



Análises Imobiliárias com Inteligência Artificial para Investidores

RÚBEN ALEXANDRE ALBUQUERQUE SANTOS

(Grau de Mestrado)

Dissertação para obtenção do Grau de Mestre em Engenharia Informática e Multimédia

Orientadora: Doutor Artur Jorge Ferreira

Júri:



Análises Imobiliárias com Inteligência Artificial para Investidores

RÚBEN ALEXANDRE ALBUQUERQUE SANTOS

(Grau de Mestrado)

Dissertação para obtenção do Grau de Mestre em Engenharia Informática e Multimédia

Orientadora: Doutor Artur Jorge Ferreira, ISEL

Júri:

Agradecimentos

Declaração de integridade

Declaro que este trabalho de projeto é o resultado da minha investigação pessoal e independente. O seu conteúdo é original e todas as fontes listadas nas referências bibliográficas foram consultadas e estão devidamente mencionadas no texto. Mais declaro que todas as referências científicas e técnicas relevantes para o desenvolvimento do trabalho estão devidamente citadas e constam das referências bibliográficas.

O autor
Lisboa, 25 de Setembro de 2024

Resumo

Neste projeto, propõe-se uma solução para o problema de previsão de preços imobiliários no mercado português, que é volátil e influenciado por múltiplas variáveis macroeconómicas. Este problema é relevante porque os preços dos imóveis dependem de fatores como as taxas de juro, inflação e o rendimento por metro quadrado, tornando a tomada de decisões de investimento uma tarefa complexa e arriscada.

A solução proposta consiste no desenvolvimento de uma aplicação web que utiliza modelos de aprendizagem automática, como Regressão Linear, Random Forest, Redes Neurais e LSTM, para prever preços imobiliários em diferentes concelhos de Portugal. A aplicação permite a visualização interativa de dados e previsões, facilitando a análise de tendências de mercado e apoiando os investidores no processo de tomada de decisão.

Como resultado da implementação desta solução, os investidores têm acesso a uma ferramenta que lhes proporciona uma análise detalhada e personalizável do mercado imobiliário, permitindo-lhes identificar áreas de investimento com maior potencial de rentabilidade. Além disso, a avaliação do feedback dos utilizadores sugere possíveis melhorias e a validação da aplicação como uma ferramenta eficaz para o apoio ao investimento.

O estudo conclui que a integração de modelos de aprendizagem automática numa ferramenta prática é viável e oferece suporte eficaz para investidores imobiliários, fornecendo previsões claras e personalizáveis para o mercado português.

Palavras-chave: Aprendizagem Automática, Indicadores de Mercado, Mercado Imobiliário, Modelos de previsão, Modelos de regressão

Abstract

This project proposes a solution to the problem of predicting real estate prices in the Portuguese market, which is volatile and influenced by multiple macroeconomic variables. This problem is relevant because real estate prices depend on factors such as interest rates, inflation, and revenue per square meter, making investment decision-making a complex and risky task.

The proposed solution involves the development of a web application that leverages machine learning models, including Linear Regression, Random Forest, Neural Networks, and LSTM, to forecast real estate prices across different municipalities in Portugal. The application provides interactive data visualizations and forecasts, facilitating market trend analysis and supporting investors in their decision-making process.

As a result of implementing this solution, investors gain access to a tool that offers detailed and customizable real estate market analysis, allowing them to identify investment areas with higher potential profitability. Additionally, user feedback evaluation suggests possible improvements and validates the application as an effective investment support tool.

The study concludes that integrating machine learning models into a practical tool is feasible and provides effective support for real estate investors, offering clear and customizable forecasts for the Portuguese market.

Keywords: Machine Learning, Market Indicators, Real Estate Market, Forecast Models, Regression Models

Índice

Ín	dice	de Fig	guras	xvii			
Ín	dice	de Ta	belas	xix			
\mathbf{G}	lossá	rio		xxi			
Si	glas		2	xxiii			
1 Introdução							
	1.1	Enqua	adramento	1			
	1.2	Objet	ivos	3			
	1.3	Abord	lagem	4			
	1.4	Organ	nização do documento	6			
2	Rev	isão d	la literatura	9			
	2.1	Carac	terização do setor imobiliário	9			
	2.2	As pri	imeiras abordagens	11			
		2.2.1	Fatores determinantes	11			
		2.2.2	Análise de séries temporais	14			
		2.2.3	Análise de Regressão	15			
		2.2.4	Modelo de Preços Hedónico	15			
		2.2.5	Algoritmos de inteligência artificial	16			
		2.2.6	Métodos de previsão econométricos	16			
		2.2.7	Métodos de previsão da análise espacial	17			
	2.3	Avalia	ıção dos métodos	18			
	2.4	Sumá	rio	19			
3	Dac	los, M	etodologias e Métricas	21			
	3.1	Fontes	s de Dados	21			
		3.1.1	Instituto Nacional de Estatística (INE)	22			
		3.1.2	Casafari	23			
		3.1.3	DoutorFinanças	23			
		3.1.4	Banco de Portugal	24			
		3.1.5	Qualidade dos Dados e Limitações	25			
	3.2	Proces	ssamento de Dados	25			
		3 2 1	Pré-processamento de Dados	25			

		3.2.2	Transformação de Dados	26
		3.2.3	Integração de Dados	27
	3.3	Anális	ses dos Dados	28
		3.3.1	Exploração de Dados (EDA)	28
		3.3.2	Correlação entre as Variáveis	30
		3.3.3	Insights Baseados em Dados	30
	3.4	Desen	volvimento de Modelos de Aprendizagem	32
		3.4.1	Principais Métricas de Avaliação	33
		3.4.2	Ajuste de Dados- Etapas Prévias ao Treino dos Modelos	33
		3.4.3		37
		3.4.4	Random Forest	39
		3.4.5	Redes Neurais Artificiais: MLP (Multi Layer Perceptron)	42
		3.4.6	Redes Neurais Recorrentes: LSTM (Long Short-Term Memory) .	45
	3.5	Comp	aração dos resultados dos Modelos	47
		3.5.1	Comparação de Modelos- Resultados Gerais	48
		3.5.2	Escolha do Melhor Modelo	49
	ъ			
4			• •	51
	4.1			51
	4.2			52
	4.3	_		53
		4.3.1		53 - 1
		4.3.2		54
	4 4	4.3.3		56
	4.4	_		57
		4.4.1		57 50
		4.4.2	-	58
	4 5	4.4.3		58 50
	4.5	Avalia	3	59
		4.5.1	Resultados da Avaliação	60
5	Ava	liação	experimental	61
	5.1	Anális	se e comparação do desempenho dos Modelos	61
	5.2	Avalia	ção da Aplicação Web	62
		5.2.1	Participantes	62
		5.2.2	Feedback de Utilizadores	62
		5.2.3	Participantes	63
		5.2.4	Feedback de Utilizadores	63
		5.2.5	Melhorias Futuras da Aplicação	63
	5.3	Discus	ssão de Resultados	64
c	C -	-1~	The ball of Buttones	- م
6				65 65
	6.1			65 ce

Bibliografia 67

Apêndices

Índice de Figuras

1.1	Evolução projetada do setor imobiliário pelo Statista (2024-2028) [5]
1.2	Etapas de Aquisição de Dados e Desenvolvimento de Modelos
1.3	Processo de Desenvolvimento da Aplicação
2.1	Caraterização do bem imobiliário [20]
2.2	Métodos de previsão e avaliação de bens imóveis [5]
3.1	Distribuição dos preços dos imóveis por região
3.2	Distribuições de preços médios de venda por concelho
3.3	A matriz de correlação
3.4	Relação Preço/Venda e Arrendamento por Região
3.5	Volatilidade dos Preços de Venda e arrendamento nas Regiões de Oeiras, Odi-
	velas, Amadora, Sintra, Loures e Cascais
3.6	Preço media de venda após a normalização $\dots 3^2$
3.7	Localização da divisão dos Dados
3.8	Divisão Temporal com TimeSeriesSplit (5 Folds)
3.9	Previsão do modelo de Regressão Linear
3.10	Top 20 variáveis mais importantes do modelo Regressão Linear
3.11	Performance do modelo Random florest
3.12	Top 20 variáveis mais importantes do modelo Random Forest
3.13	Performance do modelo Redes Neurais Artificiais (MLP)
3.14	Performance do modelo Redes Neurais Artificiais
3.15	Performance do modelo Redes Neurais Artificiais
3.16	Performance do modelo Redes Neurais Artificiais
4.1	Esboço em papel da aplicação
4.2	da aplicação
4.3	Mockups da aplicação
4.4	Arquitetura da aplicação
4.5	Primeira Pagina da aplicação implementada

Índice de Tabelas

3.1	Resultados do Modelo de Regressão Linear	37
3.2	Resultados da Regressão Linear em Diferentes Divisões Temporais	38
3.3	Desempenho do Modelo de Regressão Linear no Conjunto de Treino e Teste .	38
3.4	Resultados do modelo Random Forest	40
3.5	Resultados do <i>TimeSeriesSplit</i> com Random Forest	41
3.6	Analise do sobre-aprendizagem (overfitting)	42
3.7	Resumo da Arquitetura do Modelo	43
3.8	Resultados e análise da sobre-aprendizagem (overfitting) do modelo MLP $$	44
3.9	Resumo da Arquitetura do Modelo LSTM	46
3.10	Desempenho do modelo LSTM nos conjuntos de treino e teste	46
3.11	Comparação de Modelos: Resultados Gerais	48
5.1	Comparação de Modelos	61
.1	Dados de Preços de Venda e Arrendamento	71
.2	Endividamento dos Particulares	72
.3	Desemprego e Indicadores Econômicos	73
.4	Dados Temporais e Euribor	73

Glossário

ReactJS Uma biblioteca JavaScript para criação de interfaces de utilizador, desenvolvida pelo Facebook e amplamente utilizada para criar aplicações web interativas. 57

Vercel Uma plataforma para deploy de aplicações front-end, focada em performance e automação contínua, especialmente para aplicações JavaScript e frameworks como React. 58

Siglas

R^2	Coeficiente de Determinação 46
API	Application Programming Interface 58
CD CI	Continuous Deployment (Deploy Contínuo) 58 Continuous Integration (Integração Contínua) 58
EDA	Análise exploratória de dados 28
INE	Instituto Nacional de Estatística 21
LSTM	Long Short-Term Memory 45
MAE MLP MSE	Mean Absolute Error 21, 46 Multi-Layer Perceptron 42 Mean Squared Error 44
REST	Representational State Transfer 58
SPA	Single Page Application 58
UX	User Experience (Experiência do Utilizador) 53

1 Introdução

Este primeiro capítulo introduz o contexto e a relevância do problema abordado e a respetiva motivação bem como os objetivos e a contribuição da tese. Na Secção 1.1, é fornecida uma visão sobre o contexto e a motivação deste trabalho. A Secção 1.2 apresenta os principais objetivos da tese. A contribuição da tese para este campo de estudo é retratada na Secção 1.3. Por fim, a Secção 1.4 descreve a organização e estrutura deste documento.

1.1 Enquadramento

O setor imobiliário é dos mais relevantes para a economia e a sociedade de um país, pois envolve a produção, a comercialização e o uso de um bem essencial: a habitação. Além disso, o setor imobiliário tem um efeito multiplicador sobre outros setores, tais como por exemplo a construção civil, os serviços, o turismo, o crédito, etc., criando emprego, gerando impostos e investimentos. Isso acontece porque o setor imobiliário é um dos principais impulsionadores da atividade económica, agindo como um indicador-chave de estabilidade e crescimento, influenciando o consumo, a confiança do consumidor e os padrões de investimento numa economia.

Por outro lado, é importante lembrar que o setor imobiliário também enfrenta problemas importantes, como bolhas especulativas, flutuações de preços, escassez de habitação acessível e problemas ambientais associados ao crescimento urbano. Esses obstáculos têm o potencial de prejudicar tanto a economia quanto a sociedade, enfatizando a importância de políticas eficazes e soluções criativas para apoiar um crescimento sustentável e inclusivo neste setor vital.

O mercado imobiliário global conforme previsto pelo Statista [31], atingirá 637,80 trilião de dólares em 2024. A projeção para o setor sugere uma expansão contínua a uma taxa anual de 3,41% (CAGR 2024-2028), prevendo um volume de mercado de 729,40 triliões de dólares até 2028, como podemos ver na Figura 1.1.

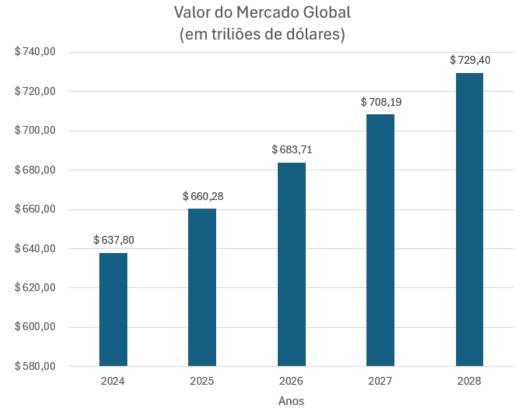


Figura 1.1: Evolução projetada do setor imobiliário pelo Statista (2024-2028) [5].

Conforme Statista [33], com um volume projetado de 518,90 triliões de dólares em 2024, o mercado imobiliário residencial destaca-se entre seus vários setores. Prevê-se uma taxa de crescimento anual (CAGR) de 3,64% neste segmento até 2028, resultando num volume de mercado de 600 triliões de dólares .

Além disso, a China destaca-se como o principal contribuinte para o valor total do setor imobiliário no mundo [30]. Em 2024, atingiu 117,40 triliões de dólares e, com uma previsão de crescimento anual (CAGR 2024–2028), atingirá um volume de mercado de 153,10 triliões de dólares até 2028.

No contexto do espaço europeu, de acordo com as projeções macroeconómicas do especialista do Banco Central Europeu sobre a contribuição do setor imobiliário para o PIB da zona Euro [12], o setor cresceu apenas 0,7% em 2023, uma queda significativa em comparação com o crescimento de 3,4% em 2022. Mas as previsões mostram uma recuperação gradual. Temos agora taxas de crescimento esperadas de 1,0 por cento em 2024 e 1,5 por cento em 2025. Estas projeções foram baixas em comparação com as previsões anteriores, indicando uma situação económica desafiadora para o setor.

Nos últimos anos, o mercado imobiliário português tem enfrentado várias tendências e dificuldades. Segundo José Cabral em doutorfinancas [9] partir do segundo trimestre de 2022, as transações de propriedade diminuíram significativamente em 22,9% em comparação com o mesmo período do ano anterior. Essa tendência de queda nas vendas tem sido contínua e teve registos regulares nos trimestres subsequentes. Ao mesmo tempo em que as

transações diminuíram, os preços das casas aumentaram 8,7% no segundo trimestre de 2022. A variação trimestral dos preços, que aumentou 3,1% em relação ao trimestre anterior, mostra ainda mais a valorização dos imóveis. No entanto, os efeitos das taxas de juro, que têm um impacto direto na dinâmica do mercado, diminuem essa perspetiva favorável. Aproximadamente 70% das transações imobiliárias em Portugal envolvem hipotecas, e o aumento das taxas de juro tem forçado algumas famílias a reconsiderar as suas opções de compra. Essas tendências mostram o mercado imobiliário em constante mudança, que é marcado pela sustentabilidade do crescimento do setor e questões de acessibilidade.

A integração da Inteligência Artificial (IA) emerge como uma tendência crescente no setor imobiliário, proporcionando novas oportunidades e eficiências tanto para agentes imobiliários quanto para compradores. A IA está a ser utilizada para análise de dados de mercado como previsão de tendências de preços, personalização de recomendações de propriedades e até mesmo para otimizar processos de marketing e vendas. Com o uso de algoritmos avançados e aprendizagem automática, as empresas imobiliárias podem oferecer uma experiência mais personalizada aos clientes, identificando melhor suas preferências e necessidades específicas. Tal melhora a eficiência do processo de compra e venda de imóveis e aumenta a satisfação do cliente, contribuindo para um mercado imobiliário mais dinâmico e competitivo [37] [11].

