tentativa de desenvolver uma abordagem sistemática na geração de regras difusas partindo de um conjunto de dados, uma regra num modelo TSK com duas variáveis independentes *x* e *y* para explicar a variável dependente *z*, costuma ter a seguinte forma:

$$Se\ x\ \text{é}\ A\ e\ y\ \text{é}\ B\ , então\ z\ \text{é}\ z = f(x, y) \tag{2}$$

onde
*x* e *y* são variáveis independentes
*z* é a variável dependente
A e B são conjuntos difusos dos antecedentes

Assim como no método FIS, todas as regras serão calculadas e é comum em sistemas com muitas regras que a maior parte das regras tenha peso zero e, portanto, não serão ativadas.

Os pesos são obtidos a partir do cálculo com t-normas, Valle (2021) explica que usualmente se utilizam o mínimo e o produto.

A etapa de defuzificação é facilitada por usar média ponderada ou soma ponderada, e qualquer destas alternativas é facilmente computada, então os

modelos TSK demandam menor custo computacional. A média ponderada ocorre conforme a equação 3:

$$z = \frac{\sum_{i=1}^{n}(w_i * f(x_1, x_2, \ldots, x_n)}{\sum_{i=1}^{n} w_i} \quad (3)$$

onde

$x_i$ são as variáveis independentes

$z$ é a variável dependente

$w_i$ são os pesos (ativação das regras)

### 2.3.4 Sistemas de Inferência Difusos Neuroadaptativos (Adaptive Neuro Fuzzy Inference System / ANFIS)

Os Sistemas de Inferência Difusos Neuroadaptativos (Adaptive Neuro Fuzzy Inference System / ANFIS) foram idealizados por Jang (1993), onde partindo de um modelo TSK cujas funções de pertinência são gaussianas, se faz uso de redes neurais artificiais para calibração destas funções.

Valle (2021) explica que dispondo-se de um conjunto de dados significativo, pode-se abrir mão de compor um dicionário e de estabelecer uma base de regras, então a partir dos dados automaticamente se determina o dicionário e as regras difusas num modelo de Takagi-Sugeno-Kang, fazendo uso da técnica ANFIS (*Adaptative Neuro Fuzzy Inference System*), que é um modelo de Takagi-Sugeno adaptativo.

A ideia é reforçada por Sarip *et al* (2016), que destacam que a tarefa de ajustar as regras e conjuntos difusos pode consumir muito tempo, razão pela qual o sistema difuso pode ser integrado com as capacidades de aprendizado da ANN, resultando no Sistema de Inferência Difuso Neuroadaptativo (ANFIS).

Deste modo, Sarip *et al* (2016) ressaltaram a potencialidade em ser usado para suprir o problema da dificuldade em ajustar as regras e os conjuntos difusos, onde a capacidade de aprendizado da RNA tem excelente desempenho, especialmente para conjuntos grandes.

Segundo Sarip *et al* (2016), uma série de estudos foram realizados para investigar a aplicação do ANFIS para prever os preços dos imóveis e concluiu-se

argumentando que mais pesquisas são necessárias para explorar a abordagem da ANFIS usando um conjunto de dados com maior dimensionalidade.

Pelli (2004) ressalta que a capacidade de aprendizado de mapeamentos complexos depende da tipologia usada e do número de funções de pertinência no caso das redes neuro-*fuzzy*. Isto significa que um incremento na complexidade do problema poderá implicar no aumento da quantidade de funções de pertinência ou até mesmo, por ser um sistema híbrido, da quantidade de neurônios.

Segundo Pelli (2004) as redes neuro-*fuzzy* seguem a configuração de acordo com a teoria dos sistemas difusos, onde as transições dos elementos em determinado conjunto ocorrem gradualmente conforme um grau de relacionamento que é chamado de função de pertinência. O autor ressalta que o uso desta estrutura é útil para representar variáveis linguísticas em regras do tipo se-então, onde tais regras podem ser combinadas produzindo um mecanismo de inferência. Ao combinar estes sistemas de regras se-então com as estruturas das redes neurais artificiais, tem-se um sistema de inferência adaptativo neuro-*fuzzy*, que a literatura trata por ANFIS (Adaptive Neuro-*Fuzzy* Inference Systems).

Ao usar tais modelos ANFIS, é necessário que preliminarmente ocorram os procedimentos de treinamento, para calibrar os parâmetros, e somente então haver a estimação.

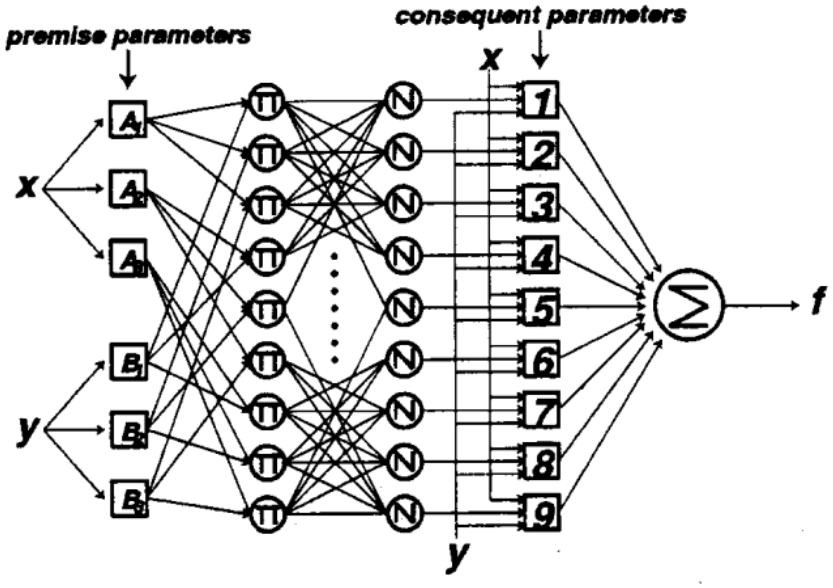
Pelli (2004) adverte que as redes neuro-*fuzzy* costumam apresentar resultados mais frágeis que outros métodos no que tange à sensibilidade das variáveis, podendo ocorrer até mesmo inversões em alguns dos pontos estimados, enquanto que as outras técnicas (ANN, por exemplo) e a MLR costumam apresentar comportamentos consistentes.

Segundo Kamire *et al* (2021), quando os dados já estão preparados, é difícil encontrar uma relação linear entre as diversas variáveis independentes do conjunto de dados, então é quando a técnica ANFIS é empregada.

Corrêa (2020) explica que o ANFIS é uma rede *feedforward* (ANN onde o fluxo avança apenas em uma direção, da camada de entrada em direção à camada de saída), cuja aprendizagem se dá pelo método da descida em gradiente e também mínimos quadrados, e onde um sistema TSK é gerado.

Quanto à arquitetura da técnica ANFIS, a figura 1 mostra as camadas.

Figura 1 - Arquitetura da técnica ANFIS




Fonte: Jang (1993)

Corrêa (2020) explica que esta rede possui cinco camadas e cada camada atua como um dos processos de inferência de um TSK.

Notam-se cinco camadas, onde a primeira recebe os valores de entrada e forma as funções de pertinência, é a etapa de fuzzificação. Em seguida a segunda camada calcula o peso da regra. A normalização dos pesos calculados na segunda camada ocorre na terceira camada (divide a força de ativação total de cada atributo). A quarta camada recebe os valores que foram normalizados na terceira camada e retorna os valores definidos pelos parâmetros resultados da defuzzificação, fornecendo-os à quinta camada que é a saída do sistema.

Valle (2021) explica que o ANFIS implementa um modelo TSK de ordem 1, com número fixo de regras e conjuntos difusos nos antecedentes. De fato, para a situação em que muitas variáveis são necessárias para explicar a variabilidade da variável dependente, a técnica ANFIS pode não ser uma alternativa adequada.

Nesta técnica, a função de sino é estabelecida para as funções de pertinência e é possível determinar seu aspecto por meio da definição de seu ponto central, da dispersão (espalhamento) e da proporção que ocorre seu decaimento.

### 2.3.5 Técnica de Subamostragem da Classe Majoritária (Majority Class Undersampling Technique/ *undersampling*)

A Técnica de Subamostragem da Classe Majoritária (Majority Class Undersampling Technique / *undersampling*) visa balancear as classes, por meio da eliminação de dados da classe majoritária. Dessa maneira a classe majoritária terá uma quantidade de elementos compatível com os elementos da classe minoritária.

Por ocasião deste procedimento informações importantes podem ser perdidas, então não necessariamente a técnica de *undersampling* irá conduzir a melhores resultados.

A ideia é que ao garantir o balanceamento (equilíbrio) na distribuição dos elementos, o problema do conjunto de dados com classes desbalanceadas ficaria resolvido.

O receio em treinar um modelo a partir de um conjunto de dados com classes desbalanceadas é que o modelo irá aprender muito mais sobre a classe majoritária do que as classes minoritárias e então a capacidade de generalização do modelo fica comprometida.

Nos trabalhos correlatos (vide capítulo 7), nenhum trabalho aplicou *undersampling*, sendo treinados os modelos com classes desbalanceadas e eventualmente colhendo péssimos resultados, reforçando a evidência de fazer uso da abordagem de *undersampling* para melhorar o conjunto de dados.

Em termos de aplicabilidade desta técnica ao escopo das avaliações imobiliárias a NBR 14.653-2 (2011) ignora o assunto.

Cabe destacar que a aplicação desta técnica sob o contexto das avaliações imobiliárias não é proibida por norma e avaliadores experientes evitam o uso de classes com muitos elementos assemelhados, sob a justificativa que o modelo reduz seu desempenho em termos de generalização. Nota-se ainda que o processo de seleção das variáveis independentes e de elementos fica facilitado.

# 3 TRABALHOS CORRELATOS

Neste capítulo será apresentada uma análise do estado da arte sobre o domínio do problema, com foco em estudos e pesquisas publicadas que utilizam as técnicas MLR, FIS, TSK e ANFIS na avaliação imobiliária. Este capítulo é fundamental para o sucesso deste trabalho em razão de apresentar possíveis abordagens que podem ser utilizadas na avaliação imobiliária, destacando os principais desafios a serem enfrentados, tais como as particularidades das técnicas, o tamanho do conjunto de dados, a escolha das métricas e as possíveis conclusões sobre as comparações entre os modelos gerados. Ao apresentar essas abordagens, o capítulo serve de suporte para atingir os objetivos estabelecidos para o trabalho, atuando como guia no desenvolvimento e análise dos modelos propostos.

Além disso, esta análise do estudo da arte estabelece critérios para a definição da base metodológica e justifica a sequência de etapas e de decisões tomadas ao longo do trabalho. Desta forma, diversos estudos realizados na área de avaliação imobiliária são descritos, enfatizando as contribuições e eventualmente também as limitações dos modelos propostos.

Diversos trabalhos correlatos foram elencados, todos eles aderentes ao domínio do problema enfrentado neste trabalho e fazendo uso de pelo menos uma dentre as técnicas elencadas na pesquisa. Eventualmente algum trabalho usou denominação própria para denominar as técnicas e seus componentes, então os rótulos que identificam as técnicas e seus componentes aqui descritos estarão uniformizados para manter um padrão nesta pesquisa, facilitando o entendimento.

Nas cinco primeiras seções tem-se a aplicação de cada uma das técnicas em conjuntos de dados distintos, então os trabalhos descritos estão altamente aderidos ao domínio do problema. Já nas últimas três seções empregaram-se técnicas híbridas e mostrou-se a aplicação sob o contexto da formação de uma planta de valores genéricos.

## 3.1 DESENVOLVIMENTO E COMPARAÇÃO DE MODELOS FIS E ANFIS BASEADOS EM CONHECIMENTO PARA AVALIAÇÃO DE IMÓVEIS RESIDENCIAIS

O artigo intitulado Desenvolvimento e comparação de modelos FIS e ANFIS baseados em conhecimento para a avaliação de imóveis residenciais (*Knowledge-based FIS and ANFIS models development and comparison for residential real state valuation*) de autoria de Yalpir e Ozkan (2018) tem como principal objetivo investigar aplicabilidade da metodologia dos sistemas baseados em lógica difusa na determinação dos valores dos imóveis em avaliações imobiliárias, comparando os resultados obtidos por meio da técnica FIS com a técnica ANFIS. Para tanto, criaram uma base amostral composta por 320 imóveis situados em 8 bairros da região metropolitana da cidade de Konya (Turquia), contendo as variáveis que poderiam explicar a variação dos preços dos imóveis, e essa base amostral serviu para compor um modelo FIS. As variáveis independentes usadas foram área, idade, condições de acesso, propriedades físicas e localização, e a variável dependente foi o valor de mercado. Dessa maneira, o modelo se constituiu através da combinação de diversos valores de entrada para obter a saída. A estrutura do modelo usou Mamdani através do *Fuzzy Toolbox* do Matlab 6.0. A qualidade do modelo foi constatada pela métrica MAPE e por haver sido constatada alta acurácia nos resultados do modelo, concluiu-se pela confiabilidade do modelo obtido para fins de avaliação imobiliária.

Segundo Yalpir e Ozkan (2018) o mercado imobiliário não é um mercado ideal, é imperfeito e a variabilidade de preços imobiliários carrega subjetivismo porque varia de pessoa a pessoa, já o valor de mercado é o resultado da intersecção entre as curvas de oferta e demanda. Nessa ótica, destacam que as diferenças são inevitáveis e que o processo avaliatório deve ocorrer baseado em critérios objetivos, se distanciando tanto quanto possível da subjetividade, sendo importante que os critérios sejam postos antes do início do modelo para possibilitar uma padronização.

Yalpir e Ozkan (2018) esclarecem que para enfrentar a problemática, a literatura apresenta diversas técnicas computacionais avançadas que têm sido sugeridas para a avaliação imobiliária, em especial por haver grandes quantidades de dados, tais como FIS, ANFIS, FLSR, ANN e algoritmos genéticos, e destacam a

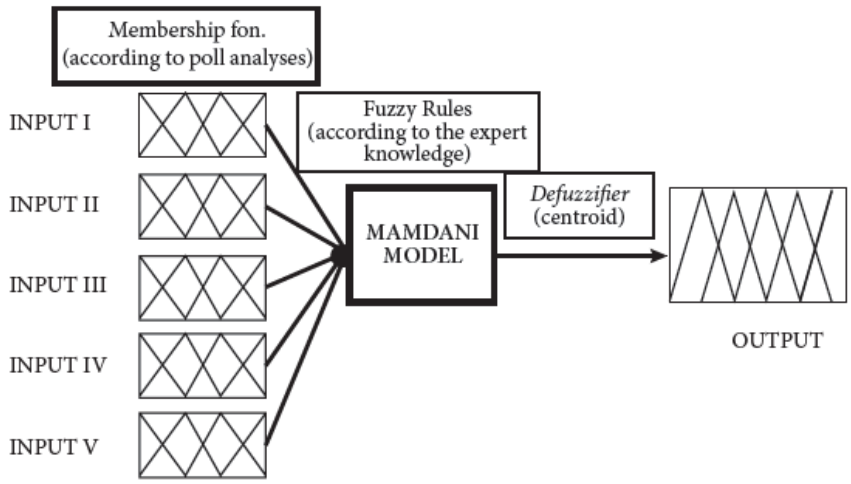
possibilidade de comparar os resultados entre uma e outra técnica. Relatam ainda a possibilidade da aplicação de abordagens híbridas.

Destacam Yalpir e Ozkan (2018) que as alternativas para usar a técnica FIS seguem as abordagens Takagi-Sugeno (menos palpável à intuição humana) e de Mamdani (requer conhecimento especialista na determinação da base de regras). Depois de delimitar a região e formar o conjunto de dados, foi aplicada a técnica FIS (usando Mamdani) e quando convergido, o conjunto de dados resultante serviu para base na modelagem de Takagi-Sugeno, onde posteriormente os resultados foram comparados.

A definição das variáveis linguísticas foram resultado de pesquisas com pessoas, para coleta de opinião. Posteriormente o conjunto de dados foi revisto conforme os resultados das pesquisas, resultando um conjunto de dados composto por 120 elementos, onde 80 elementos foram usados para treino e os 40 elementos restantes foram usados para teste.

A análise da técnica FIS conforme o modelo de Mamdani se iniciou pela determinação das variáveis de entrada e de saída, que são a estrutura FIS indicada na figura 2.

Figura 2 - Estrutura para a técnica FIS

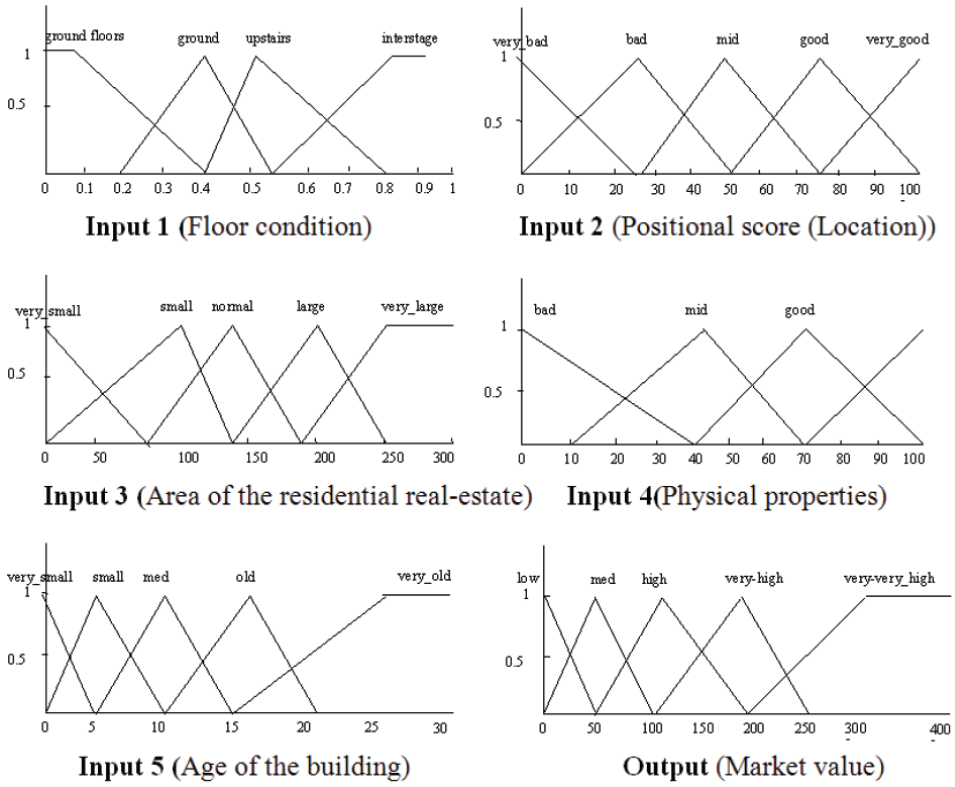



Fonte: Yalpir e Ozkan (2018)

Sob o intuito de definir os conjuntos difusos, grupos foram formados por meio da especificação de intervalos. Neste estudo as funções triangulares foram escolhidas ao invés de funções trapezoidais ou gaussianas, sendo que em alguns casos os últimos membros foram transformados em funções trapezoidais para as

variáveis condição de acesso, área, idade e valor de mercado, conforme consta na figura 3:

Figura 3 - Conjuntos difusos para a técnica FIS




Fonte: Yalpir e Ozkan (2018)

Posteriormente, considerando as cinco variáveis de entrada (*inputs*) foram escritas 85 regras, todas elas usando o operador AND. A etapa de defuzzificação traz por alternativas os métodos do centróide, do centro dos máximos, e da média dos máximos, houveram testes e concluiu-se que o melhor valor de mercado foi alcançado pelo método do centróide (COG).

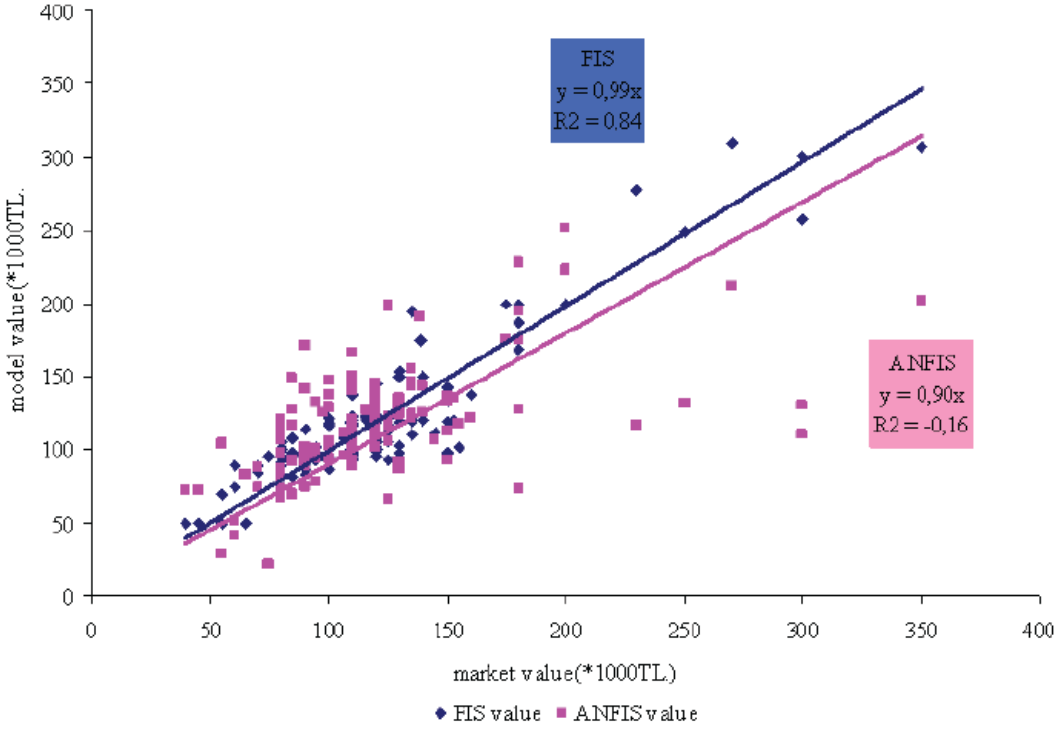
O tipo de sistema de inferência difusa segundo Mamdani estabeleceu critérios que serviram de base para a aplicação da técnica ANFIS.

A técnica ANFIS faz uso de conjuntos difusos modelados com função gaussiana nos antecedentes e implementa um modelo de Takagi-Sugeno com equações de primeiro grau sem constante nos consequentes, tendo um número fixo de regras.

A avaliação estatística do desempenho do modelo usou os índices MAPE, desvio padrão e $r^2$.

Numa análise conclusiva gerada projetando os dados no modelo (obtenção dos preços preditos) seguido de regressão linear auxiliar, ficou evidente que para estes dados e configuração de modelos, os índices de desempenho ($r^2$ e MAPE) mostraram que a faixa de estimativa do modelo FIS é mais bem-sucedida do que a do modelo ANFIS. Na figura 4 pode-se observar o gráfico de poder de predição, conforme orienta a NBR 14.653-2 (2011), onde os preços preditos pelo modelo constam no eixo das ordenadas e os valores de mercado constam no eixo das abscissas, além de mostrar o intervalo possível de dados para não haver extrapolação do modelo, o gráfico sugere possíveis valores *outliers*. O gráfico mostra os preço preditos e observados e a relação é destacada por uma equação. Como a métrica $r^2$ avalia a qualidade da predição dos dois modelos, tem-se a indicação que os resultados do modelo FIS apresentaram uma melhor explicação para a variação entre os preços preditos e observados do que o modelo ANFIS. Isso significa que o modelo FIS se ajustou melhor aos dados do que o modelo ANFIS.

Figura 4 - Gráfico do poder de predição: FIS x ANFIS

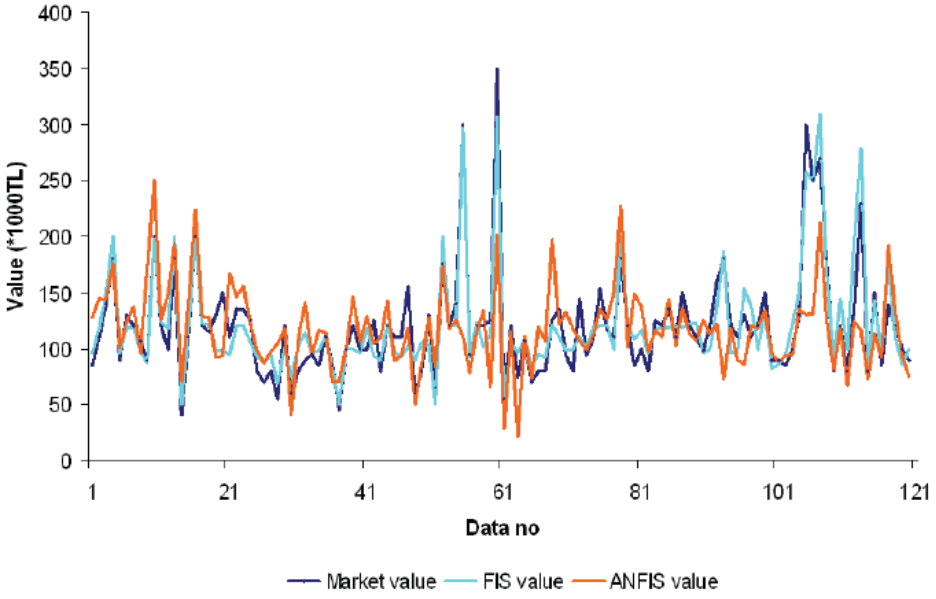



Fonte: Yalpir e Ozkan (2018)

A figura 5 ilustra o traçado de pontos quando comparando os dados com preços de mercado e os preços preditos pelos modelos FIS e ANFIS. Nela, nota-se

que o modelo ANFIS não tem boa precisão para valores muito altos, apresentando erros grandes nessas ocasiões.

Figura 5 - Comparações: FIS e ANFIS por dado




Fonte: Yalpir e Ozkan (2018)

Da figura 5 notou-se que o modelo FIS apresentou maior aderência aos dados de teste quando comparado com o ajuste gerado através da técnica ANFIS, razão pela qual se conclui que o modelo FIS é mais bem sucedido que o ANFIS. Em especial, nota-se que o modelo gerado usando a técnica FIS prediz os valores de valores mínimos e máximos com menor erro que o modelo ANFIS.

Desta forma, Yalpir e Ozkan (2018) concluíram que o sistema FIS apresenta muitas vantagens na avaliação imobiliária, mas que por não ser diretamente treinado diretamente a partir do conjunto de dados, a aplicabilidade dos modelos fica reduzida.

## 3.2 CAPACIDADE DE PREVISÃO DE VALOR DE PROPRIEDADE REAL USANDO LÓGICA DIFUSA E ANFIS

O artigo intitulado Capacidade de previsão de valor de propriedade real usando lógica difusa e ANFIS (*Real Property Value Prediction capability using Fuzzy Logic and ANFIS*), de autoria de Kamire *et al* (2021) tem como objetivo principal a comparação entre modelos FIS e ANFIS na predição de imóveis situados na cidade de Pune, na Índia, através da comparação de desempenho (MAPE e $r^2$), sendo análises estratificadas conforme cada bairro da cidade. Cabe destacar que Pune é um importante centro industrial e acadêmico da Índia, contando com uma população pouco superior a 6.200.000 pessoas. O conjunto de dados contém 1.460 elementos situados ao longo de quatro bairros. As variáveis independentes usadas foram área, idade, condições de acesso, propriedades físicas e localização, e a variável dependente foi o valor de mercado. Dessa maneira, o modelo se constituiu através da combinação de diversos valores de entrada para obter a saída. A estrutura do modelo usou Mamdani através do *Fuzzy Logic Toolbox* do Matlab 6.0. A qualidade dos modelos foi constatada pelas métricas MAPE e $r^2$, constatou-se alta acurácia nos resultados dos modelos, e concluiu-se pela confiabilidade do modelo obtido para fins de avaliação imobiliária.

Para tanto, obtiveram-se dados de transações e de avaliações, dispersos em quatro bairros da cidade por cada bairro ter suas características, as variáveis independentes (área do imóvel, potencial construtivo, classe social predominante, idade da construção, número de vagas, proximidade com equipamentos urbanos, proximidade com polo desvalorizador), foram relacionadas num conjunto de dados e a variável dependente escolhida foi o valor do imóvel. Estes dados foram a base para desenvolvimento de modelos FIS e ANFIS. A abordagem foi partir dos bairros menores, obtendo resultados para conduzir a formação dos modelos nos bairros maiores.

Na modelagem com a técnica FIS, o modelo foi melhorado incrementalmente e o alcance das funções de pertinência dos conjuntos difusos ocorreu conforme a quantidade de dados. Cada bairro recebeu regras próprias. Decidiu-se o alcance do conjunto de saída para ficar compatível com o conjunto de dados, e em todos os conjuntos difusos foi feito uso de funções triangulares. O número de regras variou de setor a setor conforme a disponibilidade de elementos

no conjunto de dados. Concluiu-se que o modelo fez predições compatíveis com os dados de teste. Inicialmente um menor conjunto de dados foi usado para testar diversas funções de pertinência, onde constatou-se que a função de pertinência triangular tinha o menor erro. Esse foi o critério para usar a função de pertinência triangular no restante da modelagem. Foram feitas regras com OR e AND, constatou-se que regras com AND tinham menores erros e o operador AND foi o padrão para o restante da pesquisa.

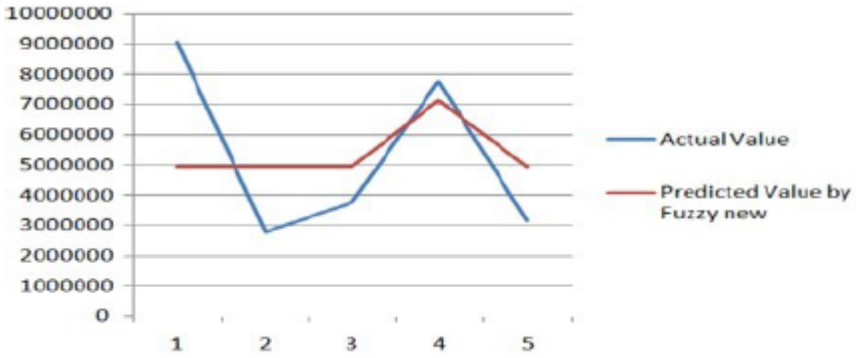
Já na modelagem com a técnica ANFIS, Kamire *et al* (2021) destacaram que foi muito mais fácil e rápida que a modelagem em FIS. A metodologia seguiu o mesmo caminho daquela que originou os modelos conforme a técnica FIS: partiram de um conjunto de dados menor (do bairro 1) e a partir dele fizeram evoluções. Para tanto, o conjunto de dados foi dividido entre treino e teste. A abordagem por clusterização conduziu resultados melhores que a abordagem por *grids*. O erro restou reduzido para processamentos em 300 épocas. A divisão do conjunto de dados em 80% dos dados para treino e 20% dos dados para teste foi a de menor erro e foi adotada para os modelos dos outros bairros.

Os modelos obtidos foram usados para a predição dos valores não incluídos no treino e a partir deles obtiveram-se as métricas de desempenho MAPE e $r^2$.

Os autores observaram que a técnica ANFIS resultou modelos cujas predições feitas nos dados de teste apresentaram valores muito próximos aos valores reais, tanto para conjuntos pequenos (setor 1 treinado com 55 dados) quanto para conjuntos grandes (setor 6 com 632 dados). Também ficou evidenciado que a técnica FIS apresentou resultados bons para os conjuntos treinados com muitos dados e resultados ruins para o conjunto com poucos dados.

A figura 6 mostra o desempenho precário do modelo FIS para poucos dados, nota-se que somente uma das cinco predições resultou próxima do valor real.

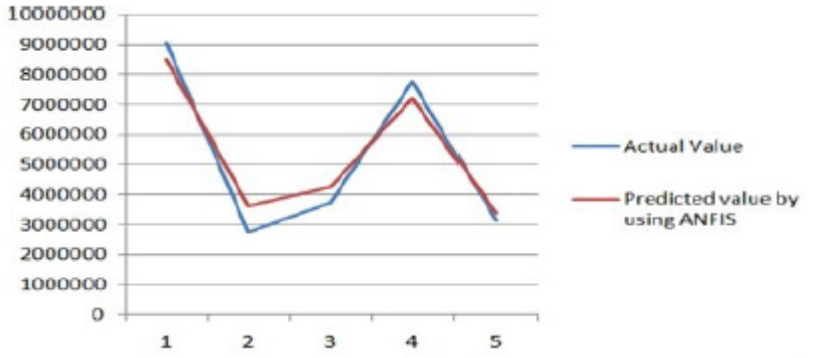
Figura 6 - Precisão de predição para o Bairro 1 usando FIS




Fonte: Kamire *et al* (2021)

A figura 7 mostra o ótimo desempenho do modelo ANFIS para poucos dados, todas as cinco predições resultaram valores próximos aos valores reais.

Figura 7 - Precisão de predição para o Bairro 1 usando ANFIS




Fonte: Kamire *et al* (2021)

Segundo Kamire *et al* (2021), os dados do bairro 3 contém particularidades regionais importantes e então nem o método FIS, nem o método ANFIS mostraram bons resultados e neste caso, o método ANFIS apresentou resultados um pouco melhores que o FIS.

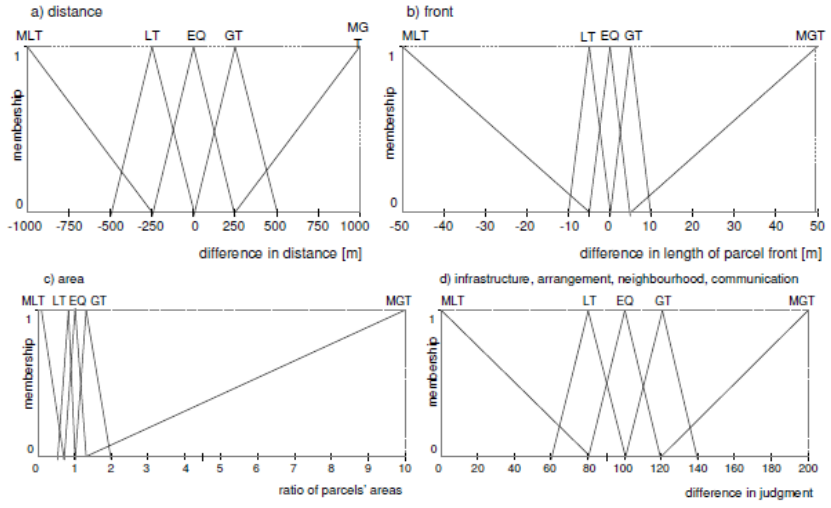
Os autores concluíram que em conjunto de dados menores, os erros dos modelos são maiores e que em conjuntos de dados maiores, os erros dos modelos são menores.

## 3.3 COMPARAÇÃO DE MODELOS FUZZY MAMDANI E TSK PARA AVALIAÇÃO IMOBILIÁRIA

O artigo intitulado Comparação dos modelos difusos Mamdani e TSK para avaliação imobiliária (*Comparison of Mamdani and TSK Fuzzy models for real state appraisal*), de autoria de Król *et al* (2007) tem como objetivo principal fazer uso de um algoritmo evolucional para a criação automática de uma base de regras para aplicação das técnicas FIS e TSK sob o escopo da avaliação imobiliária. Para tanto conduziram experimentos usando Matlab num conjunto de dados composto por oito variáveis independentes (distância ao centro, frente, área, infraestrutura da localidade, infraestrutura do imóvel, posição, vizinhança, acesso ao transporte público) e pela variável dependente uma taxa que representa a proporção do valor da propriedade (dado) em relação ao valor daquela escolhida como representativa, e 134 dados numa localidade da Polônia, sendo 60 dados para treino e 74 dados para teste. Para cada variável, cinco funções de pertinência triangulares ou trapezoidais foram estabelecidas.

