

CURSO DE CIÊNCIA DA COMPUTAÇÃO – TCC		
() PRÉ-PROJETO	(X) PROJETO	ANO/SEMESTRE: 2023/1

ANÁLISE DO MERCADO IMOBILIÁRIO: INFLUÊNCIA E IMPACTO DOS FATORES REGIONAIS

Thomas Michels Rodrigues

Prof. Aurélio Faustino Hoppe – Orientador

1 INTRODUÇÃO

O mercado imobiliário é um dos setores mais importantes da economia, tanto no Brasil quanto em todo o mundo. Ele tem-se mostrado bastante aquecido e abrange a construção, a venda e o aluguel de imóveis, desde residenciais até comerciais, movimentando uma grande quantidade de dinheiro todos os anos. No entanto, de acordo com Sidrim (2013), a dinâmica do mercado imobiliário não é homogênea em todas as regiões. A economia, a infraestrutura, as políticas públicas e o desenvolvimento urbano são alguns dos fatores regionais que influenciam a demanda e oferta de imóveis em cada localidade.

Green *et al.* (2005) ressaltam que os compradores valorizam as características de um imóvel de forma diferente, de acordo com as suas preferências particulares. Já a precificação de imóveis é um processo complexo que envolve uma série de fatores, como localização, tamanho, acabamento, idade do imóvel, o em torno e características ambientais desta edificação (ARON *et al.*, 2012). Além disso, a proximidade dos principais comércios como supermercados, farmácias e escolas tendem a ser mais valorizados, pois oferecem mais comodidade e praticidade aos moradores. Por outro lado, imóveis localizados em áreas isoladas, com poucos serviços e comércios ou com proximidade a locais com risco de deslizamentos e cotas de enchentes inferiores, podem ser fatores determinantes na escolha e precificação de um imóvel, pois podem afetar a segurança e o conforto dos moradores (PACHECO; TAVARES, 2015).

Segundo Campos (2014) para realizar a precificação de um imóvel, os corretores e avaliadores imobiliários usam diversas técnicas e metodologias, como a análise comparativa de mercado, que envolve a comparação do imóvel em questão com outras propriedades semelhantes na mesma região. Outra técnica comum é a avaliação por custo, que leva em conta o custo de construção do imóvel e adiciona uma margem de lucro para determinar o preço de venda. Além de tudo, segundo Campos (2014), a precificação de imóveis depende diretamente da necessidade do consumidor, sendo ela um dos fatores determinantes para a execução da compra e venda.

Em Blumenau, o mercado imobiliário é caracterizado por uma grande diversidade de imóveis, desde apartamentos de alto padrão até casas populares, passando por terrenos, salas comerciais e imóveis industriais. A cidade é um importante polo industrial da região e conta com uma economia diversificada, o que atrai investidores e empreendedores de diversas áreas. Em termos de preços, o mercado imobiliário de Blumenau apresenta uma grande variação, dependendo da localização, tamanho e características do imóvel (TEIXEIRA *et al.*, 2016). Em geral, os bairros mais valorizados são aqueles próximos ao centro da cidade, onde se concentram os principais serviços e comércios, além de ter fácil acesso às principais vias e rodovias. No entanto, existem outras regiões em Blumenau que apresentam preços mais acessíveis e são procuradas por pessoas que buscam imóveis mais baratos. Todavia, nos últimos anos, o mercado imobiliário de Blumenau vem passando por algumas transformações, como o aumento da oferta de imóveis, especialmente apartamentos de médio e alto padrão, e a busca por áreas mais afastadas do centro da cidade, em busca de mais tranquilidade e qualidade de vida. Contudo, a valorização imobiliária em Blumenau vem se mantendo estável, com pequenas variações de acordo com a oferta e demanda em cada região (ROVERI, 2020).

Do mesmo modo, de acordo com Cekic *et al.* (2022), através do uso de dados e técnicas de análise, é possível identificar tendências e padrões que afetam o mercado imobiliário, como o impacto das enchentes e eventos climáticos na valorização ou desvalorização dos imóveis, o crescimento econômico da região, a taxa de juros e a disponibilidade de crédito para financiamento imobiliário. Além disso, a análise do mercado imobiliário pode ser aprimorada com o uso de ferramentas de visualização de dados, como gráficos e mapas interativos, que permitem identificar padrões espaciais e geográficos em relação à oferta e demanda de imóveis.

Diante do exposto, este trabalho possui a seguinte pergunta de pesquisa: Como os fatores regionais influenciam e impactam o mercado imobiliário de Blumenau, e como esses efeitos podem ser previstos ou compreendidos por meio de análise de dados e modelos meta-heurísticos?

1.1 OBJETIVOS

O objetivo principal deste trabalho é analisar a influência e o impacto dos fatores regionais no mercado imobiliário de Blumenau por meio de um modelo computacional meta-heurístico e informações georreferenciadas.

Os objetivos específicos são:

- identificar as variáveis e os fatores regionais que afetam o mercado imobiliário de Blumenau, como infraestrutura, transportes, segurança, serviços públicos, demografia, economia e política local;
- identificar de forma georreferenciada as áreas com maior concentração de imóveis à venda e comparar os preços médios desses imóveis com outras áreas da cidade;
- avaliar o desempenho do modelo *Grey Wolf Optimizer* (GWO) com outros métodos de precificação imobiliária.

2 TRABALHOS CORRELATOS

Nesta seção serão apresentados trabalhos que possuem semelhança com o tema proposto. Na subseção 2.1 encontra-se uma abordagem que utiliza redes neurais artificiais e o otimizador Grey Wolf, para predição de preços e seleção das características mais impactantes nos preços dos imóveis (AL-GBURY; KURNAZ, 2020). A subseção 2.2 descreve o trabalho de Cekic *et al.* (2022) que utilizaram diferentes técnicas de aprendizado de máquina para predição de preços no contexto residencial. Por fim, a subseção 2.3 apresenta o trabalho de Zheng *et al.* (2022) que propuseram a utilização de uma rede neural de retropropagação otimizada por um algoritmo genético para precificação de imóveis levando em consideração as políticas imobiliárias regionais.