1.2 Objetivos

Dado que a IA tem sido um foco na evolução do mercado imobiliário, abrangendo uma variedade de aplicações e usos, o objetivo principal desta pesquisa é desenvolver uma aplicação web com análises abrangentes para o mercado imobiliário, utilizando técnicas de IA e análise de dados. Esta aplicação visa fornecer entendimentos para investidores, permitindo uma avaliação precisa das propriedades, análise das tendências do mercado e identificação de oportunidades de investimento. Pretende-se assim contribuir para o bem-estar económico de indivíduos e comunidades. uma vez que a crescente disponibilidade de dados, o avanço das técnicas de IA e análise de dados oferecem oportunidades sem precedentes para melhorar a compreensão e previsão do comportamento do mercado imobiliário.

Durante o desenvolvimento desta pesquisa, quatro questões fundamentais serão tratadas com o intuito de alcançar os objetivos delineados, como:

- 1. Q1: Quais os fatores que os investidores consideram quando pretendem investir?
- 2. Q2: Quais os melhores algoritmos para realizar previsão do mercado e para realizar uma análises comparativa?
- 3. Q3: Em que medida esses algoritmos podem ser considerados confiáveis?
- 4. Q4: Qual é a melhor maneira de apresentar as análises e entendimentos aos utilizadores da aplicação de forma clara e intuitiva?

Como a análise de dados imobiliários pode ajudar a identificar oportunidades de investimento antes que se tornem visíveis para investidores comuns? Quais foram os principais

desafios ao comparar o desempenho de diferentes algoritmos de aprendizado de máquina para previsão de preços imobiliários? Quais foram os principais desafios ao comparar o desempenho de diferentes algoritmos de aprendizado de máquina para previsão de preços imobiliários?

Como os resultados e as previsões dos modelos foram apresentados aos usuários de forma clara e intuitiva na interface da aplicação? Quais foram as abordagens mais eficazes para garantir que os usuários pudessem personalizar a análise de acordo com seus interesses de investimento? Que feedback foi coletado dos usuários sobre a usabilidade da plataforma? E como isso influenciou a melhoria contínua do design e das funcionalidades? Quais foram os principais desafios no desenvolvimento da interface do usuário (UI) e como eles foram superados para garantir uma boa experiência do usuário (UX)? Como a aplicação pode ser usada por investidores ou analistas para melhorar as decisões de investimento imobiliário no mercado atual? Quais foram as limitações enfrentadas ao integrar dados de diferentes fontes, como APIs e bases de dados externas? Como essas limitações podem ser abordadas em trabalhos futuros? Quais novos algoritmos ou abordagens podem ser explorados no futuro para melhorar a previsão de preços e identificar mais oportunidades de investimento? Quais foram as maiores dificuldades para obter dados imobiliários em tempo real ou históricos? Como as limitações no acesso a dados relevantes (por exemplo, dados restritos, privados ou pagos) influenciaram a precisão das previsões de preços?

1.3 Abordagem

A abordagem desta pesquisa será baseada no desenvolvimento de uma aplicação web voltada para o mercado imobiliário, utilizando técnicas de Inteligência Artificial e análise de dados para fornecer *insights* para investidores. A implementação da aplicação será estruturada em três camadas principais: 1) aquisição e processamento de dados, 2) desenvolvimento de modelos de IA e visualização de resultados e 3) interface com o utilizador. Essas camadas, combinadas, permitirão a construção de uma plataforma que facilite a análise de tendências do mercado e a avaliação de propriedades.

A imagem a seguir ilustra os passos referentes às camadas 1 e 2, desde a aquisição de dados até a obtenção de um modelo de aprendizagem automática treinado e validado:

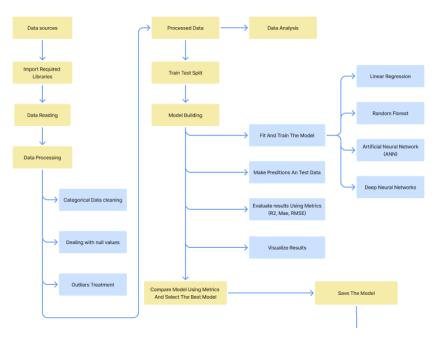


Figura 1.2: Etapas de Aquisição de Dados e Desenvolvimento de Modelos

Já a segunda imagem foca no desenvolvimento da camada 3, que detalha o processo de criação da interface da aplicação web, conectando o modelo de IA à interface para oferecer uma experiência amigável e interativa ao utilizador:

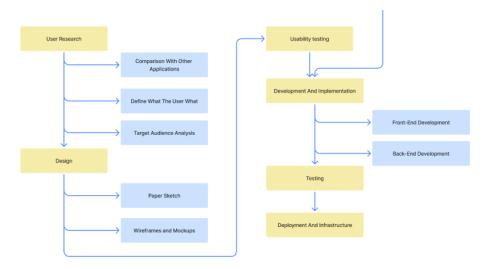


Figura 1.3: Processo de Desenvolvimento da Aplicação

Essas etapas combinadas irão proporcionar uma solução completa, da coleta de dados à visualização dos *insights* no front-end da aplicação.

A primeira etapa será a aquisição de dados de múltiplas fontes relevantes ao mercado imobiliário, incluindo bases de dados públicas e privadas, APIs de plataformas imobiliárias e dados demográficos. Esses dados fornecerão uma visão abrangente das variáveis que influenciam o valor de mercado de propriedades e a dinâmica do setor imobiliário. Para garantir a qualidade dos dados, será realizada uma etapa de pré-processamento, que inclui a limpeza de dados, remoção de *outliers*, normalização de variáveis e preenchimento de

valores que estejam em falta. Serão utilizadas tecnologias como Python, com bibliotecas especializadas em manipulação de dados como Pandas e NumPy, para realizar a integração e tratamento dos dados de maneira eficiente. Além disso, a análise exploratória de dados (EDA) será conduzida para identificar padrões iniciais, relações entre variáveis e preparar o conjunto de dados para uso nos modelos de IA.

Com os dados processados, o foco será na implementação de modelos de aprendizagem automática e aprendizagem profunda para análise de tendências e previsões de preços imobiliários. A escolha dos modelos será feita com base em testes de desempenho e precisão, utilizando técnicas como regressão linear, random forest, redes neurais artificiais, entre outros. A biblioteca Scikit-learn será utilizada para os algoritmos tradicionais de machine learning, enquanto bibliotecas como TensorFlow ou PyTorch serão empregadas para o desenvolvimento de redes neurais, caso seja necessário trabalhar com grandes volumes de dados ou identificar padrões mais complexos. A validação dos modelos será realizada através de técnicas como validação cruzada e a utilização de métricas de desempenho como erro quadrático médio (MSE) e coeficiente de determinação (R²).

Após o desenvolvimento dos modelos de IA, a aplicação web oferecerá uma interface interativa, permitindo que investidores imobiliários acessem as análises geradas e visualizem tendências do mercado de maneira intuitiva. O front-end será desenvolvido utilizando React ou Next.js, proporcionando uma interface ao utilizador moderna e otimizada para uma experiência mais fluida. A visualização dos resultados incluirá gráficos dinâmicos e relatórios detalhados sobre a valorização de propriedades e tendências de mercado, gerados por bibliotecas de visualização como D3.js ou Plotly. Essas visualizações ajudarão os utilizadores a compreenderem as previsões e análises de forma clara e objetiva, auxiliando nas tomadas de decisão.

1.4 Organização do documento

A tese está estruturada da forma que se apresenta em seguida. No Capítulo 2, é realizada uma análise abrangente da literatura relevante para o projeto. Este capítulo explora os fundamentos do mercado imobiliário, incluindo conceitos-chave, fatores influentes, tendências e desenvolvimentos recentes. Além disso, discute-se a aplicação de tecnologias associadas ao setor imobiliário, como a inteligência artificial e a análise de dados.

No Capítulo 3, são detalhadas as fontes de dados utilizadas no projeto e os métodos de tratamento e preparação desses dados para análise. Este capítulo também aborda as *Application Programming Interface* (APIs) e outras ferramentas utilizadas para aceder e manipular os dados necessários para o desenvolvimento da aplicação, de forma a preparar uma otimização das integrações dos dados.

A criação e a comparação de vários modelos de aprendizagem automática são discutidas no Capítulo 4, no qual o objetivo é escolher o modelo mais adequado e robusto para integração no desenvolvimento de aplicações web. Neste capítulo, aborda-se o processo de escolha e avaliação de modelos de inteligência artificial destinados à previsão e análise do mercado imobiliário, com o recurso a métricas de avaliação standard. Além disso, apresentam-se

considerações adicionais tais como a capacidade do modelo ser interpretado.

O Capítulo 5, aborda os aspetos técnicos envolvidos na construção de uma aplicação web. A arquitetura da aplicação é discutida com detalhes sobre a sua estrutura e seus componentes. Além disso, são abordados os procedimentos técnicos usados na criação e implementação da aplicação. Esses procedimentos incluem as tecnologias usadas, os padrões de desenho e as estratégias de implementação usadas para garantir desempenho eficiente e adequada experiência do utilizador.

Finalmente, o Capítulo 6, compila os resultados e as análises encontradas ao longo do estudo. Este capítulo discute os resultados importantes do projeto e oferece sugestões para pesquisas futuras.



Revisão da literatura

Neste capítulo, será realizada uma revisão da literatura sobre os trabalhos relacionados ao tema da pesquisa, abordando diferentes aspetos do mercado imobiliário, análise de dados e inteligência artificial. Iniciaremos com a caracterização do setor imobiliário, abordando sua complexidade e as influências macroeconómicas que o moldam (Secção 2.1). Em seguida, examinaremos as primeiras abordagens na análise de dados imobiliários, desde métodos estatísticos tradicionais, como Análise de Séries Temporais e Regressão, até modelos mais recentes, como o Modelo de Preços s e técnicas de Inteligência Artificial (Secção 2.2). Posteriormente, analisaremos a avaliação dos métodos, destacando a variedade de métricas utilizadas, desde as convencionais, como RMSE e R2, até abordagens mais sofisticadas, como análise de erro e validação cruzada (Secção 2.3). Por fim, faremos um sumário das principais conclusões e considerações levantadas na revisão da literatura, contextualizando o presente trabalho e justificando a abordagem adotada (Secção 2.4).

2.1 Caracterização do setor imobiliário

O setor imobiliário, um segmento crucial da economia que se ocupa da negociação de propriedades, possui características únicas que o distinguem de outros ativos financeiros. Estas características têm um impacto significativo no comportamento dos mercados imobiliários, influenciando a dinâmica das transações e a definição dos preços.

As características físicas e económicas das propriedades imobiliárias são elementos fundamentais que moldam o cenário dos mercados imobiliários. A interação complexa e dinâmica entre esses atributos influencia diretamente o comportamento dos mercados e os padrões de investimento. Para investidores, agentes imobiliários e governos, compreender esses atributos é crucial para tomar decisões informadas e estratégicas no contexto do mercado imobiliário. Esses atributos fundamentais incluem a imobilidade espacial, a durabilidade, a heterogeneidade, a liquidez reduzida, os altos custos de informação e transação, o alto preço por unidade, a intervenção governamental e o ciclo de oferta e demanda. Ao reconhecer a importância desses elementos e sua influência no mercado, os participantes podem melhor avaliar oportunidades, mitigar riscos e promover um desenvolvimento sustentável no setor imobiliário [3] .

A interação entre esses atributos é exemplificada na figura seguinte, que ilustra a relação entre as diversas características do mercado imobiliário:

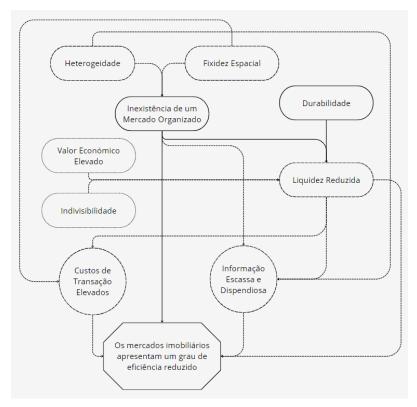


Figura 2.1: Caraterização do bem imobiliário [20].

Os bens imobiliários possuem características físicas únicas que influenciam profundamente o mercado. A fixidez espacial é uma dessas características essenciais, indicando que as propriedades são permanentes na sua localização. Esta característica está intrinsecamente ligada à sua valorização e atratividade, uma vez que a localização é um dos principais determinantes do valor imobiliário. Além disso, a durabilidade e indestrutibilidade das propriedades garantem a sua capacidade de manter o valor ao longo do tempo, perante mudanças nas condições de mercado. Por outro lado, a heterogeneidade das propriedades pode tornar a avaliação e comparação uma tarefa desafiadora. Cada propriedade é única em termos de tamanho, tipo de construção e características arquitetónicas, o que pode dificultar a análise de mercado. Além disso, a liquidez limitada é uma característica intrínseca do mercado imobiliário, indicando que as propriedades não podem ser facilmente convertidas em dinheiro devido a processos burocráticos e custos de transação significativos. Isso influencia a velocidade e facilidade de venda de propriedades, afetando a dinâmica do mercado imobiliário [2] [8] .

Do ponto de vista económico, as propriedades imobiliárias são impactadas por características como escassez de terra e recursos, além de melhorias que podem aumentar o seu valor. Esses elementos aumentam a valorização das propriedades, especialmente em locais com alta demanda por espaço e recursos limitados. Como as propriedades imobiliárias geralmente são investimentos de longo prazo, é importante que permaneçam investidos.

Para obter benefícios a longo prazo, como valorização do capital e fluxos de renda estáveis, que contribuem para a estabilidade do mercado imobiliário, os proprietários e investidores geralmente mantêm propriedades por períodos prolongados [2] [8].

2.2 As primeiras abordagens

Ao longo dos anos, têm sido realizadas diversas pesquisas sobre o mercado imobiliário, diretamente relacionado com as dinâmicas sociais, económicas e políticas. Diversos estudos procuram desvendar os mecanismos do comportamento e auxiliar na tomada de decisões neste mercado.

Nestes estudos consegue-se observar que a análise do mercado imobiliário envolve uma complexa variedade de fatores interligados. Entre eles, podemos destacar os fatores fundamentais como localização, idade e tamanho da propriedade, elementos intrínsecos a cada imóvel que impactam diretamente no seu valor. Também são importantes os fatores macroeconómicos, como taxa de juro, índices económicos e outros indicadores que refletem o contexto económico global e influenciam o mercado imobiliário. Além disso, eventos inesperados e singulares como bolhas imobiliárias, crises e pandemias podem causar perturbações e mudanças significativas no mercado, sendo estes fatores.

2.2.1 Fatores determinantes

Segundo Pai e Wang (2020) [28], os fatores que afetam o mercado imobiliário podem ser categorizados como qualitativos ou quantitativos. Os fatores qualitativos referemse à informação subjetiva sobre o imóvel e seu entorno, como qualidade da construção, segurança do bairro e infraestrutura local. Uma vez que esta informação é subjetiva (como a qualidade da vista do imóvel, ou estilo arquitetónico e ambiente do bairro) é difícil de quantificar e existe muita dificuldade de obter estes dados padronizados para análise. Já os fatores quantitativos são dados numéricos que caracterizam o mercado e as propriedades, como por exemplo indicadores macroeconómicos (ex. taxa de desemprego, PIB), ciclos dos negócios (ex. demanda e procura, crises e políticas governamentais) e atributos das próprias propriedades (ex. valor, tamanho, idade das propriedades). Uma vez que estes dados são mais fáceis de obter e podem ser facilmente analisados usando ferramentas estatísticas, contêm uma grande coleção de estudos.

Ao avançar na análise dos principais elementos que moldam o mercado imobiliário, é fundamental entender como as características intrínsecas de uma propriedade impactam diretamente seu valor. O artigo de Ferlan, Bastic e Psunder (2017) [13] examina as variáveis que impactam o valor de um imóvel residencial. O estudo concentra-se especificamente em aspetos como localização, disponibilidade de estacionamento, presença de elementos negativos no ambiente, nível do andar do imóvel e outros aspetos relacionados com a qualidade da propriedade.