A figura 8 mostra as funções de pertinência usadas para as variáveis dependentes:

Figura 8 - Funções de pertinência para as variáveis independentes




Fonte: Król *et al* (2007)

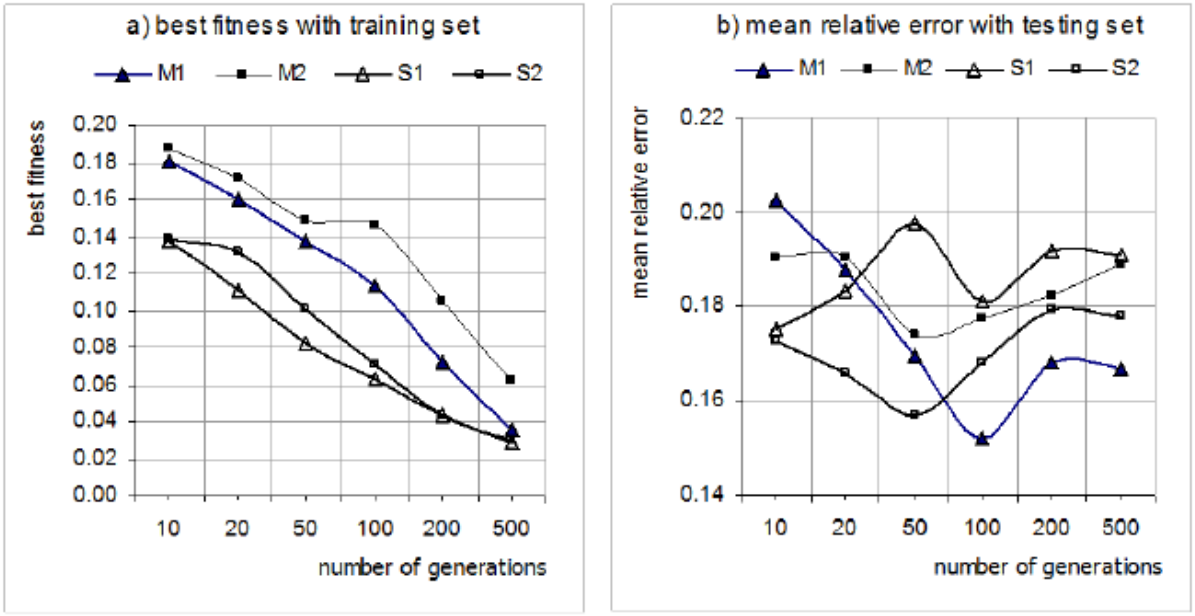
Para a saída, foram nove funções (sob o método de defuzzificação COG) foram usadas na aplicação da técnica FIS e nove funções de grau zero para a alternativa TSK, estas obtidas por média ponderada entre todas as regras de saída.

O aprendizado por regras difusas através do algoritmo evolucionário usou a ferramenta *Matlab Genetic Algorithm and Direct Search Toolbox*, parametrizada conforme o código do cromossomo, cruzamento e operações de mutação.

No treinamento do algoritmo genético, cada cromossomo continha uma regra (8 *bytes*) onde o último *byte* continha a saída (variável dependente). A inicialização se deu com os cromossomos sendo as regras possíveis de serem formadas a partir do conjunto de dados de treino, onde novos cromossomos posteriormente foram gerados a partir dos cromossomos existentes, de maneira aleatória. A função de aptidão foi calculada como sendo a média do erro relativo entre os preços dos elementos do conjunto de dados de treino conjunto e os valores daqueles elementos determinados pelo FIS usando uma base de regras produzida por uma geração subsequente do algoritmo genético. Quanto à criação de novos cromossomos, o elitismo foi fixado em 2, a taxa de cruzamento foi estabelecida em 0,8 e portanto a taxa de mutação resultou 0,2. O conjunto de dados gerados foi dividido em duas categorias (grupos 1 e 2) conforme o critério para mutação, sendo o grupo 1 formado por cromossomos que vão sendo alterados com regras aleatórias conforme probabilidade de 7% em cada cromossomo e o grupo 2 formado por cromossomos gerados inteiramente por regras aleatórias (os cromossomos anteriores são descartados).

A figura 9 mostra os resultados obtidos, onde M indica Mamdani e S indica o emprego da técnica de Takagi-Sugeno_kang, e os índices 1 e 2 indicam o grupo. Na figura 9, o melhor *fitness* com o conjunto de treinamento foi praticamente o mesmo para todas as alternativas. Nota-se também que o RMSE indica que os modelos baseados em TSK convergem mais rapidamente, independentemente do critério de mutação.

Figura 9 - Gráfico comparativo sobre a convergência entre os modelos




Fonte: Król *et al* (2007)

Quanto à convergência ao longo de 500 gerações, se deu mais rapidamente conforme a técnica de TSK (indicada por S) quando comparada com a técnica de Mamdani (indicada por M). A qualidade dos modelos se deu por meio da análise da métrica MAPE. Os autores concluíram que ambos modelos atingem praticamente os mesmos valores de melhor taxa *fitness* (sendo relativamente baixa), e destacaram a necessidade de eliminar *outliers* dos conjunto de dados de treino e teste.

## 3.4 GERAÇÃO DE REGRAS DIFUSAS POR APRENDIZADO A PARTIR DE EXEMPLOS

O artigo intitulado Geração de regras difusas por aprendizado a partir de exemplos (*Generating Fuzzy Rules by Learning from Examples*) de autoria Wang e Mendel (1992) tem como objetivo principal apresentar uma técnica que foi desenvolvida para gerar regras difusas a partir de dados numéricos, conforme o sistema de inferência de Mamdani (FIS).

O estudo se inicia destacando os problemas notados na época, onde problemas de controle e processamento de sinais por apresentarem entradas numéricas (obtidas dos sensores) e informações linguísticas (obtidas de especialistas humanos) precisavam ser tratadas, então usualmente sistemas

especialistas combinavam controles padronizados e técnicas para processamento de sinais com sistemas especialistas de uma maneira empírica, sendo necessários testes de simulação até concluir que a abordagem era suficiente para o problema enfrentado. Um ponto fraco nesta metodologia é que o funcionamento com êxito para um sistema e não garante necessariamente que irá funcionar para outro sistema e não havendo um padrão para modelar e representar diferentes aspectos dos dados de entrada, análises teóricas para essas abordagens resta prejudicada.

Os autores explicam que no projeto de um sistema não linear interessam duas informações, sendo a experiência do especialista humano (expressa por regras se-então) e elementos amostrais de entradas-saídas, que possam ser entendidos como cases de sucesso.

Wang e Mendel (1992) partem do pressuposto que é possível desenvolver uma abordagem geral que combina ambos tipos de informação num *framework* (no caso, o *framework* é a base de regras difusas), fazendo uso das informações cooperativamente e ao mesmo tempo para resolver problemas de projeto envolvendo controladores, desenvolvendo uma abordagem geral. Cinco etapas estabelecem esta técnica:

d) dividir as entradas e saídas dos dados numéricos em regiões difusas: conforme a relação entre número de variáveis (N) que uma determinada regra tem, divide-se os intervalos seguindo 2N + 1 regiões, então cada variável recebe uma função de pertinência, onde a forma e cada função é triangular então cada um dos vértices atingirá o ponto central das outras funções de pertinência, sendo o ápice do triângulo situado na metade do intervalo;

e) gerar regras difusas a partir dos dados fornecidos: determina-se os níveis das diferentes regiões e pondera-se para que a soma dos níveis seja 1, como a premissa é que nenhuma das variáveis de entrada sejam idênticas, então fez-se uso do operador AND e posteriormente as regras são criadas;

f) atribuir um nível em cada uma das regras geradas sob o propósito de resolver conflitos entre as regras geradas: os conflitos ocorrem em razão de que cada par de dados gera uma regra e há muitos pares de dados, e portanto muitas regras são geradas admitindo-se a possibilidade que mesmos antecedentes resultem consequentes

diferentes, então criaram-se níveis entre as regras e será escolhida a regra com maior nível, resolvendo o conflito por eliminação das regras que não possuem o maior nível;

g) criar uma base de regras difusas combinadas entre si baseada em ambas regras geradas e regras linguísticas de especialistas humanos: analogamente à etapa anterior, a base de regras combinadas também recebe um nível e conflitos são resolvidos elegendo-se a regra que tem o maior nível.
h) determinar um mapeamento do espaço de entradas para o espaço de saídas baseado na base de regras conjunta usando um procedimento de defuzzificação: a defuzzificação ocorre pelo método do centróide (COG) na determinação da saída.

Estabelecidas as cinco etapas, resta evidente que este método é simples e rápido para ser implementado, e por não requerer treinamento se sobressai à técnica de rede neural. Ainda, esta técnica pode ser expandida para múltiplas variáveis de entrada. Notou-se capacidade de generalização, então novas entradas conduzem a predições de saídas com êxito.

Um ponto que merece destaque é que esta técnica não precisa de um modelo matemático, que o sistema aprende por exemplos e regras especializadas, então o mapeamento de regras pode ser atualizado em tempo real quando novos exemplos estão disponíveis, tornando o sistema adaptativo.

Os autores destacam que o sistema baseado nesta técnica funcionará corretamente se houver uma base de regras que seja completa, ou seja, uma base de regras em que seja possível criar regras para todas as combinações entre as variáveis dependentes.

Concluiu-se que o mapeamento de regras é capaz de aproximar qualquer função continua real num conjunto compacto que traz resultados com precisão. Fez-se um teste sob o experimento de controlar as manobras de um veículo, comparando-se o desempenho de um controlador que faz uso desta técnica com um controlador de rede neural artificial e com um controlador difuso com poucas regras e o controlador que usa esta técnica apresentou um desempenho melhor que os outros dois. Testou-se também o poder de predição numa série temporal (*Mackey-Glass chaotic time series*), então esta técnica (denominada neste trabalho por WM,

iniciais dos autores) foi comparada com um preditor que usa redes neurais artificiais e novamente esta técnica (WM) apresentou melhor desempenho. Esta constatação se deu por meio da observação da acurácia através da plotagem dos preços preditos e os preços de teste, então quando a curva gerada com os dados de predição se afasta da curva gerada pelos dados de teste, tem-se uma menor acurácia.

## 3.5 AVALIAÇÃO DE IMÓVEIS URBANOS COM UTILIZAÇÃO DE SISTEMAS NEBULOSOS (REDES NEURO-FUZZY) E REDES NEURAIS ARTIFICIAIS

O artigo intitulado Avaliação de imóveis urbanos com utilização de sistemas nebulosos (redes neuro-fuzzy) e redes neurais artificiais, de autoria de Pelli (2004) tem como objetivo principal desenvolver três metodologias para estimação do valor de mercado de imóveis, sendo elas a rede neuro-*fuzzy* (no caso o autor usou a técnica ANFIS) e duas alternativas com redes neurais artificiais (ANN), então se dá a comparação com os estimadores dos mínimos quadrados (MLR).

O trabalho elencou nove variáveis independentes, onde nota-se variáveis quantitativas, *proxy* e qualitativas, havendo intervenção na escala, são elas: nível, setor, total de vagas, área coberta, número de dormitórios, número de sanitários, equipamentos, padrão de acabamento, estado de conservação e valor unitário do imóvel, esta última como variável dependente.

Neste artigo, Pelli (2004) fez uso de 172 (cento e setenta e dois) apartamentos situados em Belo Horizonte (MG), onde 22 (vinte e dois) elementos foram destacados da amostra para servir de teste dos modelos matemáticos. O autor explica que o maior esforço para o treinamento da rede neural se deu na coleta e pré-processamento dos dados. Antes do processamento, houveram procedimentos de normalização dos dados de entradas, então para facilitar a convergência durante o treinamento. Segundo o autor, é necessário que os dados da entrada e da saída sejam normalizados de maneira a se situarem entre 0 e 1 por requisito da função sigmóide, usada no treinamento de redes neurais. Buscando convergência, os valores da entrada e da saída eleitos ficaram sendo somente aqueles cujos elementos situavam-se entre 0,2 e 0,8, uma vez que a função

sigmóide tende a infinito para 0 e 1. Existem meios de normalizar os dados forçando que estes se situem entre 0,2 e 0,8. Em seguida se deu a desnormalização.

É importante destacar que na técnica ANN multicamadas com aprendizado supervisionado fez-se uso do sistema retro propagação do erro, onde há duas ou mais camadas de neurônios de processamento. Para o treinamento dos dados, utilizaram-se as estruturas de redes neurais multi-camadas (com 19 neurônios na camada escondida, sendo do tipo retro propagação do erro e treinada por meio de aprendizado supervisionado) e de *parallel layer perceptron* (nesta estrutura o número de neurônios para a camada escondida é 4, por sua estrutura não linear com minimização do erro pela técnica do gradiente e estrutura linear com mínimos quadrados para minimizar o erro, sendo esta também do tipo retro propagação do erro).

Quanto ao processo de treinamento das redes, os conjuntos de treinamento e de validação foram representados pelo valor de venda por metro quadrado do imóvel. O autor adverte que resíduos significantes podem surgir em aproximações lineares para estes conjuntos e assim ao avaliar um imóvel, as estimativas podem resultar erradas quanto ao seu valor.

Para a rede ANFIS o treinamento se deu em 500 épocas e taxa de aprendizado 0,1 e a base de regras foi composta por cinco regras. Houve a normalização dos dados de entrada e de saída antes do processamento da rede. Para o treinamento fez-se uso de 150 elementos.

Para evitar o *overfitting* (sobreajuste), fez-se uso do método *cross-validation*.

Finalmente, Pelli conclui da comparação entre os métodos que a alternativa com uso da ANN multicamadas foi aquela que mais se aproximou dos valores de mercado em imóveis urbanos. Quanto ao poder de predição dos modelos, todos os modelos obtiveram êxito na predição dos valores de mercado selecionados (aqueles 22 elementos destacados da amostra) para a validação.

## 3.6 AVALIAÇÃO DE IMÓVEIS URBANOS COM UTILIZAÇÃO DA LÓGICA DIFUSA

A dissertação intitulada Avaliação de imóveis urbanos com utilização da lógica difusa, de autoria de Santello (2004) tem como objetivo principal desenvolver um método e também um modelo de apoio decisório aplicável à avaliação de imóveis urbanos. Desta maneira se torna possível a aquisição do conhecimento sobre como as variáveis difusas podem impactar na composição dos valores dos imóveis. Para tanto, critérios de referências por indicadores físicos, sociais e econômicos, são agrupados e combinados em um dendograma hierárquico, de maneira que por meio de uma árvore de decisão tem-se nós cujas operações lógicas se dão por meio do uso de sistemas difusos. Com o modelo proposto, os indicadores de qualidade de construção, de localização e do mercado foram ordenados e combinados, sendo agregados em blocos de regras num dendograma tipo *top down induction of decision trees*, que operam conforme as regras de operações básicas da lógica difusa. Finalmente, o autor faz uso de um modelo que é aplicado na avaliação de um apartamento.

Quanto às regras de produção, estas são usualmente compostas pela parte SE e pela parte ENTÃO. O autor aplicou o modelo de inferência difuso de Mamdani, com funções de pertinência do tipo triangular e trapezoidal (nos limites inferior e superior) dos gráficos. Neste modelo, recebe-se um número escalar, que é convertido para um número difuso (generalização), passa-se por uma máquina de inferência, que se comunica com um banco de dados e devolve um número difuso, que é convertido para escalar (defuzzificação). Na defuzzificação, Santello (2004) observa que o método MEAN.MAX (média dos máximos) apresentou a melhor solução plausível quando comparado com os métodos COG e WAM. O autor utilizou como indicadores os acessos, cenário, conservação do apartamento e do prédio, despesas, insolação, infraestrutura, facilidades nas proximidades do lote, ocupação e outras.

Neste trabalho as regras foram organizadas e confinadas em grupos e posteriormente o modelo proposto foi disposto em árvore de decisão. As diversas características do modelo foram processadas com auxílio do programa *fuzzyTECH* ®. Obteve-se o resultado (variável dependente do preço unitário) e a análise de verificação da sensibilidade para cada indicador foi realizada isoladamente.

Santello (2004) conclui elencando seu trabalho como um diagnóstico de qualidade do imóvel em estudo e expõe fatores que limitaram sua aplicação, onde nem sempre os indicadores de mercado existem no modelo. Ainda, o uso do programa *fuzzyTECH* ® seria um complicador pela questão do custo, por ser um produto importado.

## 3.7 APLICAÇÃO DO MODELO DE REGRESSÃO DIFUSA PARA PREDIÇÃO DE VALORES IMOBILIÁRIOS

O artigo intitulado Aplicação de modelo de regressão difusa para predição de preços imobiliários (*Application of fuzzy regression model for real state price prediction*), de autoria de Sarip *et al* (2016) tem como objetivo principal propor um modelo baseado em regressão de mínimos quadrados difusos (FLSR) para prever os valores dos imóveis, visando identificar o desempenho da técnica sob este escopo. Um estudo comparativo abrangente em termos de precisão de predição e complexidade computacional entre as técnicas de ANN, ANFIS e FLSR foi realizado. Os autores iniciam explicando que o método direto de comparação de dados de mercado por inferência estatística parte do pressuposto que existe uma relação direta entre o imóvel que se deseja inferir o valor e os preços praticados no mercado imobiliário.

Sarip *et al* (2016) alertam que muitas vezes faz-se uso de novas abordagens visando avaliações rápidas sem tanto compromisso com a qualidade. Sarip *et al* (2016) explicam ainda que a técnica MLR tem sido largamente utilizada apesar de suas limitações ao lidar com conjuntos de dados que contém *outliers*, não linearidade e não normalidade e que estes desafios podem ser enfrentados com o uso de outras técnicas, como FIS, TSK e ANN.

Segundo Limsombunchai (2004, *apud* Sarip *et al*, 2006, p. 2), é possível explorar as correlações entre as variáveis independentes que explicam o valor de mercado de um imóvel e quantificar a contribuições de cada fator para o valor no setor imobiliário usando a técnica ANN com *back-propagation* disposta em duas camadas ocultas. Com base nessas informações, o impacto dos fatores de preços e a relação entre o preço da propriedade e cada um dos fatores é obtido.

Segundo Sarip *et al* (2016) as abordagens por lógica difusa têm destaque em situações onde a base de dados contém valores linguísticos ou variáveis obtidas por meio de questionários. Os autores continuam estabelecendo que a formação dos conjuntos difusos e da base de regras é um desafio que consome muito tempo e depende do conhecimento de um especialista humano, razão pela qual uma boa alternativa é deixar esta tarefa para ser resolvida com apoio das ANN, sendo esta a técnica ANFIS.

Conforme constatado por Zurada *et al* (2006, *apud* Sarip *et al*, 2006, p. 2), que realizaram uma comparação entre quatro técnicas (ou seja, MLR, ANN, raciocínio baseado em memória e lógica difusa) para estimar o valor da casa, destacou-se que nenhum desses métodos foi capaz de superar a capacidade de estimativa de MLR, embora cada método fosse bom o suficiente para servir como complemento da técnica MLR. No entanto, seu estudo destacou a força potencial desses outros métodos para melhorar a previsão de preços de venda onde os dados quantificáveis são menores.

Sarip *et al* (2016) objetivaram identificar o desempenho da previsão de valor da abordagem de lógica difusa no contexto da avaliação de imóveis. Para tanto, propuseram um modelo ANN e outros dois modelos *fuzzy* preditivos para estimar o valor da propriedade, a saber, ANFIS e FLSR.

Quanto à aplicação da técnica ANFIS, esta se baseia no sistema de inferência difuso tipo Sugeno, com regras. A camada de entrada contém nós das variáveis de entrada representando as oito características da propriedade avaliada, a camada 1 contém as funções de pertinência dos conjuntos difusos associados a cada variável de entrada, a camada 2 inclui os pesos das regras. Então, a força de ativação de cada regra é calculada pelo produto dos graus de pertinência em sua parte antecedente disponível na camada anterior. Portanto, o número de nós nessa camada é idêntico ao número de regras; a camada 3 realiza a normalização sobre os pesos resultantes da camada anterior; a camada 4 consiste em nós representando funções lineares, cujos coeficientes são ajustados pelo algoritmo dos mínimos quadrados. Finalmente, os autores destacam que a camada de saída produz um valor que é a soma ponderada das funções lineares das camadas anteriores. O processo de treinamento inclui duas passagens, a saber, o passe para frente e o passo para trás, respectivamente. Durante o passo para frente, os parâmetros antecedentes são fixos e os parâmetros consequentes são estimados

usando o algoritmo dos mínimos quadrados, uma vez que eles são linearmente relacionados à saída. Enquanto isso, durante o passo para trás, os parâmetros consequentes são fixos e os parâmetros antecedentes são otimamente estimados pelo algoritmo de gradiente descendente.

Para controlar a potencial proliferação de regras resultantes do grande número de entradas, a técnica de *clustering* subtrativo foi empregada para otimizar a seleção de regras e os valores dos parâmetros iniciais com base no conjunto de dados de treinamento fornecido.

O que há de mais relevante neste trabalho é o uso de números difusos triangulares para representar os parâmetros difusos denotados por β. Adentrando neste escopo, os autores esclarecem que para computar a distância entre dois números difusos triangulares, A1 e A2, foi usada uma métrica de distância conhecida por distância Diamond, adaptada.

O conjunto de dados deste estudo foi extraído de um registro de vendas de 352 propriedades no Distrito de Petaling, em Kuala Lumpur (capital da Malásia), restando 348 dados sem *outliers*.

Os autores elencaram oito variáveis, sendo cinco quantitativas (área do terreno, área construída, número de quartos, número de banheiros e idade da construção) e três qualitativas (estado de conservação, qualidade da mobília e localização), que foram convertidas para valores numéricos para fins de cálculo estatístico.

Em termos gerais, dentre as técnicas inovadoras consagradas tem-se a ANN e o FIS e o desafio foi combinar os métodos, já há algumas (poucas) publicações neste sentido, de onde originou-se o ANFIS.

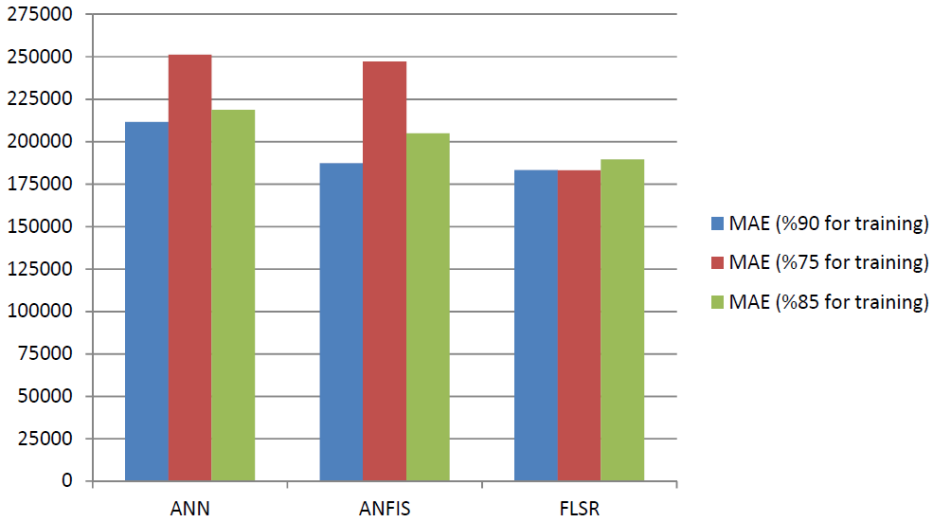
Os autores destacaram que quando comparado ao processamento da técnica ANFIS, o FLSR tem menor custo computacional e modela melhor a relação funcional entre variáveis dependentes e independentes.

Para coletar os resultados do modelo de simulação ANFIS, foi usada a ferramenta Matlab *Fuzzy Logic Toolbox* e, como houveram muitos *inputs*, não foi possível constituir uma base de regras com grande aderência à todas as possíveis permutações (em razão dos valores espúrios).

O processo de treinamento foi dividido em duas partes, sendo 90% para treinamento e 10% para teste. Houve o controle de erros e descarte de resíduos.

Quanto ao FLSR, os resultados da simulação seguiram sob a pretensão de determinar os valores ótimos para os parâmetros beta que minimizam a soma dos quadrados dos erros, foram usadas 27 equações com 27 variáveis (estas são os coeficientes de cada um dos oito parâmetros estudados).

Figura 10 - Comparativo entre as técnicas, proporções para treinamento e métrica MAE




Fonte: Sarip *et al* (2016)

A observação do gráfico da figura 10 mostra que para um maior percentual de treinamento (maior a tendência em convergir para o valor correto), comparando as técnicas, a FLSR convergiu com uma quantidade de dados muito menor quando comparada com as técnicas ANFIS e ANN. A análise dos autores revela que o método FLSR supera tanto a ANFIS quanto a ANN.

Ainda, é válido ressaltar que a abordagem difusa baseada em regressão prevê os preços dos imóveis com maior precisão em comparação com ANN e ANFIS.

Segundo os autores a única etapa do desenvolvimento do FLSR que causou alto custo computacional foi o processamento do sistema linear de equações, que tem complexidade quadrática conforme o número de variáveis *input*. Mesmo assim, a complexidade computacional do FLSR é menor que a complexidade dos outros dois (ANFIS e ANN).

Notou-se que a técnica ANN com *back-propagation* tem complexidade cúbica. Consequentemente, a ANFIS também, por usar ANN. Os resultados obtidos foram submetidos às métricas RMSE (para o treinamento dos modelos ANN e

híbrido ANN com ANFIS) e a análise final se deu pela métrica MAE, cuja interpretação indicou que a FLSR pode efetivamente prever os preços dos imóveis.

Concluindo, pode-se afirmar que o desempenho superior da técnica FLSR ora proposta se mostra promissora para a valoração imobiliária. O fato de que tal desempenho é alcançado em um tamanho de amostra relativamente pequeno ressalta ainda mais o potencial.

Obviamente, os autores advertem que torna-se temerário basear o resultado num único modelo e ressaltam que a aplicação da técnica FLSR na área imobiliária é relativamente mais recente que outras técnicas de aprendizado de máquina, então mais estudos podem ser necessários.

## 3.8 APLICAÇÃO DE LÓGICA FUZZY NA ELABORAÇÃO DE PLANTA DE VALORES GENÉRICOS

A dissertação intitulada Aplicação de lógica *fuzzy* na elaboração de planta de valores genéricos, de autoria de Malaman (2014) tem como objetivo principal propor a utilização da lógica difusa para avaliar imóveis e assim gerar uma planta de valores genéricos. Para validar o modelo gerado por lógica difusa (FIS), faz-se a comparação com a técnica MLR, e conclui-se que a técnica FIS pode substituir ou complementar a técnica MLR. O trabalho destaca que a determinação do valor imobiliário de um imóvel precisa considerar diversas variáveis, como a área, a idade e o padrão construtivo.

Quanto à escolha das funções de pertinência, a autora destaca que a escolha nem sempre é óbvia, e que as funções trapezoidais e triangulares são mais intuitivas e este é o motivo pelo qual costumam ser mais usuais que as funções gaussianas (estas dependem do valor médio e da dispersão). Segundo Malaman (2014), comumente os modelos FIS usam os métodos para defuzzificação COG e MEAN.MAX.

A aplicação da técnica proposta se deu no município de Álvares Machado (SP) onde foi usado um conjunto de dados a partir de um trabalho anterior, com 71 elementos distribuídos em torno do município. Destes, 6 elementos foram destacados para verificações, e o modelo foi treinado com 71 elementos (100% do conjunto de dados). O estudo partiu de nove variáveis independentes (área do

terreno, distância ao centro, frente, distância à via principal, topografia, padrão construtivo, zona comercial, e as coordenadas UTM X e Y) para explicar a variável dependente chamada de valor unitário da testada corrigida. O modelo MLR foi gerado tendo por ferramenta o *software* Minitab versão 14. Diversos modelos foram realizados e escolheu-se por um modelo com as variáveis independentes distância ao centro, topografia, padrão construtivo, zona comercial e as coordenadas UTM X e Y, que apresentou $r^2$ ajustado 87,2% num modelo que apresentou o menor RMSE (comparando com os outros modelos).

Posteriormente a autora usou as mesmas variáveis independentes para compor um modelo por meio da técnica FIS, obtido pela ferramenta *Fuzzy Logical Toolbox* do Matlab versão 5.3. Definidas as funções de pertinência, foram elaboradas 69 regras ao invés de 1080 regras (advindas das combinações entre as funções de pertinência), justificado pelo fato de haver poucos elementos. Usando a defuzzificação tipo COG, todos os pontos do conjunto de dados foram preditos para efeito de comparação com os pontos do conjunto de dados segundo a métrica RMSE.

O trabalho passou para a etapa final, onde os valores serviram para a criação de um modelo de superfície, que dá base à planta de valores genéricos proposta nos objetivos da autora.

## 3.9 QUADRO COMPARATIVO ENTRE OS TRABALHOS CORRELATOS

O quadro 1 traz uma breve comparação entre os trabalhos correlatos que foram anteriormente descritos .

Quadro 1 - comparativo entre os trabalhos correlatos

| Autor/ano | Local, tamanho do conjunto de dados, variáveis usadas | Técnicas | Métricas | Desempenho | Observações |
|---|---|---|---|---|---|
| **Yalpir e Ozkan (2018)** | 320 imóveis situados em 8 bairros da região metropolitana da cidade de Konya (Turquia). Variáveis: valor de mercado x área, idade, condições de acesso, propriedades físicas e localização. Atingiu MAPE 75,32% e $r^2$ igual a 69,21% | FIS, ANFIS | $r^2$ e MAPE | Ajuste usando FIS resultou melhor que ANFIS | FIS modelado com funções triangulares e trapezoidais nas bordas, 85 regras usando AND, defuzzificação COG |
| **Kamire *et al* (2021)** | 1.460 imóveis (80% treino x 20% teste) situados em 4 bairros da cidade de Pune (Índia). Variáveis: valor de mercado x área, idade, condições de acesso, propriedades físicas e localização. O bairro 3 atingiu MAPE 4,85% e $r^2$ de 29,05% na lógica difusa, enquanto que para ANFIS obteve-se MAPE 24,86% e $r^2$ de 10,19% | FIS, ANFIS | $r^2$ e MAPE | FIS apresentou bons resultados para muitos dados e péssimos resultados para poucos dados | FIS modelado com funções triangulares, não especificou a quantidade de regras, usou AND |
| **Król *et al* (2007)** | 134 imóveis (45% treino x 55% teste) situados numa localidade da Polônia. Variáveis: proporção do valor da propriedade sendo avaliada em relação ao valor daquela representativa x distância ao centro, frente, área, infraestrutura da localidade, infraestrutura do imóvel, posição, vizinhança, acesso ao transporte público. Os melhores resultados ocorreram por volta de 500 gerações. | FIS, TSK | MAPE | Ambos modelos atingiram o *fitness* aceitável, mas pode melhorar se eliminar os *outliers* | A base de regras foi treinada com AG. FIS modelado com funções triangulares e trapezoidais nas bordas, defuzzificação COG. TSK com grau zero. |
| **Wang e Mendel (1992)** | por ser um trabalho de dedução de uma técnica, não se aplicou a um conjunto de dados, mas validou seus resultados por meio de um controlador de manobras de veículos e de uma série temporal | FIS, ANN | Observ. gráfica da acurácia | a técnica FIS apresentou um desempenho melhor que a técnica ANN | A base de regras para a técnica FIS foi obtida automaticamente |
| **Pelli (2004)** | 172 apartamentos (87% treino x 13% teste) situados em Belo Horizonte (MG). Variáveis: valor unitário x nível, setor, total de vagas, área coberta, n dormitórios, n bwc, equipamentos, padrão acab., estado de conservação | MLR, ANN, ANFIS | RMSE | todas as técnicas tiveram resultado satisfatório, ANN superou os outros | ANN multicamadas e sistema de retro propagação do erro |
| **Santello (2004)** | apartamentos (não especificou quantos) situados em Florianópolis (SC), sem dividir entre treino x teste. Nós da árvore: indicadores físicos, sociais e econômicos, dispostos em dendograma | FIS em árvore de decisão | Sem métrica | MEAN.MAX melhor quando comparado com os COG e WAM | FIS foi usado para calibrar cada um dos nós da árvore hierárquica de decisão |

| Autor/ano | Local, tamanho do conjunto de dados, variáveis usadas | Técnicas | Métricas | Desempenho | Observações |
|---|---|---|---|---|---|
| **Sarip *et al* (2016)** | 348 imóveis (90% treino x 10% teste) situados em Kuala Lumpur (Malásia). Variáveis: valor do imóvel x área do terreno, área construída, número de quartos, número de banheiros e idade da construção. Para o ANFIS, MAE resultou 187,35; para FLSR resultou 183,35 e para ANN resultou 211,58. | Híbrido ANN para clusterizar a entrada ANFIS, ANN, FLSR | RMSE, MAE | Os melhores resultados foram observados na técnica FLSR, seguido por ANFIS e por ANN | FLSR converge mais rapidamente e com uma quantidade menor de dados |
| **Malaman (2014)** | 71 elementos (todos usados no treino) situados em Álvares Machado (SP). Variáveis: valor unitário da testada corrigida x distância ao centro, topografia, padrão construtivo, zona comercial e as coordenadas UTM X e Y. O modelo escolhido apresentou $r^2$ de 66,02% e RMSE de R$ 79,48. | MLR, FIS | $r^2$, RMSE | FIS pode substituir ou complementar MLR | FIS modelado com funções triangulares e trapezoidais nas bordas, 69 regras usando AND, defuzzificação COG |

Fonte: elaborado pelo autor (2023)

## 3.10 COMPARAÇÃO E DISCUSSÃO DOS TRABALHOS CORRELATOS

Ao longo da leitura dos trabalhos, todos eles bastante similares ao estudo proposto, nota-se maior aderência com os trabalhos de Yalpir e Ozkan (2018), de Kamire *et al* (2021) e de Wang e Mendel (1992). As pesquisas trazem diversas técnicas aplicáveis à avaliação imobiliária, dentre elas MLR, ANN, FIS, TSK, e ANFIS. Algumas técnicas híbridas também foram propostas. Desta maneira, o estudo proposto neste trabalho se torna relevante por comparar diversas técnicas, notadamente MLR, FIS, TSK e ANFIS.

O trabalho de Yalpir e Ozkan (2018) modelou funções triangulares e trapezoidais para a abordagem difusa, sendo uma evidência para sua aplicação nas técnicas FIS e TSK. Destaca-se desde já que a técnica ANFIS utiliza obrigatoriamente função gaussiana, não cabendo considerações sobre a possibilidade de emprego de outros tipos de funções.

Quanto ao operador para a base de regras, todos os trabalhos que divulgaram qual operador foi usado fizeram uso do operador AND.

Quanto às métricas, Król *et al* (2007) e Yalpir e Ozkan (2018) usaram $r^2$ e MAPE. Já Sarip *et al* (2016), autores do trabalho que faz uso das variáveis mais assemelhadas às deste estudo, usaram RMSE e MAE, então é razoável fazer uso de mais de uma métrica e escolher as principais dentre estas elencadas.

Em especial, o trabalho de Król *et al* (2007), que estuda FIS e TSK apresenta um conjunto de dados com 134 elementos, na mesma ordem de grandeza do conjunto de dados deste trabalho. Como o trabalho de Król *et al* (2007) faz a divisão entre dados de treino x teste em proporção 45% treino x 55% teste, sugere-se que uma distribuição assemelhada possa ser aplicável aos métodos FIS e TSK. Considerando que o método ANFIS é uma particularidade do método TSK, em que o treinamento das regras e das funções polinomiais é automatizado, resta a evidência que esta proporção pode trazer bons resultados.