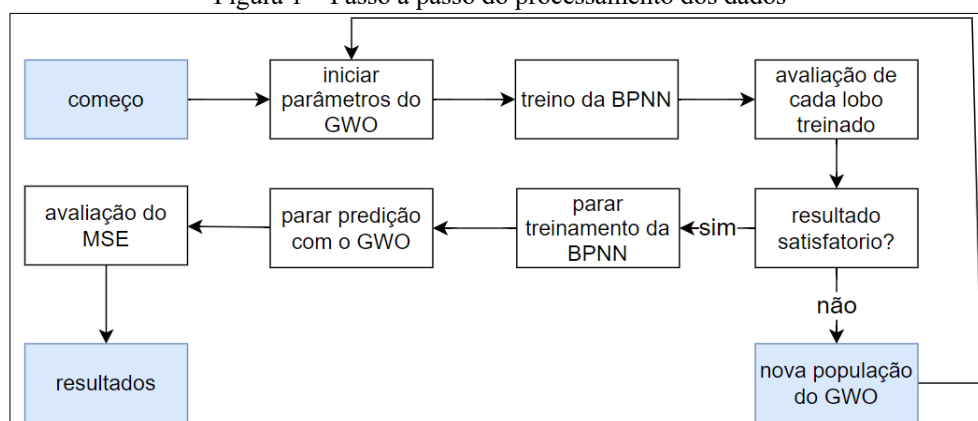
2.1 REAL ESTATE PRICE RANGE PREDICTION USING ARTIFICIAL NEURAL NETWORK AND GREY WOLF OPTIMIZER

O crescimento do mercado imobiliário, aliado à baixa popularidade do setor preditivo em comparação com o mercado de ações, motivou Al-Gbury e Kurnaz (2020) a desenvolverem uma rede neural para seleção dos melhores pares de características e, posteriormente, aplicá-las no GWO.

Segundo Al-Gbury e Kurnaz (2020), os preços de um imóvel são afetados por múltiplas características, ao qual torna difícil compreender quais são as mais importantes e com maior sensibilidade. Para entender isso, os autores utilizaram o GWO, que é um algoritmo meta-heurístico que tem funcionamento semelhante a uma alcateia, as características mais importantes são equivalentes aos lobos alpha, as com um pouco menos são as betas, já as menos relevantes são os ômegas.

De acordo com Al-Gbury e Kurnaz (2020), o processamento de dados passa por diversas etapas, iniciando pela definição dos parâmetros para aplicação do algoritmo GWO. Em seguida, o modelo é treinado em uma rede neural de retropropagação, que geralmente é composta por três camadas: camada de entrada, camada oculta (uma ou mais) e camada de saída. Posteriormente, são avaliados os pares de parâmetros e, caso o resultado seja positivo, interrompe o treinamento da rede neural e do algoritmo GWO. Neste caso, calcula-se o Erro Quadrático Médio (MSE) para avaliar a precisão do modelo e identificar a necessidade de ajustes nos parâmetros. A Figura 1 apresenta o esquema visual do processo de treinamento e seleção de parâmetros.

Figura 1 – Passo a passo do processamento dos dados



Fonte: Al-Gbury e Kurnaz (2020).

No processo de treinamento, Al-Gbury e Kurnaz (2020) utilizaram o conjunto de dados disponibilizado pela UCI do Sindian Dist. Segundo os autores, as autoridades distritais coletam o histórico da avaliação imobiliária, que é baseado na análise dos valores de mercado em Sindian Dist, localizada na cidade de New Taipei, em Taiwan.

Com base nos resultados obtidos pelo treinamento, Al-Gburi e Kurnaz (2020) concluíram que é necessário utilizar mais de uma camada oculta ou mais neurônios para obter resultados satisfatórios. Foram testadas quatro configurações diferentes de neurônios, sendo elas: 4, 5, 6 e 7. Cada um passou pelo mesmo processo e quantidade de iterações, tendo como resultados a matriz confusão apresentada na Tabela 1.

Tabela 1 – Matriz confusão dos resultados da aplicação da rede neural com determinado número de neurônios

Método	4 neurônios	5 neurônios	6 neurônios	7 neurônios
Verdadeiro Positivo	22	24	24	24
Verdadeiro Negativo	54	56	58	55
Falso Positivo	3	1	1	1
Falso Negativo	4	2	0	3

Fonte: adaptado de Al-Gburi e Kurnaz (2020).

Al-Gburi e Kurnaz (2020) concluem que para se obter a melhor performance da combinação dos algoritmos GWO e da rede neural de retropropagação são necessários 6 neurônios na camada oculta para se obter uma acurácia de 98%. Como trabalhos futuros, os autores sugerem a utilização de outros tipos de redes neurais, tais como a rede neural convolucional ou a rede neural recorrente, aos quais podem melhorar a extração de características nas etapas iniciais.

2.2 ARTIFICIAL INTELLIGENCE APPROACH FOR MODELING HOUSE PRICE PREDICTION

Para compreender quais fatores influenciam na precificação de imóveis, Cekic *et al.* (2022) utilizaram o método de seleção de características baseado na correlação. Cada característica foi analisada quanto à sua importância. Posteriormente, os autores selecionaram um conjunto com as melhores características. Os autores utilizaram duas bases de dados da cidade de Boston, sendo a primeira de instalações públicas, como escolas e parques para analisar o ambiente dos imóveis e uma outra contendo métricas como taxa de crime, número médio de habitantes, índice de acessibilidade às rodovias e taxa de imposto sobre a propriedade.

Cekic *et al.* (2022) utilizaram cinco algoritmos de aprendizado de máquina para encontrar a maneira mais eficiente de prever o preço de casas. Foram analisados os algoritmos de Regressão Linear (RL), Árvore de Decisão (AD), *XGBoost* (XGB), *Extra Trees* (ET) e Floresta Aleatória (FA).

Na avaliação do modelo, Cekic *et al.* (2022) utilizaram três métricas de avaliação, RMSE, Erro Absoluto Médio (MAE) e Coeficiente de Determinação (R^2). Nelas, segundo os autores, o preço previsto significa o preço de saída dos algoritmos e o preço real, o preço atual do imóvel. Para avaliar o desempenho do modelo, o erro para cada modelo é derivado e em seguida é feita uma média para n número de amostras.