O estudo descobriu que a localização em relação ao centro da cidade é o fator mais importante para avaliar o valor dos imóveis residenciais. O estacionamento e a falta de fatores ambientais prejudiciais, como a proximidade de instalações industriais, estão em

segundo lugar. Além disso, foi encontrado que a disposição da propriedade, o tamanho, a tranquilidade e a eficiência energética são fatores importantes na determinação do valor de um imóvel. Conclui-se que uma abordagem holística, que leva em consideração aspetos visíveis observáveis e não observáveis, é necessária para uma avaliação precisa do valor de um imóvel residencial. Isso fornece perceções úteis para profissionais do mercado imobiliário e contribui para um mercado mais transparente e eficaz.

O artigo de Pashardes et al. (2009) [29] investigou recorrendo ao modelo hedónico, os determinantes dos preços das casas no Chipre durante o período de 1988 a 2008, analisando características específicas das habitações e fatores macroeconómicos. A análise concluiu que os preços das casas são influenciados por muitos fatores. No nível das características das habitações, descobriu-se que o tipo de habitação (como o número de quartos e o estilo da casa) afeta os preços, com apartamentos maiores e casas destacadas geralmente com preços mais elevados por metro quadrado.

Alguns artigos estudam a influência de fatores macroeconómicos, onde tentam explorar de entre várias variáveis as que têm mais impacto. Por exemplo, o Grum e Govekar (2016) [15] analisaram a inter-relação entre fatores macroeconómicos e os preços do mercado imobiliário em diferentes contextos culturais e económicos, abrangendo Eslovénia, Grécia, França, Polónia e Noruega. Ao utilizar modelos de regressão, examinou-se a influência de variáveis como taxa de desemprego, índice de ações, balança comercial, produção industrial e PIB. Os resultados mostraram que, enquanto na Eslovénia os preços imobiliários se relacionavam ao índice de ações, nos outros países, como Grécia, França, Polónia e Noruega, verificava-se a relação com a taxa de desemprego.

O estudo conduzido por Rong Ma da Escola de Economia da Universidade de Finanças e Economia de Shanxi, China [24], complementa os estudos anteriores ao focar nos elementos específicos que afetam os preços das habitações na China. Este estudo visa compreender a dinâmica do mercado imobiliário chinês e fornecer diretrizes para políticas de controle de preços eficazes. Para isso, neste estudo foram examinados vários fatores, incluindo população, renda disponível per capita, preço médio das transações de terrenos e área de vendas de habitações comerciais em 30 regiões administrativas provinciais de 2013 a 2018. Os resultados mostraram que esses elementos têm impactos diferentes em várias partes da China, destacando a complexidade do mercado imobiliário chinês. Por exemplo, enquanto o aumento da renda disponível per capita aumentou os preços das habitações em áreas como o Leste e o Sul da China, outros fatores, como o preço médio das transações de terrenos e a população, também afetaram os preços das habitações em áreas como o Norte, Nordeste e Centro da China. Essas conclusões mostram que o mercado imobiliário chinês é muito complexo e que diferentes regiões precisam de políticas de controle de preços diferentes.

Diante da compreensão dos fatores fundamentais e macroeconómicos que moldam o mercado imobiliário, existem alguns eventos que são estudados por serem pontuais e que afetam tanto o imobiliário como as economias de diversos países. Eventos, como o Brexit (2016), as eleições presidenciais nos Estados Unidos (2020) e a guerra na Ucrânia (2022) e anos seguintes, destacam-se duas crises recentes que tiveram um impacto particularmente marcante no mercado imobiliário: a crise financeira sub-prime (2008) e a pandemia de

COVID-19 (2019). Esses eventos não apenas desencadearam disrupção imediata, mas também deixaram legados duradouros que continuam a moldar o mercado imobiliário até hoje.

O estudo conduzido por Grybauskas et al (2021) [16] que investigou como a pandemia de COVID-19 afetou o mercado imobiliário, identificando quais atributos dos apartamentos têm maior probabilidade de influenciar revisões de preços durante esse período. Para isso, os autores utilizaram análises preditivas e Big Data, coletando dados de 18.992 listagens de imóveis em Vilnius na Lituânia, durante a primeira vaga da pandemia e com 15 modelos diferentes. Os resultados indicaram que, apesar das expectativas de uma queda dramática nos preços, o mercado imobiliário mostrou-se relativamente resiliente durante a pandemia. Verificou-se apenas uma pequena proporção de habitações com quedas médias modestas nos preços. Além disso, o modelo de Extreme Gradient Boosting foi considerado o mais preciso, e a variável "tempo no mercado"foi identificada como a mais dominante e consistente para prever revisões de preço, destacando a importância de monitorar de perto essa variável para avaliar as condições do mercado imobiliário durante a pandemia.

Nesse contexto, o estudo de Mora-Garcia et al. (2022) [26] sobre o impacto da pandemia de COVID-19 nos preços das casas em Alicante, Espanha, oferece uma compreensão das dinâmicas do mercado imobiliário perante de eventos extraordinários. Ao analisar a performance de diferentes algoritmos de aprendizagem automática na previsão de preços de habitação e ao quantificar o efeito da pandemia nos preços das casas, o estudo conclui que o impacto da pandemia de COVID-19 foi significativo e imediato. A análise indica que durante o período inicial da pandemia de COVID-19, o mercado imobiliário em Alicante enfrentou uma queda acentuada nos preços das casas, no entanto o impacto foi localizado e de duração transitória. Verificou-se também que os primeiros trimestres após a declaração da pandemia registaram uma queda nos preços a qual não foi tão significativa nem duradoura quanto a observada durante a crise financeira de 2008. Os resultados indicaram que os preços atingiram a maior descida no quarto trimestre de 2020 e no primeiro trimestre de 2021. No entanto, a recuperação dos preços começou no terceiro trimestre de 2021 e, em aproximadamente um ano e meio, os preços ultrapassaram os níveis anteriores à pandemia.

O estudo referente à crise de 2008 efetuado por Sjöling (2012) [34] onde se analisa quatro indicadores comuns para prever o valor das propriedades imobiliárias em quatro mercados distintos durante a crise de 2008: Alemanha, Suécia, Espanha e Reino Unido. Embora os indicadores tenham mostrado correlação entre si, não conseguiram prever adequadamente o desenvolvimento do mercado imobiliário antes da recessão. Além disso, destacou-se a influência das condições de empréstimo hipotecário na dinâmica dos preços imobiliários, especialmente entre a Suécia e o Reino Unido, onde a predominância de taxas variáveis na Suécia e taxas fixas no Reino Unido influenciou significativamente a resposta dos mercados imobiliários durante a recessão. Isso ressalta a importância de considerar não apenas indicadores tradicionais, mas também as condições específicas do mercado de empréstimos ao formular políticas e modelos de previsão para o mercado imobiliário, especialmente em períodos de instabilidade económica global, tal como a crise de 2008.

Além dos fatores já mencionados, como atributos dos imóveis, fatores macroeconómicos e

eventos singulares como crises, outros estudos abordam a previsão e avaliação de imóveis que podem ser igualmente interessantes numa perspetiva mais abrangente. O artigo de Ernawati Kamal (2016) [18] aborda a perspetiva de como os agentes imobiliários olham para o mercado e os fatores que influenciam as habitações em Penang, Malásia. Identificam-se cinco fatores-chave indicados pelos agentes: 1) localização, que na perspetiva dos agentes era o fator mais importante, especialmente quando combinada com infraestrutura adequada e design de alta qualidade, 2) fatores macroeconómicos, que no estudo se destaca a influência da taxa de inflação e das taxa de juro sobre os preços das habitações, destacando a sua influência sobre as decisões dos agentes, 3) fatores demográficos, onde pode entender que mudanças na demografia, como migração populacional e padrões de crescimento, impactam os preços das habitações, 4) questões de terra/zona onde se entendeu a importância da disponibilidade de terras e o status da urbanização e 5) fatores da indústria que tais exercem uma influência significativa nos preços das habitações, demonstrando a complexidade e a interconexão entre os custos de construção, regulamentações governamentais e requisitos da indústria no mercado imobiliário.

Após uma discussão sobre as variáveis que impactam o mercado imobiliário, examinamos como as técnicas de análise de dados imobiliários evoluíram ao longo do tempo. Nas sub-secções seguintes abordamos métodos como Análise de Séries Temporais, Análise de Regressão, Modelos de Preços Hedónico e Inteligência Artificial, destacando suas aplicações e resultados.

2.2.2 Análise de séries temporais

Um dos primeiros métodos utilizados na análise de dados imobiliários foi o modelo ARIMA (AutoRegressive Integrated Moving Average). Este modelo estatístico foi criado na década de 1970 e foi pioneiro na deteção e explicação de padrões dependentes do "Time Series Analysis" [5], ou seja, era realizada uma análise que utiliza técnicas estatísticas de previsão de dados que variam ao longo do tempo. Ao ser usada na indústria imobiliária, permitiu que os analistas modelassem e previssem tendências de preços, o que resultou em decisões de investimento mais conscientes. No entanto, o modelo ARIMA tem limitações. Estas ocorrem especialmente quando o modelo se depara com dados não estacionários (aqueles cujas propriedades estatísticas, como média e variância, mudam com o tempo), ou padrões não lineares complexos, porque as restrições do modelo ARIMA residem na sua natureza linear e na assunção de estacionariedade [10].

Outros métodos com um histórico significativo na "Time Series Analysis" incluem os métodos de suavização exponencial, como SES (Simple Exponential Smoothing), Holt's Linear Exponential Smoothing e Holt-Winters' Exponential Smoothing [10]. Estes métodos são particularmente úteis para fazer previsões a curto prazo no setor imobiliário, com base em observações passadas e considerando a sazonalidade do mercado. No entanto, uma limitação importante desses métodos é que podem não ter bom desempenho com previsões de longo prazo ou quando os dados mudam de repente e inesperadamente. Portanto, embora sejam ferramentas valiosas a sua eficácia pode ser limitada em certos cenários [22] .

2.2.3 Análise de Regressão

À medida que a análise estatística se desenvolvia, os investigadores começaram a explorar modelos de regressão para previsão imobiliária. Esses modelos permitiam a inclusão de múltiplas variáveis explicativas e a avaliação de suas relações com os preços imobiliários. De acordo com o artigo [21], a introdução desses modelos permitiu uma abordagem mais abrangente na previsão de preços imobiliários, com a inclusão de múltiplas variáveis explicativas, como nível de renda, despesas per capita e área construída. O modelo de regressão linear múltipla foi amplamente utilizado devido à sua capacidade de lidar com a complexidade dos dados do mercado imobiliário e fornecer previsões mais precisas, contribuindo assim para uma melhor compreensão das dinâmicas do mercado e orientação de políticas e estratégias de investimento.

Ainda que seja um modelo popular existem alguns problemas ao aplicá-lo. Essas restrições afetam a interpretação dos resultados e a validade das previsões, dada a relação linear entre as variáveis independentes e a variável dependente. Entre elas, este modelo está sujeito à multicolinearidade, um fenómeno que ocorre quando duas ou mais variáveis independentes são muito correlacionadas, o que dificulta a interpretação individual do efeito de cada uma dessas variáveis. Também está sujeito à sensibilidade a valores atípicos, que significa que os valores extremos, ou ausências, nos dados, podem afetar significativamente os resultados do modelo [32] .

2.2.4 Modelo de Preços Hedónico

O Modelo de Preços Hedónico surgiu como resultado da evolução da análise de dados imobiliários. Ao contrário dos métodos de análise de séries temporais e de regressão, este modelo concentra-se na valorização de cada característica única de uma propriedade imobiliária. Ao contrário dos modelos de regressão tradicionais, que tratavam o imóvel como uma entidade homogénea, o modelo hedónico reconheceu a heterogeneidade dos imóveis e a influência única de cada característica no seu valor de mercado [17].

O trabalho de Lorenz et al. (2022) [23] ressalta que o Modelo de Preços Hedónico é baseado na ideia de que o preço de um imóvel é determinado tanto por suas características internas (como tamanho, aparência, condição, e características especiais como painéis solares ou torneiras de última geração) quanto por características do seu ambiente externo (como a taxa de criminalidade do bairro, proximidade a escolas e ao centro da cidade, nível de poluição da água e do ar, ou o valor de outras casas próximas).

Neste trabalho também se observa que com o avanço da tecnologia e o aumento da capacidade computacional, os métodos de aprendizagem automática, tais como gradient tree boosting (GTB), random forest regression (RFR) e support vector regression (SVR), têm sido cada vez mais utilizados para complementar o processo de estimação hedónico Estes algoritmos de aprendizagem automática são capazes de aprender artificialmente a partir dos dados subjacentes e melhorar continuamente sua performance preditiva. No entanto, a principal crítica aos modelos de aprendizagem automática é sua falta de transparência e interpretabilidade, o que pode dificultar a compreensão das relações entre as características

do imóvel e seu valor de mercado.

O artigo Abidoye e A. Chan, [35], destacou limitações significativas do modelo hedónico na valoração de propriedades imobiliárias. Através de uma análise detalhada das transações de vendas de propriedades residenciais na metrópole de Lagos, Nigéria, o estudo evidenciou que o modelo hedónico não foi capaz de produzir estimativas precisas de valoração. Essas descobertas destacam a importância de reconhecer as limitações do modelo hedónico e considerar abordagens alternativas na valoração imobiliária.

2.2.5 Algoritmos de inteligência artificial

A história da Inteligência Artificial (IA) é marcada por uma progressão constante, desde os primeiros métodos até às tecnologias mais recentes. Inicialmente, os métodos de IA, como árvores de decisão, sistemas periciais, lógica difusa e raciocínio automatizado, eram notavelmente mais interpretáveis e auto explicativos [6] .

Com o avanço da tecnologia e a disponibilidade de grandes conjuntos de dados, a aplicação de técnicas de IA têm-se destacado como uma abordagem promissora. As abordagens mais proeminentes são o uso de algoritmos de aprendizagem automática e aprendizagem profunda, que permitem aos sistemas aprenderem a partir dos dados sem a necessidade de programação explícita. Esses algoritmos capacitam os sistemas a realizar tarefas complexas, como reconhecimento de padrões, previsão de tendências e tomada de decisões, com base nos dados fornecidos.

Estes modelos de IA têm a vantagem nos mercados imobiliários por conseguirem ter a capacidade de lidar com grandes volumes de dados (como fatores económicos e sociais) e identificar padrões complexos que não podem ser detetados por métodos mais simples [4]. Embora vários estudos tenham mostrado que modelos de aprendizagem profunda, como redes neurais, são altamente precisos, também apresentam problemas significativos. Esses modelos são geralmente muito complexos, o que dificulta a transparência, de forma a que utiliza os modelos tenha dificuldade em entender as relações entre as características dos dados e os resultados previstos. Essa falta de clareza e interpretabilidade é uma crítica comum aos modelos de aprendizagem automática, especialmente em avaliações imobiliárias. A complexidade dos modelos de aprendizagem automática pode dificultar a compreensão da relação entre as características de um imóvel e seu valor de mercado. Por outro lado, métodos mais antigos, como o modelo hedónico, são mais fáceis de entender e transparentes, pois permitem uma análise mais direta das variáveis, sendo que avaliam cada caraterística do imóvel de forma individual, o que proporciona uma visão clara e intuitiva de como cada característica contribui para o valor final do imóvel [23].

2.2.6 Métodos de previsão econométricos

Os modelos econométricos são ferramentas analíticas fundamentais usadas pelos economistas para prever e compreender as tendências económicas futuras. Estes modelos criam relações quantitativas entre um parâmetro de interesse - normalmente uma variável dependente - e uma coleção de fatores que o influenciam, as variáveis explicativas, como fatores

democráticos e económicos. Nestes métodos, destacam-se a capacidade de fornecer previsões baseadas em dados quantitativos, permitindo uma avaliação mais objetiva do cenário económico e a identificação de relações complexas entre variáveis e ajudar a entender os mecanismos subjacentes às mudanças económicas [14].