O trabalho de Wang e Mendel (1992) traz uma automatização da criação dos conjuntos difusos e também da base de regras, com resultados já comprovados e ampla aplicação no meio técnico. Como sua concepção usa o operador AND e funções triangulares e trapezoidais, está alinhado às outras pesquisas e sua aplicação se faz fundamental na técnica FIS deste trabalho.

# 4 METODOLOGIA

O estudo inicia com uma revisão bibliográfica das principais técnicas para avaliação de imóveis, incluindo a avaliação mercadológica por inferência estatística usando Regressão Linear Múltipla (*Multiple Linear Regression* / MLR), e outras técnicas como Sistema de Inferência Difusos de Mamdani (*Fuzzy Inference Systems* / FIS), Sistema de Inferência Difuso de Takagi-Sugeno-Kang (*Takagi-Sugeno-Kang Inference Systems* / TSK) e Sistemas de Inferência Difusos Neuroadaptativos (*Adaptative Neuro Fuzzy Inference System* / ANFIS). Em seguida alguns modelos são recalibrados usando a base amostral depois de aplicada a Subamostragem da Classe Majoritária (*Majority Class Undersampling Technique* / *undersampling*) sob a expectativa de melhorar o conjunto de dados.

Estando consolidada a revisão bibliográfica e a análise do estado da arte, coletam-se os elementos para compor uma base amostral de imóveis do tipo apartamento na região do Bairro Trindade, em Florianópolis (SC). Esta base amostral (conjunto de dados) permite a avaliação imobiliária para determinar o valor de avaliação mercadológica dos imóveis, ocorrendo por meio de códigos em linguagem de programação R usando pacotes (*libraries*) de uso já difundido neste meio técnico.

Em relação à terminologia, existem duas correntes principais no meio técnico, sendo aqueles que consideram as alternativas MLR, FIS, TSK e ANFIS como modelos e aqueles que as rotulam como técnicas. Ambas as abordagens são amplamente aceitas. Neste trabalho, optou-se por adotar a denominação por técnica. Além disso, as equações e modelos gerados foram simplesmente chamados de modelos. Isso implica que, na perspectiva deste trabalho, o desenvolvimento de um modelo é uma solução para um problema específico, fazendo uso de uma técnica de inteligência computacional para a obtenção de um modelo.

A metodologia adotada nesta pesquisa consiste na utilização de técnicas de lógica difusa para criar modelos com objetivo de avaliar apartamentos no Bairro Trindade em Florianópolis (SC), para que comparando com técnicas convencionais seja evidenciada a validade de sua aplicação.

A comparação da qualidade dos modelos se dá por meio da análise das métricas, sob o intuito de determinar qual das técnicas apresentou o modelo com o

melhor desempenho. Para compor os modelos foi utilizado um conjunto de dados coletado em março/2023 com 132 elementos (base_trindade_mar23.csv), que contém as informações sobre os elementos da base amostral, tais como preço de oferta, área privativa, coordenadas de localização, número de dormitórios, número de suítes, número de banheiros, número de vagas, padrão de acabamento e idade.

Para a realização dos procedimentos deste trabalho, foi utilizado o *software* R, planilha eletrônica online (*google sheets*), sendo que os *scripts* constam no *link* disponível no Anexo B deste trabalho.

## 4.1 OBTENÇÃO DO CONJUNTO DE DADOS

As fontes de dados utilizadas são sites de vendas de apartamentos. A coleta se deu manualmente, em razão que um procedimento de *webscrapping* demandaria autorizações (nem todo site de anúncios em seu robots.txt permite *webscrapping*) e ainda, não necessariamente os dados estarão na mesma estrutura entre um site e outro. Outro problema que impossibilitou a amostragem automatizada foi que algumas variáveis são subjetivas, como a área, então um anunciante usa a área total e outro usa a área privativa, um anunciante inclui a vaga de garagem na área e outro não. A própria norma que traz disposições para condomínios (NBR 12.721 / 2006) admite ambas possibilidades.

Quanto aos empreendimentos com vagas rotativas, como existe a possibilidade da guarda de um único veículo automotor à proporção de cada apartamento, como há monitoramento por serviço de portaria para garantir este pressuposto e como estatisticamente não se registram falta de local para guarda, considerou-se uma única vaga na variável vaga. Apartamentos situados em empreendimentos com menos de um ano, considerou-se 1, porque 7 meses em diante se arredonda para 1 ano. Isto evita problema de indeterminação em estudos de variáveis, onde a idade pode ser transformada usando a função inversa, visto que se espera que apartamentos com mais idade tenham menor valor unitário e menor valor (estas são as duas possibilidades para variável dependente).

A definição das variáveis de interesse parte do estabelecimento da variável resposta, também chamada de variável dependente ou variável explicativa.

Em seguida, é realizada uma análise sobre quais variáveis independentes são relevantes para explicar a variação da variável dependente. Nesta etapa entra a

experiência do avaliador, sob a ótica de corretamente escolher as variáveis. Um erro nesta etapa compromete as etapas seguintes, sob a ótica de resultar num conjunto de dados com dados faltantes ou um conjunto de dados que só terá utilidade se o avaliador retornar a campo para coletar informações sobre variáveis não inclusas inicialmente no modelo. A NBR 14.653-2 (2011) estabelece meios para definir o planejamento da pesquisa.

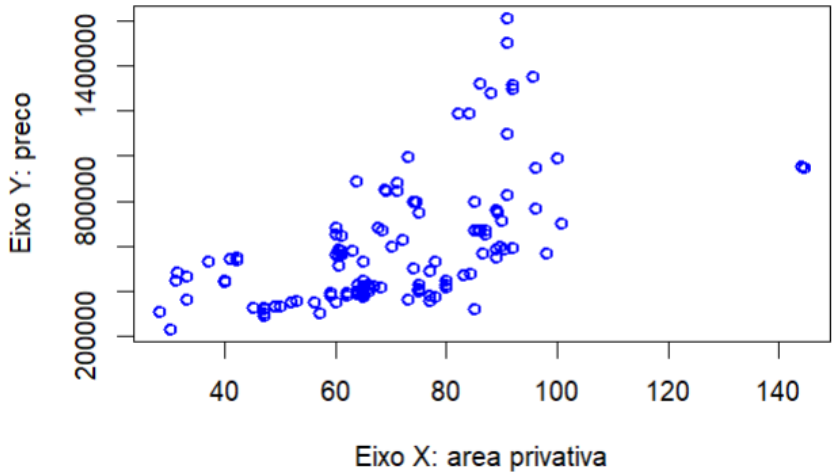
Quanto à preparação dos dados, a própria atividade de coleta da maneira como fora prevista já conduz a um conjunto de dados com resultados dados preparados (sem dados faltantes, sem duplicados e sem valores inválidos).

A análise exploratória dos dados foi iniciada pelo comando str() que mostra a estrutura dos dados, de imediato foi necessário remover R$ seguido de espaço, converter variáveis categóricas quantitativas para fator e remover variáveis não utilizadas.

A variável dependente (preço unitário) foi criada no *dataframe*.

O gráfico de dispersão mostrou existir pontos distantes dos demais, a investigação mostrou que se tratava de duas coberturas, ou seja, imóveis com tipologia diferente que foram excluídos do conjunto de dados. A Figura 11 mostra o ocorrido.

Figura 11 - Gráfico de dispersão

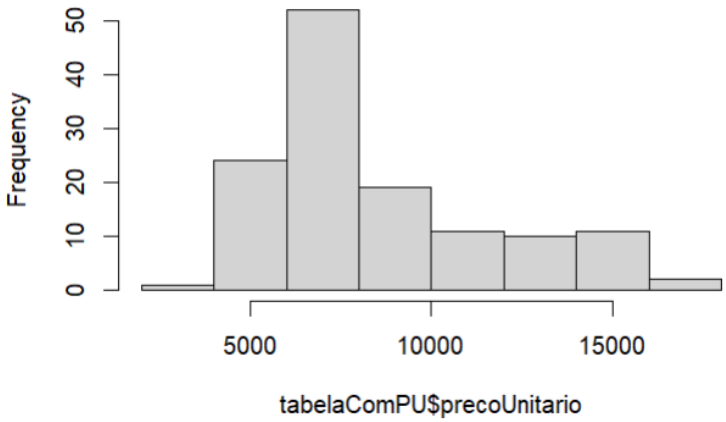



Fonte: elaborado pelo autor (2023)

Aplicando a função summary nos dados pode-se buscar por discrepâncias nos valores mínimo e máximo e pela média e mediana, ter uma ideia da concentração dos dados.

O histograma de frequência da Figura 12 mostra haver uma concentração grande de elementos por volta de R$ 7.000 / m² e indica a necessidade da aplicação de um procedimento de *undersampling*.

Figura 12 – Histograma de frequências




Fonte: elaborado pelo autor (2023)

A função boxplot identificou haver elementos *outliers* e também a distribuição irregular do conjunto de dados.

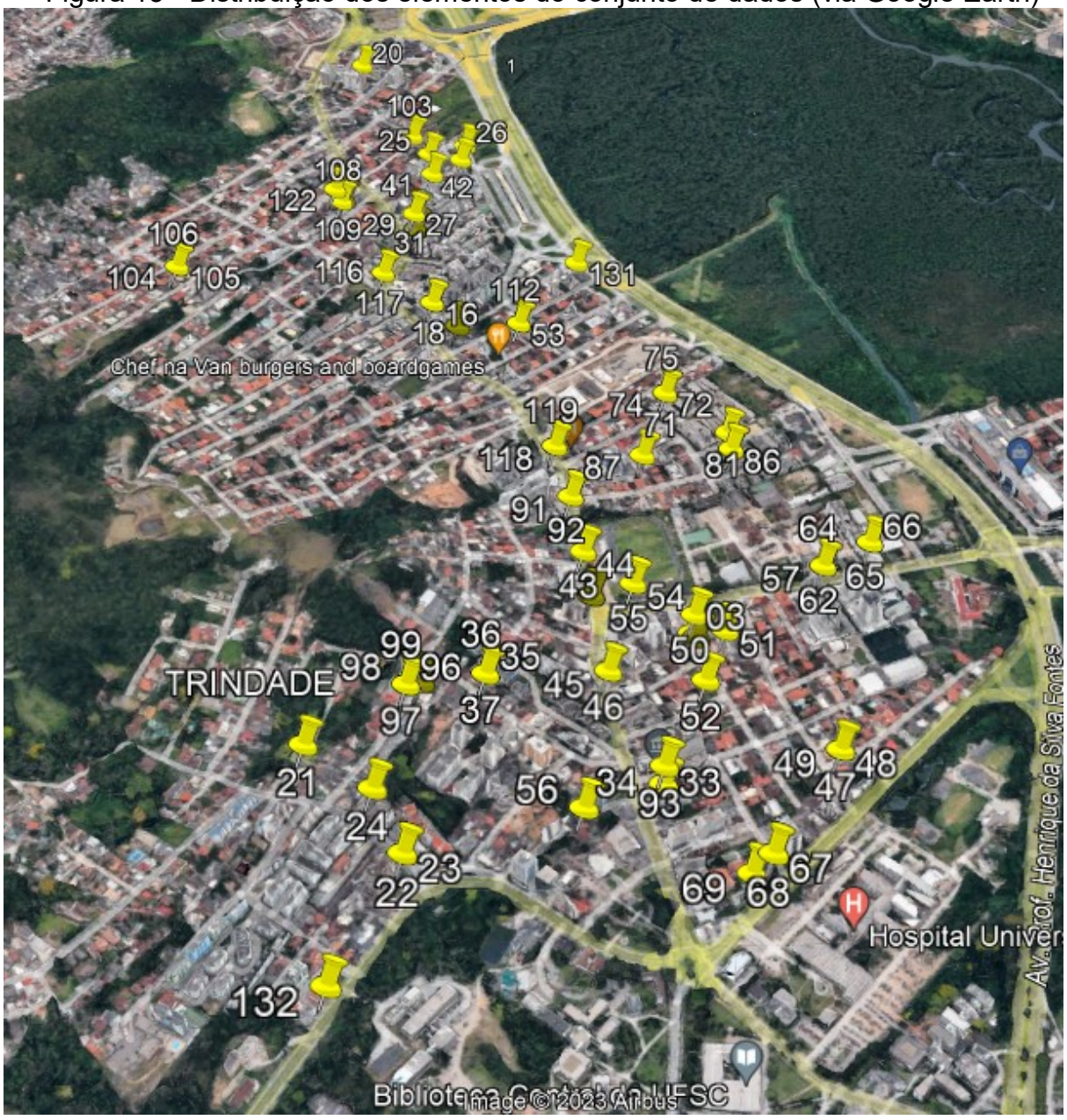
A matriz de correlação faz parte da análise exploratória e será mostrada no capítulo de aplicação do método.

Nesta etapa de obtenção do conjunto de dados inclui-se a análise exploratória dos dados. Assim, fazendo uso de técnicas estatísticas e de técnicas de visualização identificou-se a correlação entre as variáveis, a expectativa de crescimento, a eliminação de dados duplicados, a presença de *outliers*, os agrupamentos (distribuição dos dados).

Depois de obtidos, os dados são planilhados e o conjunto de dados resultante é salvo em formato CSV, facilitando a entrada de dados na linguagem R.

A figura 13 mostra a distribuição espacial dos elementos que compõe o conjunto de dados.

Figura 13 - Distribuição dos elementos do conjunto de dados (via Google Earth)




Fonte: Google Earth (2023)

## 4.2 APLICAÇÃO DAS TÉCNICAS DE IC NA AVALIAÇÃO IMOBILIÁRIA

Nesta seção, descreve-se para cada técnica os procedimentos adotados para atingir os resultados e os critérios usados na avaliação dos resultados.

Sob o intuito de tornar os resultados reproduzíveis, definiu-se um número inicial (semente) para o mecanismo de geração de números aleatórios, nota-se pelo comando set.seed(123), prática comum em *scripts* para análise de dados. Isto é importante porque as funções de divisão entre treino x teste e outras funções que envolvem amostragem (como a função *sample*, por exemplo), e a calibração da técnica ANFIS permitirão que outros usuários atinjam os mesmos resultados.

Os trabalhos correlatos trouxeram diversas proporções de treino x teste:

a) a proporção 80% x 20% conduziu a bons resultados na pesquisa de Kamire *et al* (2021), então fez-se 70% x 30% , 75% x 25%, 80% x 20% e 85% x 15%, analisando-se as métricas de cada resultado com vistas a estabelecer como modelo escolhido aquele que apresentar as melhores técnicas;
b) a proporção 90% x 10% utilizada na pesquisa de Sarip *et al* (2016) e a proporção de 87% x 13% de Pelli (2004) são inaplicáveis a este trabalho, por uma questão de limitação de dados, de onde a baixa quantidade de dados restantes para teste afetará o cálculo das métricas;
c) a proporção de 45% x 55% praticada por Król *et al* (2007) parece razoável e será utilizada em situações bastante específicas, conforme será relatado oportunamente.

Este trabalho faz uso das proporções de Kamire *et al* (2021) para todas as opções, exceto a abordagem com a técnica ANFIS que traz um comparativo sob a ótica da influência de *undersampling* e presença de *outliers* no conjunto de dados, então preferiu-se praticar a proporção de Król *et al* (2007).

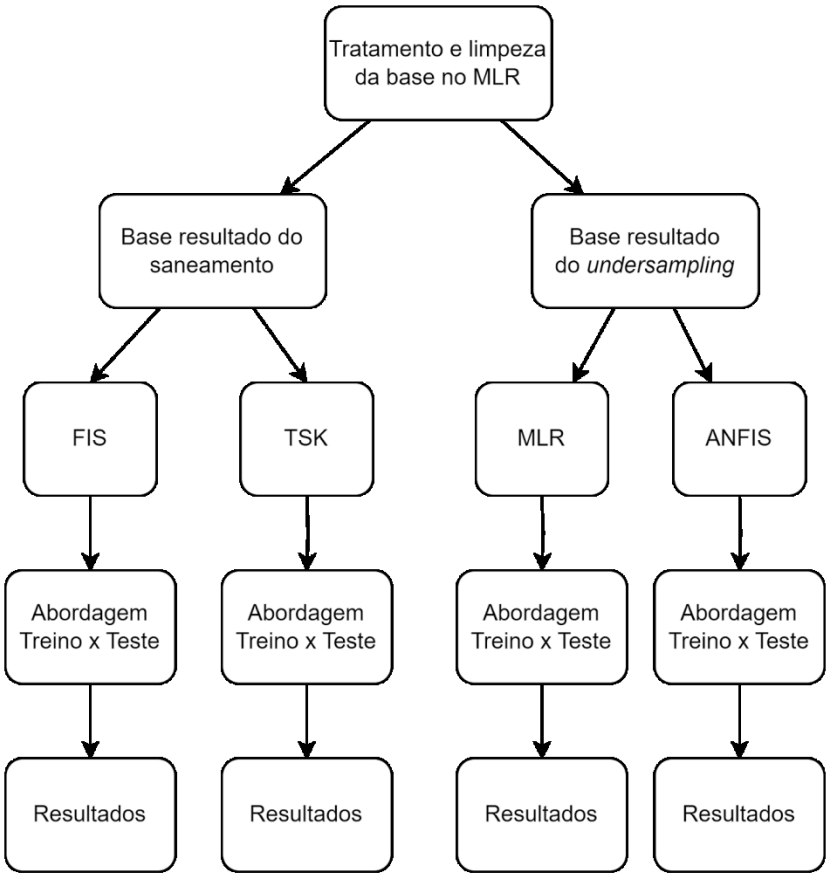
A criação dos modelos ocorreu no pacote estatístico R, bastante usual para estes tipos de procedimentos. Este pacote estatístico é performado com uma IDE chamada R Studio. Há diversos pacotes aplicáveis à lógica difusa, optou-se pelo uso do pacote gratuito frbs para executar a técnica ANFIS.

A aplicação das técnicas é possível por ser uma etapa seguinte à coleta e preparação dos dados. Através da aplicação das técnicas serão gerados modelos

cujos resultados serão melhores se os dados forem bem coletados e forem representativos do imóvel objeto da análise. Os dados são então repartidos entre treino e teste, posteriormente os dados de teste são usados para calcular as métricas. A análise das métricas é tida como determinante na escolha do modelo.

A figura 14 indica a sequência de procedimentos entre as técnicas.

Figura 14 - Fluxograma da sequência de aplicação das técnicas




Fonte: elaborado pelo autor (2023)

O modelo é validado pelos resultados das métricas, depois de confrontadas as quatro proporções de distribuições entre treino x teste (70% x 30%, 75% x 25%, 80% x 20% e 85% x 15%).

### 4.2.1 Aplicação da Regressão Linear Múltipla (MLR)

O trabalho se inicia com a aplicação da técnica MLR, conforme consta no *script* MLR_model-ABR23.r. O conjunto de dados bruto é tratado, então variáveis menos importantes são excluídas, colunas que não servem para a análise de dados (como o nome do condomínio, o endereço e o link do anúncio) são removidas, e linhas com valores faltantes (NA) são removidas.

O estudo seguiu com as alternativas sem *undersampling* sob a finalidade exploratória, então bons modelos foram encontrados e melhorados através da exclusão de elementos *outliers* e seleção das variáveis independentes que melhor explicam a variabilidade da variável dependente por meio da métrica r² ajustado, que resultou 88,32%.

Na aplicação da técnica MLR o conjunto de dados é modelado por meio de transformações nas variáveis para melhor ajustar os dados. Nessa ótica a variável distância foi modelada com função inversa, a variável preço unitário com função logarítmica, exponencial, quadrática e as outras com funções lineares, conforme o crescimento da respectiva variável independente com relação à variável dependente. Diversas combinações precisam ser executadas, cabendo a inclusão ou exclusão de variáveis menos relevantes.

Comumente, há métodos para mais rapidamente convergir no melhor modelo, como por exemplo:

a) testar todas as combinações possíveis: esta possibilidade traz complicadores e é um tanto quanto desnecessária, tendo em vista as outras alternativas;
b) eliminação passo atrás (*backward elimination*): conforme Diez et al (2019), o modelo é formado com todas as potenciais variáveis preditoras (variáveis independentes) e a cada ciclo, a variável que reduzir mais as estatísticas do modelo é eliminada (uma a uma), até atingir um estado em que a exclusão de variáveis independentes não melhora o modelo em termos de análise da métrica r² ajustado;
c) inclusão passo a frente (*forward elimination*): conforme Diez et al (2019), a estratégia é o oposto da estratégia *backward elimination*, então o modelo é constituído somente com a melhor variável independente e a cada ciclo as outras variáveis independentes são

incluídas somente se houver melhoria no modelo (também em termos de análise da métrica r² ajustado);

d) outros, como a seleção passo-a-passo (*stepwise*) e técnicas usando o AIC (Critério de informação de Akaike).

Quanto ao processamento do modelo, o método do gradiente descendente se sobressai ao método de inversão de matriz em termos de desempenho para conjunto de dados com muitas variáveis independentes. Na linguagem de programação R pode ser escolhido qual método usar. Como a natureza desse tipo de problema (avaliação imobiliária de apartamentos no bairro Trindade) envolve um conjunto de dados com poucas variáveis independentes, então pode ser usado o método da inversão de matriz, que ocorre por meio da chamada da função lm(), que faz parte do pacote padrão da linguagem R. Caso fosse intenção usar o método do gradiente descendente, poderia-se chamar a função glmnet() do pacote glmnet.

Calculado o modelo, faz-se a análise de aderência onde deve-se prestar atenção no gráfico de dispersão que relaciona os valores observados e preditos, com relação aos pontos que se distanciam da reta bissetriz. Nestes casos, convém averiguar se o dado foi corretamente coletado ou se é um valor atípico.

Ao longo destas etapas deve-se encontrar o modelo ideal e, em paralelo, sanear o conjunto de dados. Assim, quando o melhor modelo foi encontrado conforme a métrica r² ajustado, os elementos saneados foram reinseridos e reanalisados. Esta prática garantiu o aumento do tamanho de elementos do conjunto de dados e também um acréscimo no valor de r² ajustado.

Em suma, a análise se deu nas seguintes etapas:

a) recebimento da base amostral bruta e tratamento: houve a conversão dos tipos, eliminação de partes textuais em meio aos dados, de colunas desnecessárias e de dados atípicos;

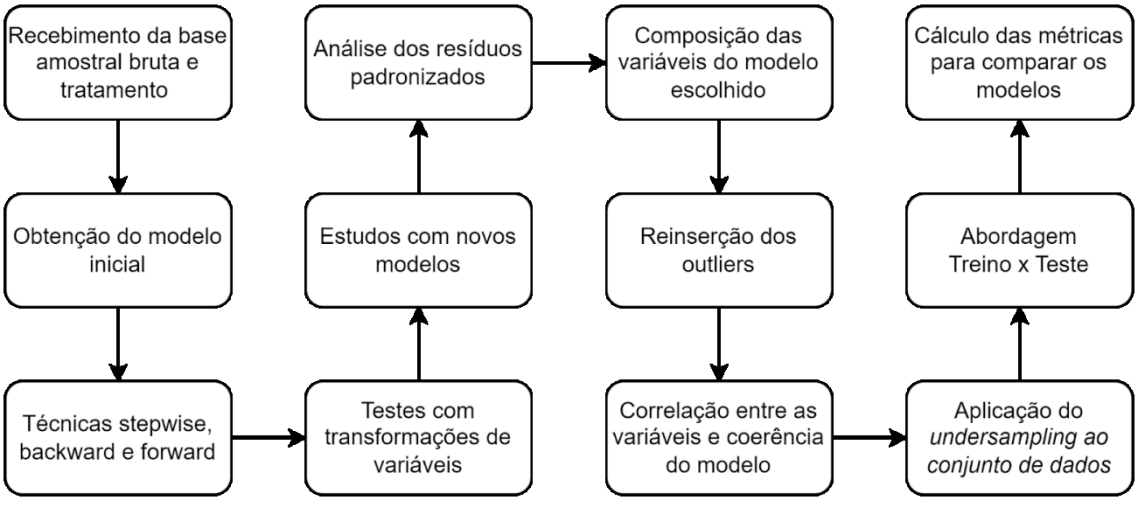
b) obtenção do modelo inicial: atingir um primeiro modelo é relevante para o entendimento das variáveis desnecessárias e de elementos *outliers*;

c) técnicas *stepwise*, *backward* e *forward*: esta etapa busca consolidar as melhores combinações de variáveis para conduzir a um modelo cujo formato permita a melhor explicação da variável dependente

d) testes com transformações de variáveis: partindo do gráfico de dispersão que relaciona as variáveis do conjunto de dados, observa-se a tendência dos dados e a partir disso as transformações ln x, ex, 1/x e $x^2$ são postas em análise;
e) estudos com novos modelos: para que seja possível analisar o impacto de outras variáveis visando explicar o comportamento da variável dependente;
f) análise dos resíduos padronizado: para eliminar *outliers*;
g) composição das variáveis do modelo: é uma etapa final que identifica quais variáveis melhor explicam o modelo;
h) reinserção dos *outliers*: para não penalizar os elementos que foram excluídos em outros modelos;
i) correlação entre as variáveis e coerência do modelo: a análise da correlação entre as variáveis através do coeficiente de Pearson permite entender a tendência de crescimento ou decrescimento de uma variável em relação à variável resposta, então o sinal do coeficiente de cada variável pode ser conferido, evitando modelos inconsistentes;
j) abordagem treino x teste: visa a aplicação das métricas.

O diagrama de setas disposto na figura 15 estabelece a sequência utilizada.

Figura 15 – MLR: sequência empregada na abordagem




Fonte: elaborado pelo autor (2023)

Nos modelos (*scripts* iniciados em MLR_model-ABR23_70x30.r) fez-se uso da variável independente preço unitário, explicada pelas variáveis dependentes área privativa, número de vagas e padrão de acabamento. As variáveis padrão de acabamento e número de vagas foram modeladas como fator.

Quanto aos pressupostos básicos, notadamente aqueles citados no item A.2 da NBR 14.653-2 (2011), por se tratar de um modelo de regressão linear, cada modelo teve a verificação da micronumerosidade (em função do número mínimo de dados de mesma característica), quanto ao equilíbrio da amostra (com dados bem distribuídos para cada variável ao longo do intervalo amostral), modelos homocedásticos (variância constante dos resíduos), distribuição normal dos resíduos, resíduos sem autocorrelação, variáveis importantes (área privativa) inclusa nos modelos, ausência de multicolinearidade importante entre variáveis independentes e ausência de pontos influenciantes.

A normalidade dos resíduos é verificada a partir da função nativa do pacote estatístico R chamada de shapiro.test(), aplicada aos resíduos do modelo, cujo p-valor maior que 0,05 indica a aceitação da hipótese nula que os resíduos seguem a distribuição normal. A aplicação do teste pode ser observada nos *scripts* dos modelos MLR.

A análise do gráfico de distância de Cook mostra que todos os elementos estão muito abaixo do limite 1,0, conforme orienta ÇETINKAYA-RUNDEL (2022), então conclui-se por não haver pontos influenciantes.

Ao conjunto de dados estabelecido aplicou-se a técnica *undersampling*, resultando um conjunto de dados que, por apresentar um incremento no $r^2$ ajustado, foi usado na formulação dos modelos segundo as distribuições treino x teste.

### 4.2.2 Aplicação do Sistema de Inferência Difuso de Mamdani (FIS)

A modelagem parte do conjunto de dados saneado que foi resultado da análise MLR. Dele são extraídas as variáveis preço unitário, área privativa, número de vagas e padrão de acabamento (os *scripts* são aqueles iniciados WM_model_MAI23_70.r). Ocorre o particionamento do conjunto de dados entre treino e teste nas proporções aproximadas de 70% x 30%, 75% x 25%, 80% x 20% e 85% x 15%.

A aplicação da técnica FIS fez uso da função frbs.learn com parâmetros para aplicação do método proposto por Wang e Mendel (1992), onde o conjunto de dados é normalizado e para a quantidade de elementos deste conjunto de dados resultaram quatro funções de pertinência por variável (dadas por 2N+1 onde N é um parâmetro que varia conforme o conjunto de dados, neste caso igual a 1,5), todas elas modeladas com a função gaussiana. Cabe destacar que o artigo original de Wang e Mendel (1992) traz a modelagem com funções triangulares (as bordas são funções trapezoidais), mas a biblioteca faz uso de funções gaussianas, provavelmente para facilitar os cálculos.

Do conjunto de dados as regras são inferidas e é atribuído um grau a cada regra, onde cada dado gera uma regra, todas com o operador AND. As regras são combinadas por importância para fins de reduzir o número de regras. A defuzzificação ocorre pelo método do centróide.

Quanto às funções de pertinência, algumas regras gerais são praticadas:

a) em geral os conjuntos *fuzzy* costumam ter sua intersecção no ponto de 0,5 de grau de pertinência, e o desalinhamento das interseções entre as funções deve ser evitado, exceto se houver alguma explicação para o fenômeno a partir do conjunto de dados. De qualquer modo, causa estranheza um elemento ter um grau de pertinência elevado a dois conjuntos ou ter um grau de pertinência que o torna pouco pertinente a dois conjuntos;
b) as bordas devem ser modeladas de maneira a evitar o erro de reduzir o grau de pertinência ao conjunto quando na verdade deveria aumentar. Desta maneira, as bordas devem ser formadas por funções trapezoidais (em modelos cujas funções intermediárias são triangulares) ou então se modelada com função gaussiana (ou sino-generalizada) devem ter o limite inferior com o valor máximo da função gaussiana com grau de pertinência 1, consideração análoga se aplica ao limite superior;
c) via de regra, as funções intermediárias precisam estar bem equilibradas para conjuntos de dados homogêneos;
d) como o número de regras para a modelagem difusa cresce exponencialmente de acordo com a quantidade de variáveis independentes, é importante manter a menor quantidade possível

> de classes, desde que coerente com a realidade do que está sendo estudado. Por esta razão, parece razoável observar a estatística t de cada variável no modelo MLR para ter um palpite sobre a necessidade (ou não) de criar subintervalos nas classes.

O modelo é validado pelos resultados das métricas, depois de confrontadas as quatro proporções de distribuições entre treino x teste (70% x 30%, 75% x 25%, 80% x 20% e 85% x 15%).

### 4.2.3 Aplicação do Sistema de Inferência Difuso de Takagi-Sugeno-Kang (TSK)

Como o trabalho partiu da técnica MLR, diversas alternativas já haviam sido rodadas e um estudo inicial já eliminou diversos elementos discrepantes do conjunto de dados, então o conjunto de dados já estava sem *outliers*. Os *scripts* são aqueles iniciados por sugeno_70x30.r. Os modelos TSK partem de regras onde cada regra faz uso de uma equação que é gerada por um subconjunto de dados, representando uma clusterização. Dessa maneira, considerando que o modelo TSK tem a clusterização embutida, seguiu-se com um conjunto de dados sem a aplicação da técnica de *undersampling*. O critério para esta decisão é que a separação do conjunto de dados em diversos subconjuntos tende a amenizar os efeitos do conjunto de dados desbalanceado, não tornando relevante a aplicação do procedimento de *undersampling*. Ainda, como diversas equações são praticadas, a redução no número de elementos é impactante, visto que equações formadas a partir de conjuntos de dados com poucos elementos tendem ao subajuste.

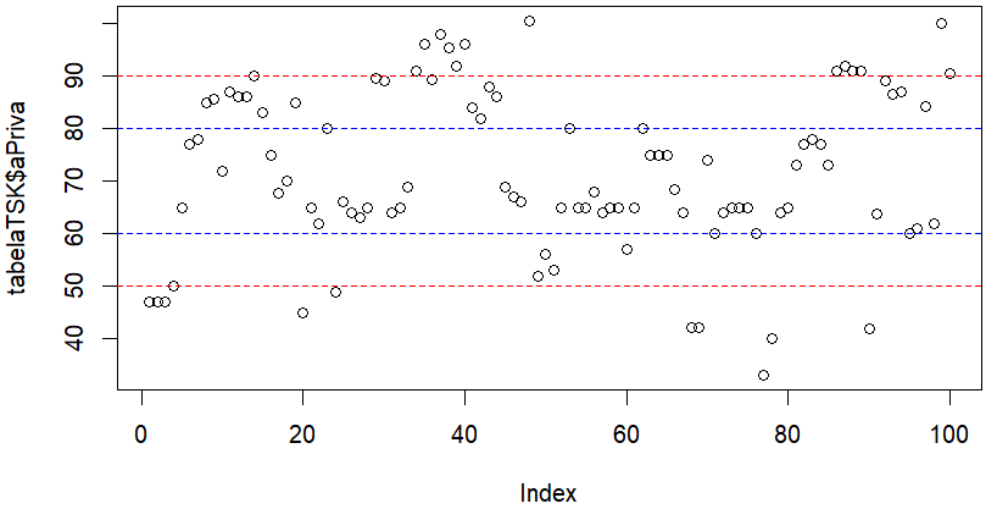
Da observação da dispersão das variáveis do conjunto de dados, as funções de pertinência foram estabelecidas com funções triangulares, assim como também fizeram os trabalhos de Yalpir e Ozkan (2018), Kamire *et al* (2021), Król *et al* (2007), Wang e Mendel (1992) e Malaman (2014).

De fato, a literatura pesquisada sobre o tema (vide o capítulo de trabalhos correlatos), todos os autores usaram funções de pertinência triangulares ou em forma de sino.

As figuras 16, 17, 18 mostram a distribuição em termos de gráfico de dispersão das variáveis área, idade e padrão, respectivamente dispostas no eixo y e

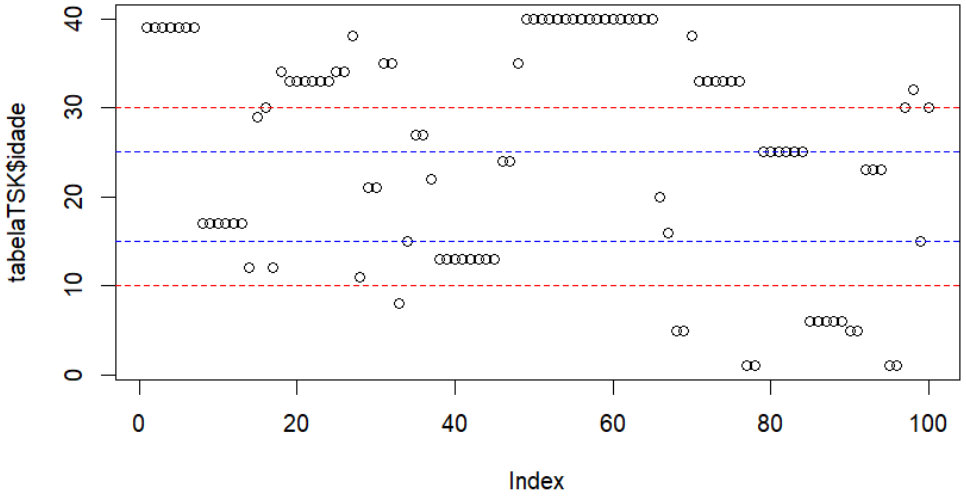
a sequência no eixo x. As linhas auxiliares facilitam a visualização. Em todas as figuras, nota-se que três termos linguísticos por variável resolvem a demanda.