Cekic *et al.* (2022) destacam que, ao utilizar todas as características disponíveis, os resultados obtidos não foram satisfatórios, uma vez que exigiu um grande tempo de treinamento e apresentou uma MAE de 76,59, 98,36, 71,87, 52,83 e 68,25 para a regressão linear, a floresta aleatória, a árvore de decisão, o *Extra Trees* e o *XGBoost*, respectivamente. Os resultados das métricas de RMSE e R^2 podem ser visualizados na Tabela 2.

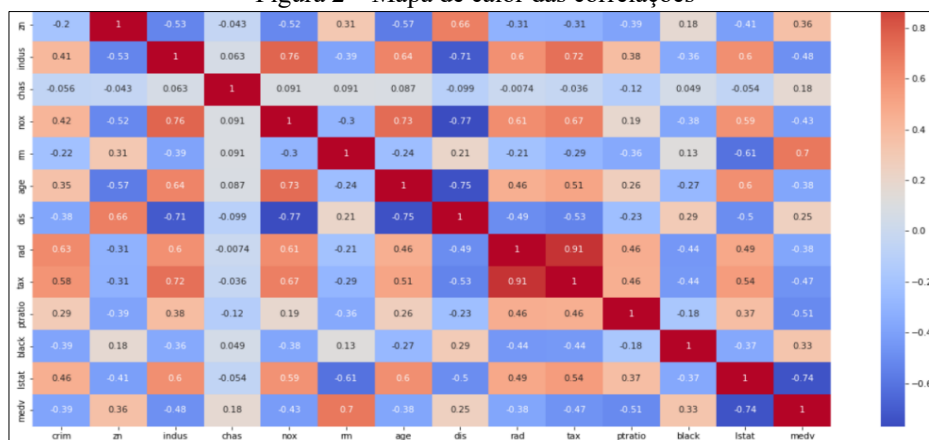
Tabela 2 – Comparativo dos resultados de cada algoritmo

Algoritmos	MAE	RMSE	R^2
Regressão Linear	76,59	56,09	0,87
Floresta Aleatória	98,36	76,98	0,70
Árvore de Decisão	71,87	59,93	0,92
<i>Extra Trees</i>	52,83	40,31	0,96
<i>XGBoost</i>	68,25	56,84	0,93

Fonte: Cekic *et al.* (2022).

Buscando melhorar a precisão do modelo, Cekic *et al.* (2022) analisaram a correlação de cada variável. Na Figura 2 é evidenciado a distribuição dos valores e a matriz de correlação das características utilizadas. No gradiente, vermelho e azul escuro representam alta correlação, enquanto azul claro e amarelo claro significam baixa correlação. A partir desses valores foi reaplicado os mesmos modelos em apenas metade das características. As características foram selecionadas com base em seus autovalores (altos).

Figura 2 – Mapa de calor das correlações



Fonte: Cekic *et al.* (2022).

Com a nova abordagem, Cekic *et al.* (2022) destacam que houve, em média, uma melhora de 8-12% na precisão. O modelo de regressão linear produziu o maior MAE, enquanto os outros modelos resultaram em um MAE semelhante. Do mesmo modo, para o RMSE, todos os modelos produziram valores semelhantes, exceto a regressão linear (35,58) e a árvore de decisão (39,65). O R^2 foi muito semelhante para todos os modelos. Os novos resultados podem ser visualizados na Tabela 3.

Tabela 3 – Comparativos dos resultados para cada algoritmo com a nova abordagem

Algoritmos	MAE	RMSE	R^2
Regressão Linear	23,87	35,58	0,90
Floresta Aleatória	10,93	39,65	0,91
Árvore de Decisão	10,44	21,75	0,91
Extra Trees	10,56	19,87	0,90
XGBoost	10,22	18,76	0,92

Fonte: Cekic *et al.* (2022).

Cekic *et al.* (2022) concluíram que os cinco algoritmos de aprendizado de máquina utilizados foram eficientes na predição de preços de imóveis. A utilização do método de seleção de características baseadas em seus autovalores resultou em métricas de erro significativamente inferiores, o que indica uma melhora na precisão das predições. Os autores também ressaltam que enquanto a primeira abordagem usou 14 características, a segunda foi mais assertiva utilizando apenas as 7 mais correlatas. Além disso, Cekic *et al.* (2022) destacam que a maior limitação do trabalho está relacionada à utilização de dados históricos, uma vez que foram utilizados dados relativamente antigos em comparação com o cenário atual da cidade.

2.3 MASS APPRAISAL OF REAL ESTATE PRICES USING IMPROVED BP NEURAL NETWORK WITH POLICY EVALUATION

Zheng *et al.* (2022) utilizaram uma *Back Propagation Neural Network* (BPNN) otimizada pelo algoritmo genético (GA) e o *Particle Swarm Optimization* (PSO), para avaliar os preços de imóveis levando em consideração as políticas imobiliárias regionais. Os algoritmos PSO e o GA possuem certa semelhança, o GA tem o foco na evolução dos conjuntos até chegar na melhor combinação. Já o PSO funciona como um conjunto de partículas que se movem em um espaço de busca multidimensional, onde cada partícula representa uma solução candidata para o problema.

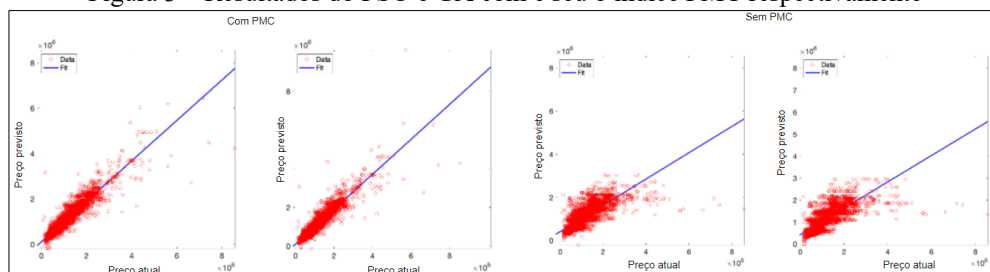
Segundo Zheng *et al.* (2022), os dados utilizados foram retirados do site do governo da cidade de Weihai - China, no qual 99 políticas públicas imobiliárias foram extraídas. Após realizar a limpeza dos dados, os autores aplicaram o índice Preço Máximo ao Consumidor (PMC), que é um modelo econométrico usado para medir o nível de política. Além disso, de acordo com os autores, o PMC é uma importante medida para avaliar a qualidade dos modelos de avaliação de preços imobiliário.