Nesta categoria existem diversos modelos, sendo os mais populares e amplamente utilizados são os Modelos VAR (Vector Autoregression) e os modelos de Dados em Painel (Panel Data Models) [1], porque estes oferecem uma abordagem robusta e flexível para analisar a interação entre diferentes variáveis económicas ao longo do tempo. Os modelos VAR são multi-variados e compreendem interdependências lineares entre várias séries temporais, enquanto os modelos de dados em painel utilizam dados que variam ao longo do tempo e entre indivíduos ou unidades.

Uma aplicação do modelos VAR foi a sua utilização na resposta do setor de construção a choques macroeconómicos, prever sua atividade e informar políticas económicas. Nesse contexto, analisou-se como o setor reagiu a mudanças em variáveis macroeconómicas como o PIB real, taxa de juro e créditos. Essa análise permitiu prever as atividades futuras do setor e forneceu informações importantes para a formulação de políticas económicas que visam apoiar a estabilidade e o crescimento sustentável da construção [1] .

Em contraste com o uso de modelos de dados de painel, um artigo recente examinou como as características específicas das empresas de construção impactam o seu desempenho em resposta a choques macroeconómicos. Essa abordagem complementa a visão macroeconómica dos modelos VAR, fornecendo uma compreensão mais detalhada das empresas do setor e informações sobre políticas mais específicas [19].

2.2.7 Métodos de previsão da análise espacial

O método de análise espacial para previsão combina métodos estatísticos sofisticados com uma compreensão da distribuição geográfica de dados. Esses métodos têm o potencial de revelar padrões complexos e fornecer perceções valiosas sobre eventos futuros em áreas específicas ao integrar conceitos de análise espacial e modelos de previsão. Essa abordagem pode ser usada em várias áreas de estudo, desde epidemiologia e urbanismo até economia regional. Assim, desempenha um papel importante na compreensão e previsão das mudanças que ocorrem no mundo [27] .

A Previsão Geoespacial, que utiliza Sistemas de Informação Geográfica (SIG) [7], é um exemplo desses métodos. Os sistemas SIG permitem-nos compreender, organizar e comunicar o mundo. Permitem-nos visualizar, questionar, analisar e interpretar dados para descobrir tendências, padrões e correlações. Além disso, a análise e organização de dados em camadas é possível com os SIG, o que facilita a visualização espacial de localizações através de mapas e cenas 3D. Esta capacidade dos SIG de fornecer informações detalhadas sobre dados, incluindo situações, padrões e relações, ajuda os utilizadores a tomar decisões informadas .

O artigo de Lei Wu et al. (2022) [36], por exemplo, utilizou este método de análise. Este artigo forneceu uma maneira inovadora e eficaz de avaliar o preço de terras de referência,

que é essencial para regular a procura de terras no mercado imobiliário chinês. O estudo conseguiu automatizar e melhorar o processo de avaliação, reduzindo a subjetividade e aumentando a precisão dos resultados, utilizando redes neuronais. Esse método pode melhorar significativamente a eficácia e a imparcialidade da avaliação do preço da terra noutras cidades chinesas e em contextos semelhantes em todo o mundo.

O artigo de Constantin Viorel Marian e Mihaela Iacob (2022) também este método de análise. Este artigo descreve um projeto que resultou na criação de um sistema integrado baseado em SIG para proteger o património arqueológico e cultural da Roménia. Um mapa digital interativo e um banco de dados com informações sobre sítios arqueológicos em todo o país foram criados. O sistema facilita a identificação e proteção do património arqueológico, permitindo melhores decisões sobre o desenvolvimento de áreas com esse património por meio de acesso aos escritórios do Ministério da Cultura e acesso a especialistas e investidores [25].

2.3 Avaliação dos métodos

De entre os diversos artigos mencionados nesta secção, os autores procuraram analisar os modelos recorrendo a uma avaliação de performance. A maioria procurou fazê-lo analisando o número de erros e alguns dos autores mais atuais procuraram categorizar os erros fazendo assim uma análise dos erros mais pormenorizada.

Os artigos que examinam modelos mais conhecidos, como Regression Analysis e Hedonic Pricing Model [35], usaram métodos de avaliação de desempenho mais tradicionais. Esses métodos incluem o Root mean Squared Error (RMSE), que calcula a média das diferenças ao quadrado entre os valores observados e os valores previstos pelo modelo; o R2 (Coeficiente de Determinação), que calcula a proporção da variabilidade nos dados que são explicados pelo modelo; e o Mean Absolute Percentual Error (MAPE), que calcula a média das percentagens absolutas de erro entre os valores observados e previstos.

No entanto, noutros artigos que usam modelos mais recentes, tais como Lorenz et al. (2022) [23] e Wu et al. (2022) [36], examinam modelos baseados em IA, a abordagem de avaliação da performance vai além das técnicas tradicionais. Os autores neste caso usaram métricas mais específicas, como acurácia, tempo de execução e número de parâmetros, além das métricas convencionais como RMSE e R2. Além da precisão das previsões, essas métricas adicionais abordam a complexidade do modelo e a eficiência computacional, fornecendo uma compreensão mais profunda do desempenho dos modelos. Além disso, a combinação dessas métricas com técnicas de validação, como a validação cruzada de dez partições e análise detalhada de erro, permite uma avaliação mais precisa e aprofundada da precisão e qualidade dos modelos desenvolvidos. Isso fornece informações valiosas sobre sua aplicabilidade prática e potencial de generalização.

2.4 Sumário

A revisão da literatura destacou a complexidade envolvida na previsão do mercado imobiliário, que abrange uma gama diversificada de fatores, desde influências macroeconómicas até características específicas das propriedades.

Ao abordar a complexidade do mercado imobiliário, várias abordagens têm sido estudadas, incluindo análises convencionais como séries temporais e regressão, bem como técnicas de inteligência artificial, sendo a escolha do método mais adequado influenciada pela precisão pretendida e pela capacidade de interpretar os resultados. Uma visão geral dos métodos aqui analisados é apresentada na Figura 4.1, destacando as diferenças entre os métodos de previsão e avaliação imobiliária .

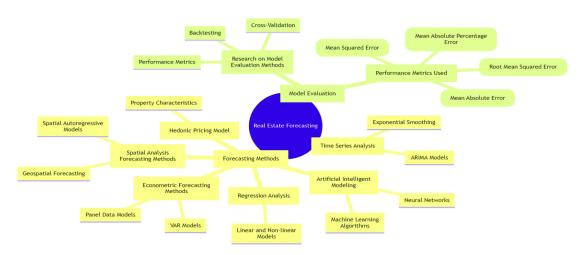


Figura 2.2: Métodos de previsão e avaliação de bens imóveis [5].

Além disso, as técnicas de avaliação desses métodos variam, indo desde métricas convencionais, como RMSE e R2, até abordagens mais sofisticadas, incluindo análise de erro e validação cruzada. Essas últimas oferecem uma compreensão mais aprofundada do desempenho dos modelos, bem como a sua aplicabilidade em contextos específicos. A escolha entre diferentes métodos e técnicas de avaliação deve ser guiada pela natureza e pelos objetivos específicos da análise de previsão do mercado imobiliário em questão.

3 Dados, Metodologias e Métricas

Neste capítulo, será apresentada a metodologia adotada para a análise do mercado imobiliário, bem como os dados utilizados e os resultados obtidos. A integração de diversas fontes de dados, como a Casafari, o Instituto Nacional de Estatística (INE), Pordata e o Banco de Portugal, permitiu uma visão abrangente e detalhada das dinâmicas do setor, combinando indicadores imobiliários com fatores macroeconómicos e financeiros.

Inicialmente, será discutido o processo de coleta, preparação e transformação dos dados, destacando as etapas de pré-processamento e a criação de novas variáveis. Na secção 3.1, abordaremos as fontes de dados utilizadas no estudo. A seguir, em 3.2, será detalhado o processamento de dados, incluindo técnicas de tratamento de valores ausentes e a criação de novas variáveis.

Na secção 3.3, será explicada a divisão dos dados em conjuntos de treino, validação e teste, para garantir uma avaliação adequada dos modelos. Em 3.4, serão descritos os diferentes modelos de aprendizagem automática aplicados à previsão dos preços imobiliários, como regressão linear, Random Forest e redes neurais. A secção 3.5 detalha as métricas utilizadas para avaliar a performance dos modelos, como Mean Absolute Error (MAE), R² e a precisão dos modelos de previsão.

Em 3.6, será analisado o impacto das principais variáveis que influenciaram os preços do mercado imobiliário, e, por fim, a secção 3.7 apresenta a implementação da aplicação web, para visualizar os resultados e fazer previsões interativas com base nos dados analisados.

3.1 Fontes de Dados

Nesta secção, são descritas as principais fontes de dados utilizadas neste estudo sobre o mercado imobiliário em Portugal. A diversidade dessas fontes é fundamental para garantir uma análise completa e aprofundada, contemplando tanto os dados específicos do setor imobiliário quanto o contexto macroeconómico e as taxas de juro, que influenciam diretamente o comportamento do mercado.

A seleção das fontes de dados foi orientada pela necessidade de captar diferentes dimensões do mercado imobiliário, combinando indicadores imobiliários com fatores macroeconómicos

e financeiros. As informações foram extraídas de instituições como o Instituto Nacional de Estatística (INE), Casafari, Doutor Finanças e Banco de Portugal, que, em conjunto, proporcionam uma visão abrangente das dinâmicas do setor.

O INE disponibiliza dados estatísticos nacionais essenciais sobre demografia e economia, a Casafari oferece dados detalhados sobre transações imobiliárias, o Doutor Finanças contribui com informações relevantes sobre condições de financiamento, e o Banco de Portugal fornece dados macroeconómicos e financeiros de longo prazo. A combinação dessas fontes permite uma análise detalhada e precisa do mercado imobiliário, capturando suas variações e influências externas ao longo do tempo.

3.1.1 Instituto Nacional de Estatística (INE)

O Instituto Nacional de Estatística (INE) desempenha um papel fundamental no panorama estatístico português, sendo a entidade responsável pela recolha, análise e difusão de dados estatísticos abrangentes. No âmbito deste estudo sobre o mercado imobiliário em Portugal, o acesso a dados precisos e atualizados fornecidos pelo INE foi crucial para oferecer uma visão detalhada das tendências e comportamentos do setor.

Como principal entidade estatística do país, o INE contribui significativamente para o planeamento, desenvolvimento e avaliação de políticas públicas, além de apoiar a tomada de decisões empresariais e de pesquisa. Este acesso a informações confiáveis e objetivas é essencial para compreender as dinâmicas do mercado imobiliário e para a formulação de estratégias baseadas em evidências.

Para a extração dos dados, foi utilizado o notebook Jupyter, API_ExtractData.ipynb. Este script facilitou a interação automatizada com a API do INE, para a descarga eficiente das seguintes variáveis-

- Avaliações Bancárias de Alojamentos Familiares- Dados mensais detalhados por tipo de construção (total, apartamentos, moradias) e localização geográfica conforme a nomenclatura das unidades territoriais para fins estatísticos (NUTS - 2024).
- Capital Médio em Dívida e Juros Totais- Informações sobre o capital médio em dívida e os juros totais em contratos de crédito à habitação, segmentados por período de celebração, regime de crédito e destino de financiamento.
- População Desempregada e Prestação Média Vencida- Estatísticas mensais sobre a população desempregada por grupo etário e a prestação média vencida nos contratos de crédito à habitação por localização geográfica (NUTS I).

Os dados foram selecionados para o período de janeiro de 2012 até junho de 2024, abrangendo diversas regiões como Continente, Amadora, Cascais, Lisboa, Loures, Mafra, Odivelas, Oeiras, Sintra e Vila Franca de Xira. A precisão e a atualidade destas informações são fundamentais para a análise do impacto de diferentes fatores económicos e demográficos sobre o mercado imobiliário nas regiões estudadas.

3.1.2 Casafari

A Casafari é uma plataforma de monitoramento do mercado imobiliário, cuja tecnologia avançada permite a agregação e análise de dados de propriedades listadas para venda e renda. Para este estudo, a plataforma foi utilizada para extrair dados relevantes ao mercado imobiliário nos concelhos de Amadora, Cascais, Loures, Odivelas, Oeiras e Sintra, durante o período de janeiro de 2018 até agosto de 2024.

Através do acesso à Casafari, foram obtidos arquivos CSV contendo dados detalhados sobre as dinâmicas de mercado nos concelhos especificados. A abrangência e a granularidade dos dados descarregados proporcionam uma base sólida para análises quantitativas rigorosas, conforme discriminado a seguir-

- Data- As entradas são marcadas com timestamps, facilitando análises temporais e de séries temporais.
- Preço Médio e Preço Médio por m²- Estes indicadores são fundamentais para avaliar o nível de valorização do mercado imobiliário e para comparações interregionais.
- Preço Médio de Fechamento (BETA) e Preço Médio de Fechamento por m² (BETA)- Fornecem estimativas sobre os valores de transação efetiva, refletindo condições de mercado fechadas.
- Subida de Preço e Redução de Preço- Variáveis que refletem a volatilidade dos preços e as estratégias de ajuste dos vendedores em resposta às condições de mercado.
- Novas Propriedades Únicas- Quantifica as novas ofertas no mercado, um indicador direto de atividade de mercado e disponibilidade de imóveis.
- Propriedades Vendidas e Retiradas- Este dado ajuda a entender a liquidez do mercado e a eficácia das estratégias de venda.
- Novas Listagens Não duplicadas- Representa o volume inicial de listagens antes da aplicação de filtros para eliminação de dados duplicados, indicando a atividade bruta de listagem..

A precisão dos dados obtidos da Casafari foi fundamental para obter análises estatísticas robustas e para a formulação de conclusões válidas dentro do contexto do estudo do mercado imobiliário nas regiões selecionadas. Esta abordagem assegura que as inferências feitas são baseadas em evidências sólidas, cumprindo com os padrões académicos.

3.1.3 DoutorFinanças

DoutorFinanças é uma plataforma financeira que fornece uma variedade de dados económicos e financeiros relevantes. Para este estudo, a plataforma foi utilizada para extrair dados históricos das taxas Euribor, que são cruciais para a análise da influência das condições

de financiamento no mercado imobiliário. As taxas Euribor são indicadores essenciais, pois influenciam diretamente as condições de crédito e, consequentemente, a atividade imobiliária.

A seleção das taxas Euribor em diferentes prazos permite avaliar como as mudanças nas condições de financiamento afetam o mercado imobiliário. Para este estudo, foram extraídas as taxas Euribor de 1, 3, 6 e 12 meses.

- Curto Prazo (1 e 3 meses) Taxas mais voláteis, influenciando linhas de crédito de curto prazo e fundos operacionais.
- Médio e Longo Prazo (6 e 12 meses) Mais relevantes para hipotecas e empréstimos de longo prazo, impactando diretamente a acessibilidade ao crédito imobiliário.

Desta forma os dados descarregados foram:

- **Tempo** As entradas são categorizadas por data, permitindo uma análise temporal detalhada das variações nas taxas Euribor ao longo de quase duas décadas.
- 1 mês (Euribor), 3 meses (Euribor), 6 meses (Euribor), 12 meses (Euribor)
 Estas categorias representam os diferentes prazos para as taxas Euribor, cada uma
 - Estas categorias representam os diferentes prazos para as taxas Euribor, cada uma influenciando de forma distinta os produtos de crédito, especialmente as hipotecas.

A análise dos dados históricos das taxas Euribor, obtidos através do Doutor Finanças, permitiu correlacionar as variações nas taxas de juro com as tendências de preços de mercado no mercado imobiliário. Essas dados foram essenciais para compreender como variações nas taxas de juros afetam a dinâmica do mercado imobiliário em Portugal. A aplicação de diferentes metodologias e métodos estatísticos permitiu examinar a relação entre as taxas de juro e o mercado imobiliário de forma quantitativa e qualitativa.