Figura 16 - TSK: dispersão da área privativa (antecedentes)



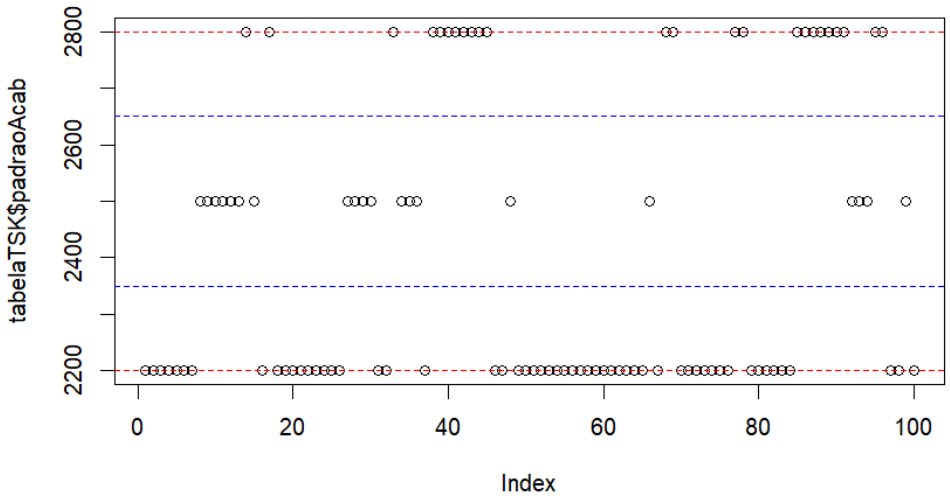

Fonte: elaborado pelo autor (2023)

Figura 17 - TSK: dispersão da idade (antecedentes)




Fonte: elaborado pelo autor (2023)

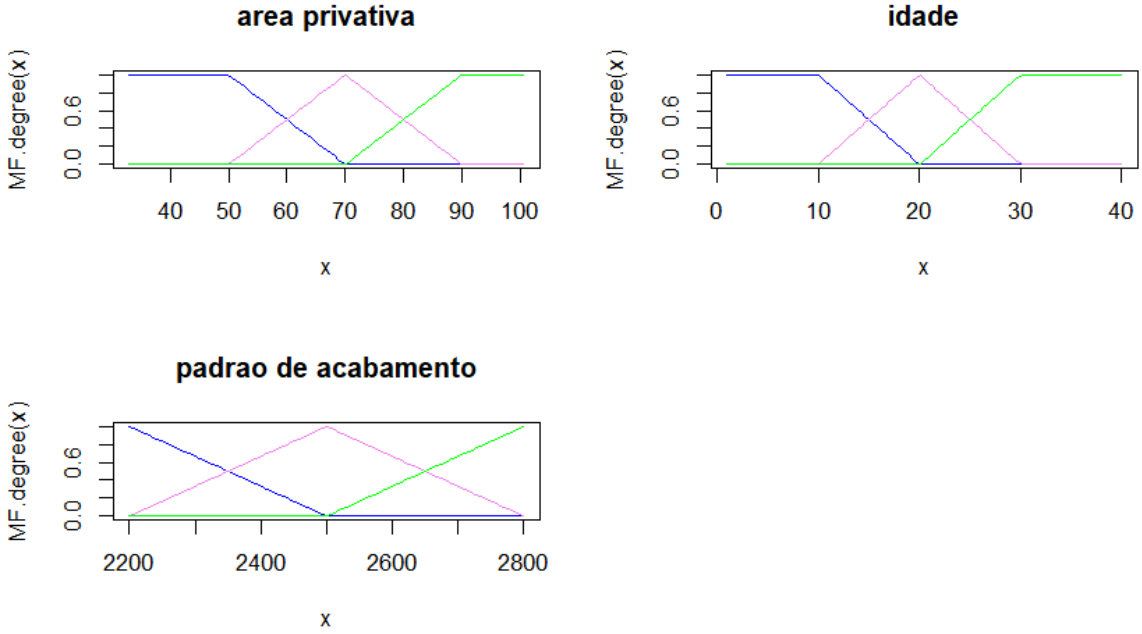
Figura 18 - TSK: dispersão do padrão de acabamento (antecedentes)




Fonte: elaborado pelo autor (2023)

A figura 19 apresenta os conjuntos difusos que são as funções de pertinência da parte antecedente da técnica. Notam-se os três termos linguísticos por variável.

Figura 19 - TSK: conjuntos difusos (antecedentes)




Fonte: elaborado pelo autor (2023)

Valle (2021) explica que sistemas TSK costumam ter no consequente polinômios e que usualmente são de ordem zero ou de ordem um. Contudo, há uma infinidade de problemas que podem ser enfrentados pela técnica TSK e já foi apresentado neste trabalho que uma só variável independente não explica a variabilidade observada na variável dependente. Mesmo assim, fez-se uso de um pacote pronto na biblioteca frbs chamado FIR.DM, que implementa o sistema de inferência difuso por regras pelo método do gradiente descendente com consequente formado por função polinomial de ordem zero. Nestes testes preliminares com modelos cujo antecedente é um polinômio de grau zero notou-se erros grandes (MAPE superior a 35%), demonstrando que nos casos em que diversas variáveis são necessárias para explicar a variação da variável dependente, não cabe a aplicação do método TSK cujo consequente tem grau zero. Desta maneira este trabalho não abordará funções com grau zero (função constante).

A variável padrão de acabamento foi modelada como variável *proxy*, em analogia ao CUB. Este tratamento é usual no ramo de engenharia de avaliações, é recomendado na NBR 14.653-2 (2011) e conduz a bons resultados.

Todas as combinações de termos linguísticos foram geradas, resultando as 27 regras, sendo que para cada regra foi calculada uma equação segundo o método dos mínimos quadrados.

O pacote frbs tem a função frbs.gen que recebe como parâmetros diversas configurações para o modelo e também a base de regras e as equações. Neste pacote o único tipo de defuzzificação usado na técnica TSK é a média ponderada, então a soma ponderada não é uma opção para parametrizar a geração dos modelos.

O modelo é validado pelos resultados das métricas, depois de confrontadas as quatro proporções de distribuições entre treino x teste (70% x 30%, 75% x 25%, 80% x 20% e 85% x 15%).

### 4.2.4 Aplicação do Sistema de Inferência Difuso Neuroadaptativo (ANFIS)

Esta técnica consiste na criação automatizada dos conjuntos difusos e na calibração do coeficiente β da função, criação automatizada da base de regras e da função polinomial do consequente num sistema TSK.

Por fazer uso de ANN, é necessário tomar o cuidado de normalizar os valores das variáveis independentes entre 0 e 1, por causa da função sigmóide.

Como o método ANFIS da biblioteca frbs é fechado, então fez-se um modelo aplicando normalização e outro não aplicando, mantendo todos os parâmetros, conjunto de dados e taxas iguais, então como o processamento conduziu a resultados idênticos entre os conjuntos de dados normalizados e não normalizados, conclui-se que a função de treinamento já possui os procedimentos de normalização inclusos, então esta etapa restou dispensada da análise.

Na aplicação da técnica ANFIS, diversas alterações na taxa de erro (fixada em step.size = 0,01) foram realizadas e não foi observada uma melhoria nos resultados, mas somente variação no tempo de treinamento.

Por outro lado, o número máximo de iterações (partindo de max.iter = 50) foi alterado e cada modelo passou por várias tentativas, sendo mantida aquela que conserva melhor resultado na análise das métricas.

Deste modo, a melhor calibração para cada alternativa de modelo usando a técnica ANFIS se deu com o número máximo de iterações variando caso a caso e taxa de erro de 0,01 para todos os modelos que usaram a técnica ANFIS.

Para subsidiar a aplicação da técnica ANFIS, uma abordagem inicial destacou a expectativa de modelos melhores quando usadas as variáveis área privativa, número de vagas, padrão de acabamento para explicar a variável dependente preço unitário. Esta configuração de variáveis foi praticada na técnica ANFIS.

Na técnica ANFIS, o modelo também foi validado pelos resultados das métricas, mas ao contrário das outras técnicas, manteve-se somente o melhor resultado da divisão treino x teste, sendo estudadas a resistência à elementos *outliers* e a melhoria do modelo em fazer uso de um conjunto de dados com ou sem a aplicação do *undersampling*.

### 4.2.5 Aplicação da Subamostragem da Classe Majoritária (Majority Class Undersampling Technique / *undersampling*)

A aplicação dos procedimentos de *undersampling* sobre o conjunto de dados ocorreu nas técnicas MLR e ANFIS.

O conjunto de dados é modelado com as classes desbalanceadas e, em seguida, obtém-se as métricas que são analisadas. Posteriormente aplica-se undersampling e novos modelos são obtidos, permitindo uma análise comparativa entre os resultados das métricas confrontando antes da aplicação com depois da aplicação do *undersampling*.

Neste trabalho, aplicou-se o balanceamento das classes na variável dependente (preço unitário), por meio da sua divisão em cinco intervalos:

a) até R$ 5.000,00 / m²;
b) entre R$ 5.000,01 / m² e R$ 8.333,00 / m²;
c) entre R$ 8.333,01 / m² e R$ 11.666,00 / m²;
d) entre R$ 11.666,01 / m² e R$ 15.000,00 / m²;
e) acima de R$ 15.000,01.

As classes citadas em a) e em e) possuem pela natureza do estudo menos elementos, então objetivou-se reduzir a classe majoritária que ocorreu em b) para a maior quantidade de elementos em c) ou d), justamente pela divisão de classes ser aquela que tendeu a igualar a quantidade de elementos em c) e d).

Este procedimento partiu de um comando de amostragem para escolher aleatoriamente elementos e depois formou um novo *dataframe* que foi usado como conjunto de dados para compor modelos em outras técnica.

## 4.3 MÉTRICAS PARA A COMPARAÇÃO DOS RESULTADOS ENTRE AS TÉCNICAS

Quando se comparam modelos, deseja-se identificar aquele modelo que melhor se ajusta aos dados, sendo o modelo com melhor desempenho. Para avaliar o desempenho dos modelos gerados (e indiretamente, a eficiência das técnicas), utilizam-se as métricas de avaliação de modelos. Por meio de uma medida quantitativa da qualidade do modelo, as métricas possibilitam a identificação do quão próximos os valores previstos por cada técnica estão dos valores reais, ou seja, o quão precisas são as predições atingidas por meio das técnicas. Isto significa que a eficiência de cada técnica é dada pelos resultados das métricas calculadas para cada modelo. A literatura consultada mostra que as métricas relatadas neste trabalho já foram empregadas com sucesso em outros estudos, sendo sugestivo utilizá-las.

### 4.3.1 Algumas métricas aplicáveis ao domínio do problema

Como este trabalho tem a variável resposta (variável dependente) contínua, então algumas métricas se aplicam na avaliação da qualidade do modelo, tais como:

a) MSE: Erro Quadrático Médio (*Mean Squared Error*)

Esta métrica é definida por

$$MSE = \left(\frac{1}{n}\right) * \sum_{i=1}^{n} [y_i - \hat{y}_i]^2 \qquad (4)$$

onde

*n* é o número de dados

$y_i$ é o preço (vide conjunto de testes)

$\hat{y}_i$ é o valor predito

Note que deve ser aplicada às predições realizadas aplicando o modelo (obtido dos dados de treino) aos dados do conjunto de teste. Da fórmula pode ser

observado que ocorrerá uma penalização quando houver erros grandes, em razão de existir o quadrado.

Para efeitos de comparação entre os modelos pode-se afirmar que como o MSE é sensível a *outliers* e valores elevados, não deve ser usada esta métrica sem ser acompanhada de outra.

b) $r^2_{\text{ajustado}}$: Coeficiente de Determinação Ajustado (*Adjusted r-squared*)

Esta métrica é definida por

$$r^2 = 1 - \left(\frac{\sum_{i=1}^{n}[y_i - \hat{y}_i]^2}{\sum_{i=1}^{n}[y_i - \bar{y}]^2}\right) \quad (5)$$

seguido de

$$r^2_{ajustado} = 1 - \left(\frac{(1-r^2)*(n-1)}{n-p-1}\right) \quad (6)$$

onde
*n* é o número de dados
p é o número de variáveis independentes
$y_i$ é o preço (vide conjunto de testes)
$\hat{y}_i$ é o valor predito
$\bar{y}$ é o valor médio (vide conjunto de testes)

É importante destacar que a inclusão de variáveis independentes aumenta o valor de r², e por esta razão deve ser utilizado $r^2_{ajustado}$ . Seu valor indica o percentual da variação da variável dependente que é explicada pelas variáveis independentes.

Note que o coeficiente de determinação ajustado será sempre um valor entre 0 e 1, e que quanto mais próximo de 1, melhor será o modelo.

Segundo Kamire *et al* (2021), esta métrica avalia quão profunda é a relação de linearidade entre as variáveis independentes e a variável dependente, sendo usualmente entendida como a medida da qualidade do ajuste. Como a técnica MLR

parte do pressuposto da linearidade, então esta métrica será usada apenas entre os modelos MLR, de maneira complementar às outras métricas.

c) MAE: Erro Absoluto Médio (*Mean Absolute Error*)

Esta métrica é definida por

$$MAE = \left(\frac{1}{n}\right) * \sum_{i=1}^{n} |y_i - \hat{y}_i| \quad (7)$$

onde
*n* é o número de dados
$y_i$ é o preço (vide conjunto de testes)
$\hat{y}_i$ é o valor predito

Note que deve ser aplicada às predições realizadas aplicando o modelo (obtido dos dados de treino) aos dados do conjunto de teste. Da fórmula, pode-se afirmar que a aplicabilidade da métrica MAE é mais adequada que o MSE quando houver *outliers* significativos ou quando a penalização por erros maiores precisa ser evitada. Escolhe-se aquele com menor MAE.

d) MAPE: Erro Médio Percentual Absoluto (*Mean Absolute Percentage Error*)

Esta métrica é definida por

$$MAPE = \left(\frac{1}{n}\right) * \sum_{i=1}^{n} \left|\frac{y_i - \hat{y}_i}{y_i}\right| \quad (8)$$

onde
*n* é o número de dados
$y_i$ é o preço (vide conjunto de testes)
$\hat{y}_i$ é o valor predito

Note que deve ser aplicada às predições realizadas aplicando o modelo (obtido dos dados de treino) aos dados do conjunto de teste. Da leitura da fórmula

percebe-se que esta métrica é aplicável em situações onde os dados não contém valores extremos e zeros. A presença do módulo evita o problema que desvios positivos venham a anular os desvios negativos. O MAPE expressa o erro em termos percentuais, então é aplicável em contextos onde a proporção do erro em relação ao valor real é importante. Na comparação entre modelos, aquele com menor MAPE indica um ajuste com melhor desempenho.

e) RMSE: Raiz do Erro Quadrático Médio (*Root Mean Squared Error*)

Esta métrica é definida por

$$RMSE = \sqrt{\left(\frac{1}{n}\right) * \sum_{i=1}^{n} (\hat{y}_i - y_i)^2} \quad (9)$$

onde

*n* é o número de dados

$y_i$ é o preço (vide conjunto de testes)

$\hat{y}_i$ é o valor predito

Note que deve ser aplicada às predições realizadas aplicando o modelo (obtido dos dados de treino) aos dados do conjunto de teste. Da leitura da fórmula percebe-se que a presença do quadrado evita o problema que desvios positivos venham a anular os desvios negativos. É uma métrica melhor que MSE. Na comparação entre modelos, aquele com menor RMSE indica um ajuste com melhor desempenho.

### 4.3.2 Outras métricas aplicáveis ao domínio do problema

A literatura traz também outras métricas, como *Akaike information criterion* (AIC), *Bayesian information criterion* (BIC), e a estatística $C_p$ ($C_p$ de Mallows) que não serão usadas neste trabalho.

Para constar, o AIC relaciona a complexidade do modelo com a qualidade do ajuste, para efeitos de comparação escolhe-se sempre o modelo com menor AIC. O critério BIC penaliza a complexidade do modelo, então é um pouco mais conservador que o AIC, para efeitos de comparação escolhe-se sempre o modelo com menor BIC. Já o $C_p$ está relacionado com a capacidade de generalização do modelo e com a qualidade do ajuste, para efeitos de comparação escolhe-se sempre o modelo $C_p$ próximo de 1. Ainda, quanto mais os valores de $C_p$ diminuem, têm-se evidências de um modelo com *underfitting* (subajuste) e quanto mais os valores de $C_p$ aumentam, têm-se evidências de um modelo com *overfitting* (sobreajuste).

### 4.3.3 Critérios para escolha das métricas

A métrica r² ajustado será usada para comparar os modelos que usam a técnica MLR numa etapa inicial de definição das variáveis do modelo. Como esta métrica é amplamente utilizada pelo meio técnico na avaliação de imóveis urbanos usando inferência estatística, então será a métrica usada na comparação entre os modelos que usam a técnica MLR, de maneira a escolher o melhor entre os ajustes. Complementarmente, serão empregadas as métricas MSE, MAE, RMSE e MAPE.

Os modelos gerados através das outras técnicas serão avaliadas conjuntamente pelas métricas MAE e MAPE, MSE e RMSE, então inicia-se por MAPE e MAE, seguido da avaliação de impacto de eventual elemento *outlier* por meio das métricas RMSE e MSE. Vale destacar que a métrica RMSE é muito sensível a variações, então seu uso se dará dependendo do caso concreto, mediante justificativa.

# 5 RESULTADOS E ANÁLISES

Esta seção apresenta os resultados obtidos neste trabalho, sob o intuito de discutir e através das métricas determinar qual das técnicas de inteligência artificial conduz a melhores modelos para o objeto de estudo e tecer comentários sobre cada particularidade.

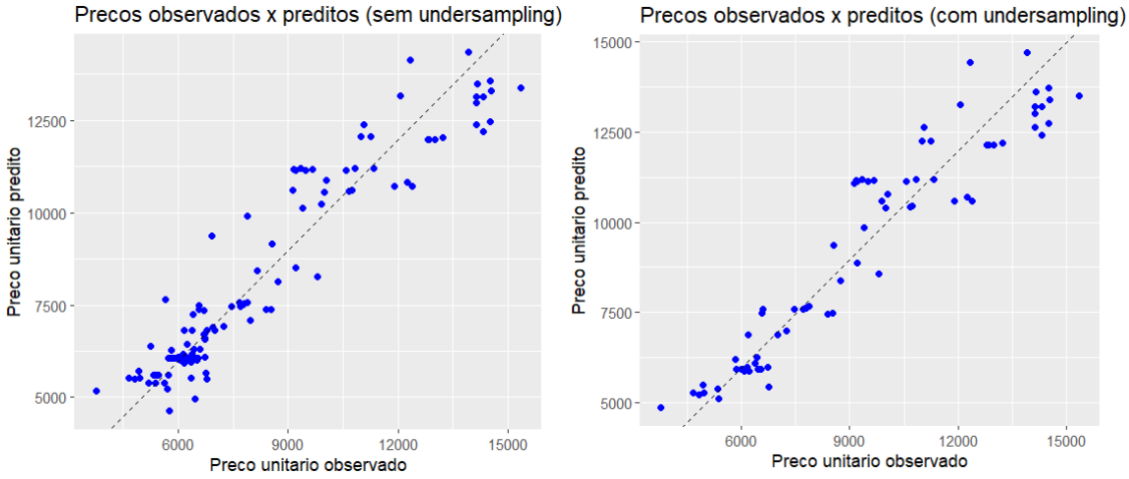
Para tanto constam tabelas e gráficos relacionando os resultados das métricas com as distribuições de treino x teste por técnica. O item 8.2.1.4.1 da NBR 14.653-2 (2011) cita que o poder de predição do modelo deve ser verificado a partir do gráfico de preços observados (no eixo x) versus valores preditos pelo modelo (no eixo y), então como comparação final (verificação do treinamento) constará também o gráfico do poder de predição, onde quanto mais próximos estiverem os pontos da linha da bissetriz, melhor será a qualidade do ajuste.

A determinação de qual técnica gera o modelo com melhor desempenho na avaliação imobiliária através do conjunto de dados proposto se dá pela comparação entre as métricas.

## 5.1 MODELOS MLR

O modelo inicial (disponível no *script* MLR_model-ABR23.r) não contemplou a divisão entre treino x teste, servindo de ponto de partida para obtenção do conjunto de dados saneado e identificação das principais variáveis independentes. A figura 20 traz os gráficos do poder de predição permitindo a comparação em termos de aderência à linha bissetriz, confrontando as alternativas sem *undersampling* (121 dados) e com *undersampling* (75 dados), de onde nota-se melhoria na homogeneização da distribuição entre os pontos, em razão do intervalo entre preço unitário de R$ 5.000 / m² e R$ 8.333 / m² passar de 72 para 25 elementos.

Figura 20 - MLR: comparação sobre a aplicação do *undersampling*




Fonte: elaborado pelo autor (2023)

Superada a etapa de análise das variáveis, o estudo dos efeitos da divisão entre treino e teste ocorreu com o conjunto resultado do *undersampling*. A tabela 1 indica o êxito na divisão do conjunto de dados sob a proporção de 80% para treino e 20% para teste, conforme as quatro métricas estudadas. Comparativamente, o menor resultado para a métrica RMSE indica a escolha da divisão 80% (treino) x 20% (teste). As outras métricas seguem a mesma análise e conclusões análogas se aplicam a este contexto.

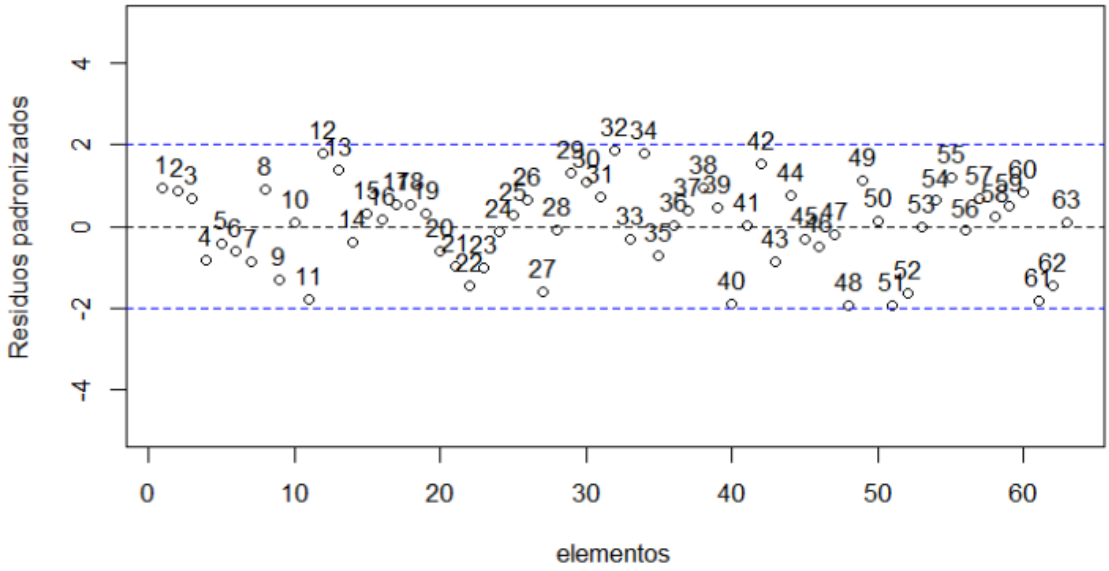
Tabela 1 – MLR: comparativo entre as proporções de treino x teste

| Divisão | Modelo MLR (PU ~ aPriva + nVagas + padraoAcabamento) | | | |
|---|---|---|---|---|
| | MSE | MAE | RMSE | MAPE |
| 70% x 30% | 1.261.330,00 | 917,28 | 1.123,09 | 9,74% |
| 75% x 25% | 948.470,20 | 806,26 | 973,89 | 9,00% |
| 80% x 20% | 793.825,70 | 724,24 | 890,97 | 7,39% |
| 85% x 15% | 1.644.782,00 | 1.103,05 | 1.282,49 | 11,09% |

Fonte: elaborado pelo autor (2023)

Tomando-se por melhor modelo aquele gerado pela distribuição 80% x 20%, não notaram-se *outliers*, conforme indica a figura 21 (resíduos padronizados). Destaca-se ainda que é possível perceber a aleatoriedade na distribuição dos resíduos padronizados.

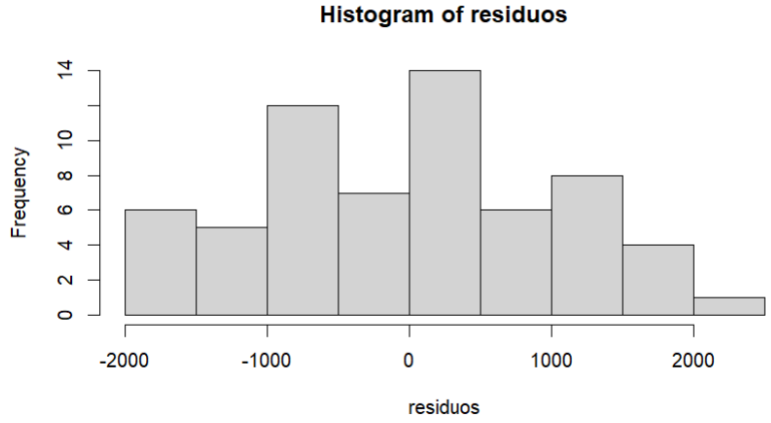
Figura 21 - MLR 80% x 20%: resíduos padronizados para o modelo




Fonte: elaborado pelo autor (2023)

Quanto à distribuição dos resíduos, o histograma de resíduos da figura 22 conserva aspecto de quase-normalidade, uma vez que a avaliação de imóveis envolve resíduos quase normais devido à imperfeição do mercado imobiliário, que é influenciado de forma não linear por fatores econômicos e eventos específicos, como mudanças na conjuntura econômica e subjetividade na definição dos preços dos imóveis. Ainda, os elementos do conjunto de dados possuem muitas desigualdades onde nem todas são abrangidas pelas variáveis e as classes são desbalanceadas. No entanto, mesmo com essa não normalidade, é possível utilizar técnicas estatísticas para obter preços confiáveis na avaliação dos imóveis.

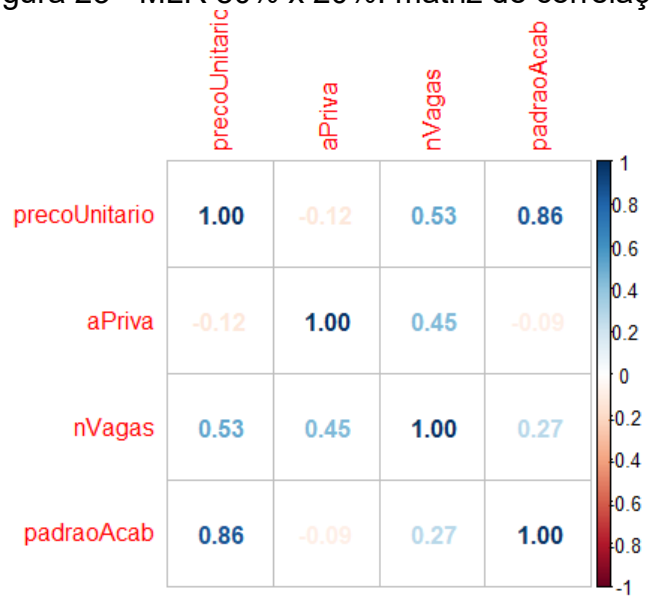
Figura 22 - MLR 80% x 20%: histograma de resíduos




Fonte: elaborado pelo autor (2023)

A matriz de correlações da figura 23 indica correlação positiva moderada a forte para as variáveis número de vagas e padrão de acabamento com a variável dependente, o que é importante. Ainda quanto à variável dependente (preço unitário), a área privativa apresentou correlação negativa (concordando com o comportamento do mercado) e fraca. Apesar de ser fraca, é importante que o modelo contenha a inclusão desta variável.

Figura 23 - MLR 80% x 20%: matriz de correlações

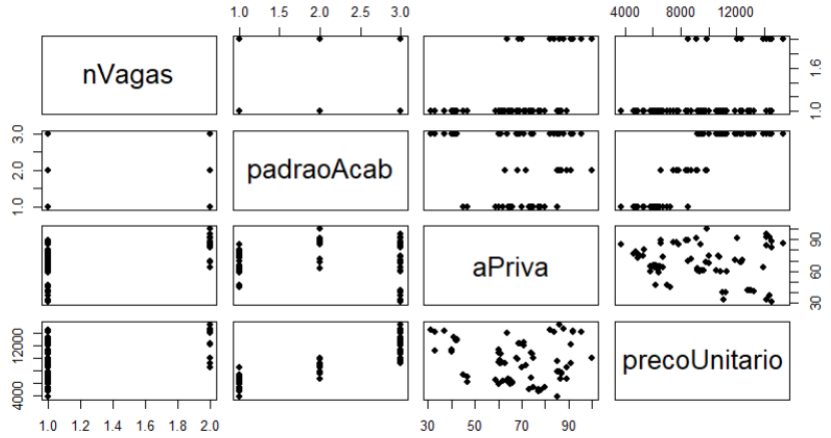



Fonte: elaborado pelo autor (2023)

A figura 23 também indica correlação fraca entre as variáveis independentes, o que vai ao encontro dos pressupostos da regressão linear múltipla. Vale destacar que a NBR 14.653-2 (2011) em seu anexo A recomenda atenção quando a correlação entre variáveis independentes ultrapassa o limite 0,80.

Quanto aos dados, muito do que se observa no gráfico de correlações também pode ser observado no gráfico de dispersão simples entre as variáveis, mostrado na figura 24:

Figura 24 - MLR 80% x 20%: gráfico de dispersão entre variáveis

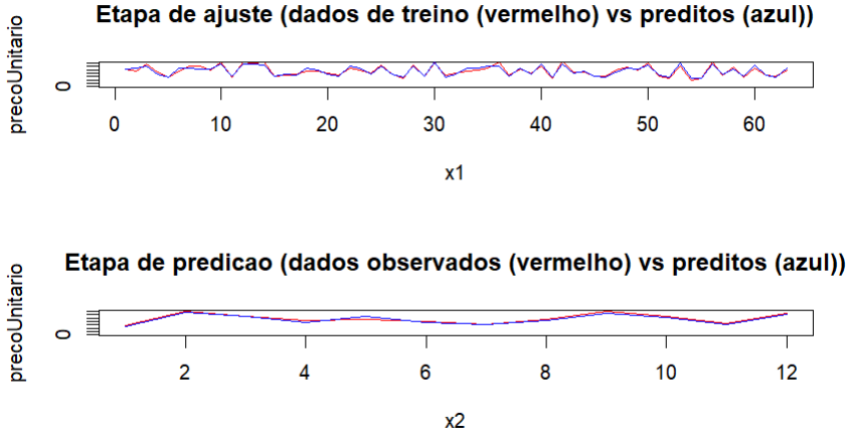



Fonte: elaborado pelo autor (2023)

Na figura 24, a dispersão quase que aleatória entre a área privativa e o preço unitário indica que somente a variável área privativa não é suficiente para explicar o comportamento da variável preço unitário.

A figura 25 mostra a qualidade do treinamento e o êxito na predição dos preços unitários de teste, indicando ser um excelente modelo.

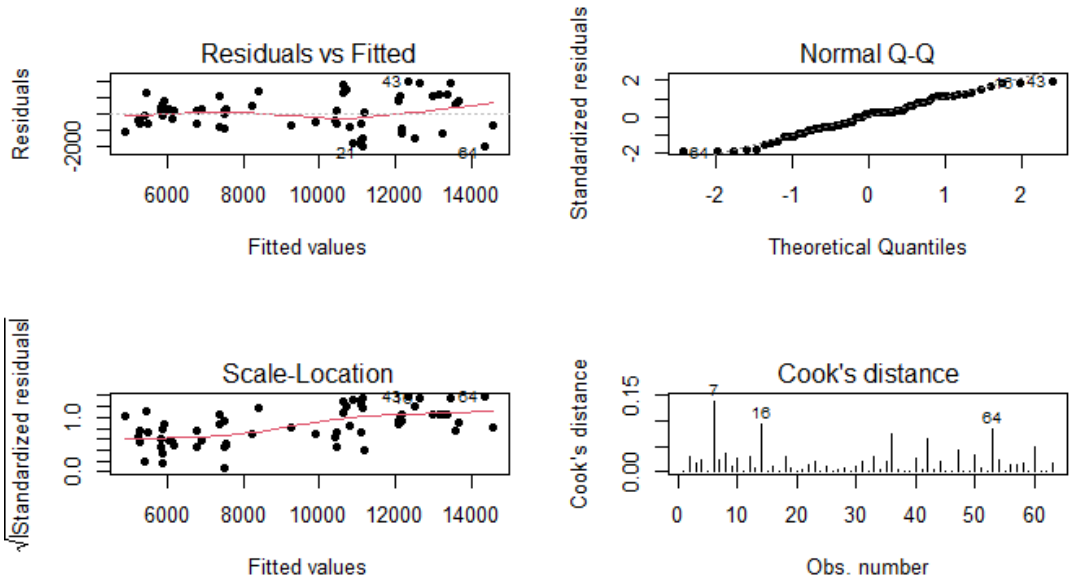
Figura 25 - MLR 80% x 20%: poder de predição (ajuste e predição) por sequência




Fonte: elaborado pelo autor (2023)

Por fim, a técnica MLR faz necessária algumas observações, tais como estas da figura 26 a aleatoriedade dos resíduos, a aderência à linha para concluir pela normalidade dos resíduos, e a distância de Cook.

Figura 26 - MLR 80% x 20% conjunto de gráficos para diagnóstico do modelo




Fonte: elaborado pelo autor (2023)

Em paralelo foi executado um estudo com as variáveis área privativa, idade (transformada com a função inversa) e padrão de acabamento (como fator) para explicar a variável preço unitário, obtendo-se os resultados da tabela 2:

Tabela 2 – MLR: comparativo entre as proporções de treino x teste

| | Modelo MLR (PU ~ aPriva + 1 / idade + padraoAcabamento) | | | |
|---|---|---|---|---|
| **Divisão** | **MSE** | **MAE** | **RMSE** | **MAPE** |
| **70% x 30%** | 1.211.689,00 | 708,31 | 1.100,77 | 8,59% |
| **77% x 23%** | 909.170,00 | 686,99 | 953,50 | 8,19% |
| **80% x 20%** | 607.681,00 | 535,71 | 779,54 | 6,75% |
| **87% x 13%** | 770.523,20 | 497,27 | 877,79 | 5,51% |

Fonte: elaborado pelo autor

A análise comentada deste modelo está no Anexo C. Cabe destacar que tratou-se de um modelo sem a aplicação da técnica de *undersampling* e que manteve-se alguns elementos cujo resíduo padronizado ultrapassou 1,96, limitado a 2,64, sob a justificativa que tratavam-se de sete pontos e que o mercado imobiliário não é perfeito. Por haver superado 1,96, foi excluído da comparação dos resultados.

## 5.2 **MODELOS FIS**

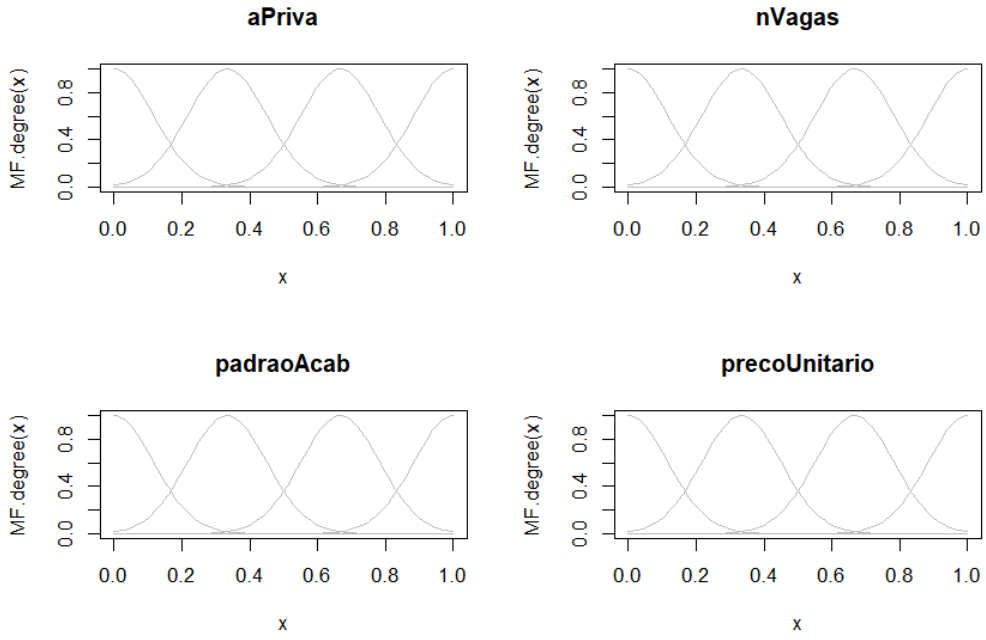
Os *scripts* de cada modelagem (de cada proporção treino x teste) estão em WM_model_MAI23_proporcao.R.