De acordo com Zheng *et al.* (2022), os resultados do índice Preço Máximo ao Consumidor (PMC) forneceram uma classificação das políticas públicas em dois conjuntos de variáveis: primárias (escopo de regulação e o nível de eficácia) e secundárias (serviços sociais e impostos). Além disso, Zheng *et al.* (2022) subdividiram a área urbana da cidade de Weihai em 31 partes com base em indicadores de ruas, pontos de referência e divisões administrativas. Em seguida, os autores escolheram um ou dois indicadores de cada parte da cidade, totalizando um total de 33 amostras com cerca de 27.802 dados de transação de 2010 a 2021.

O modelo proposto por Zheng *et al.* (2022) consiste em processar os dados através da BPNN, que os encaminha para os algoritmos GA e PSO, tendo como objetivo encontrar a melhor solução para o problema

apresentado. Ao término do processamento de cada algoritmo, a BPNN é aplicada novamente para a obtenção das métricas de avaliação R^2 , MSE e MAPE. Zheng *et al.* (2022) também consideraram a possibilidade de utilizar o desempenho do índice PMC para avaliar o desempenho dos algoritmos de GA e PSO durante o treinamento do modelo. Os resultados obtidos pelos algoritmos PSO e GA considerando o índice PMC podem ser observados na Figura 3. Ao qual, pode-se perceber que os valores estimados com PMC (em vermelho) ficam próximos dos valores atuais previstos para os imóveis (linha azul). E, sem utilizar o índice, ficam mais dispersos.

Figura 3 – Resultados do PSO e GA com e seu o índice PMC respectivamente



Fonte: Zhang *et al.* (2022).

Os resultados obtidos por Zheng *et al.* (2022), apresentados na Tabela 4, mostram que o algoritmo GA juntamente com a aplicação do índice PMC e a BPNN apresentaram, em média, menores índices de erro (0,00154) e com o R^2 de 0,9078 do que o PSO com o índice PMC e BPNN. O algoritmo PSO apresentou bons resultados na predição do preço de imóveis com apenas um treino. Já o GA, obteve a sua menor taxa de erro com três treinos. Em média, o algoritmo GA superou o PSO em todos os aspectos medidos para predição dos preços de imóveis.

Tabela 4 – Performance do GA-BPNN e PSO-BPNN

Quantidade de treinos	GA-BPNN			PSO-BPNN		
	MAPE	MSE	R^2	MAPE	MSE	R^2
1	0.1274	0.00182	0.9122	0.1197	0.00141	0.9001
2	0.1221	0.00153	0.9265	0.1320	0.00172	0.89
3	0.1164	0.00126	0.8912	0.1321	0.00187	0.8993
4	0.1233	0.00148	0.9114	0.1338	0.00161	0.8818
5	0.1252	0.00151	0.9150	0.1303	0.00168	0.88
6	0.1344	0.00156	0.8972	0.1328	0.00156	0.8671
7	0.1252	0.00148	0.9014	0.1361	0.0018	0.9097
8	0.1279	0.00171	0.9078	0.1319	0.00185	0.9052
Média	0,1252	0,00154	0,9078	0,1311	0,00169	0,8917

Fonte: Zheng *et al.* (2022).

Zheng *et al.* (2022) concluem que os algoritmos GA e o PSO minimizaram os problemas de *overfitting* na rede neural, fornecendo uma maior rapidez no treino devido a quantidade mais assertiva dos dados entregues ao modelo e melhores acurácias decorridas da implementação do índice PMC. Além disso, os autores afirmam que o modelo construído consegue se adaptar corretamente as políticas locais, índices de vizinhança, características dos imóveis e refletir em predições de preços mais precisas. Zheng *et al.* (2022) também ressaltam que a separação das políticas em duas categorias eliminou a influência de fatores artificiais como a especulação imobiliária.

3 PROPOSTA

A seguir é apresentada a justificativa para o desenvolvimento desse trabalho, os principais requisitos e a metodologia de desenvolvimento que será utilizada. Também são relacionados os assuntos e as fontes bibliográficas que irão fundamentar o estudo proposto.

3.1 JUSTIFICATIVA

No Quadro 1 é possível observar as principais características de cada trabalho correlato apresentado, onde as colunas são cada trabalho e as linhas as características, com elas pode-se entender como será solucionado os problemas propostos por este trabalho.

Quadro 1 - Comparativo dos trabalhos correlatos

Trabalhos Correlatos Características	Al-Gburi e Kurnaz (2020)	Cekic <i>et al.</i> (2022)	Zheng <i>et al.</i> (2022)
Problema a ser resolvido	Facilitar a tomada de decisões ao comprar imóveis	Facilitar a tomada de decisões ao comprar imóveis	Avaliação em massa de preços de imóveis
Objetivo	Utilizar redes neurais e o GWO para predição de preços de imóveis	Predição de preços de imóveis	Apresentar uma abordagem para avaliação em massa de preços de imóveis
Fonte das características	Localização	Índices da cidade	Políticas públicas
Datasets	Cidades de <i>New Taipei</i> e <i>Sindian Dist</i> , Taiwan	Cidade de Boston, EUA	Síte do governo da cidade de Weihai
Técnicas utilizadas	Redes neurais e GWO	Regressão linear, árvore de decisão, XGBoost, ExtraTrees e FA.	Redes neurais, algoritmo genético e algoritmo PSO
Métricas de avaliação	Acurácia e MSE	RMSE, MAE e R ²	R ² , MSE e MAPE
Resultados	98% de acurácia	MAE para RL: 23,87 MAE para FA: 10,44 MAE para XGBoost: 10,22	MAPE: 0,1311 MSE: 0,00169 R ² : 0,8917

Fonte: elaborado pelo autor.