3.1.4 Banco de Portugal

O Banco de Portugal, como autoridade monetária nacional, oferece um vasto repositório de dados económicos e financeiros. Para a análise da influência das condições económicas no mercado imobiliário, foi essencial aceder a dados abrangentes e históricos fornecidos pelo Banco de Portugal. Através da sua plataforma, foram obtidos dados cruciais que abarcam desde fevereiro de 1948 até julho de 2024, permitindo uma análise longitudinal das variáveis económicas sobre o mercado imobiliário.

A plataforma do Banco de Portugal foi utilizada para descarregar um arquivo CSV contendo uma ampla gama de indicadores económicos e financeiros. Estes dados são fundamentais para entender como variáveis macroeconómicas influenciam o setor imobiliário, além de proporcionar entendimento sobre o comportamento financeiro dos clientes particulares.

Os dados considerados são:

• **Período**- As entradas são categorizadas por data (ano, mês, dia), facilitando análises temporais detalhadas.

- Número de Beneficiários de Subsídio de Desemprego e Desemprego Registado- Estes indicadores fornecem uma medida direta da saúde económica e do mercado de trabalho, os quais estão intrinsecamente ligados à capacidade de compra de imóveis.
- Indicadores Coincidentes para a Atividade Económica e Consumo Privado-Refletem o estado geral da economia e do consumo, oferecendo uma perspetiva sobre o potencial impacto no mercado imobiliário.
- Endividamento dos Particulares- Segmentado por setor credor, esses dados ajudam a entender o nível de exposição financeira dos consumidores, o que pode afetar a sua capacidade de financiar novas compras de imóveis.
- Taxas de Inflação (IPC) e Taxas de Juro- Cruciais para avaliar o custo do crédito e o poder de compra, influenciando diretamente a demanda por imóveis.
- TAEG de Novos Empréstimos ao Consumo e à Habitação- Taxas de juro aplicadas a novos empréstimos, indicando as condições de financiamento disponíveis para os consumidores.

3.1.5 Qualidade dos Dados e Limitações

Embora as fontes de dados utilizadas sejam amplamente reconhecidas por sua qualidade e precisão, alguns desafios foram encontrados durante a coleta e processamento dos dados. Em alguns casos, como na Casafari, a granularidade excessiva dos dados resultou em dificuldades para padronizar informações entre diferentes períodos. Além disso, algumas fontes apresentaram valores ausentes ou inconsistências em determinadas variáveis, exigindo a aplicação de técnicas de imputação e pré-processamento de dados.

3.2 Processamento de Dados

O processamento de dados desempenhou um papel essencial na garantia da qualidade e consistência das informações utilizadas neste estudo sobre o mercado imobiliário. Devido à diversidade de fontes e formatos dos dados coletados, foi necessário aplicar uma série de técnicas de pré-processamento, transformação e integração, de forma a preparar o conjunto de dados para as fases de análise e modelação preditiva. Esta secção descreve as principais etapas do processamento, desde o tratamento de valores ausentes até a criação de novas variáveis, essenciais para enriquecer as análises.

3.2.1 Pré-processamento de Dados

Os dados utilizados neste estudo foram coletados a partir de várias fontes, incluindo diferentes arquivos, cada um com seus próprios formatos e características. Para permitir a integração dessas informações, foi necessário aplicar uma série de transformações e tratamentos de dados específicos para cada fonte, a fim de padronizar as estruturas e garantir a consistência das análises.

Cada fonte de dados passou por tarefas de pré-processamento específicas, dependendo de suas particularidades. Uma das primeiras etapas envolveu a transformação de formatos de data. As fontes de dados apresentavam datas em diferentes formatos, o que exigiu a conversão de todas para o padrão datetime, garantindo a coerência temporal entre os datasets. Esse passo foi importante, pois a data foi uma das variáveis que foi utilizada como chave primária para a junção dos diferentes arquivos.

Após padronizar as datas, foi realizada a verificação e remoção de dados duplicados. Essa etapa foi relevante nos dados da plataforma Casafari, onde ocorrências de múltiplas linhas com a mesmo informação do mercado imobiliário. Os dados duplicados foram removidas para garantir que os dados fossem únicos e que as análises não fossem distorcidas por entradas repetidas.

Outro aspeto importante no pré-processamento foi o tratamento de valores ausentes. Para variáveis temporais, como o preço médio de vendas e a subida de preços, foi aplicada a técnica de interpolação linear. A interpolação linear preenche os espaços vazios entre dois pontos conhecidos ao calcular valores intermediários com base numa progressão linear. Esse método foi escolhido por preservar a continuidade dos dados ao longo do tempo, o que é crucial para séries temporais. Outras técnicas, como o preenchimento por média ou mediana, foram evitadas nessas variáveis temporais, pois poderiam alterar as tendências naturais dos dados, suavizando variações que poderiam ser importantes para a análise.

Para algumas variáveis temporais, como Preço médio de fechamento e Preço médio de fechamento por m2, adotou-se o preenchimento com valores "zero"(fillna(0)), quando a ausência de dados indicava a falta de uma observação relevante, como um evento que não ocorreu ou uma medição que não foi feita. Essa abordagem simplificou o tratamento de valores nulos sem comprometer a integridade das análises.

Além disso, foi aplicado um filtro temporal aos dados para garantir que apenas informações referentes ao período de 2018 a 2024 fossem consideradas. Isso foi feito para focar a análisenum intervalo de tempo mais recente e relevante, conforme o objetivo do estudo. O código utilizado para esse filtro foi o seguinte-

Por fim, como os dados da Casafari abrangiam o período de 2018 a 2024, foi necessário restringir os dados das outras fontes para o mesmo intervalo de tempo. Isso garantiu que todas as informações estivessem alinhadas temporalmente, permitindo uma análise coerente e consistente.

3.2.2 Transformação de Dados

A transformação dos dados foi uma etapa essencial para garantir a coerência e facilitar a análise após o pré-processamento. As transformações visaram criar novas variáveis e ajustar os dados de forma a permitir uma melhor compreensão das tendências e padrões do mercado imobiliário. Esse processo também ajudou a refinar o conjunto de dados para a modelação preditiva e a análise estatística.

Após o pré-processamento, o conjunto de dados foi ordenado por localização, ano e mês. Essa ordenação foi importante para garantir que as transformações, especialmente as

relacionadas a séries temporais, fossem aplicadas corretamente, considerando a sequência cronológica dos dados.

Para enriquecer a análise e capturar a evolução do mercado imobiliário, foram criadas diversas novas variáveis com base nos dados extraídos das diferentes fontes. Entre essas variáveis, destacam-se as médias móveis e exponenciais dos preços de venda e de rendas, que ajudaram a suavizar variações de curto prazo e a identificar tendências de longo prazo. A média exponencial, em particular, atribuiu mais peso aos valores mais recentes, sendo útil para captar mudanças rápidas nos preços. Além disso, também foram calculadas médias móveis para as reduções de preço, tanto em vendas quanto em rendas. Essas médias forneceram uma visão mais clara sobre a volatilidade dos preços e as estratégias adotadas pelos vendedores para ajustar os valores, contribuindo para uma análise mais aprofundada do comportamento do mercado.

A relação preço/renda (price-to-rent ratio) foi outra variável gerada, importante para avaliar a viabilidade de investimentos ao comparar o preço de compra de um imóvel com seu potencial de retorno na perspetiva da renda. Além disso, foram criadas variáveis que medem a diferença entre as propriedades vendidas/retiradas e as novas propriedades listadas, tanto para venda quanto para as rendas, oferecendo entendimentos sobre a liquidez e a dinâmica do mercado.

Devido às transformações, especialmente nas médias móveis e exponenciais, surgiram valores nulos (NaN) nos primeiros períodos onde essas médias não puderam ser calculadas. Para evitar que esses valores impactassem negativamente a análise, foi utilizado o método backfill, que preencheu os valores nulos com o próximo valor disponível, garantindo a continuidade dos dados.

3.2.3 Integração de Dados

A integração dos dados foi essencial para juntar as informações de diferentes fontes sobre o mercado imobiliário em Portugal. Primeiro, os dados do INE foram combinados com as taxas de juros da Euribor, utilizando o mês e o ano como ponto comum. Depois, as informações do Banco de Portugal foram adicionadas, completando o quadro económico.

Para incluir os dados da Casafari, foi necessário garantir que as localizações estivessem padronizadas. Com isso, os dados de cada fonte foram reunidos, formando um único conjunto que contém tanto as informações sobre os preços médio de venda dos imóveis quanto os fatores macroeconómicos, como taxas de juros e inflação. Esse conjunto de dados unificado permite uma visão mais completa e detalhada do mercado imobiliário, facilitando a análise e a previsão de tendências futuras.

Com essas etapas de pré-processamento finalizadas, os dados ficaram prontos para as fases subsequentes de análise, garantindo que as informações fossem consistentes, completas e adequadas para a modelação preditiva e análise estatística.

3.3 Análises dos Dados

Nesta secção, são aplicadas técnicas de análise exploratória e correlação para examinar os fatores que influenciam os preços dos imóveis, as taxas de juros e outros indicadores económicos, com o objetivo de extrair entendimentos sobre a evolução dos preços em diferentes regiões e o comportamento de compra e venda. Antes de construir os modelos preditivos, foi necessário investigar e compreender como os dados gerados nos processos anteriores se comportam ao longo do tempo nas diferentes regiões, além de analisar as conexões entre os diversos indicadores. Esse entendimento é crucial para avaliar de que maneira os indicadores podem impactar o mercado e, assim, construir modelos preditivos mais precisos e alinhados à realidade do setor imobiliário.

3.3.1 Exploração de Dados (EDA)

A Análise exploratória de dados (EDA) é uma etapa que visa investigar e resumir as principais características dos dados através de métodos visuais e estatísticos. O principal objetivo da EDA nesta análise é identificar padrões nos preços dos imóveis, taxas de juros e outros indicadores económicos que influenciam o mercado imobiliário. Ao realizar a EDA, é possível entender como os preços de venda e renda variam entre diferentes concelhos e a dispersão desses valores, permitindo identificar áreas com maior volatilidade e potencial de investimento.

O conjunto de dados utilizado, resultante do processo de pré-processamento descrito no capítulo anterior, contém 1.979 linhas e 79 colunas, abrangendo dados temporais de janeiro de 2018 a agosto de 2024. Para organizar e facilitar a análise, as variáveis foram agrupadas em quatro categorias principais. No Apêndice, encontram-se amostras desses agrupamentos, classificadas por temas relevantes: Preços de Venda e Arrendamento (.1), Endividamento dos Particulares (.2), Desemprego e Indicadores Económicos (.3), e Dados Temporais e Euribor (.4). Em termos de estrutura, o conjunto de dados é composto por 52 colunas com valores decimais (float64), 24 com valores inteiros (int64), 2 variáveis categóricas (object), e 1 coluna de dados temporais (datetime64[ns]). Não há valores ausentes, uma vez que o processo de limpeza de dados garantiu que o conjunto estivesse completo e pronto para análise.

A Figura seguinte apresenta a distribuição dos preços dos imóveis por região, representada por um gráfico de boxplot. Nesta visualização, podemos observar que Cascais se destaca por ter os preços mais altos e uma grande variabilidade nos valores, enquanto outras regiões, como Loures e Sintra, apresentam preços mais baixos e menos dispersos. Isso indica que Cascais, além de ser uma região mais cara, possui uma maior variação nos preços, o que pode representar oportunidades ou riscos para os investidores, dependendo do tipo de imóvel e localização específica.

Em continuidade à análise exploratória, é examinou-se como os preços dos imóveis estão distribuídos entre as diferentes regiões. Essa abordagem permite não apenas identificar padrões gerais de preços, mas também a dispersão dentro de cada concelho, ajudando a compreender melhor a dinâmica de valorização imobiliária em áreas específicas. A Figura a

seguir ilustra essas distribuições através de um gráfico de boxplot, facilitando a visualização das variações regionais.

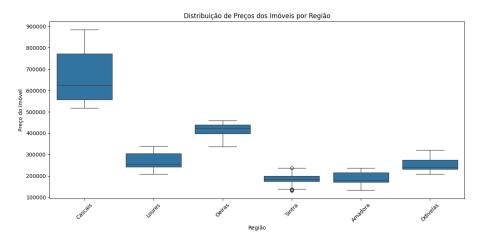


Figura 3.1: Distribuição dos preços dos imóveis por região

Observando o gráfico, destacam-se regiões como Cascais, que exibe os preços mais elevados e uma significativa variabilidade, enquanto concelhos como Loures e Sintra apresentam valores mais modestos e com menor dispersão. Esse comportamento indica que, em Cascais, além dos altos preços, a amplitude de variação é maior, sugerindo tanto oportunidades quanto riscos para investidores, conforme o tipo de imóvel e a localização.

Para complementar essa análise e obter uma visão mais precisa da distribuição dos preços, optou-se por observar as distribuições de preços médios de venda por concelho utilizando um histograma com curvas de densidade, conforme mostrado na figura a seguir. Esses gráficos permitem uma análise mais detalhada da concentração de preços em cada região.

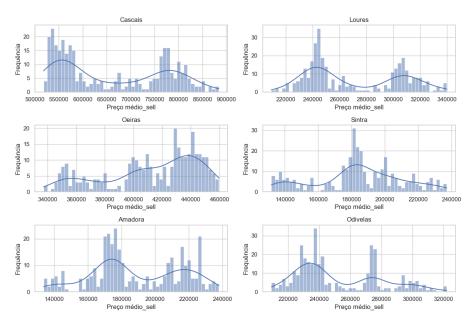


Figura 3.2: Distribuições de preços médios de venda por concelho

Por exemplo, nota-se que a maior parte dos imóveis em Cascais se encontra na faixa de 550.000€ a 750.000€, enquanto em Loures e Odivelas, a maioria dos preços está entre

220.000€ e 300.000€. Esse tipo de visualização facilita a compreensão da distribuição interna dos preços em cada concelho, permitindo uma melhor comparação entre as regiões e a identificação de oportunidades de investimento.

3.3.2 Correlação entre as Variáveis

Um mapa de correlação que podemos ver na figura seguinte foi gerado para identificar as relações entre diferentes variáveis, destacando as conexões entre indicadores económicos e o mercado imobiliário.

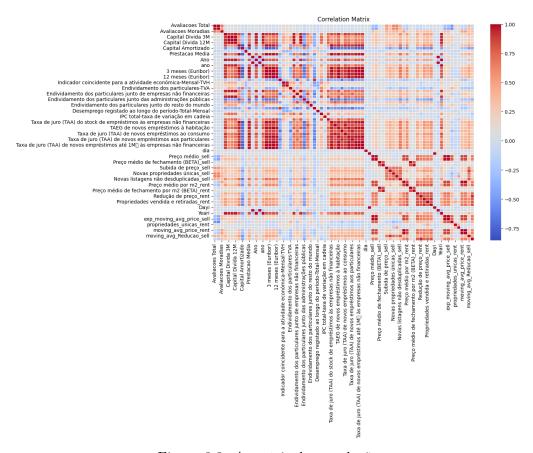


Figura 3.3: A matriz de correlação

A matriz de correlação mostra, uma forte correlação entre a TAEG e os preços de venda de imóveis, indicando que as taxas de juros afetam diretamente o comportamento dos preços. Além disso, a análise com desfasagem temporais (lags) revelou que o impacto das taxas de juros sobre os preços dos imóveis tende a ser mais significativo após alguns meses, mostrando uma resposta retardada do mercado imobiliário às variações nas condições macroeconómicas.

3.3.3 Insights Baseados em Dados

A análise de dados revelou tendências significativas, especialmente em relação à evolução dos preços dos imóveis em diferentes regiões. Em Cascais, por exemplo, observou-se um aumento contínuo dos preços ao longo dos anos (Figura 6), evidenciando uma forte valorização imobiliária na região. Esse comportamento reflete o alto interesse do mercado e a

demanda crescente por imóveis nesta área, consolidando-a como uma das mais caras do país.