Os estudos partiram do conjunto de dados sem a aplicação da técnica de subamostragem (*undersampling*), mas já saneado.

Todas as proporções treino x teste mostraram bom desempenho, notado pela métrica MAPE. Algumas distribuições apresentaram valores um pouco maiores para a métrica RMSE, por efeito da fórmula da métrica que penaliza grandes desvios.

Fez uso das variáveis área privativa, número de vagas e padrão de acabamento para explicar a variabilidade na variável preço unitário. Esta é a mesma distribuição de variáveis praticada nos modelos MLR. A figura 27 mostra as funções de pertinência.

Figura 27 - FIS: funções de pertinência




Fonte: elaborado pelo autor (2023)

O sumário do objeto (modelo FIS) criado informa que as funções de pertinência são gaussianas, que a t-norma foi mínimo, que a s-norma foi a padrão da biblioteca, que a defuzzificação foi pelo método da média ponderada (abreviado por WAM na descrição da técnica FIS), que a função de implicação de Zadeh e que

usaram-se quatro nomes linguísticos para cada uma das variáveis de entrada. Como todos os dados foram normalizados antes da aplicação dos métodos, então o cálculo da posição da média restou simplificado. A função gaussiana teve a média posta no centro do intervalo, desvio padrão 5,0 e espalhamento 0,1167. A variável da saída do sistema foi discretizada em quatro termos linguísticos e seguiu a mesma formatação das variáveis de entrada. Do mecanismo de aprendizado automático, 26 regras foram criadas. Poderia-se verificar que todas as combinações resultariam 64 regras (4x4x4), mas ocorre que existe uma simplificação por meio de uma tabela de pontuação entre as regras, que decorre do algoritmo de Wang e Mendel (1992). Em sistemas FIS, é comum que algumas regras nunca ou raramente sejam ativadas. De fato, o sistema não precisa conter todas as regras para funcionar.

O modelo FIS trouxe os resultados da tabela 3:

Tabela 3 – FIS: comparativo entre as proporções de treino x teste

| | Modelo FIS | | | |
|---|---|---|---|---|
| Divisão | MSE | MAE | RMSE | MAPE |
| 70% x 30% | 1.061.433,00 | 837,02 | 1.030,26 | 10,80% |
| 75% x 25% | 1.071.559,00 | 803,10 | 1.035,16 | 10,40% |
| 80% x 20% | 1.172.636,00 | 876,33 | 1.082,88 | 11,07% |
| 85% x 15% | 1.261.787,00 | 965,88 | 1.123,29 | 11,39% |

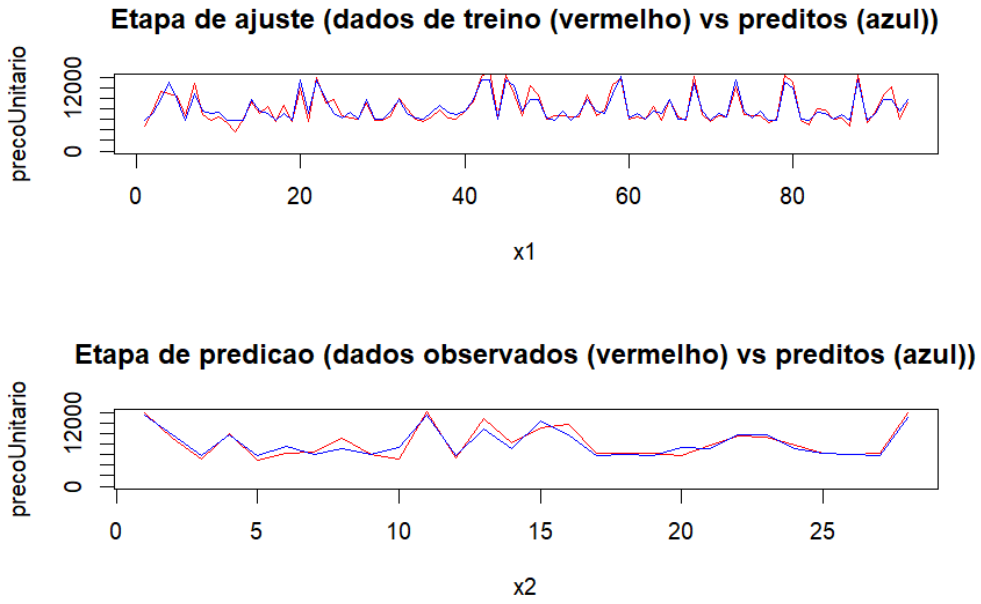
Fonte: elaborado pelo autor

As métricas indicam tanto para MAE quanto para MAPE que a divisão entre 75% e 25% apresenta os melhores resultados. As métricas MSE e RMSE indicam a divisão 70% x 30%, vale destacar que são métricas que penalizam erros maiores então algum ponto se sobressaiu e penalizou o modelo.

Estabeleceu-se o modelo 75% x 25% como melhor modelo por causa das métricas MAE e MAPE, ignorando-se a métrica RMSE justamente em razão da diferença ser de pequena magnitude. De fato, o mercado imobiliário é uma concorrência imperfeita então eventualmente pode haver alguma distorção, então a penalização por erros maiores (ocorre na métrica RMSE) não é justa.

Para fins de simplificação, os resultados adiante serão todos da proporção escolhida (75% x 25%). Do treinamento, nota-se na figura 28 grande aderência entre os dados observados e os dados preditos.

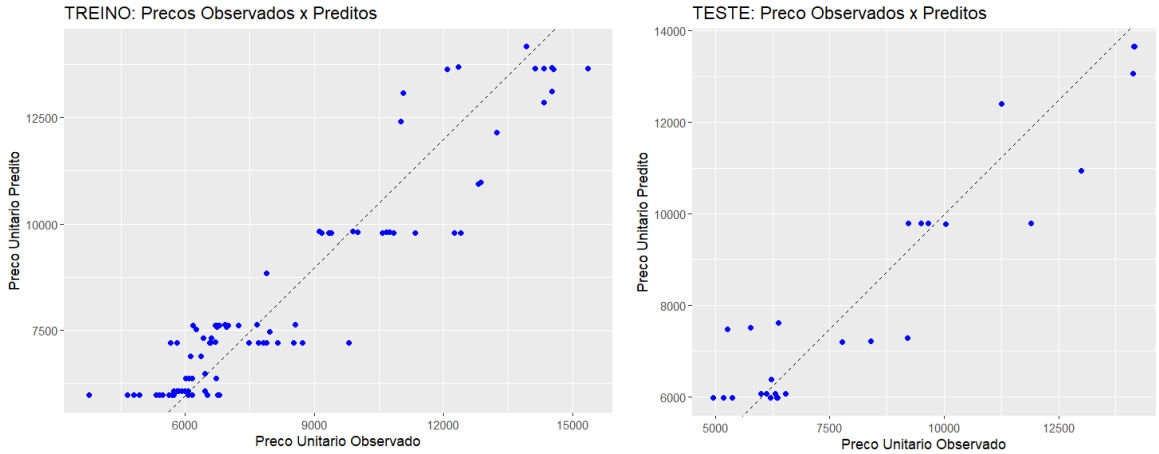
Figura 28 - FIS 75% x 25%: poder de predição (ajuste e predição) por sequência




Fonte: elaborado pelo autor (2023)

A figura 29 mostra no gráfico da esquerda que os dados foram melhor treinados para preços unitários menores, que são maioria no conjunto de dados. Já o gráfico da direita mostra que os maiores acertos (menor magnitude dos erros) ocorrem nesta faixa de preços unitários menores.

Figura 29 - FIS 75% x 25%: poder de predição (ajuste e predição) por dispersão




Fonte: elaborado pelo autor (2023)

## 5.3 MODELOS TSK

Os *scripts* de cada modelagem (de cada proporção treino x teste) estão em sugeno_proporcao.R. Diversas proporções de treino x teste foram rodadas por meio da técnica TSK. Os estudos partiram de um conjunto de dados sem *outliers* e sem aplicar a técnica de *undersampling*, resultando um conjunto de dados com 100 elementos (parcialmente saneado).

Os modelos foram realizados com as variáveis área privativa, idade e padrão de acabamento para explicar a variabilidade na variável resposta preço unitário. Destaca-se que a variável padrão de acabamento foi modelada como variável *proxy*, conforme adiantado no tópico de aplicação.

Não há biblioteca integralmente automatizada para rodar a técnica TSK, então os conjuntos foram concebidos manualmente com funções triangulares e trapezoidais nas bordas para as variáveis área privativa, padrão de acabamento e idade. Cada variável independente teve três divisões. Das divisões, infere-se ser necessário 27 regras (3x3x3). As regras que mostraram insuficiência de regras para constituir uma função de consequente foram removidas. Tem-se a proporção de uma regra para cada equação gerada. O tópico sobre a aplicação da técnica TSK explica a definição das funções de pertinência. O tipo da t-norma é mínimo, da s-norma é máximo, é usada a função de implicação de Zadeh e o operador é o AND.

Ainda, a tabela 4 traz o comparativo com as métricas, destaca-se o menor MAPE para a divisão 72%  x 28%. Como a biblioteca caret distribui os dados nas proporções aproximadas segundo um critério de aleatoriedade, então não é exato haver 75%, mas sim algo próximo disto. Na tabela 4, apesar do MAPE indicar melhor resultado para a 72% x 28%, resultados muito parecidos com a alternativa 77% x 23%, o RMSE pouco se alterou, então a segunda alternativa será praticada como efetiva neste trabalho.

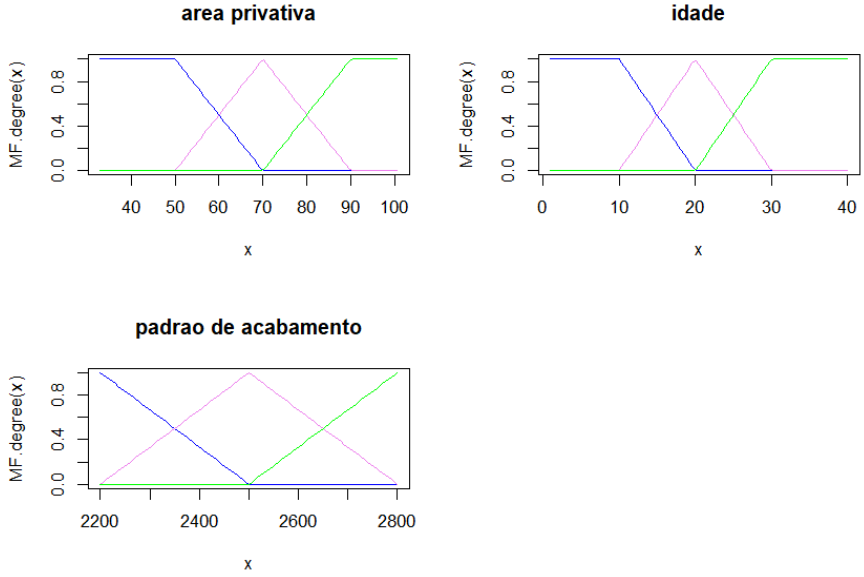
Tabela 4 – TSK: comparativo entre as proporções de treino x teste

| | **Modelo TSK: sem *outliers* e sem *undersampling*** | | | |
|---|---|---|---|---|
| **Divisão** | **MSE** | **MAE** | **RMSE** | **MAPE** |
| **72% x 28%** | 1.145.532,00 | 773,09 | 1.070,30 | 8,51% |
| **77% x 23%** | 1.158.785,00 | 778,28 | 1.076,47 | 9,74% |
| **82% x 18%** | 2.338.628,00 | 1.247,54 | 1.529,26 | 15,56% |
| **87% x 13%** | 1.799.465,00 | 1.017,96 | 1.341,44 | 12,03% |

Fonte: elaborado pelo autor

A figura 30 mostra a distribuição das funções de pertinência, todas de acordo com as recomendações expostas no tópico sobre a técnica FIS.

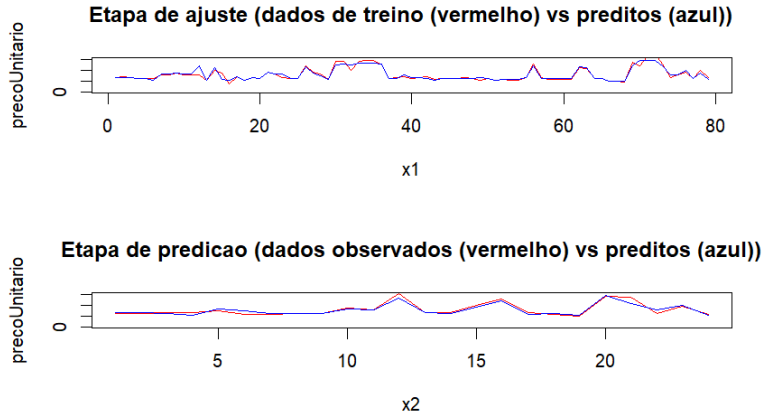
Figura 30 - TSK 72% x 28%: funções de pertinência dos antecedentes




Fonte: elaborado pelo autor (2023)

A figura 31 mostra que o treino apresentou alguns erros localizados. De fato, na distribuição entre treino x teste, admite-se certa aleatoriedade, então eventuais pontos desassemelhados do restante do conjunto ocorrem naturalmente.

Figura 31 - TSK 72% x 28%: poder de predição (ajuste e predição) por sequência

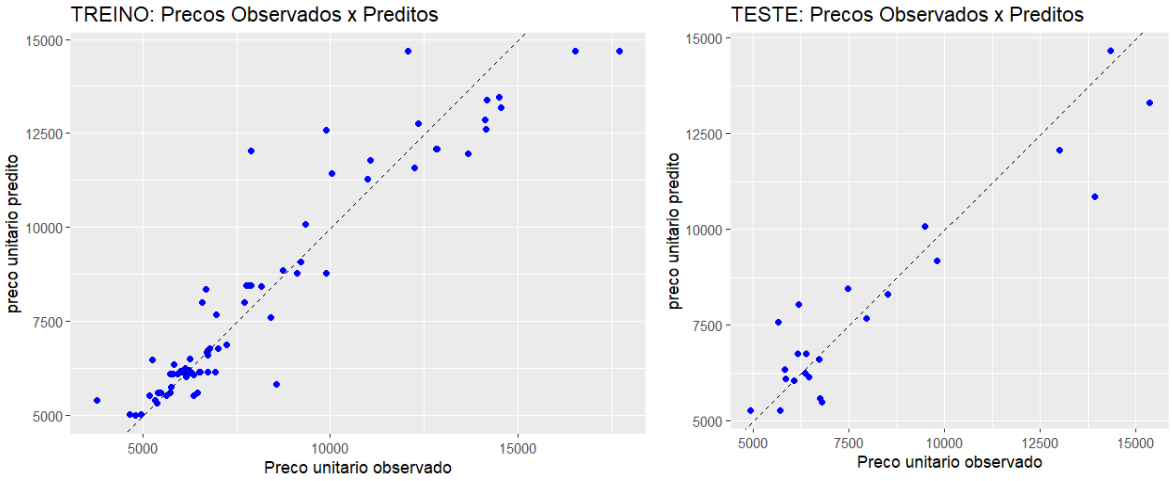



Fonte: elaborado pelo autor (2023)

A figura 32 apresenta pontos com treinamento bom e teste razoável para os elementos na faixa de preços unitários mais representativa (com maior quantidade de elementos), apresentando para a faixa de preços unitários mais elevados

resultados para treino um pouco piores (havendo erros grandes) e resultados com desvios maiores para o teste. A análise gráfica (vide figura 30, gráfico da direita) mostrou que o modelo gerado tem uma tendência em subavaliar imóveis cujos preços unitários observados eram maiores, possivelmente em razão de haver sido treinado preponderantemente com elementos de menor preço unitário.

Figura 32 - TSK 72% x 28%: poder de predição (ajuste e predição) por dispersão




Fonte: elaborado pelo autor (2023)

Nas combinações sem elementos manteve-se os coeficientes da função zerados e nestes casos tomou-se o cuidado que o conjunto de teste não contenha elementos que ativam essas regras.

Apesar de algumas equações apresentarem distorções nos sinais, os resultados foram bons.

## 5.4 **MODELOS ANFIS**

Diversas alternativas foram executadas por meio da técnica ANFIS. Os estudos partiram de um conjunto de dados parcialmente saneado, ou seja, o conjunto de dados resultado do tratamento inicial dos dados. Cabe ressaltar que o conjunto de dados saneado é resultado da eliminação dos valores cujos pontos apresentaram resíduos padronizados abaixo de -1,96 e acima de + 1,96 e será objeto de uma análise posterior.

A intenção em usar o conjunto de dados parcialmente saneado é obter evidências se a técnica ANFIS é resistente a *outliers*.

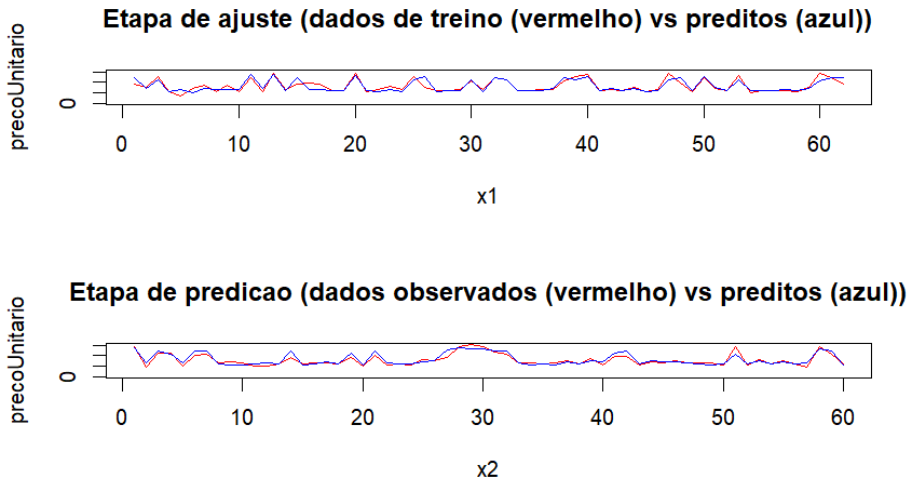
### 5.4.1 *Modelo ANFIS com outliers e sem undersampling*

A primeira abordagem segundo a técnica ANFIS (vide *script* anfis_modelo_50.r) apresentou como melhor modelo aquele obtido pela combinação de variáveis área privativa, número de vagas, padrão de acabamento e preço unitário, na proporção treino 51% x teste 49%, então o conjunto de dados com 122 elementos ficou dividido entre 62 elementos para treino e 60 elementos para teste, proporção semelhante ao estudo de Król *et al* (2007).

O número máximo de iterações que conduziu a melhores métricas para esta alternativa foi 68.

Notou-se na etapa de predição a tendência em resultar o mesmo preço unitário para diversos dados (pode-se verificar no agrupamentos pontos que aparecem na figura 33) provavelmente por uma das regras do ANFIS ser ativada muitas vezes (sugere má divisão na base de regras), então uma determinada faixa agrupou muitos dados, sugerindo um número maior de regras ou até mesmo um procedimento de *undersampling* poderiam melhorar os resultados. Cabe destacar que no ANFIS a base de regras é gerada automaticamente e não permite intervenção, sendo a biblioteca frbs fechada neste sentido.

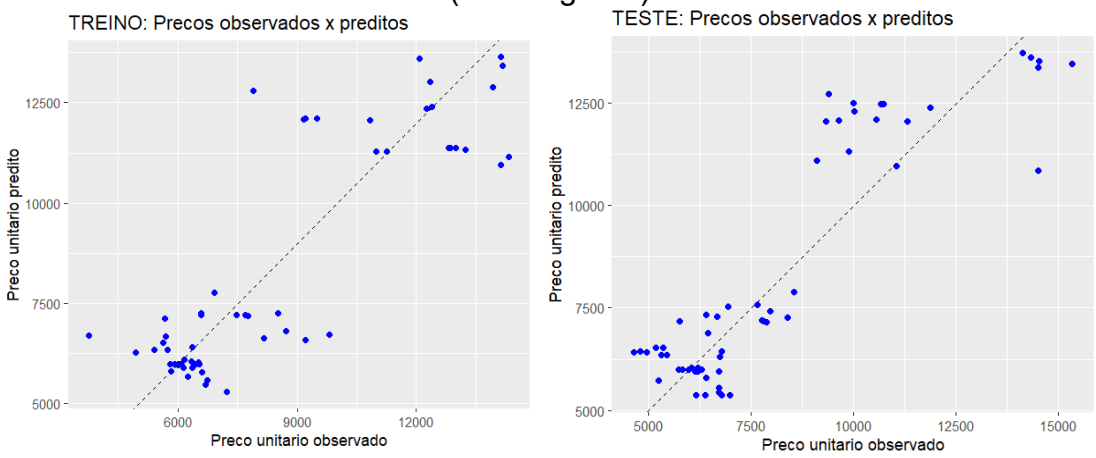
Figura 33 - ANFIS: poder de predição (ajuste e predição) por sequência (abordagem 1)




Fonte: elaborado pelo autor (2023)

A figura 34 mostra o ajuste obtido.

Figura 34 - ANFIS: poder de predição (ajuste e predição) por dispersão (abordagem1)




Fonte: elaborado pelo autor (2023)

Na figura 34, nota-se a limitação do modelo em inferir resultados para valores unitários entre R$ 7.879,26 / m² e R$ 10.865,49 / m². A inspeção do objeto resultado da predição mostra que para esta natureza de conjunto de dados somente 13 regras foram criadas. Do modelo, a figura 34 faz entender que o treino não está suficientemente bom para preços unitários observados muito altos, notadamente em razão do gráfico de treino apresentar maior dispersão nesta faixa de valores.

Ainda, a tabela 5 traz o comparativo com as métricas, destaca-se RMSE cerca do dobro do RMSE encontrado para a técnica MLR 80 x 20, e o MAPE pouco superior a 13%, indicando que alguns dos resultados do modelo possuem erros grandes.

Tabela 5 – ANFIS: comparativo entre proporções de treino x teste (abordagem 1)

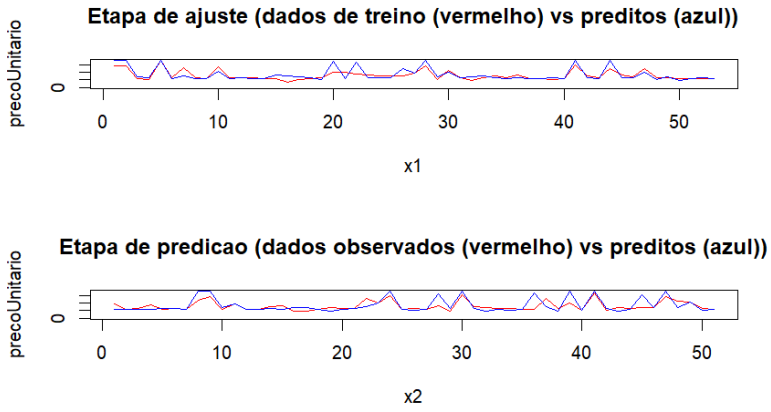
| | **Modelo ANFIS: conjunto com *outliers* e sem *undersampling*** | | | |
|---|---|---|---|---|
| **Divisão** | **MSE** | **MAE** | **RMSE** | **MAPE** |
| **51% x 49%** | 1.712.735,00 | 1.030,55 | 1.308,72 | 13,05% |

Fonte: elaborado pelo autor

### 5.4.2 *Modelo ANFIS sem outliers e sem undersampling*

A análise da técnica ANFIS continua com uma segunda abordagem (vide *script* anfis_modelo_50_semOutliers.r) que, mantendo as mesmas variáveis do comparativo anterior, apresentou como melhor modelo aquele obtido pelo número máximo de iterações igual a 25. A proporção treino 51% x teste 49% foi mantida, então o conjunto de dados com 104 elementos (é o conjunto de dados obtido depois de eliminar os elementos com resíduo padronizado abaixo de -1,96 e acima de +1,96) ficou dividido entre 53 elementos para treino e 51 elementos para teste, proporção semelhante ao estudo de Król et al (2007). A figura 35 mostra o resultado do ajustamento, em especial se destacam alguns pontos com tendência a se afastar da linha vermelha, predizendo preços unitários muito altos.

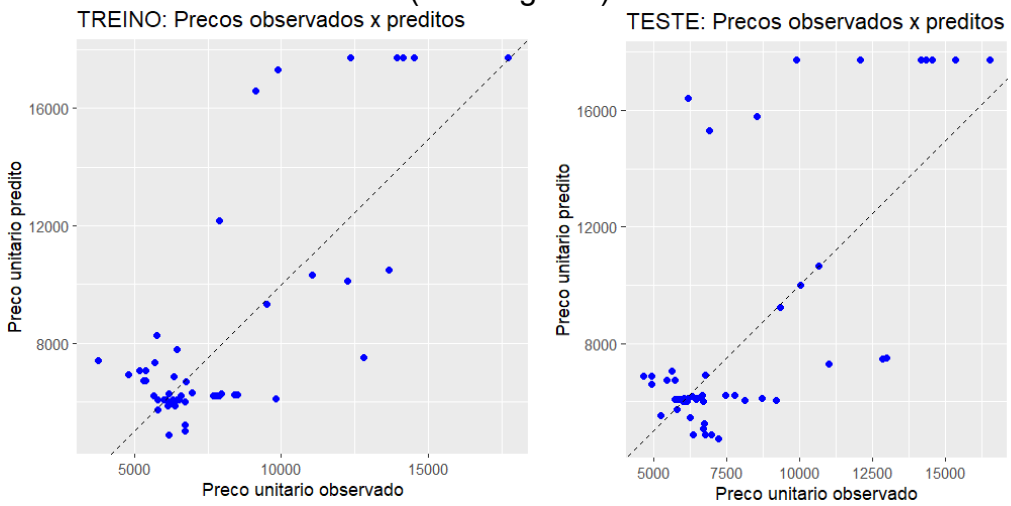
Figura 35 - ANFIS: poder de predição (ajuste e predição) por sequência (abordagem 2)




Fonte: elaborado pelo autor (2023)

Notou-se na etapa de predição a tendência em inferir preços unitários preditos abaixo de R$ 8.000,00 / m² (pode-se verificar no agrupamentos pontos que aparecem na figura 36) provavelmente porque esta faixa agrupa a maioria dos elementos de treino, que são bastante heterogêneos em termos das variáveis independentes. Assim como na análise anterior, esta é uma evidência da necessidade de um número maior de regras e até mesmo da necessidade da aplicação de um procedimento de *undersampling* para melhorar os resultados. No ANFIS a base de regras é gerada automaticamente (sem opção de intervenção ou personalização) e foram geradas 13 regras.

Figura 36 - ANFIS: poder de predição (ajuste e predição) por dispersão (abordagem2)




Fonte: elaborado pelo autor (2023)

A tabela 6 traz o comparativo com as métricas, destaca-se RMSE mais de dez vezes aquele encontrado para a técnica MLR 80 x 20, indicando haver vários elementos cuja predição se afasta do valor observado e o MAPE pouco superior a 25%, uma evidência que os resultados do modelo precisam ser usados com cautela. De fato, desde a etapa de treino ficou evidente que preços unitários acima de R$ 10.000,00 / m² conduzem a erros muito maiores, então conclui-se que o modelo não aprendeu suficientemente sobre elementos de preço unitário maior.

Tabela 6 – ANFIS: comparativo entre proporções de treino x teste (abordagem 1)

| Divisão | Modelo ANFIS: conjunto sem *outliers* e sem *undersampling* | | | |
|---|---|---|---|---|
| | MSE | MAE | RMSE | MAPE |
| **51% x 49%** | 9.789.929,00 | 2.032,08 | 3.128,89 | 24,67% |

Fonte: elaborado pelo autor

Quanto à resistência à *outliers*, restou inconclusivo, uma vez que os efeitos da heterogeneidade da distribuição dos elementos nas classes com a presença de *outliers* se misturam, não havendo possibilidade de isolamento.

Devido ao fato dos *outliers* terem sido removidos, este treinamento fez uso de um modelo com menos dados e isso afeta a capacidade de aprendizado.

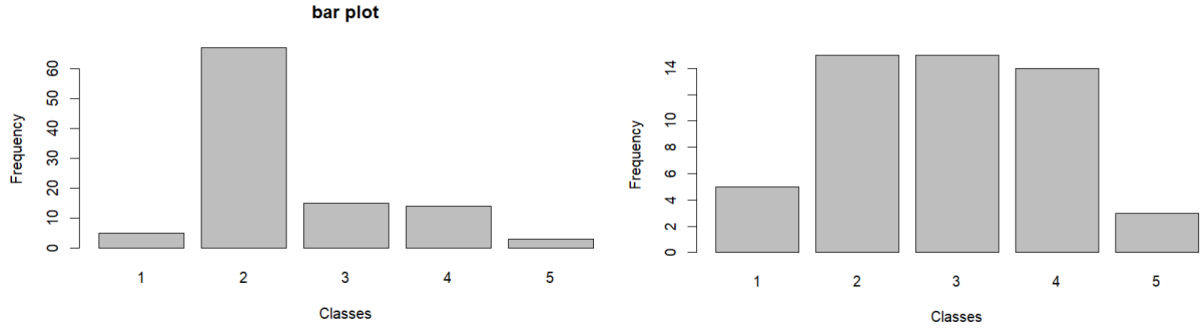
### 5.4.3 *Modelo ANFIS sem outliers e com undersampling*

Esta terceira abordagem segundo a técnica ANFIS (vide *script* anfis_modelo_50_undersampling_semOutliers.r) manteve a combinação de variáveis dos outros modelos (área privativa, número de vagas, padrão de acabamento e preço unitário) para preservar a comparação. Na proporção treino 54% x teste 46%, o conjunto de dados (saneado) com 104 elementos ficou reduzido para 52 elementos. Um dos objetivos desta alternativa é averiguar se a aplicação da técnica de *undersampling* ao conjunto de dados melhora o resultado do modelo.

Cabe destacar que o conjunto de dados sem saneado é aquele obtido depois de eliminar os elementos com resíduos padronizados abaixo de -1,96 e acima de +1,96 no modelo MLR inicial.

Na aplicação da técnica de *undersampling* a distribuição da segunda classe (entre R$ 5.000,00 / m² e R$ 8.333,00 / m²) apresentou 67 elementos, em desproporção com as outras classes majoritárias (14 e 15 elementos), então houve a redução para 15 elementos, e consequentemente o conjunto de dados passou para 52 elementos, conforme ilustra a figura 37.

Figura 37 - ANFIS: gráfico de barras antes e depois de aplicar *undersampling*

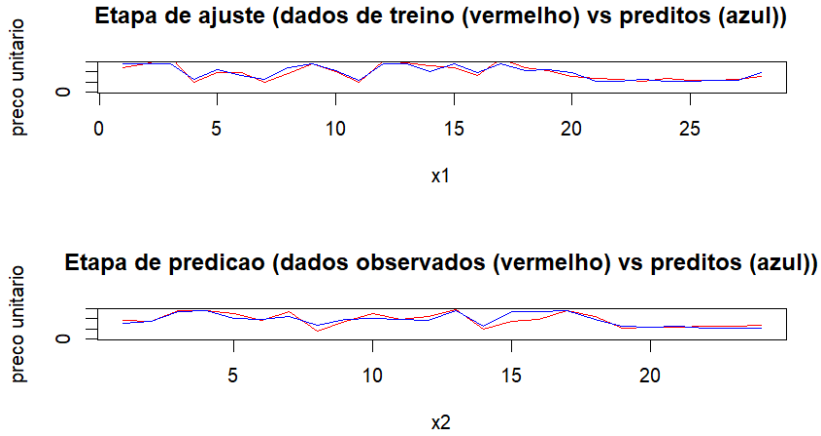



Fonte: elaborado pelo autor (2023)

Obtido o conjunto de dados (com 52 elementos) depois de aplicada a técnica de *undersampling*, houve a divisão entre dados de treino e de teste, onde a biblioteca caret resultou para a probabilidade p = 0,50 a proporção 54% para treino (28 elementos) x 46% para teste (24 elementos), proporção semelhante ao estudo de Król et al (2007). Vale destacar que mesmo solicitado p = 0,50, a função createDataPartition tem por objetivo aplicar uma técnica de probabilidade cuja aleatoriedade pode ser afetada por diversos fatores, tais como a presença de dados muito assemelhados. De fato, a intenção é balancear os dados.

O objeto resultado do aprendizado com número máximo de iterações igual a 10, com as mesmas taxas, mesmas variáveis e mesma proporção da alternativa anterior resultou 10 regras (contra 13 regras da opção anterior) e ainda notam-se grandes erros (elementos 15 e 16 do conjunto de dados de treino), conforme figura 38:

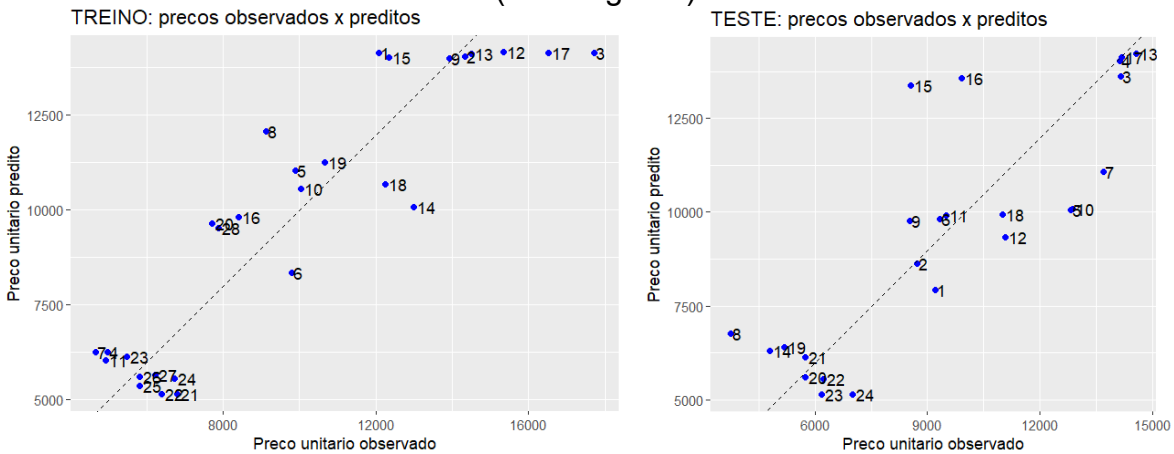
Figura 38 - ANFIS: poder de predição (ajuste e predição) por sequência (abordagem 3)




Fonte: elaborado pelo autor (2023)

A figura 39 mostra que a baixa quantidade de regras resultou na limitação do modelo em predizer preços unitários em algumas faixas, notadamente entre R$ 11.085,50 / m² e R$ 13.371,41 / m². O próprio conjunto de treino juntamente com o fato de haverem somente dez regras já havia sugerido a estrutura com três agrupamentos.