A partir do Quadro 1 é possível observar que os três trabalhos utilizam técnicas diferentes para estimar o preço dos imóveis. Al-Gburi e Kurnaz (2020) propõem o uso de redes neurais juntamente com o algoritmo GWO para conseguir se adaptar a diversos parâmetros da definir a melhor opção de compra. Já Cekic *et al.* (2022) utilizaram cinco métodos de aprendizado de máquina, tendo como o intuito encontrar o mais eficiente afim de auxiliar melhor os compradores. Zheng *et al.* (2022) propuseram uma abordagem com a utilização BPNN otimizada pelos algoritmos GA e PSO para se trabalhar com uma quantidade massiva de dados de preços de imóveis.

Zheng *et al.* (2022) utilizaram dados de políticas públicas da cidade de Weihai, na China. Cekic *et al.* (2022) utilizaram dados da cidade de Boston. Já Al-Gburi e Kurnaz (2020) buscaram entender como a localização afeta os preços em cidades de Taiwan. Todos os trabalhos possuem um objetivo comum, que é a predição de preços de imóveis, porém cada um com um contexto específico.

Dentre as técnicas apresentadas pelos autores, destacasse o GWO que consegue se adaptar corretamente ao contexto das características fornecidas e se regular para alcançar melhores resultados. Cekic *et al.* (2022) também apresentam uma abordagem inovadora na análise de imagens do ambiente urbano da cidade de Boston, ao utilizar o XGBoost. Esse modelo de aprendizado de máquina permitiu obter uma gama mais ampla de características, contribuindo para uma análise mais precisa e abrangente das informações obtidas.

Al-Gburi e Kurnaz (2020) obtiveram uma taxa de acurácia de 98%. Cekic *et al.* (2022) alcançaram o erro médio absoluto (MAE) de 10,22 para o algoritmo *XGBoost*, um resultado bom em comparação com os 23,87 da regressão linear. Zheng *et al.* (2022) conseguiram atingir um MSE abaixo de 0,00169 com o algoritmo PSO.

Diante deste cenário, este trabalho também compartilha do mesmo objetivo, todavia visando utilizar características regionais, buscando entender como elas afetam os preços e quais características são mais importantes. O foco será contribuir com o mercado imobiliário, investidores, corretores e vendedores de imóveis, disponibilizando um modelo de precificação de imóveis que tenha um embasamento em dados concretos para os seus resultados. Sua diferenciação será na maneira como os dados serão extraídos dos sites, o contexto com foco central na cidade de Blumenau e no uso de um modelo computacional meta-heurístico, como o *Grey Wolf Optimizer* (GWO), que é um método de otimização desenvolvido recentemente, que tem sido aplicado com sucesso em diferentes áreas de pesquisa. Nesse sentido, o uso deste modelo pode fornecer *insights* úteis e inovadores sobre como os fatores regionais influenciam o mercado imobiliário de forma mais precisa e eficiente do que métodos tradicionais. Além disso, a análise georreferenciada das áreas com maior concentração de imóveis à venda e a comparação de preços médios com outras áreas da cidade permite uma compreensão mais completa do mercado imobiliário local e das dinâmicas que o afetam. Essas informações são valiosas tanto para investidores e desenvolvedores imobiliários quanto para órgãos governamentais e urbanistas que buscam entender e planejar o desenvolvimento urbano da cidade, no caso, de Blumenau.

Por fim, compreende-se que um dos principais desafios deste trabalho está relacionado ao uso de múltiplas variáveis e à extração dos dados, uma vez que as informações sobre as características dos imóveis podem não estar organizadas nos portais, o que resulta em menos informações relevantes disponíveis para possíveis predições. Para transpor tais dificuldades, serão empregadas técnicas de análise de dados e inteligência artificial. Com isso, espera-se que o artefato computacional proposto possa auxiliar os vendedores de imóveis a definir um preço mais preciso,

levando em consideração diversas características do imóvel, sem a necessidade de um intermediário para a precificação.

3.2 REQUISITOS PRINCIPAIS DO PROBLEMA A SER TRABALHADO

Os Requisitos Funcionais (RF) e Requisitos Não Funcionais (RNF) do artefato computacional foram divididos em três partes: (i) módulo para extração e tratamento dos dados; (ii) modelo de precificação e; (iii) módulo de visualização georreferenciada.

O módulo de extração e tratamento de dados deverá:

- a) acessar os portais de imóveis e extrair dados relevantes, como preço, localização, tamanho etc. (RF);
- b) correlacionar, limpar e normalizar as informações de múltiplas bases de dados (RF);
- c) garantir que a extração de dados não seja bloqueada ou proibida pelos portais de imóveis (RNF);
- d) ser feito utilizando a linguagem Python e as bibliotecas Pandas e *Scikit-Learn* (RNF);
- e) utilizar o banco de dados MongoDB para persistir os dados (RNF).

O modelo de precificação deverá:

- a) realizar uma análise exploratória dos dados para entender as características dos imóveis e como a região do imóvel afeta os preços (RF);
- b) utilizar algoritmos meta-heurísticos, como o GWO, para estabelecer a precificação imobiliária (RF);
- c) estabelecer a acurácia, R^2 e MSE do modelo (RF);
- d) processar grandes volumes de dados rapidamente e de forma eficiente (RNF);
- e) ser desenvolvido utilizando a linguagem Python e as bibliotecas Pandas e *Scikit-Learn* (RNF).

O módulo de visualização georreferenciada deverá:

- a) disponibilizar um mapa de calor com a precificação de cada região (RF);
- b) fornecer o histórico de densidade demográfica para que o usuário visualize o crescimento imobiliário e da região nos últimos anos (RF);
- c) permitir que o usuário possa informar a localização ou características do imóvel antes de visualizar a precificação (RF);
- d) utilizar a API Leaflet para a disponibilização do mapa de pontos ou de calor (RNF);
- e) desenvolver a aplicação web utilizando Python e FastAPI para o *backend* e React no *frontend* (RNF).