Figura 3.4: Relação Preço/Venda e Arrendamento por Região

O price-to-rent ratio, conforme mostrado na Figura , forneceu informações valiosas sobre a atratividade relativa entre comprar ou alugar imóveis em diferentes regiões. Regiões como Amadora e Odivelas apresentaram índices mais baixos, indicando uma maior atratividade para a compra, enquanto regiões como Cascais continuam a mostrar um alto price-to-rent ratio, o que sugere que, para muitos investidores, o arrendamento pode ser uma opção mais vantajosa nesses locais, dependendo das condições económicas e dos objetivos de investimento.

Além disso, conforme mostrado nos gráficos de volatilidade (Figura 3.5), observamos comportamentos distintos no mercado de venda e arrendamento de imóveis em várias regiões. As variações de volatilidade de venda (linha azul) e arrendamento (linha verde) refletem a instabilidade e a sensibilidade do mercado imobiliário a fatores externos.

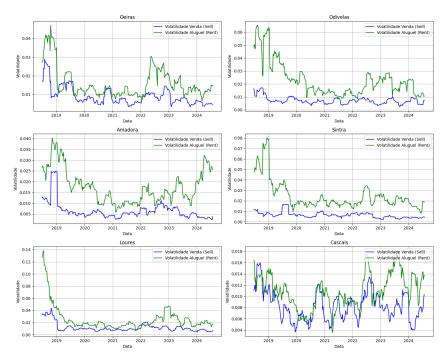


Figura 3.5: Volatilidade dos Preços de Venda e arrendamento nas Regiões de Oeiras, Odivelas, Amadora, Sintra, Loures e Cascais

Em regiões como Oeiras e Loures, a volatilidade no mercado de venda foi mais pronunciada no início do período analisado (2019), mas estabilizou significativamente ao longo do tempo. No entanto, o mercado de arrendamento nessas regiões continuou a apresentar oscilações mais frequentes, sugerindo que o mercado de arrendamento é mais sensível a mudanças nas condições económicas e na demanda.

Cascais, por outro lado, mostra uma convergência maior entre a volatilidade de venda e arrendamento, especialmente nos últimos dois anos, o que indica um equilíbrio entre os dois mercados. Já em Sintra, embora a volatilidade de venda tenha diminuído ao longo do tempo, o mercado de arrendamento ainda apresenta variações significativas, possivelmente refletindo uma maior demanda por arrendamentos na região.

Em resumo, o comportamento da volatilidade dos preços de venda e arrendamento varia de acordo com a região, e essa análise reforça a importância de entender as dinâmicas locais para decisões de investimento mais informadas.

3.4 Desenvolvimento de Modelos de Aprendizagem

Nesta secção, são apresentados os diferentes modelos de aprendizagem automática utilizados para a previsão de preços de imóveis. Foram aplicadas tanto abordagens simples, como a Regressão Linear, quanto técnicas mais avançadas, como o *Random Forest* e as Redes Neurais. O objetivo principal foi avaliar diversas metodologias para identificar o modelo mais eficaz na previsão dos preços médio de venda de imóveis, considerando um amplo conjunto de variáveis económicas e do mercado imobiliário. Cada abordagem foi avaliada com base em métricas de desempenho específicas e na sua capacidade de capturar as nuances dos dados.

3.4.1 Principais Métricas de Avaliação

Para garantir uma avaliação dos modelos de aprendizagem automática aplicados à previsão de preços médio de venda, foram utilizadas várias métricas de desempenho. Essas métricas ajudam a medir a precisão das previsões e a identificar qual modelo oferece os melhores resultados em termos de capacidade de capturar as variações dos dados reais. Cada uma delas oferece uma perspetiva diferente sobre o comportamento dos modelos, desde a medição do erro médio até a capacidade de explicar a variabilidade nos preços dos imóveis. A seguir, detalhamos as principais métricas utilizadas para essa avaliação.

- Erro Absoluto Médio (MAE)- Mede a média das diferenças absolutas entre os valores previstos e os reais. É fácil de interpretar, pois representa o erro médio em unidades monetárias. Quanto menor o MAE, mais precisas são as previsões.
- Erro Quadrático Médio (MSE)- Penaliza os erros grandes com mais intensidade, uma vez que os erros são elevados ao quadrado antes de serem somados. Ele é útil quando se deseja reduzir previsões com grandes desvios.
- Coeficiente de Determinação (R2)- Mede a proporção de variação explicada
 pelas variáveis independentes. Ele varia de −∞ a 1, sendo que valores próximos de
 1 indicam um ajuste mais preciso aos dados reais. Valores negativos indicam que o
 modelo está se saindo pior do que uma simples média dos dados.
- Acurácia- Refere-se à proporção de previsões corretas em relação ao total de previsões feitas. Embora seja mais comum em problemas de classificação, a acurácia
 pode ser aplicada a problemas de regressão transformando as previsões contínuas em
 categorias ou intervalos, representando a percentagem de previsões "suficientemente
 corretas". Quanto maior a acurácia, melhor o desempenho do modelo.

3.4.2 Ajuste de Dados- Etapas Prévias ao Treino dos Modelos

Antes do treino dos modelos de aprendizagem automática, foi necessário realizar uma série de procedimentos para melhorar a qualidade e organização dos dados, facilitando assim o processo de aprendizagem dos modelos. Nesta secção, foram executadas várias etapas importantes, como a normalização dos dados, a divisão em subconjuntos de treino, teste e validação, e a aplicação de técnicas de validação cruzada, para garantir que os modelos generalizem adequadamente em novos dados e evitem problemas como por exemplo a sobre-aprendizagem.

3.4.2.1 Normalização dos Dados

A normalização dos dados é uma etapa que serve para garantir que todas as variáveis contínuas estejam na mesma escala antes de serem utilizadas no modelos de aprendizagem automática. O objetivo da normalização é ajustar os valores das variáveis para que fiquem dentro de um intervalo comum, evitando que variáveis com magnitudes diferentes tenham impactos desproporcionais no processo de aprendizagem, especialmente em algoritmos sensíveis à escala, como Redes Neurais e modelos baseados em distância, como KNN.

Frequentemente, variáveis com valores significativamente maiores, como o preço media de venda dos imóveis, podem dominar outras com valores menores, como taxas de juros ou áreas dos imóveis. Isso pode comprometer o desempenho do modelo, dificultando a aprendizagem das relações entre as variáveis e afetando a precisão das previsões.

Neste estudo, foi utilizado o método de padronização com o *StandardScaler*, da biblioteca *scikit-learn*, que ajusta os dados para que tenham média zero e desvio-padrão igual a um. Esse método foi escolhido porque, além de garantir que as variáveis estejam na mesma escala, ele também ajusta os dados com base em sua distribuição, o que é benéfico quando os dados seguem uma distribuição aproximadamente normal.

O StandardScaler foi aplicado tanto ao conjunto de treino quanto ao de teste. Primeiro, o scaler foi ajustado no conjunto de treino com a função fit_transform, e em seguida, a transformação foi aplicada ao conjunto de teste com a função transform. Isso garantiu que ambos os conjuntos de dados estivessem padronizados, facilitando o processo de aprendizagem dos modelos e prevenindo que diferenças na magnitude das variáveis influenciassem negativamente o desempenho geral. A Figura seguinte 3.6 apresenta o gráfico do preço médio de venda dos imóveis por localização após a aplicação da normalização com o StandardScaler.

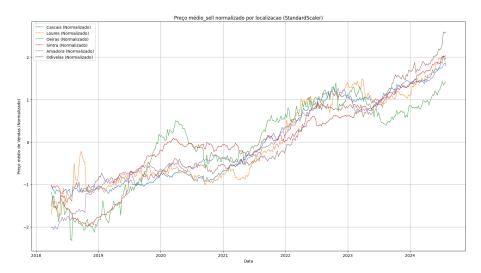


Figura 3.6: Preço media de venda após a normalização

3.4.2.2 Estrutura da Divisão dos Dados

Antes do treino dos modelos foi necessário fazer uma divisão adequada do conjunto de dados tornou-se uma etapa fundamental para garantir que os modelos de aprendizagem automática fossem treinados e avaliados de maneira eficiente e confiável. Essa divisão foi realizada com o objetivo de separar os dados em três subconjuntos- treino, teste e validação. Cada um desses subconjuntos desempenha um papel específico no desenvolvimento de modelos preditivos, garantindo que o desempenho observado durante o treino se generalize para novos dados, evitando problemas tais como a sobre-aprendizagem.

Dividir os dados em diferentes subconjuntos é essencial para avaliar o desempenho real dos modelos preditivos. Se um modelo fosse treinado e testado com o mesmo conjunto

de dados, ele poderia se ajustar demais aos padrões específicos daquele conjunto (sobreaprendizagem), resultando numa alta precisão nos dados de treino, mas falhando ao generalizar para novos dados. Desta forma a divisão de treino, teste e validação permite segue os seguintes princípios:

- Conjunto de treino (70%)- Usado para ajustar o modelo, ou seja, para que o modelo "aprenda" a partir dos dados. Este conjunto foi responsável por ensinar ao algoritmo as relações entre as variáveis independentes (features) e o preço médio de venda de imóveis, que é a variável dependente.
- Conjunto de Teste (20%)- Utilizado para avaliar a performance do modelo avaliando a aprendizagem realizada no treino.
- Conjunto de Validação (10%)- Após o teste inicial, o conjunto de validação foi empregado para uma última etapa de avaliação e ajuste dos hiperparâmetros dos modelos. Isso assegurou que a avaliação final do modelo fosse feita com um conjunto de dados totalmente independente.

A abordagem utilizada para dividir os dados foi realizada sequencialmente, preservando a ordem temporal, o que é particularmente importante em séries temporais ou dados cronológicos, como os preços de imóveis analisados neste estudo. Essa preservação temporal garante que o modelo seja treinado com dados passados e testado em dados futuros, simulando um cenário realista. Na Figura 4.3 a seguir, é ilustrada a divisão do conjunto de dados de acordo com as proporções adotadas- 70% dos dados foram utilizados para treino, 20% para teste e os 10% restantes para validação. A linha gráfica do preço médio de venda dos imóveis também evidencia os pontos de divisão entre os diferentes subconjuntos, permitindo uma visualização clara das proporções utilizadas.

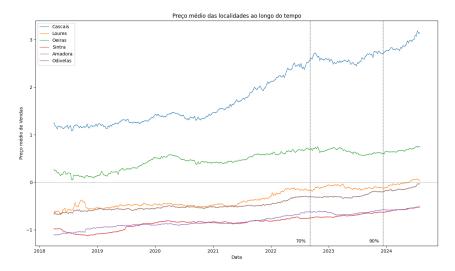


Figura 3.7: Localização da divisão dos Dados

Essas proporções de 70% para treino, 20% para teste e 10% para validação são amplamente utilizadas em estudos de aprendizagem automática, oferecendo um bom equilíbrio entre

a quantidade de dados disponíveis para treinar o modelo e a quantidade necessária para avaliá-lo de forma justa e confiável.

3.4.2.3 Validação Cruzada para Séries Temporais

Além da tradicional divisão dos dados em treino e teste, foi utilizada a técnica de validação cruzada durante o ajuste dos modelos. Essa abordagem é fundamental para garantir que os modelos possam generalizar bem para novos dados, e que o desempenho não dependa exclusivamente de uma única partição do conjunto de dados.

Para alguns modelos, como a Regressão Linear, além do treino convencional, foi utilizada uma validação cruzada específica para séries temporais, a *TimeSeriesSplit*. Essa técnica foi aplicada para garantir que as previsões levassem em consideração a natureza temporal dos dados, onde a ordem cronológica dos dados é respeitada.

No caso do *TimeSeriesSplit*, o conjunto de dados foi dividido em 5 partes, onde o modelo foi treinado em diferentes subconjuntos de treino e validado em subconjuntos de teste de forma incremental, sempre utilizando dados anteriores para treinar e os dados subsequentes para testar. Esse método assegura que o modelo seja testado de forma robusta em diferentes partes da série temporal, sem violar a independência temporal dos dados.

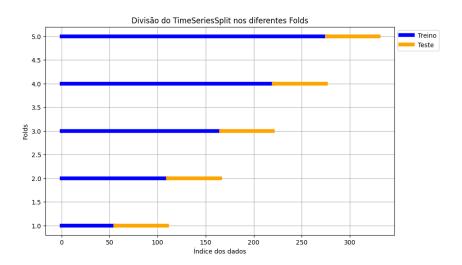


Figura 3.8: Divisão Temporal com TimeSeriesSplit (5 Folds)

A Figura ?? acima ilustra a divisão de uma série temporal em 5 conjuntos de treino e teste, com o treino representado em azul e o teste em laranja. Em cada *fold*, o conjunto de treino cresce progressivamente, e o conjunto de teste é deslocado para os dados subsequentes. Isso garante que o modelo seja testado com dados futuros, mantendo a cronologia e simulando um cenário de previsão real.

Essa abordagem foi especialmente útil para modelos como a regressão linear, pois permitiu que o modelo fosse treinado e avaliado em diferentes partes da série temporal, levando em consideração as características temporais dos dados. A validação cruzada com *Time-SeriesSplit* ajudou a melhorar a capacidade de generalização do modelo ao avaliar seu desempenho de forma incremental e consistente ao longo do tempo.

3.4.3 Regressão Linear

A Regressão Linear é um dos métodos mais simples e amplamente utilizados para prever variáveis contínuas, como o preço médio de venda de imóveis. O objetivo do modelo é ajustar uma linha reta (ou um hiperplano no caso de múltiplas variáveis) que minimize a diferença entre os valores previstos e os valores reais, usando o método de mínimos quadrados. Esta técnica é bastante útil quando há uma relação linear clara entre as variáveis explicativas (independentes) e a variável-alvo (dependente).

Embora seja uma abordagem simples, a Regressão Linear oferece uma boa base para análises iniciais e pode ser comparada com modelos mais avançados para validar a qualidade das previsões. No entanto, sua simplicidade pode ser uma limitação em cenários mais complexos, especialmente quando os dados apresentam não linearidades ou padrões mais difíceis de modelar.

Para otimizar o modelo de Regressão Linear, utilizamos o método GridSearchCV da biblioteca *scikit-learn*, que nos permite testar diferentes combinações de hiperparâmetros e escolher a melhor configuração com base na validação cruzada. O processo de otimização envolveu dois hiperparâmetros principais-

- fit_intercept- determina se uma intercetação (termo constante) será ajustada ou não no modelo.
- copy_X- define se os dados originais de entrada serão copiados ou modificados.

Após o método GridSearchCV , identificamos o melhor modelo e parâmetros. O resultado foi um modelo resultou nos parametros fit_intercept=False e copy_X=True, o que significa que o modelo não incluiu uma intercetação no ajuste e os dados de entrada não foram modificados.

Após identificar o melhor modelo com o GridSearchCV, o modelo foi treinado e testado numa divisão inicial dos dados. A divisão foi feita utilizando 70% dos dados para treino e 20% para teste, o que permitiu uma avaliação inicial do desempenho do modelo.

Os resultados obtidos com esse método foram os seguintes-

Métrica	Valor
MAE Linear	8270
R ² Linear	-0.105
Acurácia	96.25

Tabela 3.1: Resultados do Modelo de Regressão Linear

Esses resultados indicaram um desempenho insatisfatório, com um valor de R^2 negativo, sugerindo que o modelo não foi capaz de explicar a variação nos dados de teste.

Dado que estamos a lidar com dados temporais (preços de imóveis ao longo do tempo), foi aplicado o método TimeSeriesSplit para garantir que a sequência temporal dos dados fosse respeitada. O TimeSeriesSplit divide os dados em subconjuntos crescentes, onde o modelo é treinado nos dados mais antigos e testado nos dados mais recentes. Isso simula melhor o

cenário de previsão do mundo real, simulando melhor o cenário de previsão do mundo real. Os resultados para as diferentes divisões temporais estão apresentados na tabela abaixo-

Tamanho do Treino	Tamanho do Teste	MAE Linear	R ² Linear	Acurácia (%)
370	368	14861361	-17729837	-8859
738	368	70972875	-3324259672	-40950
1106	368	1492	0.583	99.15
1474	368	730	0.943	99.59
1842	368	2284	0.812	98.85
2210	368	22114	-42.724	89.77
2578	368	3093	0.481	98.55
2946	368	2455	0.418	98.92

Tabela 3.2: Resultados da Regressão Linear em Diferentes Divisões Temporais

Os resultados variaram significativamente, com alguns valores de R^2 negativos, sugerindo que o modelo estava ajustando pior do que uma simples previsão da média. No entanto, conforme a quantidade de dados de treino aumentava, o modelo apresentou melhorias, especialmente nas últimas divisões, onde o R^2 foi positivo e o MAE mais baixo.