Figura 39 - ANFIS: poder de predição (ajuste e predição) por dispersão (abordagem3)




Fonte: elaborado pelo autor (2023)

Pode-se observar o mesmo problema já reparado no conjunto de dados sem *undersampling*, onde houve uma tendência em inferir mesmos resultados para alguns dados (23 e 24 do conjunto de teste). A hipótese de que a aplicação da técnica de *undersampling* ao conjunto de dados melhoraria os resultados se confirmou para o caso do conjunto de dados sem *outliers* e a hipótese que a insuficiência na base de regras piora os resultados se manteve. Persiste a evidência que a redução da quantidade de dados de treino impactou negativamente no poder de predição do modelo.

A tabela 7 traz as métricas encontradas, o MSE e o RMSE fazem entender que o MAPE da alternativa com dados sem *outliers* que passou por *undersampling* traz melhores resultados que a alternativa com *outliers* sem *undersampling*, comparando as duas alternativas pode-se afirmar que os erros antes encontrados em alguns elementos diminuíram. De fato, mesmo com um número menor de elementos, o modelo conseguiu aprender mais por estar com classes mais bem equilibradas.

Tabela 7 – ANFIS: comparativo entre proporções de treino x teste (abordagem 3)

| | Modelo ANFIS: conjunto sem *outliers* e com *undersampling* | | | |
|---|---|---|---|---|
| **Divisão** | **MSE** | **MAE** | **RMSE** | **MAPE** |
| **54% x 46%** | 3.523.135,00 | 1.403,98 | 1.877,00 | 17,67% |

Fonte: elaborado pelo autor

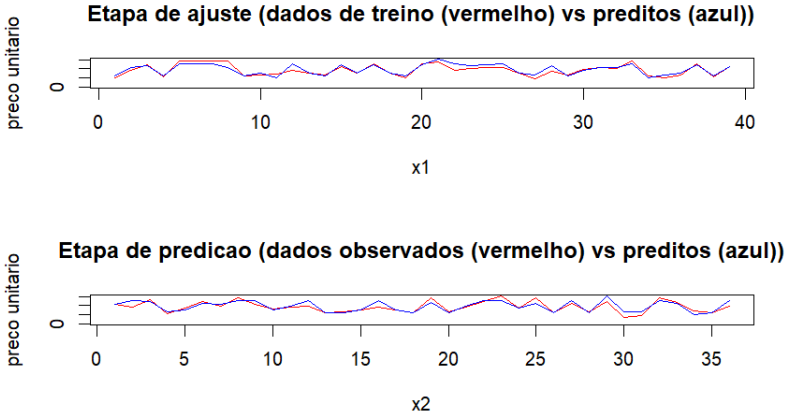
Conclui-se que para o objeto de estudo o modelo ANFIS apresenta certa resistência a *outliers*. De fato, o mercado carrega nuances que causam não linearidades que os modelos MLR não conseguem tratar e que são aprendidas pelos modelos ANFIS.

### 5.4.4 *Modelo ANFIS com outliers e com undersampling*

A quarta abordagem usando a técnica ANFIS (vide *script* anfis_modelo_50_undersampling_comOutliers.r) manteve a combinação de variáveis dos outros modelos (área privativa, número de vagas, padrão de acabamento e preço unitário) para preservar a comparação. Repartiu-se o conjunto na proporção treino 52% x teste 48%, então o conjunto de dados com 122 elementos foi reduzido para 75 elementos sendo 39 elementos para treino e 36 elementos para teste, proporção semelhante ao estudo de Król *et al* (2007).

Para a abordagem com *undersampling* e fazendo uso do conjunto não saneado, as variáveis área privativa, número de vagas, padrão de acabamento foram mantidas para explicar a variável dependente (preço unitário), visando preservar a comparação entre as alternativas da técnica ANFIS. Entre todas as proporções estudadas, aquelas com aproximadamente 50% para treino e 50% para teste resultaram nas melhores métricas. A figura 40 indica que tanto o treino quanto o teste apresentaram bons resultados.

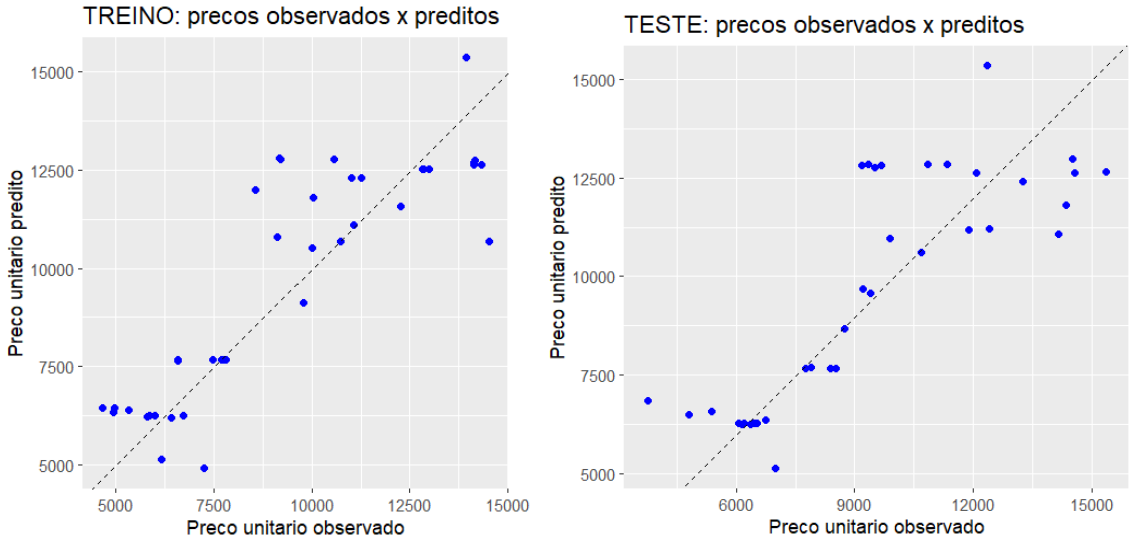
Figura 40 - ANFIS: poder de predição (ajuste e predição) por sequência (abordagem 4)




Fonte: elaborado pelo autor (2023)

A figura 41 mostra que existem duas faixas (entre R$ 8.271,46 / m² e R$ 10.891,21 / m² e entre R$ 12.955,00 / m² e R$ 13.849,90 / m²) em que não há preços preditos. Provavelmente deve-se ao fato de somente 14 regras terem sido criadas.

Figura 41 - ANFIS: poder de predição (ajuste e predição) por dispersão (abordagem4)




Fonte: elaborado pelo autor (2023)

Comparativamente com a alternativa sem *outliers* a tabela 8 indica que os resultados para o primeiro modelo ANFIS foram um pouco melhores, principalmente em relação à predições com desvios elevados. Alguns pontos não mostraram ser um ajuste ideal, principalmente para faixas maiores de preços unitários. Conclui-se que o método MLR (aplicado ao mesmo conjunto de dados) resultou em predições mais confiáveis que qualquer um dos métodos ANFIS, superando-os em todas as métricas.

Tabela 8 – ANFIS: comparativo entre proporções de treino x teste (abordagem 4)

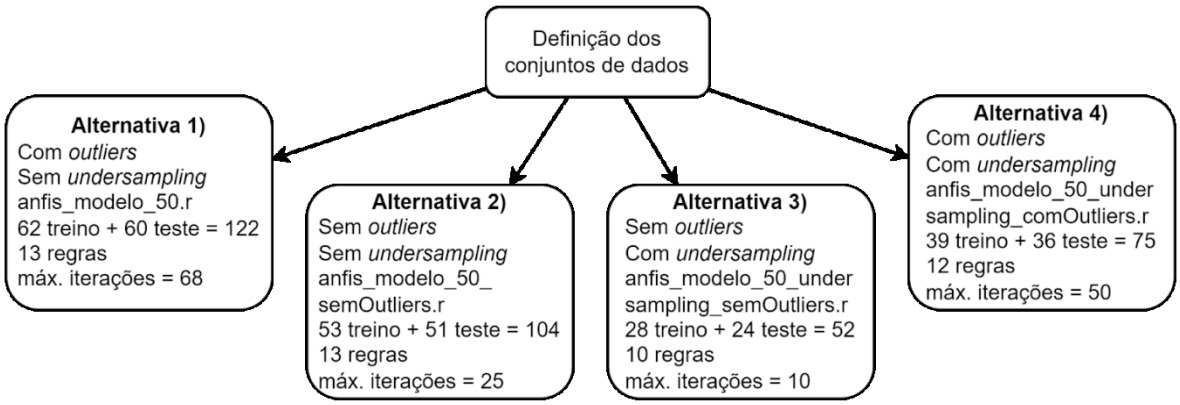
| | **Modelo ANFIS: conjunto com *outliers* e com *undersampling***| | | |
|---|---|---|---|---|
| **Divisão** | **MSE** | **MAE** | **RMSE** | **MAPE** |
| **52% x 48%** | 3.166.301,00 | 1.334,93 | 1.779,41 | 14,91% |

Fonte: elaborado pelo autor

Não é possível descartar que a evidência da aplicação do procedimento de *undersampling* sobre o conjunto de dados impacte positivamente os resultados. Embora a comparação dos resultados dos conjuntos com *outliers* entre a primeira e a quarta abordagem (respectivamente sem *undersampling* e com *undersampling*) mostre os impactos dos desvios grandes entre valores observados e preditos (vide métrica MSE), a quarta alternativa foi treinada somente com 39 dados, enquanto que a alternativa sem undersampling foi treinada com 62 dados.

A figura 42 sintetiza as abordagens trabalhadas.

Figura 42 - ANFIS: resumo das abordagens




Fonte: elaborado pelo autor (2023)

## 5.5 CONSIDERAÇÕES FINAIS DA ANÁLISE COMPARATIVA

Ao longo deste trabalho diversas alternativas foram analisadas e a tabela 9 traz a melhor alternativa de cada técnica.

Tabela 9 - Resumo dos principais resultados por técnica

| Técnica | Modelo TSK: sem *outliers* e sem *undersampling* | | | | |
|---|---|---|---|---|---|
| | Divisão | MSE | MAE | RMSE | MAPE |
| MLR (*undersampling*) | 80% x 20% | 793.825,70 | 724,24 | 890,97 | 7,39% |
| FIS (saneado) | 75% x 25% | 1.071.559,00 | 803,10 | 1.035,16 | 10,40% |
| TSK (saneado) | 72% x 28% | 1.145.532,00 | 773,09 | 1.070,30 | 8,51% |
| ANFIS (com *outliers*) | 51% x 49% | 1.712.735,00 | 1.030,55 | 1.308,72 | 13,05% |
| ANFIS (com *outliers* e *undersampling*) | 54% x 46% | 3.523.135,00 | 1.403,98 | 1.877,00 | 17,67% |

Fonte: elaborado pelo autor

O fato dos modelos MLR se destacarem não deve ser entendido como uma regra geral, então outras amostragens em outros lugares podem fazer outra técnica se sobressair sobre as demais.

A técnica FIS mostrou que a distribuição com mais elementos para treinamento se destacou, trouxe resultados aceitáveis (comparativamente com a técnica consagrada MLR) e sugeriu que a inclusão de elementos *outliers* pode prejudicar a formação das regras e consequentemente afetando os resultados, então poderia-se realizar modelos FIS incluindo *outliers* para comparar o desempenho por meio das métricas.

O modelo obtido com a técnica TSK foi aquele que mais se aproximou da técnica MLR, então por haver usado um conjunto de dados parcialmente saneado, mostrou certa resistência à *outliers*, notadamente em razão das equações não usarem todos os elementos do conjunto de dados e haver mais de uma equação que é ativada conforme o grau de pertinência ao conjunto. A análise destacou o potencial da técnica, uma vez que conservou bons resultados mesmo com equações cujo sinal dos coeficientes estava invertido.

Quanto ao ANFIS, a eliminação dos elementos *outliers* diminuiu o conjunto de dados e isso reduziu a capacidade de aprendizado. A criação automática de regras em conjuntos menores diminui a quantidade de regras e as consequências são a redução do poder de predição dos modelos e a hipótese de vários elementos do conjunto de dados ativarem as mesmas regras, predizendo valores idênticos (repetidos). Em termos de aplicabilidade da técnica ANFIS ao escopo das avaliações imobiliárias, os resultados mostraram certa resistência a *outliers*, apresentando melhores resultados para conjuntos de dados maiores e para conjuntos de dados balanceados (por *undersampling*).

Como a técnica ANFIS implementa um modelo TSK de ordem 1, situações em que muitas variáveis são necessárias para explicar a variabilidade da variável dependente, o ANFIS pode não ser uma alternativa adequada nestes casos.

Em todos os modelos os elementos com preços unitários observados mais elevados mostraram preços unitários preditos com maior dispersão. Este efeito se deve ao fato que conforme o preço do imóvel aumenta, este perde liquidez e a pressa do ofertante que o imóvel seja absorvido mais rapidamente pelo mercado (venda) é um fator que influencia sobremaneira.

# 6 CONCLUSÕES E TRABALHOS FUTUROS

Este item apresenta as principais descobertas da pesquisa, destacando qual das técnicas de inteligência computacional se mostrou mais eficiente para a avaliação imobiliária.

O objetivo deste trabalho foi analisar modelos para avaliação de imóveis do tipo apartamento utilizando técnicas de inteligência computacional. Foram utilizadas as técnicas MLR, FIS, TSK e ANFIS. Nesta ocasião ficam evidentes as possibilidades de utilização destas técnicas na prática cotidiana.

A técnica MLR apresentou os melhores resultados e esta conclusão não deve ser entendida como sendo absoluta, visto que outras amostragens sob outros contextos podem fazer com que outra técnica se sobressaia sobre a MLR. À medida que a relação entre as variáveis independentes e a variável dependente passa a carregar não linearidades importantes e o conjunto de dados traz disponíveis somente variáveis qualitativas, as técnicas derivadas da lógica difusa passam a se destacar, uma vez que a natureza destes modelos contempla a vagueza. No conjunto de dados do domínio deste problema, haviam variáveis independentes numéricas que explicavam muito da variabilidade da variável dependente, então esta foi a principal razão pela qual a técnica MLR se destacou sobre as demais. O objeto de estudo deste trabalho mostrou que, devido à representatividade do conjunto de dados, mesmo quando contiver variáveis que podem ser facilmente expressas de forma difusa (área privativa, idade e padrão de acabamento) a técnica MLR conduz a bons resultados.

A técnica FIS treinada com um conjunto de dados maior se destacou trazendo resultados aceitáveis e sugeriu que a inclusão de elementos *outliers* pode prejudicar a formação das regras e consequentemente afetando os resultados, então trabalhos futuros poderiam realizar modelos FIS incluindo *outliers* para comparar o desempenho por meio das métricas.

O modelo obtido com a técnica TSK foi aquele que mais se aproximou da técnica MLR, mesmo usando um conjunto de dados parcialmente saneado e mostrou certa resistência à *outliers*. Este fato pode ser explicado porque as equações não usam todos os elementos do conjunto de dados e de acordo com o objeto predito pode haver mais de uma equação que é ativada conforme o grau de

pertinência ao conjunto. Destaca-se o potencial da técnica, uma vez que conservou bons resultados mesmo com equações cujo sinal dos coeficientes estava invertido.

Tinha-se de antemão a expectativa que a técnica ANFIS poderia não ser uma alternativa adequada em casos como deste trabalho, onde muitas variáveis são necessárias para explicar a variabilidade da variável dependente, então como a técnica ANFIS implementa um modelo TSK de ordem 1, esta expectativa se confirmou, apresentando os resultados com maior imprecisão quando comparado com as outras técnicas.

Quanto à aceitação das técnicas derivadas da lógica difusa para a aplicação nas atividades de avaliação imobiliária, fica consolidada a necessidade de testar diversas distribuições de treino x teste e uso da técnica MLR como apoio para sanear os elementos *outliers*.

Em especial o trabalho serviu para destacar as contribuições e implicações para o campo da avaliação imobiliária sob a ótica do uso de outras técnicas além da MLR, que por vezes podem conduzir a resultados mais precisos.

Este trabalho mostrou que cada técnica tem suas particularidades e para o caso específico deste escopo (apartamentos no bairro Trindade), o uso de qualquer das técnicas isoladamente e sem a abordagem treino x teste seguida do cálculo de métricas conduz a modelos que podem conter erros importantes, vindo a comprometer os resultados.

Visando aprofundar o conhecimento sobre o assunto, sob a finalidade de sugerir pesquisas futuras, pode-se expandir este trabalho incluindo outras técnicas, como as técnicas bayesianas, as análises multicritério, redes neurais artificiais, mínimos quadrados difusos, modelos baseados em árvore, o emprego de modelos lineares generalizados e até mesmo a aplicação das técnicas para estudos em outros mercados e outras tipologias de imóveis. Outro comparativo válido é um estudo variando as funções de pertinência para averiguar o impacto no desempenho dos modelos por meio da análise das métricas. Ainda, é possível aplicar a técnica de validação cruzada no conjunto de dados para avaliar a capacidade de generalização do modelo.

Por fim, espera-se que este trabalho traga subsídios para estimular futuras pesquisas na área de avaliação imobiliária e para direcionar aplicações práticas na área.

# REFERÊNCIAS


ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **NBR 12.721**: Avaliação de custos de construção para incorporação imobiliária e outras disposições para condomínios edilícios. Rio de Janeiro, 2006.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **NBR 14.653-1**: Avaliação de bens - procedimentos gerais. Rio de Janeiro, 2019.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **NBR 14.653-2**: Avaliação de bens - imóveis urbanos. Rio de Janeiro, 2011.

BOLFARINE, H., SANDOVAL, M. J.; **Introdução à Inferência Estatística**. Notas de aula, 2000.

**BROGNOLI**. Disponível em: www.brognoli.com.br. Acesso em: fev. 2023.

BUSSAB, WO; MORETTIN, PA. **Estatística Básica**. 9ª ed. São Paulo: Editora Saraiva, 2017.

CAIXA. **Coletânea de artigos de avaliação de imóveis**. Brasília: CAIXA, 2018. 125p.

ÇETINKAYA-RUNDEL, M. **Introduction to Probability and Data with R**. Coursera, 2022. Disponível em: https://www.coursera.org/learn/probability-intro. Acesso em: 16 de abril de 2022.

ÇETINKAYA-RUNDEL, M. **Linear Regression and Modeling**. Coursera, 2022. Disponível em: https://www.coursera.org/learn/probability-intro. Acesso em: 16 de abril de 2022.

ÇETINKAYA-RUNDEL, M.; DIEZ, D. M.; BARR, C. D. **OpenIntro Statistics**. 3. ed. OpenIntro, 2019.

ÇETINKAYA-RUNDEL, M. **Statistical Inference**. Coursera, 2022. Disponível em: https://www.coursera.org/learn/probability-intro. Acesso em: 16 de abril de 2022.

CORDEIRO, L. G. T. **Geração automática de questões através de análise de texto.** 2016. 148 p. Trabalho de Conclusão de Curso (Graduação) - Universidade Federal de Santa Catarina, Florianópolis (SC), 2016.

CORRÊA, C. E. **Método para aumento de interpretabilidade em modelos de Takagi-Sugeno no sistema INFGMN**. 2020. 103 p. Trabalho de Conclusão de Curso (Graduação) - Universidade Federal de Santa Catarina, Florianópolis (SC), 2020.

FERNANDES, A. R.; MOREIRA, D. S.; SILVA, R. S. **Avaliação de imóveis utilizando análise multicritério e redes neurais artificiais**. In: XXX ENCONTRO

NACIONAL DE ENGENHARIA DE PRODUÇÃO, São Carlos, SP, Brasil, 12 a 15 de outubro de 2010.

FERNANDES, A. R.; MOREIRA, D. S.; SILVA, R. S. **Engenharia de avaliações de imóveis apoiada em técnicas de análise multicritério e redes neurais artificiais**. In: Trilha Estudantil: Revista de Sistemas de Informação da FSMA, Rio de Janeiro, n. 6, p. 49 – 58, 2010.

FERNANDES, A. R.; MOREIRA, D. S.; SILVA, R. S. **Utilização de uma rede neural artificial e análise multicritério para determinação do valor de aluguel de apartamentos**. In: VII Simpósio de Excelência em Gestão e Tecnologia, 15 p. 2010, Resende (RJ). Anais. São José: UNIVALI.

**GLMNet: Lasso and Elastic-Net Regularized Generalized Linear Models**. Stanford University, mar/2023. Disponível em: https://glmnet.stanford.edu. Disponível em: https://cran.r-project.org/web/packages/glmnet/index.html Acesso em: 05 fev 2023.

INTERNATIONAL VALUATION STANDARDS COUNCIL (IVSC). **Normas IVS: Normas Internacionais de Avaliação**. Traduzido por Carlos Eduardo Cardoso. 2. ed. São Paulo: IBAPE, 2012. ISBN 978-0-9569313-0-6.

JANG, J.-S. R. **ANFIS: Adaptive-Network-Based Fuzzy Inference System**. IEEE Transactions on Systems, Man, and Cybernetics, v. 23, n. 3, p. 665-685, mai. 1993.

KAMIRE, A.; CHAPHALKAR, N.; SANDBHOR, S. **Real Property Value Prediction Capability Using Fuzzy Logic and ANFIS**. In: International Conference on Smart Data Intelligence (ICSMDI 2021).

KRÓL, D.; LASOTA, T.; TRAWINSKI, B.; TRAWINSKI, K. **Comparison of Mamdani and TSK Fuzzy Models for Real Estate Appraisal**. In: Knowledge-Based Intelligent Information and Engineering Systems, vol. 4694. Berlin, Heidelberg: Springer, 2007. Disponível em: https://link.springer.com/chapter/10.1007/978-3-540-74829-8_123. Acesso em: 12/janeiro/2023.

LUGER, G. F. **Inteligência artificial - estruturas e estratégias para a solução de problemas complexos.** 4.ed. Porto Alegre: Bookman, 2004.

NASSAR, S. M. **Sistemas Especialistas Difusos.** Notas de aula, Curso de Pós Graduação em Ciências da Computação. Universidade Federal de Santa Catarina, 2018.

NORVIG, P.; RUSSEL, S. **Inteligência artificial.** 2. ed. Rio de Janeiro: Elsevier, 2004.

PELLI, A. **Avaliação de imóveis urbanos com utilização de sistemas nebulosos (redes neuro-fuzzy) e de redes neurais artificiais.** In: XXI Congreso Panamericano de Valuación Cartagena, 18 p. 2004, Colômbia.

R CORE TEAM. **R: A Language and Environment for Statistical Computing**, R Foundation for Statistical Computing, Vienna, Austria. Disponível em: https://www.R-project.org. Acesso em: 12 jan, 2023.

RIZA, L. S.; BERGMEIR, C.; HERRERA, F.; BENITEZ, J. M. **frbs: Fuzzy Rule-Based Systems for Classification and Regression Tasks**. 3.2-0. Disponível em: https://cran.r-project.org/package=frbs. Acesso em: 11 maio 2023.

ROSS, T. J. **Fuzzy Logic with Engineering Applications**. 4 ed. , 2017.
**R STUDIO: ambiente integrado de desenvolvimento para R**. Boston, MA: RStudio, PBC Posit Software, 2022. Disponível em: https://www.rstudio.com/. Acesso em: 12/janeiro/2023.

SANTELLO, R. **Avaliação de imóveis urbanos com utilização da lógica difusa -** Dissertação. Programa de Pós Graduação em Engenharia de Produção. Florianópolis: UFSC, 2004. 149 p.

SANTOS, E. R. **Sistemas inteligentes**. Notas de aula, Curso de Graduação em Sistemas de Informação. Universidade Federal de Santa Catarina, 2020.

SARIP, A. G., HAFEZ, M. B., DAUD, M. N.. **Application Of Fuzzy Regression Model For Real Estate Price Prediction**. Malaysian Journal of Computer Science, v. 29, n. 1, p. 15-27, 13 p, mar. 2016. Malásia.

THOFEHRN, R. **Avaliação de terrenos urbanos por fórmulas matemáticas.** 2. ed. São Paulo: PINI, 2008.

TOGGWEILER, J. G.; MARQUES, L. S. **Automação residencial para conservação e eficiência energética por meio de técnicas de inteligência artificial**. 2017. 122 p. Trabalho de Conclusão de Curso (Graduação) - Universidade Federal de Santa Catarina, Florianópolis (SC), 2017.

VALLE, M. **Sistema Tipo Mamdani e Takagi-Sugeno**. Apresentação para o curso de pós graduação, disciplina de lógica difusa na UNESP, 09 set. 2021. Disponível em: https://www.youtube.com/watch?v=T-CupbBcK-M. Acesso em: 10 fev. 2021.

**VIVAREAL**. Disponível em: www.vivareal.com.br. Acesso em: fev. 2023.

YALPIR, S.; OZKAN, G. **Knowledge-based FIS and ANFIS models development and comparison for residential real estate valuation**. International Journal of Strategic Property Management, v. 22, p. 110-118, 2018. DOI: 10.3846/ijspm.2018.442.

WANG, L. X.; MENDEL, J. M. . **Generating Fuzzy Rules by Learning from Examples**. IEEE Transactions on Systems, Man, and Cybernetics, ENova Jersey, n. 6, p. 1414 - 1428, 1992.

| ANEXO A: | endereco | nomeCondominio | nDorm | nSuites | nBWCs | nVagas | padraoAcab | aPriva | ano | idade | preco | latitude_m_S | longitude_m_S | link | data |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| EI_01 | Travessa Antenor Cardos | Absolut Residence | 2 | 2 | 3 | 1 | 3 | 75 | 2018 | 5 | R$ 750.000,00 | 6945367,90 | 744781,74 | https://www.vivareal.com.br/imovel/apartamento-2-q | 23FEV23 |
| EI_02 | Travessa Antenor Cardos | Absolut Residence | 2 | 2 | 3 | 1 | 3 | 74 | 2018 | 5 | R$ 795.000,00 | 6945367,90 | 744781,74 | https://www.vivareal.com.br/imovel/apartamento-2-q | 23FEV23 |
| EI_03 | Rua Antenor Cardoso da | Belvedere | 2 | 2 | 3 | 1 | 3 | 74,53 | 2005 | 18 | R$ 795.000,00 | 6945375,11 | 744776,06 | https://www.brognoli.com.br/imovel/comprar-florianop | 16MAR23 |
| EI_04 | Rua Lauro Linhares, 689 | Granville | 1 | 0 | 1 | 1 | 1 | 47 | 1984 | 39 | R$ 290.000,00 | 6946287,58 | 744361,39 | https://www.vivareal.com.br/imovel/apartamento-1-q | 24FEV23 |
| EI_05 | Rua Lauro Linhares, 689 | Granville | 1 | 0 | 1 | 1 | 1 | 47 | 1984 | 39 | R$ 300.000,00 | 6946287,58 | 744361,39 | https://www.vivareal.com.br/imovel/apartamento-1-q | 24FEV23 |
| EI_06 | Rua Lauro Linhares, 689 | Granville | 1 | 0 | 1 | 1 | 1 | 47 | 1984 | 39 | R$ 319.000,00 | 6946287,58 | 744361,39 | https://www.vivareal.com.br/imovel/apartamento-1-q | 24FEV23 |
| EI_07 | Rua Lauro Linhares, 689 | Granville | 1 | 0 | 1 | 1 | 1 | 47 | 1984 | 39 | R$ 329.000,00 | 6946287,58 | 744361,39 | https://www.vivareal.com.br/imovel/apartamento-1-q | 24FEV23 |
| EI_08 | Rua Lauro Linhares, 689 | Granville | 1 | 0 | 1 | 1 | 1 | 50 | 1984 | 39 | R$ 335.000,00 | 6946287,58 | 744361,39 | https://www.vivareal.com.br/imovel/apartamento-1-q | 24FEV23 |
| EI_09 | Rua Lauro Linhares, 689 | Granville | 2 | 0 | 1 | 1 | 1 | 65 | 1984 | 39 | R$ 395.000,00 | 6946287,58 | 744361,39 | https://www.vivareal.com.br/imovel/apartamento-2-q | 13MAR23 |
| EI_10 | Rua Lauro Linhares, 689 | Granville | 2 | 0 | 1 | 1 | 1 | 66 | 1984 | 39 | R$ 410.000,00 | 6946287,58 | 744361,39 | https://www.vivareal.com.br/imovel/apartamento-2-q | 12MAR23 |
| EI_11 | Rua Lauro Linhares, 689 | Granville | 2 | 0 | 1 | 1 | 1 | 65 | 1984 | 39 | R$ 420.000,00 | 6946287,58 | 744361,39 | https://www.vivareal.com.br/imovel/apartamento-2-q | 14MAR23 |
| EI_12 | Rua Lauro Linhares, 829 | Granville | 3 | 0 | 2 | 1 | 1 | 77 | 1984 | 39 | R$ 490.000,00 | 6946287,58 | 744361,39 | https://www.vivareal.com.br/imovel/apartamento-3-q | 12MAR23 |
| EI_13 | Rua Lauro Linhares, 829 | Granville | 3 | 0 | 2 | 1 | 1 | 78 | 1984 | 39 | R$ 530.000,00 | 6946287,58 | 744361,39 | https://www.vivareal.com.br/imovel/apartamento-3-q | 05MAR23 |
| EI_14 | Rua Lauro Linhares, 897 | Porto da Trindade | 3 | 1 | 2 | 1 | 2 | 85 | 2006 | 17 | R$ 670.000,00 | 6946093,05 | 744394,38 | https://www.vivareal.com.br/imovel/apartamento-3-q | 24FEV23 |
| EI_15 | Rua Lauro Linhares, 897 | Porto da Trindade | 3 | 1 | 2 | 1 | 2 | 85,7 | 2006 | 17 | R$ 670.000,00 | 6946093,05 | 744394,38 | https://www.vivareal.com.br/imovel/apartamento-3-q | 24FEV23 |
| EI_16 | Rua Lauro Linhares, 897 | Porto da Trindade | 2 | 1 | 2 | 1 | 2 | 72 | 2006 | 17 | R$ 628.536,00 | 6946093,05 | 744394,38 | https://www.vivareal.com.br/imovel/apartamento-2-q | 24FEV23 |
| EI_17 | Rua Lauro Linhares, 897 | Porto da Trindade | 3 | 1 | 2 | 1 | 2 | 87 | 2006 | 17 | R$ 650.000,00 | 6946093,05 | 744394,38 | https://www.vivareal.com.br/imovel/apartamento-3-q | 12MAR23 |
| EI_18 | Rua Lauro Linhares, 897 | Porto da Trindade | 3 | 1 | 2 | 1 | 2 | 86 | 2006 | 17 | R$ 668.000,00 | 6946093,05 | 744394,38 | https://www.vivareal.com.br/imovel/apartamento-3-q | 24FEV23 |
| EI_19 | Rua Lauro Linhares, 897 | Porto da Trindade | 3 | 1 | 2 | 1 | 2 | 86 | 2006 | 17 | R$ 669.000,00 | 6946093,05 | 744394,38 | https://www.vivareal.com.br/imovel/apartamento-3-q | 27FEV23 |
| EI_20 | Rua Alba Dias Cunha, 19 | Portal da Trindade | 3 | 1 | 2 | 1 | 3 | 90 | 2009 | 12 | R$ 710.000,00 | 6946779,56 | 744244,51 | https://www.vivareal.com.br/imovel/apartamento-3-q | 24FEV23 |
| EI_21 | Rua Douglas Seabra Levi | Belluno | 3 | 1 | 2 | 1 | 2 | 83 | 1994 | 29 | R$ 470.000,00 | 6945172,1 | 744250,42 | https://www.vivareal.com.br/imovel/apartamento-3-q | 27FEV23 |
| EI_22 | Rua Douglas Seabra Levi | Luciana | 1 | 0 | 1 | 1 | 1 | 30 | 1994 | 29 | R$ 230.000,00 | 6945009,43 | 744386,03 | https://www.vivareal.com.br/imovel/kitnet-1-quartos-t | 12MAR23 |
| EI_23 | Rua Douglas Seabra Levi | Luciana | 1 | 0 | 1 | 1 | 1 | 28,12 | 1994 | 29 | R$ 310.000,00 | 6945009,43 | 744386,03 | https://www.attria.com.br/imovel/apartamento-a-venc | 12MAR23 |
| EI_24 | Rua Douglas Seabra Levi | Rosana | 2 | 0 | 1 | 1 | 1 | 75 | 1993 | 30 | R$ 400.000,00 | 6945104,07 | 744343,51 | https://www.vivareal.com.br/imovel/apartamento-2-q | 12MAR23 |
| EI_25 | Rua Professor Milton Roq | Das Américas | 2 | 1 | 2 | 1 | 3 | 67,72 | 2011 | 12 | R$ 680.000,00 | 6946506,95 | 744377,14 | https://www.vivareal.com.br/imovel/apartamento-2-q | 27FEV23 |
| EI_26 | Rua Professor Milton Roq | Sanford | 3 | 1 | 2 | 2 | 1 | 70 | 1989 | 34 | R$ 599.000,00 | 6946531,06 | 744440,56 | https://www.vivareal.com.br/imovel/apartamento-3-q | 12MAR23 |
| EI_27 | Rua Lauro Linhares, 635 | Arquipélago | 2 | 0 | 1 | 1 | 1 | 85 | 1990 | 33 | R$ 320.000,00 | 6946330,22 | 744356,87 | https://www.vivareal.com.br/imovel/apartamento-2-q | 12MAR23 |
| EI_28 | Rua Lauro Linhares, 635 | Arquipélago | 2 | 0 | 1 | 1 | 1 | 45 | 1990 | 33 | R$ 325.741,00 | 6946330,22 | 744356,87 | https://www.vivareal.com.br/imovel/apartamento-2-q | 12MAR23 |
| EI_29 | Rua Lauro Linhares, 635 | Arquipélago | 3 | 0 | 1 | 1 | 1 | 65 | 1990 | 33 | R$ 380.000,00 | 6946330,22 | 744356,87 | https://www.vivareal.com.br/imovel/apartamento-3-q | 12MAR23 |
| EI_30 | Rua Lauro Linhares, 635 | Arquipélago | 2 | 0 | 1 | 1 | 1 | 62 | 1990 | 33 | R$ 395.000,00 | 6946330,22 | 744356,87 | https://www.vivareal.com.br/imovel/apartamento-2-q | 12MAR23 |
| EI_31 | Rua Lauro Linhares, 635 | Arquipélago | 3 | 0 | 2 | 1 | 1 | 80 | 1990 | 33 | R$ 430.000,00 | 6946330,22 | 744356,87 | https://www.vivareal.com.br/imovel/apartamento-3-q | 12MAR23 |
| EI_32 | Rua Lauro Linhares, 635 | Arquipélago | 2 | 0 | 1 | 1 | 1 | 49 | 1990 | 33 | R$ 330.000,00 | 6946330,22 | 744356,87 | https://www.vivareal.com.br/imovel/apartamento-2-q | 12MAR23 |
| EI_33 | Rua Lauro Linhares, 359 | Lauro Linhares | 3 | 0 | 1 | 1 | 1 | 66 | 1989 | 34 | R$ 400.000,00 | 6945113,05 | 744722,65 | https://www.vivareal.com.br/imovel/apartamento-3-q | 12MAR23 |
| EI_34 | Rua Lauro Linhares, 359 | Lauro Linhares | 3 | 0 | 1 | 1 | 1 | 64 | 1989 | 34 | R$ 390.000,00 | 6945113,05 | 744722,65 | https://www.vivareal.com.br/imovel/apartamento-3-q | 12MAR23 |
| EI_35 | Ogê Fortkamp, 119 | Ogê Studios | 2 | 2 | 3 | 1 | 3 | 60,52 | 2023 | 1 | R$ 554.696,00 | 6945293,45 | 744489,01 | https://www.vivareal.com.br/imovel/apartamento-2-q | 05MAR23 |
| EI_36 | Ogê Fortkamp, 119 | Ogê Studios | 2 | 2 | 3 | 1 | 3 | 61 | 2023 | 1 | R$ 562.000,00 | 6945293,45 | 744489,01 | https://www.vivareal.com.br/imovel/apartamento-2-q | 05MAR23 |
| EI_37 | Ogê Fortkamp, 119 | Ogê Studios | 2 | 2 | 2 | 1 | 3 | 60,52 | 2023 | 1 | R$ 554.697,00 | 6945293,45 | 744489,01 | https://www.vivareal.com.br/imovel/apartamento-2-q | 05MAR23 |
| EI_38 | Ogê Fortkamp, 119 | Ogê Studios | 2 | 1 | 2 | 1 | 3 | 60,52 | 2023 | 1 | R$ 584.697,00 | 6945293,45 | 744489,01 | https://www.vivareal.com.br/imovel/apartamento-2-q | 05MAR23 |
| EI_39 | Ogê Fortkamp, 119 | Ogê Studios | 2 | 1 | 2 | 1 | 3 | 60,52 | 2023 | 1 | R$ 514.897,00 | 6945293,45 | 744489,01 | https://www.vivareal.com.br/imovel/apartamento-2-q | 05MAR23 |
| EI_40 | Ogê Fortkamp, 119 | Ogê Studios | 1 | 0 | 1 | 1 | 3 | 41 | 2023 | 1 | R$ 542.550,00 | 6945293,45 | 744489,01 | https://www.vivareal.com.br/imovel/apartamento-1-q | 12MAR23 |
| EI_41 | Rua Procópio Manoel Pire | Ilha de Majorca | 2 | 0 | 1 | 1 | 2 | 63 | 1985 | 38 | R$ 580.000,00 | 6946442,81 | 744384,82 | https://www.vivareal.com.br/imovel/apartamento-2-q | 15MAR23 |
| EI_42 | Rua Procópio Manoel Pire | Acquafredda | 2 | 1 | 2 | 1 | 2 | 65 | 2012 | 11 | R$ 530.000,00 | 6946488,92 | 744434,11 | https://www.vivareal.com.br/imovel/apartamento-2-q | 05MAR23 |
| EI_43 | Rua Lauro Linhares, 1520 | Dona Augusta | 3 | 1 | 2 | 1 | 2 | 89,65 | 2002 | 21 | R$ 600.000,00 | 6945524,66 | 744628,45 | https://www.brognoli.com.br/imovel/comprar-florianop | 16MAR23 |
| EI_44 | Rua Lauro Linhares, 1520 | Dona Augusta | 3 | 1 | 2 | 1 | 2 | 89 | 2002 | 21 | R$ 759.000,00 | 6945524,66 | 744628,45 | https://www.vivareal.com.br/imovel/apartamento-3-q | 05MAR23 |
| EI_45 | Rua Lauro Linhares, 1730 | Trindade | 2 | 0 | 1 | 1 | 1 | 64 | 1988 | 35 | R$ 385.000,00 | 6945305,25 | 744656,54 | https://www.vivareal.com.br/imovel/apartamento-2-q | 12MAR23 |
| EI_46 | Rua Lauro Linhares, 1730 | Trindade | 2 | 0 | 1 | 1 | 1 | 65 | 1988 | 35 | R$ 410.000,00 | 6945305,25 | 744656,54 | https://www.vivareal.com.br/imovel/apartamento-2-q | 14MAR23 |
| EI_47 | Rua João de Deus Macha | La Belle | 1 | 0 | 1 | 1 | 3 | 31,45 | 2023 | 1 | R$ 482.000,00 | 6945171,18 | 744961,58 | https://www.brognoli.com.br/imovel/comprar-florianop | 16MAR23 |
| EI_48 | Rua João de Deus Macha | La Belle | 2 | 1 | 2 | 1 | 3 | 71,06 | 2023 | 1 | R$ 845.000,00 | 6945171,18 | 744961,58 | https://www.brognoli.com.br/imovel/comprar-florianop | 16MAR23 |
| EI_49 | Rua João de Deus Macha | La Belle | 2 | 1 | 2 | 1 | 3 | 71 | 2023 | 1 | R$ 880.000,00 | 6945171,18 | 744961,58 | https://www.vivareal.com.br/imovel/apartamento-2-q | 05MAR23 |
| EI_50 | Rua João de Deus Macha | Montreal | 2 | 1 | 2 | 1 | 3 | 69 | 2015 | 8 | R$ 845.514,00 | 6945409,40 | 744781,94 | https://www.vivareal.com.br/imovel/apartamento-2-q | 05MAR23 |
| EI_51 | Rua João de Deus Macha | Savoya | 3 | 1 | 2 | 2 | 2 | 91 | 2008 | 15 | R$ 830.000,00 | 6945379,91 | 744822,66 | https://www.vivareal.com.br/imovel/apartamento-3-q | 05MAR23 |
| EI_52 | Rua Professor José Brasil | Trinità | 2 | 1 | 2 | 1 | 3 | 85 | 2019 | 4 | R$ 799.000,00 | 6945291,47 | 744790,32 | https://www.vivareal.com.br/imovel/apartamento-2-q | 05MAR23 |
| EI_53 | Rua Juvêncio Costa,97 | Ripresa | 1 | 0 | 1 | 1 | 3 | 31 | 2017 | 5 | R$ 450.000,00 | 6946046,37 | 744535,90 | https://www.vivareal.com.br/imovel/apartamento-1-q | 05MAR23 |
| EI_54 | Rua Salomé Damazio Jac | Real Trindade | 3 | 1 | 2 | 1 | 2 | 96 | 1996 | 27 | R$ 765.000,00 | 6945469,63 | 744698,23 | https://www.vivareal.com.br/imovel/apartamento-3-q | 05MAR23 |
| EI_55 | Rua Salomé Damazio Jac | Real Trindade | 3 | 1 | 2 | 1 | 2 | 89,3 | 1996 | 27 | R$ 750.000,00 | 6945469,63 | 744698,23 | https://www.brognoli.com.br/imovel/comprar-florianop | 16MAR23 |
| EI_56 | Praça Santos Dumont, 16 | Santos Dumont | 3 | 1 | 2 | 1 | 1 | 98 | 2001 | 22 | R$ 565.000,00 | 6945074,88 | 744617,58 | https://www.vivareal.com.br/imovel/apartamento-3-q | 05MAR23 |
| EI_57 | Avenida Madre Benvenu | Plaza du Soleil | 3 | 1 | 3 | 2 | 3 | 95,45 | 2010 | 13 | R$ 1.350.000,00 | 6945502,69 | 744974,33 | https://www.vivareal.com.br/imovel/apartamento-2-q | 05MAR23 |
| EI_58 | Avenida Madre Benvenu | Plaza du Soleil | 3 | 1 | 2 | 2 | 3 | 92 | 2010 | 13 | R$ 1.300.000,00 | 6945502,69 | 744974,33 | https://www.vivareal.com.br/imovel/apartamento-3-q | 05MAR23 |
| EI_59 | Avenida Madre Benvenu | Plaza du Soleil | 3 | 1 | 2 | 2 | 3 | 96 | 2010 | 13 | R$ 950.000,00 | 6945502,69 | 744974,33 | https://www.vivareal.com.br/imovel/apartamento-3-q | 05MAR23 |
| EI_60 | Avenida Madre Benvenu | Plaza du Soleil | 3 | 1 | 2 | 2 | 3 | 84 | 2010 | 13 | R$ 1.190.000,00 | 6945502,69 | 744974,33 | https://www.vivareal.com.br/imovel/apartamento-3-q | 05MAR23 |
| EI_61 | Avenida Madre Benvenu | Plaza du Soleil | 3 | 2 | 2 | 2 | 3 | 82 | 2010 | 13 | R$ 1.190.000,00 | 6945502,69 | 744974,33 | https://www.vivareal.com.br/imovel/apartamento-3-q | 05MAR23 |
| EI_62 | Avenida Madre Benvenu | Plaza du Soleil | 3 | 1 | 2 | 2 | 3 | 88 | 2010 | 13 | R$ 1.280.000,00 | 6945502,69 | 744974,33 | https://www.vivareal.com.br/imovel/apartamento-3-q | 12MAR23 |
| EI_63 | Avenida Madre Benvenu | Plaza du Soleil | 3 | 1 | 2 | 2 | 3 | 86 | 2010 | 13 | R$ 1.320.000,00 | 6945502,69 | 744974,33 | https://www.vivareal.com.br/imovel/apartamento-3-q | 15MAR23 |
| EI_64 | Avenida Madre Benvenu | Plaza di Mônaco | 2 | 1 | 2 | 2 | 3 | 68,85 | 2010 | 13 | R$ 850.000,00 | 6945502,69 | 744974,33 | https://www.vivareal.com.br/imovel/apartamento-2-q | 05MAR23 |
| EI_65 | Avenida Madre Benvenu | São Conrado | 3 | 0 | 1 | 1 | 1 | 67 | 1999 | 24 | R$ 425.000,00 | 6945548,24 | 745047,44 | https://www.vivareal.com.br/imovel/apartamento-3-q | 05MAR23 |
| EI_66 | Avenida Madre Benvenu | São Conrado | 3 | 0 | 1 | 1 | 1 | 66 | 1999 | 24 | R$ 430.000,00 | 6945548,24 | 745047,44 | https://www.vivareal.com.br/imovel/apartamento-3-q | 05MAR23 |
| EI_67 | Rua Prof Maria Flora Pau | Viva Trindade | 1 | 0 | 1 | 1 | 3 | 37 | 2023 | 1 | R$ 530.547,00 | 6945005,83 | 744858,24 | https://www.vivareal.com.br/imovel/apartamento-1-q | 05MAR23 |