3.3 METODOLOGIA

O trabalho será desenvolvido observando as seguintes etapas:

- a) levantamento bibliográfico: pesquisar e estudar sobre mercado imobiliário, algoritmos meta-heurísticos, georreferenciamento e trabalhos correlatos;
- b) busca por bases dados: pesquisar bases de dados que possam estabelecer o perfil da região (imobiliário ou socioeconômico);
- c) coleta de dados: coletar dados de portais imobiliários realizando a raspagem dos dados (*webscrapping*) utilizando a linguagem Python e a biblioteca *beautifulsoup*;
- d) montagem da base de dados: com as bases nos itens (b) e (c) fazer a correlação dos dados, utilizando a linguagem Python, biblioteca Pandas e banco de dados MongoDB;
- e) tratamento de dados: limpar a base de dados deixando apenas os documentos consistentes (sem informações faltantes), utilizando a linguagem Python e a biblioteca *Scikit-Learn*;
- f) definição das variáveis preditoras: definir, de forma *ad hoc*, quais variáveis possuem maior relevância e caracterizam a região e o mercado imobiliário de Blumenau;
- g) definição dos algoritmos: definir quais algoritmos meta-heurísticos serão utilizadas no desenvolvimento do modelo, com base na etapa (f). Inicialmente, pretende-se usar o GWO;
- h) desenvolvimento do modelo: a partir do que foi formalizado nos itens (e) (f) e (g), realizar o carregamento e preparação dos dados, assim como a implementação do algoritmo meta-heurístico, utilizando a linguagem Python e a biblioteca Pandas e *Scikit-Learn*;
- i) validação do modelo: avaliar o modelo considerando os resultados das previsões e os indicadores de performance, como a sensibilidade, a especificidade e a taxa de assertividade geral em relação a precificação feitas por corretores. Para isso, pretende-se utilizar as métricas como MSE e R^2 .
- j) implementação da visualização georreferenciada: realizar a implementação da aplicação, permitindo que o usuário possa interagir com a API de mapa, recebendo *feedback* na forma de um mapa de calor. A aplicação será implementada utilizando Python e FastAPI para o *backend* e React no *frontend*;
- k) testes da visualização: avaliar em conjunto com corretores se a indicação dos locais e das precificações estão coerentes com a realidade local. Elaborar testes para validar a usabilidade da aplicação.

As etapas serão realizadas nos períodos relacionados no Quadro 2.

Quadro 2 - Cronograma de atividades a serem realizadas

etapas / quinzenas	2023									
	jul.		ago.		set.		out.		nov.	
	1	2	1	2	1	2	1	2	1	2
levantamento bibliográfico										
busca por bases dados										
coleta de dados										
montagem da base de dados										
tratamento de dados										
definição das variáveis preditoras										
definição dos algoritmos										
desenvolvimento do modelo										
validação do modelo										
implementação da visualização georreferenciada										
testes da visualização										

Fonte: elaborado pelo autor.

4 REVISÃO BIBLIOGRÁFICA

Esta seção tem como objetivo explorar os principais assuntos que fundamentarão o estudo a ser realizado neste projeto. A subseção 4.1 aborda mercado imobiliário e a precificação de imóveis. Já a seção 4.2 discute georreferenciamento. Por fim, a seção 4.3 discorre sobre algoritmos meta-heurísticos.

4.1 MERCADO IMOBILIÁRIO E PRECIFICAÇÃO DE IMÓVEIS

O mercado imobiliário refere-se a um dos maiores setores econômicos de um país, ele engloba a compra, venda, aluguel e administração de imóveis, onde cada um possui um valor agregado. Os seus preços podem ser influenciados por diversos fatores, como a economia, as políticas governamentais, as tendências de mercado, localização dos imóveis e métricas como a criminalidade regional. Além disso, Pindyck (1994, p. 29) destaca que a demanda e oferta pode interferir no preço de um produto que está sendo comercializado, este mesmo pensamento pode ser aplicado ao contexto imobiliário na avaliação de imóveis.

Segundo Sales (2023), os corretores de imóveis utilizam vários métodos para precificar imóveis, sendo normalmente levado em consideração vários fatores. Ainda de acordo com o autor, as metodologias mais comuns podem ser separadas em duas abordagens (i) direta no qual é feito um comparativo dos dados de mercado disponíveis ou também tendo como base o custo de reprodução de benfeitorias; (ii) indireto que utiliza os métodos de renda, involutivo e o residual.

De acordo com González (1997, p. 36), o método mais utilizado é o de comparação direta com imóveis similares. Neste contexto, torna-se necessário a existência de dados de transações semelhantes em número e especificações razoáveis para se ter confiabilidade. Outro método que se destaca é o involutivo que, segundo a ABNT NBR 14653-2 (2011, p.17), é dividido em etapas como vistoria, projeto hipotético, pesquisa de valores, previsão de receitas, levantamento de custo, previsão de despesas, margem de lucro, prazos, taxas e modelo do imóvel que ao fim chegam ao valor de mercado.

Wheaton (1999) destaca que a experiência do corretor de imóveis pode influenciar significativamente no preço final de um imóvel. Um corretor experiente e bem-informado sobre o mercado local pode fornecer uma avaliação mais precisa do valor de um imóvel, levando em consideração todos os fatores relevantes, como localização, características do imóvel, tendências de mercado e demanda. Ainda segundo Wheaton (1999), a experiência do corretor pode ser especialmente importante em mercados imobiliários dinâmicos e em constante mudança, onde a capacidade de acompanhar as tendências e prever mudanças pode ser crucial para a precificação precisa de imóveis.

Com o avanço das tecnologias nos últimos anos, é possível obter análises e conclusões que auxiliam nas decisões acerca de um imóvel utilizando de sistemas georreferenciados (FELIPETTO; ELIAS, 2010, p. 80). Também segundo Carminato (2018), essa estratégia permite a pessoa compradora ou vendedora entender como o mercado está ao seu redor, aplicando filtros sobre a sua região é possível identificar as forças e fraquezas e então poder tomar as melhores decisões.

4.2 GEORREFERENCIAMENTO

Nos últimos anos, o mercado imobiliário tem passado por mudanças significativas impulsionadas pelo crescimento acelerado da sociedade contemporânea. Nesse contexto, a demanda por informações georreferenciadas tem se tornado cada vez mais relevante. De acordo com Andrade *et al.* (2020), essa necessidade

decorre da busca por uma produção mais eficiente e da compreensão dos fatores geográficos que influenciam o valor e a atratividade de uma propriedade.