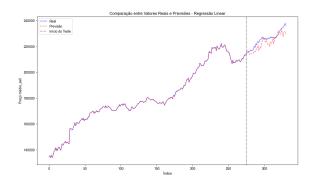


Figura 3.9: Previsão do modelo de Regressão Linear

Os resultados indicam que o modelo de Regressão Linear conseguiu capturar bem os padrões dos dados de treino, com um excelente valor de R² no conjunto de treino, mas o desempenho foi inferior no conjunto de teste, sugerindo a possibilidade de sobre-aprendizagem (overfitting).

Métrica	Valor
MAE (Treino)	213
R ² (Treino)	0.999
MAE (Teste)	3117
R ² (Teste)	0.600

Tabela 3.3: Desempenho do Modelo de Regressão Linear no Conjunto de Treino e Teste

O gráfico a seguir mostra as 20 variáveis mais influentes no modelo de Regressão Linear, de acordo com seus coeficientes-

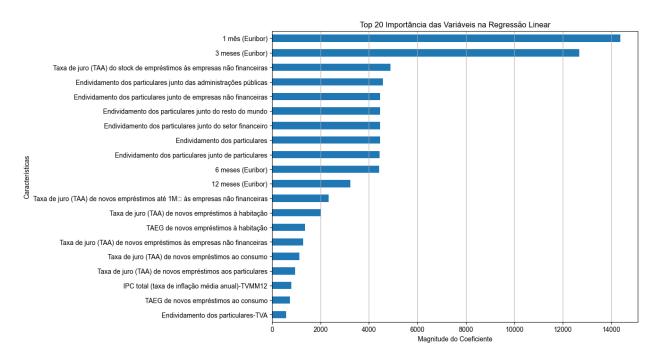


Figura 3.10: Top 20 variáveis mais importantes do modelo Regressão Linear

Essas variáveis desempenharam um papel crucial na previsão do preço médio de venda de imóveis, mostrando a importância de se considerar múltiplas características do mercado imobiliário ao criar modelos preditivos.

3.4.4 Random Forest

O Random Forest é um algoritmo de aprendizagem supervisionado baseado na combinação de múltiplas árvores de decisão. Cada árvore é construída a partir de uma amostra aleatória dos dados e de um subconjunto das variáveis, resultandonum modelo que reduz a variabilidade e aumenta a precisão das previsões. Essa robustez faz com que o Random Forest seja uma excelente escolha para dados complexos e com relações não lineares, como os dados de preços do mercado imobiliário.

Neste estudo, o Random Forest foi escolhido pela pela sua capacidade de lidar com interações complexas entre as variáveis imobiliárias e macroeconómicas. Além disso, o modelo possui uma alta resistência a sobre-aprendizagem (overfitting), característica especialmente útil ao trabalhar com dados temporais, onde as relações entre variáveis podem mudar ao longo do tempo. A flexibilidade desse modelo permitiu capturar padrões não lineares que outros modelos, como a Regressão Linear, poderiam não identificar adequadamente.

Para maximizar o desempenho do Random Forest, foi utilizada a técnica de GridSearchCV para a otimização dos hiperparâmetros. Através desta técnica, foram testadas várias combinações de parâmetros com o objetivo de encontrar a configuração que proporcionasse o menor erro de previsão. A configuração de hiperparâmetros incluía:

• n_estimators- número de árvores na floresta, testado com valores entre 400 e 700.

- max_features- número máximo de variáveis consideradas ao dividir cada nó (teste realizado com os parametros None, sqrt e log2).
- max_depth- profundidade máxima das árvores, variando entre valores de 10, 20, 30 ou sem limite.
- min_samples_split- número mínimo de amostras necessárias para dividir um nó (10, 15 ou 20)
- min_samples_leaf- número mínimo de amostras numa folha (1, 2 ou 4).

Após testar várias combinações de hiperparâmetros, o melhor modelo encontrado foi o RandomForestRegressor com os seguintes parâmetros- max_features=None, min_samples_leaf=4, min_samples_split=15, n_estimators=500 e random_state=42. Essa configuração mostrouse eficaz para modelar os dados, proporcionando um bom equilíbrio entre a complexidade do modelo e a capacidade de generalização dos resultados. Essa configuração mostrou-se a mais eficaz para modelar o comportamento dos preços médio de venda dos imóveis, proporcionando um bom equilíbrio entre a complexidade do modelo e a generalização dos resultados.

Após a otimização com o GridSearchCV, o modelo Random Forest foi treinado e testado com base na divisão tradicional dos dados (70% para treino e 20% para teste). A seguir, estão as métricas de avaliação calculadas:

Métrica	Valor
MAE Linear	9742.66
R ² Linear	-2.89
Acurácia	95.68%

Tabela 3.4: Resultados do modelo Random Forest

Esses resultados indicam que o modelo foi capaz de capturar padrões complexos, embora tenha apresentado um valor negativo de R^2 em algumas divisões, o que pode sugerir um sub-ajuste em determinadas partes dos dados.

Para avaliar a robustez do modelo em diferentes períodos de tempo, foi utilizada a técnica de validação cruzada com séries temporais, o *TimeSeriesSplit*. Esta técnica é apropriada para séries temporais, pois respeita a ordem dos dados, garantindo que as previsões sejam feitas com base em informações passadas. O modelo foi treinadonuma parte dos dados e testado em outra, sempre respeitando a sequência temporal.

Os resultados da validação cruzada demonstraram que o modelo Random Forest teve um desempenho estável em diferentes períodos. Em cada uma das divisões temporais, o erro de previsão variou, mas a acurácia média permaneceu alta. Abaixo estão os resultados das três divisões temporais-

Esses resultados indicam que o modelo Random Forest conseguiu capturar a maior parte da variação nos preços médio de venda dos imóveis, com um erro absoluto médio relativamente baixo, em torno de 4.939,85, o que é considerado um bom desempenho para este tipo de previsão.

Tamanho do Treino	Tamanho do Teste	MAE Linear	${ m R}^2$ Linear	Acurácia (%)
56	55	9901.53	-20.59	94.24%
111	55	2574.55	-0.53	98.54%
166	55	14468.96	-2.30	92.52%
221	55	14868.41	-10.03	93.06%
276	55	9742.66	-2.89	95.68%

Tabela 3.5: Resultados do TimeSeriesSplit com Random Forest

Para ilustrar visualmente o desempenho do modelo, foi gerado um gráfico comparando os valores reais dos preços médio de venda dos imóveis com as previsões do modelo. A linha azul no gráfico representa os valores reais ao longo do tempo, enquanto a linha vermelha tracejada representa as previsões feitas pelo modelo Random Forest. A linha pontilhada preta indica o ponto onde o conjunto de treino termina e o conjunto de teste começa.

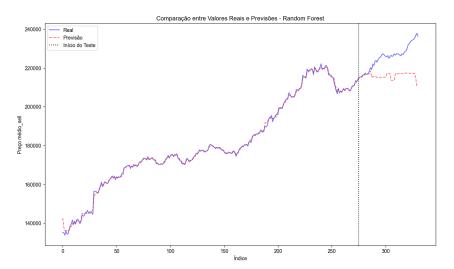


Figura 3.11: Performance do modelo Random florest

O gráfico mostra que as previsões seguem de forma bastante próxima os valores reais, especialmente no conjunto de treino. No entanto, há uma pequena divergência no início do conjunto de teste, onde o modelo subestima os preços. Apesar disso, no geral, as previsões se mantêm bastante próximas dos valores reais, validando a eficácia do Random Forest na previsão dos preços do mercado imobiliário.

O modelo Random Forest foi avaliado quanto a possível sobre-aprendizagem (overfitting). Os resultados a seguir mostram o desempenho tanto no conjunto de treino quanto no conjunto de teste. Observa-se que o desempenho no treino foi muito superior, sugerindo que o modelo se ajustou bem aos dados de treino, mas não conseguiu generalizar adequadamente para o conjunto de teste.

Esses resultados sugerem que o modelo se ajustou bem aos dados de treino, mas o desempenho no conjunto de teste foi consideravelmente inferior, indicando uma possível sobre-aprendizagem (overfitting).

O gráfico a seguir mostra as 20 variáveis mais influentes no modelo Random Forest, de

Métrica	Valor
MAE (Treino)	520.7
R ² (Treino)	0.998
Accurary (Treino)	99.97
MAE (Teste)	9742.6
R ² (Teste)	-2.89
Accurary (Teste)	95.68

Tabela 3.6: Analise do sobre-aprendizagem (overfitting)

acordo com suas importâncias. Essas variáveis influenciaram significativamente o desempenho do modelo, sendo fatores-chave para a previsão de preços imobiliários.

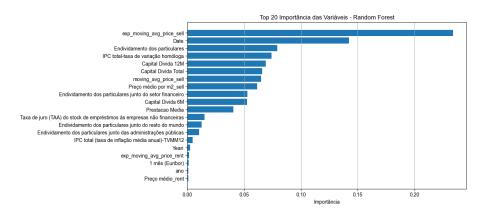


Figura 3.12: Top 20 variáveis mais importantes do modelo Random Forest

3.4.5 Redes Neurais Artificiais: MLP (Multi Layer Perceptron)

As Redes Neurais Artificiais, mais especificamente o Perceptron Multi-camadas Multi-Layer Perceptron (MLP), foram implementadas utilizando o TensorFlow para prever os preços médio de venda dos imóveis. Esta abordagem foi escolhida pela sua capacidade de capturar relações não-lineares complexas entre as variáveis, oferecendo flexibilidade no tratamento de conjuntos de dados desafiadores, como o mercado imobiliário. A arquitetura do modelo e o processo de treino foram projetados para evitar o sobre-ajuste e otimizar o processo de aprendizagem.

Implementação do Modelo

A otimização do modelo foi feita manualmente, em vez de usar o GridSearchCV, com o objetivo de explorar diferentes combinações de hiperparâmetros e identificar a melhor configuração. Foram testados diferentes números de camadas ocultas, neurónios em cada camada, funções de ativação e regularização para minimizar o erro de previsão e prevenir o sobreajuste.

A arquitetura final do modelo MLP foi configurada com:

- Camada de Entrada- Corresponde ao número de variáveis do conjunto de dados.
- Primeira Camada Oculta- 320 neurónios, ativação tanh, com regularização L2 (0.0015).

- Segunda Camada Oculta- 64 neurónios, ativação tanh, com regularização L2 (0.003).
- Camada Final Oculta- Tamanho do batch, ativação tanh.
- Camada de Saída- Um único neurónio sem função de ativação para prever valores contínuos (preços médio de venda dos imóveis).

O modelo foi compilado usando o otimizador Adam com uma taxa de aprendizagem adaptativa, configurada através de um decaimento exponencial. Além disso, foi utilizada a função de perda Huber, devido à sua robustez em relação a *outliers*, que são comuns neste tipo de dados.

A tabela ?? apresenta um resumo detalhado da arquitetura do modelo MLP (Perceptron Multicamadas) utilizado para prever os preços médios de venda dos imóveis que inclui informações sobre o tipo de cada camada, o formato de saída e o número de parâmetros treináveis em cada camada.

Layer (type)	Output Shape	Param #	
dense_276 (Dense)	(None, 320)	24,960	
dense_277 (Dense)	(None, 64)	20,544	
dense_278 (Dense)	(None, 32)	2,080	
dense_279 (Dense)	(None, 1)	33	

Tabela 3.7: Resumo da Arquitetura do Modelo

Treino com 70%, 20% dos dados

O modelo foi treinado por 650 épocas, com um tamanho de lote de 32. Para garantir que o modelo não sofresse de sobreajuste, foi utilizado o Early Stopping, monitorando a perda de validação e restaurando os melhores pesos após 150 épocas sem melhoria significativa. Os gráficos resultantes do treino e validação do modelo estão ilustrados na Figura 4.3.

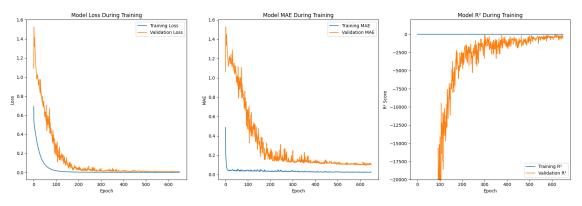


Figura 3.13: Performance do modelo Redes Neurais Artificiais (MLP)

Esses gráficos mostram que a perda e o Erro Absoluto Médio (MAE) tanto no conjunto de treino quanto no de validação convergiram de maneira satisfatória. Observa-se que o modelo aprendeu de forma eficaz sem sinais evidentes de sobreajuste. Além disso, a métrica

R² demonstra que o modelo conseguiu capturar aproximadamente 63,6% da variabilidade dos dados de teste, o que é um resultado razoável, dado a complexidade dos dados.

Avaliação e Métricas

Após o treino, o modelo foi avaliado no conjunto de teste. Os resultados da avaliação, incluindo o MAE, Mean Squared Error (MSE) e o coeficiente de determinação R^2 , estão resumidos na Tabela $\ref{eq:Result}$?

Métrica	Valor
MAE (Treino)	1853.86
R ² (Treino)	0.9987
MSE (Treino)	6350173.89
MAE (Teste)	6775.02
R ² (Teste)	0.636
MSE (Teste)	71710290.39

Tabela 3.8: Resultados e análise da sobre-aprendizagem (overfitting) do modelo MLP

Os valores de MAE, MSE e R^2 foram calculados para avaliar o desempenho do modelo no conjunto de teste, indicando que o modelo conseguiu capturar aproximadamente 53,35% da variabilidade dos dados de teste, o que é um resultado razoável, considerando a complexidade do conjunto de dados.

Demonstração dos Dados Previstos em Gráfico de Série Temporal

A seguir está a demonstração visual das previsões feitas pelo modelo MLP, comparadas com os valores reais dos preços médio de venda dos imóveis no conjunto de teste. O gráfico também exibe os valores reais do conjunto de treino para referência.

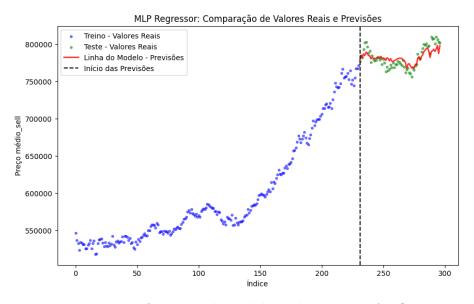


Figura 3.14: Performance do modelo Redes Neurais Artificiais

Apesar de uma pequena divergência no início do conjunto de teste, as previsões do modelo seguem de forma bastante próxima os valores reais, especialmente no conjunto de treino.

A linha preta tracejada indica o ponto em que o conjunto de teste começa, evidenciando a eficácia do modelo na previsão dos preços médio de venda dos imóveis.

3.4.6 Redes Neurais Recorrentes: LSTM (Long Short-Term Memory)

O modelo Long Short-Term Memory (LSTM) foi desenvolvido para capturar dependências temporais nos dados, o que o torna particularmente eficaz para prever os preços médios de venda dos imóveis ao longo do tempo. Essas redes são amplamente usadas em séries temporais devido à sua capacidade de "lembrar"informações relevantes durante longos períodos de tempo, o que é essencial para capturar padrões e tendências no mercado imobiliário.

A otimização dos hiperparâmetros do modelo LSTM foi realizada manualmente, sem o uso de técnicas automatizadas como GridSearchCV. Isso permitiu uma análise mais cuidadosa do impacto de diferentes configurações, como o número de camadas LSTM, neurónios, funções de ativação e regularizadores, sobre o desempenho preditivo. Diversas iterações foram realizadas para encontrar a melhor configuração possível, garantindo que o modelo fosse capaz de generalizar bem, evitando o subajuste.