| ANEXO A: | endereco | nomeCondominio | nDorm | nSuites | nBWCs | nVagas | padraoAcab | aPriva | ano | idade | preco | latitude_m_S | longitude_m_S | link | data |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| EI_68 | Rua Prof Maria Flora Pau | Viva Trindade | 1 | 0 | 1 | 1 | 3 | 33 | 2023 | 1 | R$ 466.417,00 | 6945005,83 | 744858,24 | https://www.vivareal.com.br/imovel/apartamento-1-q | 05MAR23 |
| EI_69 | Rua Prof Maria Flora Pau | Dom Afonso | 3 | 1 | 3 | 1 | 2 | 100,6 | 1988 | 35 | R$ 700.000,00 | 6944976,42 | 744828,36 | https://www.brognoli.com.br/imovel/comprar-florianop | 16MAR23 |
| EI_70 | Rua Luiz Oscar de Carval | Itambé | 2 | 0 | 1 | 1 | 1 | 52 | 1983 | 40 | R$ 350.000,00 | 6945873,67 | 744764,53 | https://www.vivareal.com.br/imovel/apartamento-2-q | 12MAR23 |
| EI_71 | Rua Luiz Oscar de Carval | Itambé | 2 | 0 | 1 | 1 | 1 | 56 | 1983 | 40 | R$ 350.000,00 | 6945873,67 | 744764,53 | https://www.vivareal.com.br/imovel/apartamento-2-q | 15MAR23 |
| EI_72 | Rua Luiz Oscar de Carval | Itambé | 2 | 0 | 1 | 1 | 1 | 53 | 1983 | 40 | R$ 357.000,00 | 6945873,67 | 744764,53 | https://www.vivareal.com.br/imovel/apartamento-2-q | 15MAR23 |
| EI_73 | Rua Luiz Oscar de Carval | Itambé | 3 | 0 | 1 | 2 | 1 | 65 | 1983 | 40 | R$ 450.000,00 | 6945873,67 | 744764,53 | https://www.vivareal.com.br/imovel/apartamento-3-q | 05MAR23 |
| EI_74 | Rua Luiz Oscar de Carval | Itambé | 3 | 0 | 1 | 1 | 1 | 80 | 1983 | 40 | R$ 450.000,00 | 6945873,67 | 744764,53 | https://www.vivareal.com.br/imovel/apartamento-3-q | 05MAR23 |
| EI_75 | Rua Luiz Oscar de Carval | Itambé | 3 | 0 | 1 | 1 | 1 | 65 | 1983 | 40 | R$ 398.000,00 | 6945873,67 | 744764,53 | https://www.vivareal.com.br/imovel/apartamento-3-q | 12MAR23 |
| EI_76 | Rua Luiz Oscar de Carval | Itambé | 3 | 0 | 1 | 1 | 1 | 65 | 1983 | 40 | R$ 420.000,00 | 6945873,67 | 744764,53 | https://www.vivareal.com.br/imovel/apartamento-3-q | 12MAR23 |
| EI_77 | Rua Luiz Oscar de Carval | Itambé | 3 | 0 | 1 | 1 | 1 | 68 | 1983 | 40 | R$ 419.000,00 | 6945873,67 | 744764,53 | https://www.vivareal.com.br/imovel/apartamento-3-q | 12MAR23 |
| EI_78 | Rua Luiz Oscar de Carval | Itambé | 3 | 0 | 1 | 1 | 1 | 64 | 1983 | 40 | R$ 398.000,00 | 6945873,67 | 744764,53 | https://www.vivareal.com.br/imovel/apartamento-3-q | 12MAR23 |
| EI_79 | Rua Luiz Oscar de Carval | Itambé | 3 | 0 | 1 | 1 | 1 | 65 | 1983 | 40 | R$ 425.000,00 | 6945873,67 | 744764,53 | https://www.vivareal.com.br/imovel/apartamento-3-q | 12MAR23 |
| EI_80 | Rua Luiz Oscar de Carval | Itambé | 3 | 0 | 1 | 1 | 1 | 65 | 1983 | 40 | R$ 390.000,00 | 6945873,67 | 744764,53 | https://www.vivareal.com.br/imovel/apartamento-3-q | 12MAR23 |
| EI_81 | Rua Luiz Oscar de Carval | Verde Mar | 2 | 0 | 1 | 1 | 1 | 57 | 1983 | 40 | R$ 300.000,00 | 6945780,46 | 744856,17 | https://www.vivareal.com.br/imovel/apartamento-2-q | 12MAR23 |
| EI_82 | Rua Luiz Oscar de Carval | Solar Santa Paula | 3 | 0 | 1 | 1 | 1 | 65 | 1983 | 40 | R$ 390.000,00 | 6945744,33 | 744859,01 | https://www.vivareal.com.br/imovel/apartamento-3-q | 12MAR23 |
| EI_83 | Rua Luiz Oscar de Carval | Solar Santa Paula | 3 | 0 | 1 | 1 | 1 | 80 | 1983 | 40 | R$ 415.000,00 | 6945744,33 | 744859,01 | https://www.vivareal.com.br/imovel/apartamento-3-q | 12MAR23 |
| EI_84 | Rua Luiz Oscar de Carval | Solar Santa Paula | 3 | 0 | 1 | 1 | 1 | 75 | 1983 | 40 | R$ 430.000,00 | 6945744,33 | 744859,01 | https://www.vivareal.com.br/imovel/apartamento-3-q | 12MAR23 |
| EI_85 | Rua Luiz Oscar de Carval | Solar Santa Paula | 3 | 0 | 1 | 1 | 1 | 75 | 1983 | 40 | R$ 410.000,00 | 6945744,33 | 744859,01 | https://www.vivareal.com.br/imovel/apartamento-3-q | 12MAR23 |
| EI_86 | Rua Luiz Oscar de Carval | Solar Santa Paula | 3 | 0 | 1 | 1 | 1 | 75 | 1983 | 40 | R$ 405.000,00 | 6945744,33 | 744859,01 | https://www.vivareal.com.br/imovel/apartamento-3-q | 12MAR23 |
| EI_87 | Rua Sergio Lopes Falcao | Ilha do Sol | 2 | 1 | 2 | 1 | 2 | 68,36 | 2003 | 20 | R$ 670.000,00 | 6945733,11 | 744722,64 | https://www.brognoli.com.br/imovel/comprar-florianop | 16MAR23 |
| EI_88 | Rua Lauro Linhares, 970 | Império do Sol | 2 | 0 | 1 | 1 | 1 | 64 | 2007 | 16 | R$ 430.000,00 | 6946034,85 | 744437,08 | https://www.vivareal.com.br/imovel/apartamento-2-q | 14MAR23 |
| EI_89 | Rua Lauro Linhares, 970 | Império do Sol | 2 | 1 | 2 | 1 | 2 | 144,63 | 2007 | 16 | R$ 950.000,00 | 6946034,85 | 744437,08 | https://balzerimoveis.com.br/imovel/457/apartamento | 12MAR23 |
| EI_90 | Rua Lauro Linhares, 970 | Império do Sol | 2 | 1 | 3 | 2 | 2 | 144 | 2007 | 16 | R$ 953.741,00 | 6946034,85 | 744437,08 | https://www.vivareal.com.br/imovel/cobertura-2-quar | 05MAR23 |
| EI_91 | Rua Lauro Linhares, 1390 | Rodrik Residence | 1 | 0 | 1 | 1 | 3 | 42,15 | 2018 | 5 | R$ 540.000,00 | 6945637,39 | 744611,54 | https://www.vivareal.com.br/imovel/apartamento-1-q | 15MAR23 |
| EI_92 | Rua Lauro Linhares, 1390 | Rodrik Residence | 1 | 0 | 1 | 1 | 3 | 42,15 | 2018 | 5 | R$ 548.000,00 | 6945637,39 | 744611,54 | https://www.vivareal.com.br/imovel/apartamento-1-q | 15MAR23 |
| EI_93 | Rua Lauro Linhares, 1921 | São Francisco | 3 | 0 | 2 | 1 | 1 | 74 | 1985 | 38 | R$ 500.000,00 | 6945144,13 | 744725,53 | https://www.vivareal.com.br/imovel/apartamento-3-q | 14MAR23 |
| EI_94 | Rua João Marçal, 176 | Pereque | 2 | 0 | 2 | 1 | 1 | 60 | 1990 | 33 | R$ 349.000,00 | 6945276,52 | 744380,54 | https://www.vivareal.com.br/imovel/apartamento-2-q | 12MAR23 |
| EI_95 | Rua João Marçal, 176 | Pereque | 2 | 1 | 2 | 1 | 1 | 64 | 1990 | 33 | R$ 430.000,00 | 6945276,52 | 744380,54 | https://www.vivareal.com.br/imovel/apartamento-2-q | 05MAR23 |
| EI_96 | Rua João Marçal, 176 | Pereque | 2 | 0 | 2 | 1 | 1 | 65 | 1990 | 33 | R$ 377.741,00 | 6945276,52 | 744380,54 | https://www.vivareal.com.br/imovel/apartamento-2-q | 12MAR23 |
| EI_97 | Rua João Marçal, 176 | Pereque | 2 | 0 | 2 | 1 | 1 | 65 | 1990 | 33 | R$ 373.000,00 | 6945276,52 | 744380,54 | https://www.vivareal.com.br/imovel/apartamento-2-q | 12MAR23 |
| EI_98 | Rua João Marçal, 176 | Pereque | 2 | 0 | 2 | 1 | 1 | 65 | 1990 | 33 | R$ 385.000,00 | 6945276,52 | 744380,54 | https://www.vivareal.com.br/imovel/apartamento-2-q | 12MAR23 |
| EI_99 | Rua João Marçal, 176 | Pereque | 2 | 0 | 2 | 1 | 1 | 60 | 1990 | 33 | R$ 349.000,00 | 6945276,52 | 744380,54 | https://www.vivareal.com.br/imovel/apartamento-2-q | 12MAR23 |
| EI_100 | Rua João Marçal, 216 | Spazio Trindade | 1 | 0 | 1 | 1 | 3 | 33 | 2022 | 1 | R$ 365.000,00 | 6945281,47 | 744400,53 | https://www.vivareal.com.br/imovel/apartamento-1-q | 12MAR23 |
| EI_101 | Rua João Marçal, 216 | Spazio Trindade | 1 | 0 | 1 | 1 | 3 | 40 | 2022 | 1 | R$ 440.000,00 | 6945281,47 | 744400,53 | https://www.vivareal.com.br/imovel/apartamento-1-q | 12MAR23 |
| EI_102 | Rua João Marçal, 216 | Spazio Trindade | 1 | 0 | 1 | 1 | 3 | 40 | 2022 | 1 | R$ 450.000,00 | 6945281,47 | 744400,53 | https://www.vivareal.com.br/imovel/apartamento-1-q | 12MAR23 |
| EI_103 | Rua Edison Areas, 132 | Mariely | 3 | 1 | 2 | 1 | 2 | 89 | 2002 | 21 | R$ 585.000,00 | 6946560,72 | 744346,18 | https://www.vivareal.com.br/imovel/apartamento-3-q | 12MAR23 |
| EI_104 | Rua Europa, 370 | Munique | 2 | 0 | 1 | 1 | 1 | 65 | 1998 | 25 | R$ 380.000,00 | 6946136,81 | 743977,73 | https://www.vivareal.com.br/imovel/apartamento-2-q | 12MAR23 |
| EI_105 | Rua Europa, 370 | Munique | 2 | 0 | 1 | 1 | 1 | 64 | 1998 | 25 | R$ 394.000,00 | 6946136,81 | 743977,73 | https://www.vivareal.com.br/imovel/apartamento-2-q | 12MAR23 |
| EI_106 | Rua Europa, 370 | Munique | 2 | 0 | 1 | 1 | 1 | 65 | 1998 | 25 | R$ 395.000,00 | 6946136,81 | 743977,73 | https://www.vivareal.com.br/imovel/apartamento-2-q | 14MAR23 |
| EI_107 | Rua Europa, 150 | Coimbra | 2 | 0 | 1 | 1 | 1 | 73 | 1998 | 25 | R$ 360.000,00 | 6946370,06 | 744234,89 | https://www.vivareal.com.br/imovel/apartamento-2-q | 12MAR23 |
| EI_108 | Rua Europa, 128 | Barcelona | 3 | 1 | 2 | 1 | 1 | 77 | 1998 | 25 | R$ 382.000,00 | 6946364,13 | 744230,29 | https://www.vivareal.com.br/imovel/apartamento-3-q | 05MAR23 |
| EI_109 | Rua Europa, 128 | Barcelona | 3 | 1 | 2 | 1 | 1 | 78 | 1998 | 25 | R$ 375.000,00 | 6946364,13 | 744230,29 | https://www.vivareal.com.br/imovel/apartamento-2-q | 05MAR23 |
| EI_110 | Rua Europa, 106 | Bruxelas | 3 | 0 | 2 | 1 | 1 | 77 | 1998 | 25 | R$ 359.000,00 | 6946368,55 | 744233,8 | https://www.vivareal.com.br/imovel/apartamento-2-q | 12MAR23 |
| EI_111 | Rua Trajano Margarida, 1 | Back Green Towers | 2 | 1 | 2 | 1 | 3 | 73,00 | 2017 | 6 | R$ 998.000,00 | 6946046,37 | 744535,9 | https://www.vivareal.com.br/imoveis-lancamento/resi | 12MAR23 |
| EI_112 | Rua Trajano Margarida, 1 | Back Green Towers | 3 | 1 | 2 | 2 | 3 | 91 | 2017 | 6 | R$ 1.099.000,00 | 6946046,37 | 744535,9 | https://www.vivareal.com.br/imovel/apartamento-3-q | 12MAR23 |
| EI_113 | Rua Trajano Margarida, 1 | Back Green Towers | 3 | 1 | 2 | 2 | 3 | 92 | 2017 | 6 | R$ 1.319.000,00 | 6946046,37 | 744535,9 | https://www.vivareal.com.br/imoveis-lancamento/resi | 12MAR23 |
| EI_114 | Rua Trajano Margarida, 1 | Back Green Towers | 3 | 1 | 2 | 2 | 3 | 91 | 2017 | 6 | R$ 1.504.000,00 | 6946046,37 | 744535,9 | https://www.vivareal.com.br/imovel/apartamento-3-q | 12MAR23 |
| EI_115 | Rua Trajano Margarida, 1 | Back Green Towers | 3 | 1 | 2 | 2 | 3 | 91 | 2017 | 6 | R$ 1.611.814,00 | 6946046,37 | 744535,9 | https://www.vivareal.com.br/imovel/apartamento-3-q | 12MAR23 |
| EI_116 | Rua José Batista Rosa | Inês de Castro | 1 | 0 | 1 | 1 | 3 | 42 | 2018 | 5 | R$ 540.000,00 | 6946170,14 | 744310,59 | https://www.vivareal.com.br/imovel/apartamento-1-q | 12MAR23 |
| EI_117 | Rua José Batista Rosa | Inês de Castro | 2 | 1 | 2 | 2 | 3 | 63,75 | 2018 | 5 | R$ 887.707,00 | 6946170,14 | 744310,59 | https://www.brognoli.com.br/imovel/comprar-florianop | 16MAR23 |
| EI_118 | Rua Lauro Linhares, 1288 | Stoneville | 3 | 1 | 2 | 2 | 2 | 89 | 2000 | 23 | R$ 550.000,00 | 6945749,88 | 744589,93 | https://www.vivareal.com.br/imovel/apartamento-3-q | 12MAR23 |
| EI_119 | Rua Lauro Linhares, 1288 | Stoneville | 3 | 1 | 2 | 1 | 2 | 86,6 | 2000 | 23 | R$ 570.000,00 | 6945749,88 | 744589,93 | https://www.brognoli.com.br/imovel/comprar-florianop | 16MAR23 |
| EI_120 | Rua Lauro Linhares, 1288 | Stoneville | 3 | 1 | 2 | 1 | 2 | 87 | 2000 | 23 | R$ 670.000,00 | 6945749,88 | 744589,93 | https://www.vivareal.com.br/imovel/apartamento-3-q | 15MAR23 |
| EI_121 | Rua Luiz Pasteur, 89 | Living Trindade | 2 | 1 | 2 | 1 | 3 | 60 | 2022 | 1 | R$ 560.000,00 | 6946399,02 | 744217,26 | https://www.vivareal.com.br/imovel/apartamento-2-q | 12MAR23 |
| EI_122 | Rua Luiz Pasteur, 89 | Living Trindade | 2 | 0 | 1 | 1 | 3 | 61 | 2022 | 1 | R$ 579.000,00 | 6946399,02 | 744217,26 | https://www.vivareal.com.br/imovel/apartamento-2-q | 15MAR23 |
| EI_123 | Rua Luiz Pasteur, 89 | Living Trindade | 2 | 0 | 1 | 1 | 3 | 61 | 2022 | 1 | R$ 645.000,00 | 6946399,02 | 744217,26 | https://www.vivareal.com.br/imovel/apartamento-2-q | 15MAR23 |
| EI_124 | Rua Luiz Pasteur, 89 | Living Trindade | 2 | 0 | 1 | 1 | 3 | 60 | 2022 | 1 | R$ 650.000,00 | 6946399,02 | 744217,26 | https://www.vivareal.com.br/imovel/apartamento-2-q | 15MAR23 |
| EI_125 | Rua Luiz Pasteur, 89 | Living Trindade | 2 | 0 | 1 | 1 | 3 | 60 | 2022 | 1 | R$ 680.000,00 | 6946399,02 | 744217,26 | https://www.vivareal.com.br/imovel/apartamento-2-q | 15MAR23 |
| EI_126 | Rua Lauro Linhares, 657 | Vilamares | 2 | 0 | 1 | 1 | 1 | 59 | 1993 | 30 | R$ 379.000,00 | 6946334,72 | 744360,11 | https://www.vivareal.com.br/imovel/apartamento-2-q | 12MAR23 |
| EI_127 | Rua Lauro Linhares, 657 | Vilamares | 2 | 0 | 1 | 1 | 1 | 59 | 1993 | 30 | R$ 390.000,00 | 6946334,72 | 744360,11 | https://www.vivareal.com.br/imovel/apartamento-2-q | 12MAR23 |
| EI_128 | Rua Lauro Linhares, 657 | Vilamares | 3 | 1 | 2 | 1 | 1 | 84,25 | 1993 | 30 | R$ 480.000,00 | 6946334,72 | 744360,11 | https://www.brognoli.com.br/imovel/comprar-florianop | 16MAR23 |
| EI_129 | Rua Procópio Manoel Pire | Pedro Scheidt | 2 | 0 | 1 | 1 | 1 | 62 | 1991 | 32 | R$ 380.000,00 | 6945444,19 | 744636,64 | https://www.vivareal.com.br/imovel/apartamento-2-q | 14MAR23 |
| EI_130 | Lauro Linhares, 1600 | Piemonte | 3 | 1 | 2 | 2 | 2 | 100 | 2006 | 15 | R$ 990.000,00 | 6945444,19 | 744636,64 | https://www.vivareal.com.br/imovel/apartamento-3-q | 15MAR23 |
| EI_131 | Presidente Gama Rosa, 2 | Firenze | 3 | 1 | 2 | 1 | 1 | 90,5 | 1993 | 30 | R$ 585.000,00 | 6946200,41 | 744632,91 | https://www.brognoli.com.br/imovel/comprar-florianop | 16MAR23 |
| EI_132 | Desembargador Vitor Lim | Scarlett | 3 | 1 | 3 | 1 | 2 | 92 | 1994 | 29 | R$ 590.000,00 | 6944820,1 | 744300,93 | https://www.brognoli.com.br/imovel/comprar-florianop | 16MAR23 |

## ANEXO B – *Scripts* deste trabalho

Os *scripts* dos modelos, o conjunto de dados e outros códigos estão disponíveis no seguinte repositório:

https://github.com/arturdalpra/TCC-avaliacaoImobiliariaIC/

*Script* em Python para obtenção dos elementos da base amostral em PDF

*Script* em Python para gerar arquivo kml a partir de dados num arquivo csv

*Script* em R para Regressão Linear Múltipla (MLR)

*Script* em R para Regressão Linear Múltipla (MLR) com undersampling

*Script* em R para Sistemas de Inferência Difusos de Mamdani (FIS)

*Script* em R para Sistemas de Inferência Difusos de Takagi-Sugeno (TSK)

*Script* em R para Sistemas de Inferência Difusos Neuroadaptativos (ANFIS)

## ANEXO C - Modelo MLR alternativo

Em paralelo foi executado um estudo com as variáveis área privativa, idade (transformada com a função inversa) e padrão de acabamento (como fator) para explicar a variável preço unitário, obtendo-se os resultados da tabela 10:

Tabela 10 – MLR: comparativo entre as proporções de treino x teste

| | Modelo MLR (PU ~ aPriva + 1/idade + padraoAcabamento) | | | |
|---|---|---|---|---|
| Divisão | MSE | MAE | RMSE | MAPE |
| 70% x 30% | 1.211.689,00 | 708,31 | 1.100,77 | 8,59% |
| 77% x 23% | 909.170,00 | 686,99 | 953,50 | 8,19% |
| 80% x 20% | 607.681,00 | 535,71 | 779,54 | 6,75% |
| 87% x 13% | 770.523,20 | 497,27 | 877,79 | 5,51% |

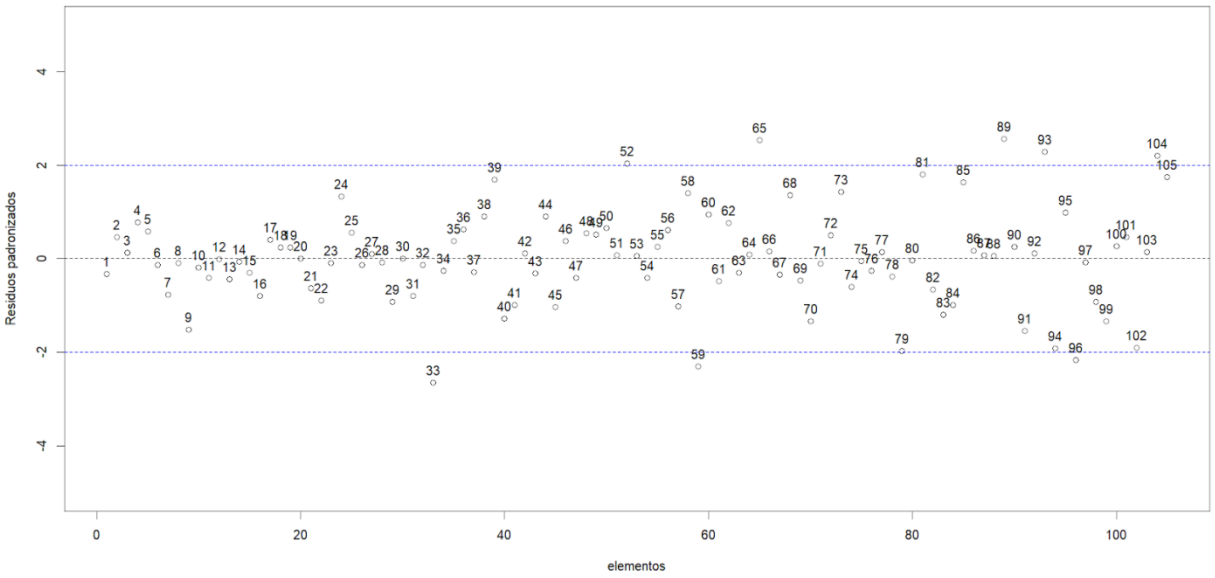
Fonte: elaborado pelo autor

Cabe destacar que tratou-se de um modelo sem a aplicação da técnica de *undersampling* e que manteve-se alguns elementos cujo resíduo padronizado ultrapassou 1,96, limitado a 2,64, sob a justificativa que tratavam-se de sete pontos e que o mercado imobiliário não é perfeito. Por haver superado 1,96, foi excluído da comparação dos resultados.

A tabela 10 o êxito na divisão do conjunto de dados sob a proporção de 87% para treino e 23% para teste, conforme as quatro métricas estudadas.

Tomando-se por melhor modelo aquele gerado pela distribuição 87% x 23%, notaram-se alguns *outliers*, conforme indica o gráfico da figura 43 (resíduos padronizados). Estes *outliers* foram mantidos, visto que o mercado não é perfeito. Destaca-se ainda que é possível notar aleatoriedade na distribuição dos resíduos padronizados.

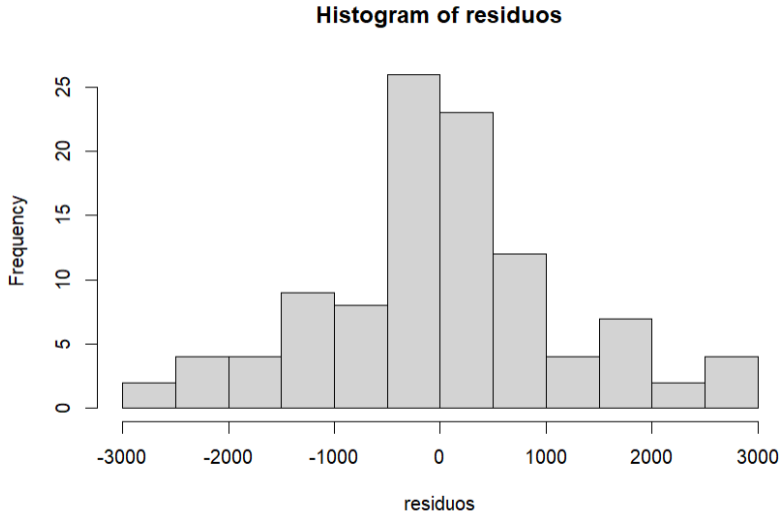
Figura 43 - MLR 87% x 23%: resíduos padronizados para modelo




Fonte: elaborado pelo autor (2023)

Quanto à distribuição dos resíduos, esta conserva aspecto de quase-normalidade, em especial se considerado que o mercado imobiliário contém particularidades, os elementos do conjunto de dados possuem muitas desigualdades onde nem todas são abrangidas pelas variáveis e as classes são desbalanceadas, sendo um mercado de concorrência imperfeita, então o histograma de resíduos da figura 44 mostra aspecto de quase-normalidade.