O georreferenciamento, por meio de ferramentas e técnicas específicas, permite mapear e analisar a localização dos imóveis, fornecendo informações valiosas sobre sua proximidade a serviços essenciais, como escolas, parques, transporte público e centros comerciais. Esses fatores geográficos desempenham um papel crucial na determinação do valor de uma propriedade, influenciando a decisão tanto de compradores quanto de vendedores. De acordo com Câmara e Ortiz (1998, p. 2), uma aplicação georreferenciada deve ter interface com usuário, entrada e integração de dados, funções de processamento gráfico e de imagens, visualização e plotagem e banco de dados geográficos. Através do georreferenciamento, é possível acessar dados precisos sobre imóveis comparáveis, facilitando a comparação de preços, características e tendências de mercado. Isso possibilita uma avaliação mais precisa do valor de uma propriedade e uma melhor orientação na definição de preços competitivos (CARMINATO, 2018).

Também segundo Carminato (2018) georreferenciar uma propriedade permite também entender como e quais fatores regionais interferem naquele imóvel, com as coordenadas é possível saber se o imóvel está em uma área perigosa em relação a criminalidade o que pode ser um ponto de atenção a um possível comprador. Além disso, ainda de acordo com os autores, a georreferenciação de propriedades fornece uma visão abrangente dos fatores regionais que podem afetar um imóvel. Ao analisar as coordenadas geográficas, é possível identificar se a propriedade foi previamente afetada por eventos como enchentes e deslizamentos, o que pode desencadear preocupações legítimas para potenciais compradores.

4.3 META-HEURÍSTICA

A meta-heurística está contida na categoria de algoritmos de aproximação, que segundo Blum e Roli (2003), visa combinar métodos heurísticos básicos em estruturas de nível superior com o objetivo de explorar eficiente e efetivamente um espaço de busca. O termo foi introduzido pela primeira vez por Glover, Laguna e Marti (1986), que deriva da composição de duas palavras gregas, no qual Heurística significa "encontrar", enquanto o sufixo meta significa "além, em um nível superior".

Segundo Garey e Johnson (1979), os algoritmos de aproximação são comumente utilizados em problemas de otimização combinatória, sendo classificados de acordo com a classe de complexidade computacional. Tal classificação é definida de acordo com o tempo de processamento polinomial de um problema. Neste caso, algoritmos meta-heurísticos tem classificação difícil, ou com a notação padrão *NP-Completo*.

Para Cekic *et al.* (2022), Meta-heurística é um procedimento que tem como foco problemas de otimização combinatória, sendo capazes de gerar ou selecionar uma heurística que pode fornecer uma solução suficientemente boa para um problema que possua uma grande quantidade de características ou parâmetros. Ainda segundo os autores, os algoritmos meta-heurísticos podem ser úteis no mercado imobiliário, pois eles são capazes de lidar com problemas de otimização complexa, como é o caso da precificação de imóveis. Esses algoritmos são amplamente utilizados em diferentes áreas da ciência, incluindo a análise de dados imobiliários.

De acordo com Liakos *et al.* (2018), os algoritmos de Aprendizado de Máquina (AM) são tipicamente classificados em diferentes categorias, dependendo do tipo de aprendizado, técnica e dos modelos de AM utilizados, juntamente com a base de dados empregada. Algumas das classificações comuns de um algoritmo de AM são: aprendizado supervisionado que ocorre quando se tem uma base com os dados já rotulados e é possível fazer a comparação com os valores gerados pelo algoritmo, outro modelo é o aprendizado não supervisionado que ao contrário do supervisionado não possui uma base de dados já rotulada, o último modelo seria o aprendizado por reforço que consiste no algoritmo receber estímulos positivos ou negativos de acordo com a sua resposta e então analisar se sua decisão foi boa ou não (LUDERMIR, 2021).

Nos últimos anos vários estudos têm aplicado técnicas que abordam os conceitos de aprendizado de máquina e meta-heurística para problemas combinatórios (FONSECA, 2018). O GWO é um algoritmo relativamente novo que tem ganhado popularidade nos últimos anos, devido à sua capacidade de otimização e adaptação a diferentes conjuntos de dados. A combinação do GWO com outros algoritmos de aprendizado de máquina como as redes neurais, pode se levar a resultados mais precisos e confiáveis (AL-GBURY; KURNAZ, 2020).

De acordo com Al-Gburi e Kurnaz (2020), o GWO possui três estágios. No primeiro, os lobos seguem e se aproximam da presa, nela os lobos de camadas inferiores avaliam se os lobos *beta* e *alpha* tem as melhores soluções. Na segunda etapa, chegou a hora de perseguir e cercar a presa até que ela congele, onde os lobos assumem as melhores posições e encurralam a presa até que sua posição não mude mais. A última etapa é a de ataque, ocasionada pelos melhores resultados encontrados para a situação. Contudo, através do uso de algoritmos meta-heurísticos, é possível encontrar um conjunto de características que influenciam o preço dos imóveis e, em seguida, determinar o preço adequado para uma determinada propriedade.