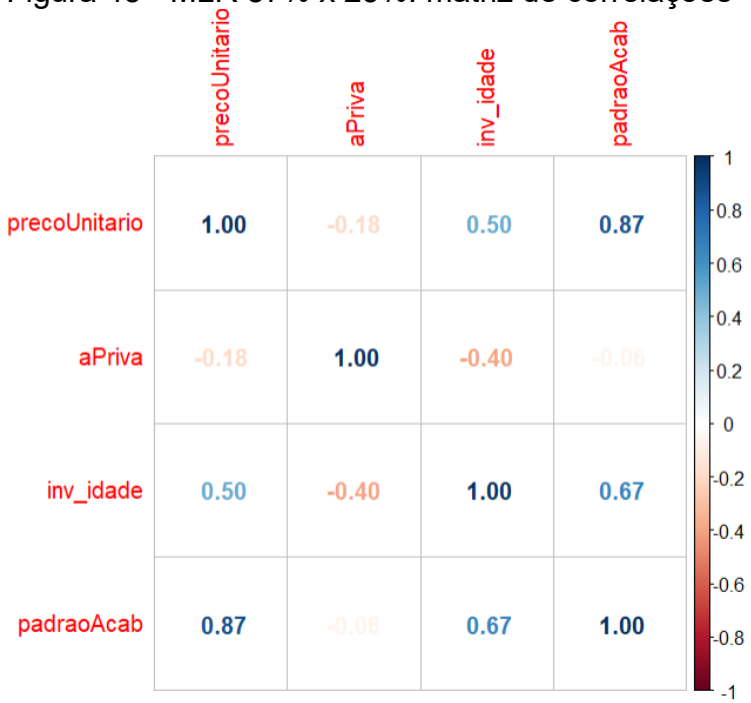
Figura 44 - MLR 87% x 23%: histograma de resíduos




Fonte: elaborado pelo autor (2023)

A matriz de correlações da figura 45 indica correlação positiva média para a variável idade (em sua forma inversa) e moderada a forte para a variável padrão de acabamento com a variável dependente, o que é importante. Ainda quanto à variável dependente (preço unitário), a área privativa apresentou correlação negativa (concordando com o comportamento do mercado) e fraca. Apesar de ser fraca, é importante que o modelo contenha a inclusão desta variável, conforme recomenda a NBR 14.653-2 (2011).

Figura 45 - MLR 87% x 23%: matriz de correlações

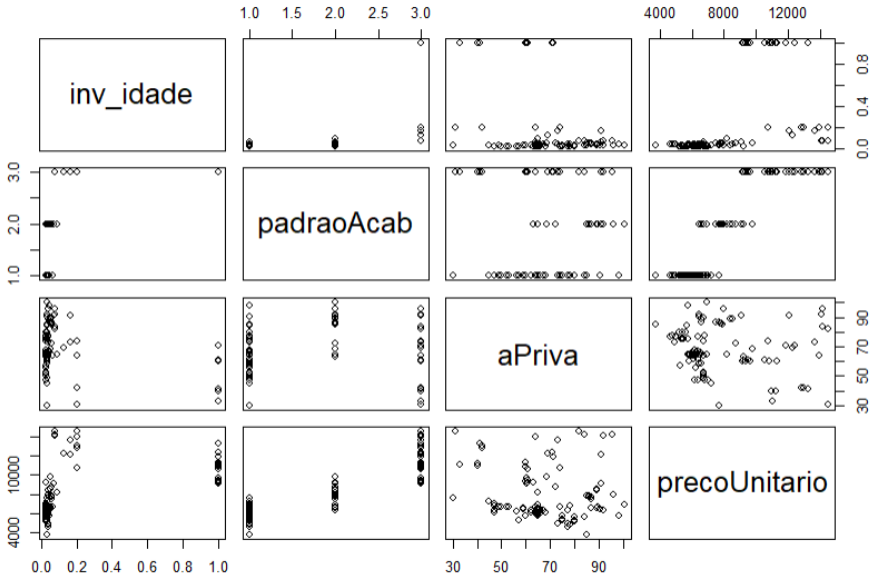



Fonte: elaborado pelo autor (2023)

Vale destacar que a NBR 14.653-2 (2011) em seu anexo A recomenda atenção quando a correlação entre variáveis independentes ultrapassa o limite 0,80.

Quanto aos dados, muito do que se observa no gráfico de correlações também pode ser observado no gráfico de dispersão simples entre as variáveis, mostrado na figura 46.

Figura 46 - MLR 87% x 23%: gráfico de dispersão entre variáveis

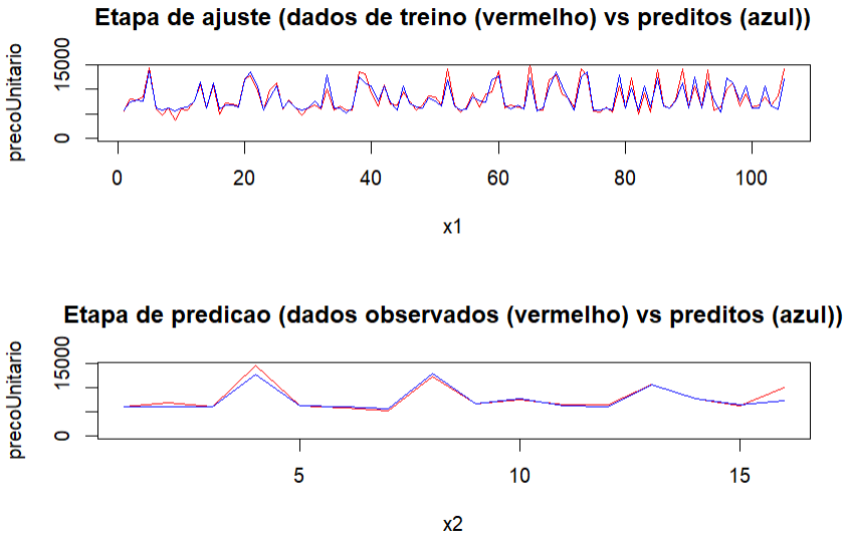



Fonte: elaborado pelo autor (2023)

Na figura 46, a dispersão quase que aleatória entre a área privativa e o preço unitário indica que somente a variável área privativa não é suficiente para explicar o comportamento da variável preço unitário.

A figura 47 mostra a qualidade do treinamento e o êxito na predição dos preços unitários de teste, indicando ser um excelente modelo.

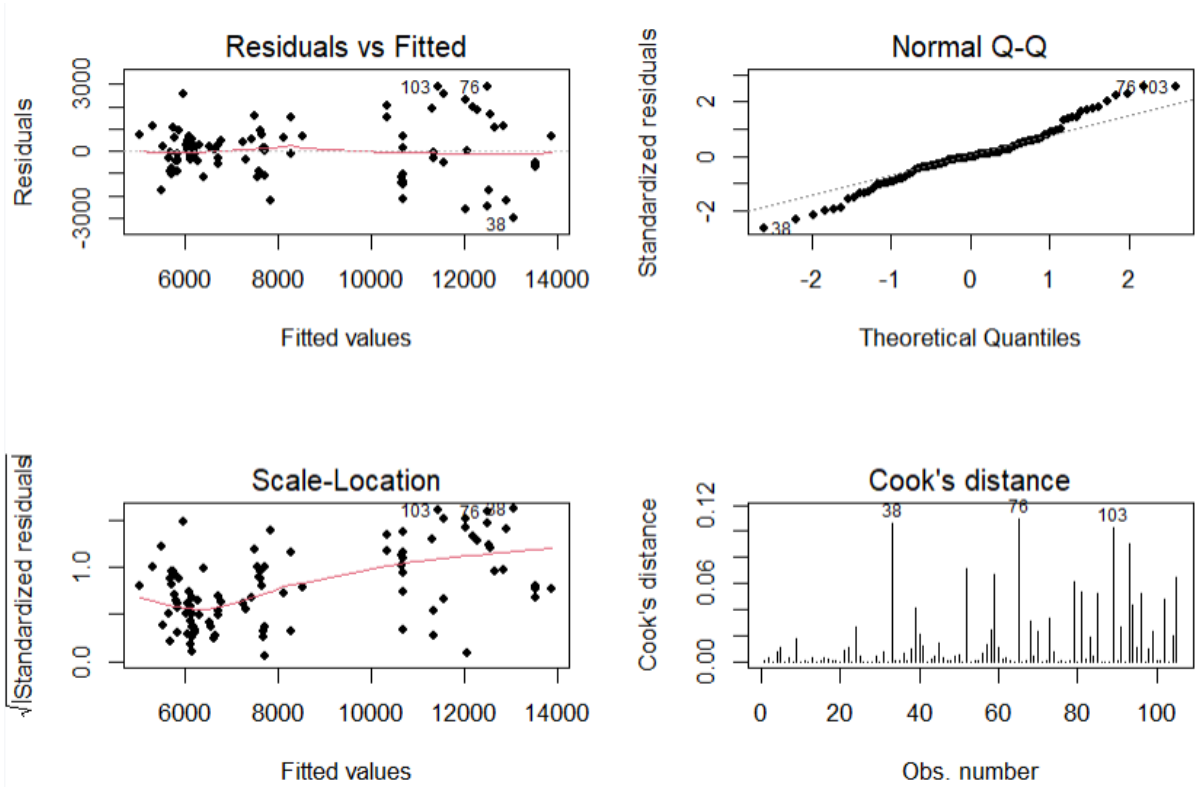
Figura 47 - MLR 87% x 23%: poder de predição (ajuste e predição) por sequência




Fonte: elaborado pelo autor (2023)

Por fim, a técnica MLR faz necessária algumas observações, tais como estas da figura 48 a aleatoriedade dos resíduos, a aderência à linha para concluir pela normalidade dos resíduos, e a distância de Cook.

Figura 48 - MLR 87% x 23%: conjunto de gráficos para diagnóstico do modelo




Fonte: elaborado pelo autor (2023)

**ANEXO D - Artigo**

# Avaliação imobiliária através de inteligência computacional

**Artur A. Dal Prá[1], Élder R. Santos[2]**

[1,2] Depto. de Informática e Estatística - Universidade Federal de Santa Catarina (UFSC)
88.036-001 – Florianópolis - SC - Brasil

{artur.antonio,elder.santos}@ufsc.br

***Abstract.*** *This study provides an overview of the state-of-the-art in computational intelligence applied to real estate valuation and its implementation through scripts using R libraries for predicting the selling price of apartments located in the Trindade neighborhood as of the reference date of march 2023. A comparative study is conducted, comparing the well-established technique of direct comparison method using statistical inference through Multiple Linear Regression (MLR) with other techniques such as Mamdani's fuzzy inference system (FIS), Takagi-Sugeno's fuzzy inference system (TSK), and adaptive neuro-fuzzy inference system (ANFIS). Additionally, the work briefly addresses class balancing through undersampling of the majority class and highlights important considerations for the successful application of fuzzy logic-based models in real estate valuation. The study concludes that MLR yields the best results, followed by the TSK technique. The FIS technique shows reasonable performance, while the ANFIS technique demonstrates potential but requires a careful application. Metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE) were implemented and jointly analyzed using various train-test set ratios. From the perspective of the MAPE metric, the best MLR model resulted in 7.39%, TSK achieved 8.51%, followed by FIS with 10.40%, and ANFIS with 13.05%.*

***Resumo.*** *Este trabalho apresenta uma visão geral sobre o estado da arte da inteligência computacional aplicada à avaliação imobiliária e sua implementação através de scripts usando pacotes em linguagem R para a predição do preço de venda de apartamentos situados no Bairro Trindade na data base março de 2023. Um estudo comparativo é realizado confrontando a técnica consagrada do método comparativo direto de dados de mercado por inferência estatística usando Regressão Linear Múltipla (MLR), comparada com outras técnicas, como o sistema de inferência difuso de Mamdani (FIS), o sistema de inferência difuso de Takagi-Sugeno (TSK), o sistema de inferência difuso neuroadaptativo (ANFIS). Além disso, o trabalho aborda brevemente o balanceamento de classes por subamostragem da classe majoritária*

*(undersampling), e destaca considerações importantes para o êxito na aplicação de modelos baseados em lógica difusa na avaliação imobiliária. O estudo conclui que a MLR apresenta os melhores resultados, seguido da técnica TSK. A técnica FIS apresenta um desempenho razoável, enquanto que a técnica ANFIS mostra potencial, mas requer ser aplicada sob certos cuidados. Foram implementadas as métricas Erro Absoluto Médio (MAE), Erro Quadrático Médio (MSE), Raiz do Erro Quadrático Médio (RMSE) e Erro Médio Percentual Absoluto (MAPE), analisadas conjuntamente sob a aplicação de conjuntos treino x teste com diversas proporções. Sob a ótica da métrica MAPE, o melhor modelo MLR resultou 7,39%, o TSK 8,51%, seguido pelo FIS 10,40% e ANFIS com 13,05%.*

## 1. Considerações iniciais

O mercado imobiliário traz uma realidade que está correlacionada com diversos outros setores da economia. Uma variável impactante à estabilidade do mercado imobiliário é a precificação dos imóveis. No julgamento dos possíveis preços de mercado que podem ser atribuídos a um imóvel urbano, muitas interpretações e técnicas coexistem.

Usualmente os avaliadores aplicam a técnica de inferência estatística com regressão linear múltipla quando o mercado disponibilizar uma quantidade de elementos suficiente para compor uma amostra representativa e enquadramento nos graus de fundamentação e de precisão citados na NBR 14.653-2 (2011), que é a norma brasileira de avaliação de bens, parte de imóveis urbanos. A tendência da norma brasileira é progredir para outros campos da IC como lógica difusa (abordagens de Mamdani, Takagi-Sugeno), sistemas híbridos (como o ANFIS), e aprofundar o tratamento com redes neurais artificiais.

Este artigo objetiva analisar modelos para avaliação de imóveis o tipo apartamento utilizando técnicas de inteligência computacional, a saber, MLR, FIS, TSK e ANFIS.

## 2. Técnicas alternativas para avaliação imobiliária

Segundo Pelli (2004), grandes avanços na área de engenharia são notados, no que tange às modelagens de sistemas reais por meio de mecanismos de inferência. Nesta ótica, este autor cita a restrição em uso de Estimadores dos Mínimos Quadrados e elenca a possibilidade do uso de sistemas híbridos, como por exemplo, as redes neuro-difusas e redes neurais artificiais, que são aplicáveis às avaliações de imóveis urbanos. O uso de muitas alternativas conduz a resultados diferentes, e então surge a possibilidade de usar um método para validar os resultados do outro. A seguir, discorre-se brevemente sobre cada alternativa.

Os modelos deste trabalho fazem uso da variável preço unitário como variável dependente e área privativa, padrão de acabamento, número de vagas e idade como variáveis independentes.

As técnicas aqui descritas são aplicadas sob um conjunto de dados com 132 elementos, coletado no Bairro Trindade, em Florianópolis (SC), em fevereiro/2023.

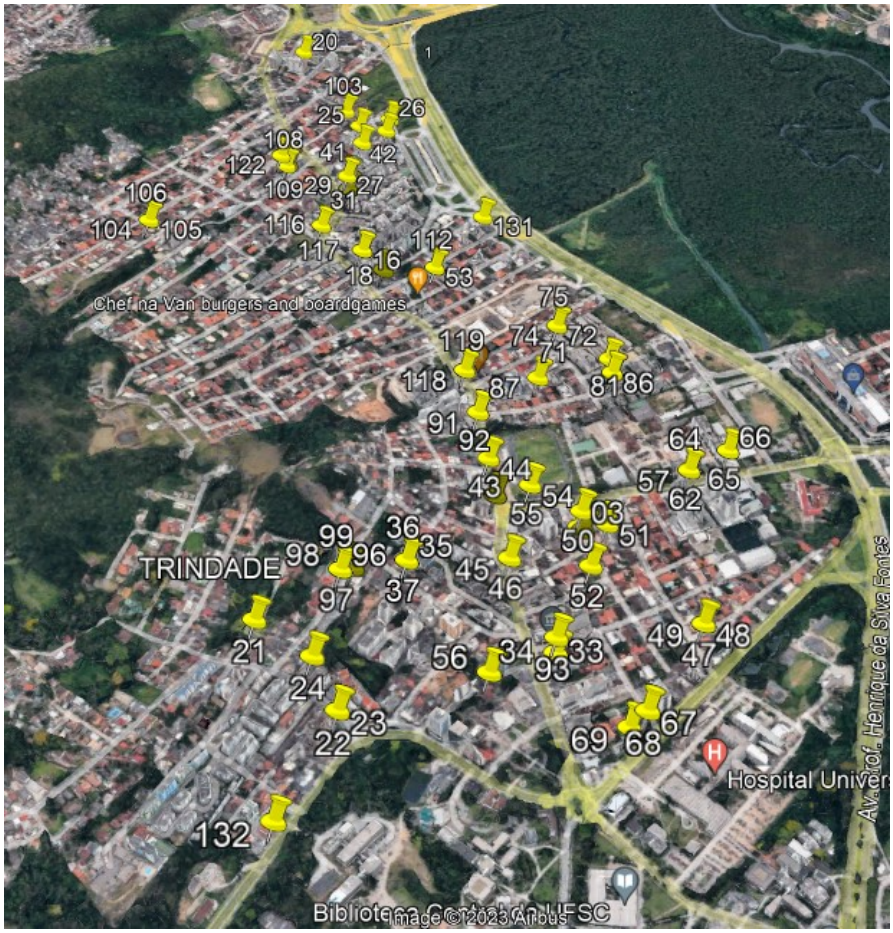



**Figura 1. Distribuição espacial dos elementos do conjunto de dados**

## 2.1. Regressão linear múltipla (MLR)

A Regressão Linear Múltipla trata-se de uma técnica estatística para modelar a relação entre a variável dependente e as variáveis independentes. Eventualmente esta técnica é citada na bibliografia internacional por Análise de Regressão Múltipla.

Segundo ÇETINKAYA-RUNDEL (2019), a MLR é uma extensão da regressão simples entre duas variáveis para o caso em que permanece havendo uma variável dependente, mas são diversos os preditores (variáveis independentes).

Ao ajustar um modelo, deve-se levar em conta a significância alcançada, então havendo rejeição da hipótese nula, considera-se que uma ou mais variáveis independentes podem explicar a variação na variável dependente. Neste trabalho, nenhum ajuste de MLR acima de 0,05 será empregado.

Quanto à distribuição dos resíduos, o histograma de resíduos apresenta aspecto de quase-normalidade, uma vez que a avaliação de imóveis envolve resíduos quase normais devido à imperfeição do mercado imobiliário, que é influenciado de forma não linear por fatores econômicos e eventos específicos, como mudanças na conjuntura econômica e subjetividade na definição dos preços dos imóveis.

A modelagem envolveu diversas etapas e distribuições e convergiu nos resultados da tabela 1:

**Tabela 1. Tabela com a alternativa MLR**

**Modelo MLR** (PU ~ aPriva + nVagas + padraoAcabamento)

| Divisão | MSE | MAE | RMSE | MAPE |
|---|---|---|---|---|
| 70% x 30% | 1.261.330,00 | 917,28 | 1.123,09 | 9,74% |
| 75% x 25% | 948.470,20 | 806,26 | 973,89 | 9,00% |
| 80% x 20% | 793.825,70 | 724,24 | 890,97 | 7,39% |
| 85% x 15% | 1.644.782,00 | 1.103,05 | 1.282,49 | 11,09% |

## 2.2. Sistema de inferência difuso de Mamdani (FIS)

O sistema de inferência difuso de Mamdani foi uma primeira abordagem para resolver um sistema de inferência baseado em lógica difusa.

Segundo Caixa (2018), aspectos qualitativos causam subjetividades, imprecisões, incertezas e ambiguidades e o tratamento costuma se dar pelo discernimento e bom senso do avaliador. Muitas das variáveis tais como pequena (área), alto (padrão), perto (distância), bastante (vagas), velho (idade), e outros podem ser tratadas por sistemas difusos. Caixa (2018) destaca que os advérbios podem ser diminuidores ou aumentadores conforme a área de pertinência dos conjuntos difusos é ampliada ou reduzida.

Os estudos partiram do conjunto de dados sem a aplicação da técnica de subamostragem (*undersampling*), mas já saneado. Os modelos foram executados conforme o algoritmo de Wang (1992), disponível na biblioteca frbs que está funciona no pacote estatístico R.

Todas as proporções treino x teste mostraram bom desempenho, notado pela métrica MAPE. Algumas distribuições apresentaram valores um pouco maiores para a métrica RMSE, por efeito da fórmula da métrica que penaliza grandes desvios.

Fez uso das variáveis área privativa, número de vagas e padrão de acabamento para explicar a variabilidade na variável preço unitário. Esta é a mesma distribuição de variáveis praticada nos modelos MLR.

Depois de alguns procedimentos para escolha de variáveis, obtiveram-se os resultados da Tabela 2.

**Tabela 2. Tabela com a alternativa FIS**

**Modelo FIS**

| Divisão | MSE | MAE | RMSE | MAPE |
|---|---|---|---|---|
| 70% x 30% | 1.061.433,00 | 837,02 | 1.030,26 | 10,80% |
| 75% x 25% | 1.071.559,00 | 803,10 | 1.035,16 | 10,40% |
| 80% x 20% | 1.172.636,00 | 876,33 | 1.082,88 | 11,07% |
| 85% x 15% | 1.261.787,00 | 965,88 | 1.123,29 | 11,39% |

## 2.3. Sistema de inferência difuso de Takagi-Sugeno (TSK)

O Sistema de Inferência Difuso de Takagi-Sugeno é uma técnica que partiu do sistema Mamdani, mantendo o antecedente e substituindo o consequente por uma equação polinomial que resulta na própria saída (*output*) do sistema.

A técnica TSK se constitui num misto entre uma regressão linear convencional e um sistema difuso. Na regressão linear convencional, o método de mínimos quadrados é usado para encontrar os parâmetros βi que minimizam a soma dos erros quadrados. Em geral, a diferença entre os valores estimados e observados surge da incerteza da estrutura do modelo ou de observações imprecisas.

A imprecisão incorporada (que é bastante comum em situações de previsões de preços de venda) torna importante a aplicação da técnica TSK. A técnica TSK consiste em usar as regras para constituir filtros e a partir deles clusterizar os dados para criar cada função polinomial. Note que a cada regra corresponderá uma única função polinomial.

Como no TSK as equações são ajustadas diretamente aos dados, então a modelagem é mais precisa que aquela obtida através da abordagem por Mamdani (FIS) por capturar mais precisamente a relação funcional entre a variável dependente e as variáveis independentes do setor imobiliário.

Os modelos foram realizados com as variáveis área privativa, idade e padrão de acabamento para explicar a variabilidade na variável resposta preço unitário. Destaca-se que a variável padrão de acabamento foi modelada como variável *proxy*, conforme adiantado no tópico de aplicação.

Não há biblioteca integralmente automatizada para rodar a técnica TSK, então os conjuntos foram concebidos manualmente com funções triangulares e trapezoidais nas bordas para as variáveis área privativa, padrão de acabamento e idade. Cada variável independente teve três divisões. Das divisões, infere-se ser necessário 27 regras (3x3x3). As regras que mostraram insuficiência de regras para constituir uma função de consequente foram removidas. Tem-se a proporção de uma regra para cada equação gerada. O tópico sobre a aplicação da técnica TSK explica a definição das funções de pertinência. O tipo da t-norma é mínimo, da s-norma é máximo, é usada a função de implicação de Zadeh e o operador é o AND.

Testadas diversas combinações e regras, a Tabela 3 foi obtida.

**Tabela 3. Tabela com a alternativa TSK**

**Modelo TSK: sem *outliers* e sem *undersampling***

| Divisão | MSE | MAE | RMSE | MAPE |
|---|---|---|---|---|
| 72% x 28% | 1.145.532,00 | 773,09 | 1.070,30 | 8,51% |
| 77% x 23% | 1.158.785,00 | 778,28 | 1.076,47 | 9,74% |
| 82% x 18% | 2.338.628,00 | 1.247,54 | 1.529,26 | 15,56% |
| 87% x 13% | 1.799.465,00 | 1.017,96 | 1.341,44 | 12,03% |

## 2.4. Sistema de inferência difuso neuroadaptativo (ANFIS)

Os Sistemas de Inferência Difusos Neuroadaptativos foram idealizados por Jang (1993), onde partindo de um modelo TSK cujas funções de pertinência são gaussianas, se faz uso de redes neurais artificiais para calibração destas funções.

Valle (2021) explica que dispondo-se de um conjunto de dados significativo, pode-se abrir mão de compor um dicionário e de estabelecer uma base de regras, então a partir dos dados automaticamente se determina o dicionário e as regras difusas num modelo de Takagi-Sugeno, fazendo uso da técnica ANFIS que é um modelo de Takagi-Sugeno adaptativo.

A ideia é reforçada por Sarip et al (2016), que destacam que a tarefa de ajustar as regras e conjuntos difusos pode consumir muito tempo, razão pela qual o sistema difuso pode ser integrado com as capacidades de aprendizado da ANN, resultando no Sistema de Inferência Difuso Neuroadaptativo (ANFIS).

Segundo Kamire et al (2021), quando os dados já estão preparados, é difícil encontrar uma relação linear entre as diversas variáveis independentes do conjunto de dados, então é quando a técnica ANFIS é empregada.

Corrêa (2020) explica que o ANFIS é uma rede *feedforward* (ANN onde o fluxo avança apenas em uma direção, da camada de entrada em direção à camada de saída), cuja aprendizagem se dá pelo método da descida em gradiente e também mínimos quadrados, e onde um sistema TSK é gerado.

Depois das diversas combinações, concluiu-se que o conjunto de dados em sua totalidade (com os elementos *outliers*) enriquecem o modelo, então fez-se uma abordagem sem *undersampling* (Tabela 4) e outra abordagem com *undersampling* (Tabela 5), notando-se como diferença que o conjunto de dados sem *undersampling* apresenta resultados mais estáveis, notadamente em razão do desempenho medido pela métrica RMSE.

**Tabela 4. Tabela com a alternativa ANFIS**

**Modelo ANFIS: conjunto com outliers e sem *undersampling***

| Divisão | MSE | MAE | RMSE | MAPE |
|---|---|---|---|---|
| 51% x 49% | 1.712.735,00 | 1.030,55 | 1.308,72 | 13,05% |

Os resultados obtidos do conjunto de dados com *undersampling* seguem na Tabela 5.

**Tabela 5. Tabela com a alternativa TSK**

**Modelo ANFIS: conjunto de dados com *outliers* e com *undersampling***

| Divisão | MSE | MAE | RMSE | MAPE |
|---|---|---|---|---|
| 52% x 48% | 3.166.301,00 | 1.334,93 | 1.779,41 | 14,91% |

## 3. Análise comparativa dos resultados

O fato dos modelos MLR se destacarem não deve ser entendido como uma regra geral, então outras amostragens em outros lugares podem fazer outra técnica se sobressair sobre as demais. A Tabela 6 resume os resultados.

**Tabela 6. Tabela resumo**

**Tabela resumo**

| Técnica | Divisão | MSE | MAE | RMSE | MAPE |
|---|---|---|---|---|---|
| MLR (*undersampling*) | 80% x 20% | 793.825,70 | 724,24 | 890,97 | 7,39% |
| FIS (saneado) | 75% x 25% | 1.071.559,00 | 803,10 | 1.035,16 | 10,40% |
| TSK (saneado) | 72% x 28% | 1.145.532,00 | 773,09 | 1.070,30 | 8,51% |
| ANFIS (com *outliers*) | 51% x 49% | 1.712.735,00 | 1.030,55 | 1.308,72 | 13,05% |
| ANFIS (com *outliers* e *undersampling*) | 54% x 46% | 3.523.135,00 | 1.403,98 | 1.877,00 | 17,67% |

A técnica FIS mostrou que a distribuição com mais elementos para treinamento se destacou, trouxe resultados aceitáveis (comparativamente com a técnica consagrada MLR) e sugeriu que a inclusão de elementos *outliers* pode prejudicar a formação das regras e consequentemente afetando os resultados, então poder-se-ia realizar modelos FIS incluindo *outliers* para comparar o desempenho por meio das métricas.

O modelo obtido com a técnica TSK foi aquele que mais se aproximou da técnica MLR, então por haver usado um conjunto de dados parcialmente saneado, mostrou certa resistência à *outliers*, notadamente em razão das equações não usarem todos os elementos do conjunto de dados e haver mais de uma equação que é ativada conforme o grau de pertinência ao conjunto. A análise destacou o potencial da técnica, uma vez que conservou bons resultados mesmo com equações cujo sinal dos coeficientes estava invertido.

Quanto ao ANFIS, a eliminação dos elementos *outliers* diminuiu o conjunto de dados e isso reduziu a capacidade de aprendizado. A criação automática de regras em conjuntos menores diminui a quantidade de regras e as consequências são a redução do poder de predição dos modelos e a hipótese de vários elementos do conjunto de dados ativarem as mesmas regras, predizendo valores idênticos (repetidos). Em termos de aplicabilidade da técnica ANFIS ao escopo das avaliações imobiliárias, os resultados mostraram certa resistência a *outliers*, apresentando melhores resultados para conjuntos de dados maiores e para conjuntos de dados balanceados (por *undersampling*).

Como a técnica ANFIS implementa um modelo TSK de ordem 1, situações em que muitas variáveis são necessárias para explicar a variabilidade da variável dependente, o ANFIS pode não ser uma alternativa adequada nestes casos.

Em todos os modelos os elementos com preços unitários observados mais elevados mostraram preços unitários preditos com maior dispersão. Este efeito se deve ao fato que conforme o preço do imóvel aumenta, este perde liquidez e a pressa do ofertante que o imóvel seja absorvido mais rapidamente pelo mercado (venda) é um fator que influencia sobremaneira.

## 4. Conclusões

Este item apresenta as principais descobertas da pesquisa, destacando qual das técnicas de inteligência computacional se mostrou mais eficiente para a avaliação imobiliária.

O objetivo deste trabalho foi analisar modelos para avaliação de imóveis do tipo apartamento utilizando técnicas de inteligência computacional. Foram utilizadas as técnicas MLR, FIS, TSK e ANFIS. Nesta ocasião ficam evidentes as possibilidades de utilização destas técnicas na prática cotidiana.

A técnica MLR apresentou os melhores resultados e esta conclusão não deve ser entendida como sendo absoluta, visto que outras amostragens sob outros contextos podem fazer com que outra técnica se sobressaia sobre a MLR. À medida que a relação entre as variáveis independentes e a variável dependente passa a carregar não linearidades importantes e o conjunto de dados traz disponíveis somente variáveis qualitativas, as técnicas derivadas da lógica difusa passam a se destacar, uma vez que a natureza destes modelos contempla a vagueza. No conjunto de dados do domínio deste problema, haviam variáveis independentes numéricas que explicavam muito da variabilidade da variável dependente, então esta foi a principal razão pela qual a técnica MLR se destacou sobre as demais

A técnica FIS treinada com um conjunto de dados maior se destacou trazendo resultados aceitáveis e sugeriu que a inclusão de elementos *outliers* pode prejudicar a formação das regras e consequentemente afetando os resultados, então trabalhos futuros poderiam realizar modelos FIS incluindo *outliers* para comparar o desempenho por meio das métricas.

O modelo obtido com a técnica TSK foi aquele que mais se aproximou da técnica MLR, mesmo usando um conjunto de dados parcialmente saneado e mostrou certa resistência à *outliers*. Este fato pode ser explicado porque as equações não usam todos os elementos do conjunto de dados e de acordo com o objeto predito pode haver mais de uma equação que é ativada conforme o grau de pertinência ao conjunto. Destaca-se o potencial da técnica, uma vez que conservou bons resultados mesmo com equações cujo sinal dos coeficientes estava invertido.

Tinha-se de antemão a expectativa que a técnica ANFIS poderia não ser uma alternativa adequada em casos como deste trabalho, onde muitas variáveis são necessárias para explicar a variabilidade da variável dependente, então como a técnica ANFIS implementa um modelo TSK de ordem 1, esta expectativa se confirmou, apresentando os resultados com maior imprecisão quando comparado com as outras técnicas.

Quanto à aceitação das técnicas derivadas da lógica difusa para a aplicação nas atividades de avaliação imobiliária, fica consolidada a necessidade de testar diversas distribuições de treino x teste e uso da técnica MLR como apoio para sanear os elementos *outliers*.

Em especial o trabalho serviu para destacar as contribuições e implicações para o campo da avaliação imobiliária sob a ótica do uso de outras técnicas além da MLR, que por vezes podem conduzir a resultados mais precisos.

## Referências bibliográficas


ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. NBR 14.653-2: Avaliação de bens - imóveis urbanos. Rio de Janeiro, 2011.

CAIXA. Coletânea de artigos de avaliação de imóveis. Brasília: CAIXA, 2018. 125p.

ÇETINKAYA-RUNDEL, M.; DIEZ, D. M.; BARR, C. D. OpenIntro Statistics. 3. ed. OpenIntro, 2019.

CORRÊA, C. E. Método para aumento de interpretabilidade em modelos de Takagi-Sugeno no sistema INFGMN. 2020. 103 p. Trabalho de Conclusão de Curso (Graduação) - Universidade Federal de Santa Catarina, Florianópolis (SC), 2020.

JANG, J.-S. R. ANFIS: Adaptive-Network-Based Fuzzy Inference System. IEEE Transactions on Systems, Man, and Cybernetics, v. 23, n. 3, p. 665-685, mai. 1993.

KAMIRE, A.; CHAPHALKAR, N.; SANDBHOR, S. Real Property Value Prediction Capability Using Fuzzy Logic and ANFIS. In: International Conference on Smart Data Intelligence (ICSMDI 2021).

PELLI, A. Avaliação de imóveis urbanos com utilização de sistemas nebulosos (redes neuro-fuzzy) e de redes neurais artificiais. In: XXI Congreso Panamericano de Valuación Cartagena, 18 p. 2004, Colômbia.

R CORE TEAM. R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria. Disponível em: https://www.R-project.org. Acesso em: 12 jan, 2023.

RIZA, L. S.; BERGMEIR, C.; HERRERA, F.; BENITEZ, J. M. frbs: Fuzzy Rule-Based Systems for Classification and Regression Tasks. 3.2-0. Disponível em: https://cran.r-project.org/package=frbs. Acesso em: 11 maio 2023.

R STUDIO: ambiente integrado de desenvolvimento para R. Boston, MA: RStudio, PBC Posit Software, 2022. Disponível em: https://www.rstudio.com/. Acesso em: 12/janeiro/2023.

SARIP, A. G., HAFEZ, M. B., DAUD, M. N.. Application Of Fuzzy Regression Model For Real Estate Price Prediction. Malaysian Journal of Computer Science, v. 29, n. 1, p. 15-27, 13 p, mar. 2016. Malásia.

VALLE, M. Sistema Tipo Mamdani e Takagi-Sugeno. Apresentação para o curso de pós graduação, disciplina de lógica difusa na UNESP, 09 set. 2021. Disponível em: https://www.youtube.com/watch?v=T-CupbBcK-M. Acesso em: 10 fev. 2021.

VIVAREAL. Disponível em: www.vivareal.com.br. Acesso em: fev. 2023.

WANG, L. X.; MENDEL, J. M. . Generating Fuzzy Rules by Learning from Examples. IEEE Transactions on Systems, Man, and Cybernetics, ENova Jersey, n. 6, p. 1414 - 1428, 1992.