REFERÊNCIAS

- ABNT NBR 14653-2 - **Avaliação de bens - Parte 2: Imóveis urbanos**. Rio de Janeiro: Associação Brasileira de Normas Técnicas, 2011.
- ANDRADE, Thayllon Silva. Et al. **O Georreferenciamento como ferramenta de auxílio para agricultura brasileira**. Revista Científica Multidisciplinar Núcleo do Conhecimento. Ano 05, Ed. 02, Vol. 04, pp. 155-164. Fevereiro de 2020. ISSN: 2448-0959, Link de acesso: <https://www.nucleodoconhecimento.com.br/engenharia-de-producao/agricultura-brasileira>.
- AL-GBURY, Omar; KURNAZ, Sefer. Real Estate Price Range Prediction Using Artificial Neural Network And Grey Wolf Optimizer. **IEEE: 2020 4th International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT)**, Istanbul, Turkey, ed. 4, p. 1-5, 2020.
- ARON, Janine *et al.* **Credit, housing collateral, and consumption**: evidence from Japan, the UK, and the US. Review of Income and Wealth, v. 58, n. 3, p. 397-423, 2012. DOI:10.1111/j.1475-4991.2011.00466.x.
- BLUM, Christian; ROLI, Andrea **Metaheuristics in combinatorial optimization**: Overview and conceptual comparison. ACM Computing Surveys, v. 35, n. 3, p. 268-308, set. 2003.
- CÂMARA, Gilberto; ORTIZ, Manoel. J. Sistemas de informação geográfica para aplicações ambientais e cadastrais: uma visão geral. In: CONGRESSO BRASILEIRO DE ENGENHARIA AGRÍCOLA: CARTOGRAFIA, SENSORIAMENTO E GEOPROCESSAMENTO, 27. 1998, Poços de Caldas, MG. **Anais [...]**. Poços de Caldas – MG: DPI/INPE, 1998. Disponível em: <http://www.dpi.inpe.br/gilberto/papers/analise.pdf>. Acesso em: 16 Abr. 2023.
- CAMPOS, Suelen Ferreira. **Precificação De Imóveis E Seus Elementos Agregadores De Valor Sob A Visão Do Consumidor**: Uma análise do mercado imobiliário de João Pessoa-PB. Orientador: Prof. Pierre Lucena. 2014. 1 f. Projeto de Dissertação de Mestrado (Mestrado em Administração) - Universidade Federal de Pernambuco, Pernambuco, 2014.
- CARMINATO, Larissa Porteiro. Uso de Geoprocessamento no Mercado Imobiliário. **Revista Mackenzie**, [s. l.], 2018. Disponível em: https://www.mackenzie.br/fileadmin/ARQUIVOS/Public/1-mackenzie/universidade/laboratorios/labgeo/2018/Ajustes/ARTIGO_-_USO_DE_GEOPROCESSAMENTO_NO_MERCADO_IMOBILIARIO.pdf. Acesso em: 19 jun. 2023.
- CEKIC, Melih Sah *et al.* Artificial Intelligence Approach for Modeling House Price Prediction. 2022. **2022 3rd International Conference on Artificial Intelligence and Data Sciences (AiDAS)**, Istanbul, Turkey, 15-16 July 2022. pp. 1-5, doi: 10.1109/ICMI55296.2022.9873784.
- FELIPETTO, Henrique dos Santos; ELIAS, Adão Robson. O que é Geomarketing e suas aplicações. In: SEMANA DA GEOMÁTICA, 3.. 2010, Santa Maria. **Anais Eletrônicos...** Santa Maria: UFSM, 2010. p.79-85. Disponível em: <[http://www.alice.cnptia.embrapa.br/bitstream/doc/883379/1/Anais3SemanaDaGeomati ca.pdf#page=91](http://www.alice.cnptia.embrapa.br/bitstream/doc/883379/1/Anais3SemanaDaGeomati%20ca.pdf#page=91)>. Acesso em: 13 mar. 2016.
- FONSECA, Thiago H. Heurística para sintonização de meta-heurísticas aplicada ao Problema de FlowShop Permutacional. Dissertação (Mestrado em Ciências da Computação) — Universidade Federal do Maranhão, São Luís, 2018.
- GAREY, Michael R; JOHNSON, David S. Computers and Intractability: **A Guide to the Theory of NP-Completeness**. San Francisco: W.H. Freeman and Company, 1979.
- GLOVER, Fred; LAGUNA, Manuel; MARTI, Rafael Scatter Search and Path Relinking: Advances and Applications. In **Handbook of Metaheuristics**, F. Glover and G. Kochenberger, Eds. International Series in Operations Research & Management Science, vol. 57, 1986.
- GONZÁLEZ, Marcos Aurélio Stumpf. **A Engenharia de Avaliações na Visão Inferencial**. São Leopoldo: Unisinos, 1997. 142 p.
- GREEN, Richard K. *et al.* Metropolitan-specific estimates of the price elasticity of supply of housing, and their sources. **The American Economic Review**, v. 95, n. 2, p. 334-339, 2005.
- LIAKOS, Konstantinos G. *et al.* Machine learning in agriculture: A review. **Sensors**, v. 18, n. 8, 2018.
- LUDERMIR, Teresa B.. Inteligência Artificial e Aprendizado de Máquina: estado atual e tendências. **Estudos Avançados**, v. 35, n. 101, p. 85–94, jan. 2021.
- PINDYCK, Robert S. **Microeconomia**. São Paulo: Makron, 1994. 968 p.
- ROVERI, Claudia. Proposta de Implantação de um Observatório do Mercado Imobiliário em Blumenau. Prefeitura Municipal de Blumenau, Blumenau, 2020.
- SALES, Francisco Janilson Lopes. **Precificação de Imóveis**: Um estudo de Caso. Orientador: Dr. Francisco Isidro Pereira. 2013. 56 p. Trabalho de Conclusão de Curso (Graduação em Administração) - Universidade Federal do Ceará, Fortaleza - CE, 2013. Disponível em: https://repositorio.ufc.br/bitstream/riufc/29751/1/2013_tcc_fjsales.pdf. Acesso em: 16 jun. 2023.
- SIDRIM, Pedro Daniel BRAZ. **O mercado imobiliário Brasileiro**: Discussão sobre a existência de uma possível bolha imobiliária. Orientador: Prof. Paulo Neto. 2013. 1 f. Trabalho de Conclusão de Curso (Bacharel em Ciências Econômicas) - Universidade Federal do Ceará, Fortaleza, 2013.

TEIXEIRA, Clarissa., *et al.* (2016). Ecosistema de inovação na educação de Santa Catarina. Vieira, M. S.; Teixeira, C. S. T.; Ehlers, A. C.T.(Orgs). **Educação fora da caixa**, volume 2, p. 11-30. Disponível em: <<https://goo.gl/JtXxUb>>. Acesso em: 09 de maio 2023.

WHEATON, William C. Real estate “cycles”: some fundamentals. *Real Estate Economics*, **v. 27, n. 2**, p. 209-230, 1999.

ZHENG, Yikai *et al.* Mass Appraisal of Real Estate Prices Using Improved BP Neural Network with Policy Evaluation. **2022 IEEE Conference on Telecommunications: Optics and Computer Science (TOCS)**, Dalian, China, p. 1036-1041, 11 dez. 2022.