Arquitetura do Modelo LSTM

A arquitetura final do modelo LSTM foi selecionada após várias iterações de ajustes manuais. A configuração que apresentou os melhores resultados consistiu nas seguintes camadas-

- Primeira Camada LSTM- Com 126 unidades e ativação tanh, configurada para retornar as sequências, permitindo que a próxima camada LSTM processe essas informações. Foi aplicada uma regularização L2 para evitar sobreajuste.
- Segunda Camada LSTM- Também com 126 unidades e ativação tanh, sem retorno de sequência, para fornecer uma saída condensada que alimenta as camadas densas subsequentes.
- Camada Densa- Com número de unidades igual ao tamanho do lote (32), usando tanh como função de ativação.
- Camada de Saída- Um único neurónio sem função de ativação, responsável pela previsão do valor contínuo dos preços médios de venda dos imóveis.

A regularização L2 foi aplicada para reduzir o risco de sobreajuste, e o otimizador Adam foi utilizado com uma taxa de aprendizagem adaptativa, controlada por um decaimento exponencial. O decaimento da taxa de aprendizagem foi configurado para reduzir gradualmente a learning rate após cada 10 épocas, com um fator de decaimento de 0.9644, a fim de estabilizar o processo de treinamento.

A tabela está o resumo da arquitetura do modelo:

Layer (type)	Output Shape	Param #	
lstm_81 (LSTM)	(None, 1, 126)	102,816	
lstm_82 (LSTM)	(None, 126)	127,512	
dense_274 (Dense)	(None, 32)	4,064	
dense_275 (Dense)	(None, 1)	33	

Tabela 3.9: Resumo da Arquitetura do Modelo LSTM

Essa arquitetura resultou num número significativo de parâmetros treináveis, distribuídos pelas camadas LSTM e densas, permitindo ao modelo capturar as complexidades dos dados temporais e prever as variações nos preços.

Treino e Avaliação

O modelo foi treinado durante 450 épocas, utilizando um tamanho de lote de 32. Para evitar sobreajuste, o early stopping foi aplicado, interrompendo o treinamento caso a perda de validação não melhorasse após 150 épocas consecutivas. A função de perda escolhida foi a Huber Loss, conhecida por sua robustez em relação a valores atípicos (outliers). As métricas utilizadas para avaliar o desempenho do modelo foram o MAE e o Coeficiente de Determinação (R^2) .

Após o treinamento, o desempenho do modelo foi avaliado nos conjuntos de treino e teste. A tabela a seguir resume os resultados obtidos:

Métrica	Valor
MSE (Treino)	5.868.810
MAE (Treino)	1.820
R ² (Treino)	0.998
MSE (Teste)	5.7885.163
MAE (Teste)	6.257
R ² (Teste)	0.706

Tabela 3.10: Desempenho do modelo LSTM nos conjuntos de treino e teste

Os resultados indicam que o modelo apresentou um excelente desempenho no conjunto de treino, com um erro muito baixo e um coeficiente de determinação próximo de 1. No entanto, o desempenho no conjunto de teste foi inferior, sugerindo que o modelo pode ter sobreajustado aos dados de treino, necessitando ajustes adicionais para melhorar a generalização.

Visualizações dos Resultados

Abaixo está o gráfico que mostra a evolução da perda e do MAE durante o treinamento, tanto para o conjunto de treino quanto para o de validação.

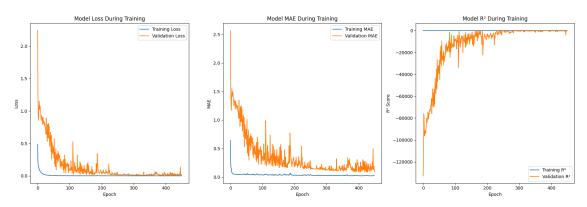


Figura 3.15: Performance do modelo Redes Neurais Artificiais

O gráfico seguinte ilustra a comparação entre os valores reais dos preços médio de venda dos imóveis e as previsões feitas pelo modelo LSTM para o conjunto de teste-

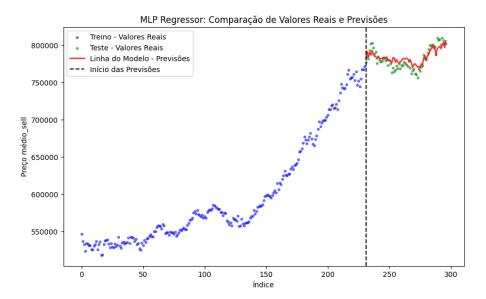


Figura 3.16: Performance do modelo Redes Neurais Artificiais

Esses gráficos mostram que o modelo LSTM foi capaz de capturar a tendência geral dos dados de teste, mas apresentou dificuldades em prever picos e quedas mais acentuadas nos preços, sugerindo que, embora a tendência principal tenha sido identificada, eventos de curto prazo ainda não foram modelados com precisão.

3.5 Comparação dos resultados dos Modelos

Nesta secção, comparamos o desempenho de diferentes modelos de aprendizagem automática treinados e avaliados com base num conjunto de dados de treino e teste. Utilizamos métricas amplamente aceitas para medir a precisão preditiva e a capacidade de generalização de cada modelo. As métricas de avaliação escolhidas para este estudo incluem o Erro Absoluto Médio (MAE), o Erro Quadrático Médio (MSE) e o Coeficiente de Determinação (\mathbb{R}^2).

Essas métricas são essenciais para avaliar a qualidade dos modelos e oferecer uma visão completa do desempenho, com destaque para a precisão (MAE), a sensibilidade a erros maiores (MSE e RMSE) e a explicação da variabilidade dos dados pelo modelo (R²).

3.5.1 Comparação de Modelos- Resultados Gerais

Após o treinamento e avaliação, os resultados das principais métricas foram observados e comparados entre os modelos testados. A tabela a seguir resume os resultados de cada modelo-

Modelo	MAE	\mathbb{R}^2	Acurácia (%)	MSE
Regressão Linear	59 152	-156.46	72.59	4 435 166 683
Regressão Linear TimeSeriesSplit	3 117	0.600	98.61	13 695 927
Random Forest	9 742	-2.89	95.68	133 518 709
Random Forest TimeSeriesSplit	9 742	-2.89	95.68	133 518 709
Dense Neural Network (MLP)	6 775	0.636	_	71 710 290
Redes Neurais Recorrentes (LSTM)	6 257	0.706	_	57 885 163

Tabela 3.11: Comparação de Modelos: Resultados Gerais

A comparação das métricas MAE, MSE, RMSE e R² entre os diferentes modelos permite identificar as principais forças e fraquezas de cada abordagem. Efetuando uma análise destes resultados temos que:

- Erro Absoluto Médio (MAE)- O modelo LSTM foi o mais preciso, com um MAE de 6.257, seguido pelo MLP com 6.775. A Regressão Linear apresentou o maior MAE, com 59.152, indicando que suas previsões foram, em média, muito distantes dos valores reais. O MAE indica que os modelos de redes neurais (MLP e LSTM) foram capazes de capturar padrões complexos com maior precisão, enquanto os modelos tradicionais (Regressão Linear e Random Forest) tiveram um desempenho inferior.
- Erro Quadrático Médio (MSE)- O LSTM teve o menor MSE, seguido de perto pelo MLP. O menor MSE do LSTM reflete sua capacidade de minimizar os grandes erros, o que é importante em aplicações com dados sensíveis, como os preços de imóveis. A Regressão Linear teve o pior desempenho, sugerindo que este modelo foi menos eficaz em evitar grandes desvios.
- Coeficiente de Determinação (R²)- O LSTM apresentou o maior valor de R² (0.706), sugerindo que este modelo foi capaz de explicar a maior parte da variação dos dados. O MLP também teve um R² positivo (0.636), demonstrando um bom desempenho. A Regressão Linear, com R² negativo, mostrou-se inadequada para o problema, incapaz de explicar a variabilidade nos dados.
- Acurácia (%)- A acurácia relativa também foi calculada para alguns modelos. Aqui, a Regressão Linear no *TimeSeriesSplit* mostrou uma precisão muito alta (98.61%), enquanto os modelos mais sofisticados, como Random Forest, tiveram uma acurácia ligeiramente inferior (95.68%).

3.5.2 Escolha do Melhor Modelo

Com base nas métricas analisadas, o LSTM (Long Short-Term Memory) foi selecionado como o modelo de melhor desempenho geral. Esse modelo, especializado para capturar dependências temporais, teve o menor erro (MAE), o melhor MSE e o maior R², mostrando que foi capaz de entender as tendências nos dados temporais de forma mais eficaz do que os outros modelos.

Embora o MLP tenha tido um desempenho muito próximo do LSTM, com um R² positivo e um MAE relativamente baixo, ele não superou o LSTM em termos de precisão e capacidade de generalização.

Por outro lado, o Random Forest demonstrou bom desempenho, especialmente ao lidar com dados que não têm uma estrutura temporal clara. No entanto, sua limitação foi mais evidente nos dados de séries temporais, onde modelos como o LSTM se destacam.

A Regressão Linear, como esperado, apresentou o pior desempenho em todas as métricas. Isso se deve à sua incapacidade de capturar relações não lineares e interações complexas entre as variáveis.

4

Processo de criação da aplicação Web

O objetivo desta fase da tese é desenvolver uma aplicação web que integre modelos de aprendizagem automática para prever preços imobiliários por concelho, oferecendo aos investidores uma ferramenta de fácil utilização para analisar tendências de mercado. A aplicação deverá permitir a visualização de dados preditivos, oferecendo uma interface que possibilite a personalização de filtros e o acesso a informações detalhadas sobre os mercados imobiliários de diferentes regiões.

Além disso, esta tese procura explorar o processo de construção da aplicação, desde o design da interface de utilizador, até à implementação das funcionalidades necessárias para a visualização dos resultados dos modelos preditivos. A aplicação deverá proporcionar uma experiência intuitiva, permitindo que os investidores compreendam rapidamente os dados e utilizem essas informações para apoiar as suas decisões de investimento.

4.1 Plataformas Similares

Plataformas internacionais como o Zillow e o Redfin já utilizam dados preditivos para ajudar investidores a tomar decisões baseadas em tendências imobiliárias. Por exemplo, o Zillow oferece a ferramenta Zestimate, que usa modelos preditivos para estimar o valor de propriedades com base em dados históricos e comparativos de mercado, permitindo prever flutuações de preço. Além disso, fornece uma análise de mercado regional, possibilitando a comparação de tendências entre diferentes áreas e a identificação de regiões com potencial de valorização.

Da mesma forma, o Redfin oferece uma ferramenta de análise de preço de venda, que prevê o valor de venda de uma propriedade com base em fatores como localização e estado do imóvel, auxiliando na estimativa de retorno de investimento. Outra funcionalidade relevante é o uso de *heatmaps* de mercado, que visualizam o nível de competição imobiliária em diferentes áreas, ajudando a identificar regiões com maior potencial de investimento.

Embora essas plataformas sejam eficazes em seus mercados, focadas principalmente nos Estados Unidos, não contem dados nem previsões do mercado português. Por isso, a aplicação desenvolvida nesta fase da tese foi projetada para procurar explorar modelos

preditivos ao contexto local e oferecendo funcionalidades ajustadas às particularidades do mercado de Portugal.

4.2 Descrição da Aplicação

A aplicação desenvolvida foi concebida para fornecer aos investidores imobiliários uma ferramenta prática e interativa que lhes permite analisar o mercado imobiliário de Portugal com diferentes níveis de detalhe. A plataforma está organizada em três páginas principais, cada uma delas abordando um aspeto específico do processo de investimento imobiliário, desde uma análise macroeconómica nacional até uma avaliação detalhada das propriedades individuais. Através de uma interface intuitiva, a aplicação disponibiliza dados preditivos que auxiliam os utilizadores a tomar decisões de investimento com base em tendências e previsões do mercado.

Com funcionalidades como a visão geral do mercado nacional, a análise detalhada por concelho e as métricas financeiras específicas por propriedade, a aplicação diferencia-se pela sua abordagem progressiva e adaptada ao contexto local, permitindo uma análise informada e detalhada do mercado imobiliário de Portugal.

Aqui estão as principais funcionalidades que destacam a aplicação:

- Dashboard (Análise Geral do País): A primeira página, o Dashboard, oferece uma visão geral dos principais indicadores do mercado imobiliário em Portugal. Nesta secção, os utilizadores podem consultar métricas nacionais chave, como o preço médio dos imóveis, o índice de custos de construção, taxas de juro, inflação, entre outros indicadores económicos e imobiliários. Esta análise macroeconómica permite aos investidores terem uma noção clara do estado atual do mercado e da sua evolução a nível nacional. Além disso, a aplicação também oferece uma análise comparativa entre as principais cidades, permitindo uma comparação rápida de variáveis como preços médios, rendimento por metro quadrado e potencial de rentabilidade em várias regiões do país.
- Market Analysis (Análise do Mercado por Concelho): A segunda página, Market Analysis, foca-se numa análise mais granular e específica do mercado de um determinado concelho. Ao selecionar uma cidade ou concelho de interesse, o utilizador tem acesso a gráficos detalhados que mostram as tendências históricas de mercado, como a evolução dos preços dos imóveis, fases de expansão e retração do mercado imobiliário e indicadores como densidade populacional e ocupação. Estes dados ajudam os investidores a compreender as dinâmicas do mercado local e prever possíveis movimentos futuros. O design gráfico e os indicadores foram criados de forma visualmente apelativa, permitindo uma rápida interpretação dos dados e facilitando o processo de decisão.
- Investment Analysis (Análise de Investimento por Propriedade): A terceira página, é dedicada a análises financeiras detalhadas de propriedades individuais. Esta secção permite aos investidores analisar os custos e rendimentos projetados

para uma propriedade específica, calculando o retorno esperado do investimento. A aplicação apresenta informações como o preço de compra, os custos de reparação e financiamento, além de simulações de fluxo de caixa, rendimentos de arrendamento e despesas. Gráficos interativos como diagramas de pizza permitem visualizar a distribuição dos custos de compra e renovação, assim como as proporções de receitas e despesas previstas. Esta página foi desenvolvida para providenciar as métricas necessárias para calcular a viabilidade e rentabilidade de um imóvel específico.

A aplicação destaca-se pela sua capacidade de oferecer uma visão abrangente e ao mesmo tempo detalhada do mercado imobiliário, desde uma análise macroeconómica nacional, passando por uma análise de mercado por concelho, até às métricas financeiras específicas de cada propriedade.

4.3 Design da Aplicação

A fase de design de uma aplicação que serve para garantir que o produto final atende às necessidades e expectativas dos utilizadores. O objetivo é criar uma User Experience (Experiência do Utilizador) (UX) que seja intuitiva, funcional e esteticamente agradável. Este processo envolve várias etapas, que começam com a exploração de ideias por meio de esboços em papel, e evoluem até à criação de wireframes, mockups e protótipos interativos. Cada uma dessas etapas permite iterar e melhorar a interface da aplicação, incorporando feedback dos utilizadores e ajustando as soluções de design para otimizar a usabilidade.

O design do wireframe e do mockup pode ser visualizado no seguinte link do Figma: Projeto no Figma

4.3.1 Esboços em Papel (Paper Sketch)

A criação de esboços em papel, é uma abordagem inicial rápida e eficaz para visualizar e iterar sobre ideias de design. Esta etapa é fundamental para dar forma às primeiras ideias, permitindo a exploração de diferentes *layouts* e conceitos de forma ágil, sem o compromisso de ferramentas digitais. Além disso, ao criar esboços em papel, é possível receber *feedback* de utilizadores e potenciais investidores de maneira simples e imediata, permitindo a identificação precoce de problemas ou oportunidades de melhoria.

Na Figura 4.3, é possível visualizar um exemplo de um esboço em papel criado durante essa fase. O esboço organizou as principais funcionalidades da aplicação de maneira lógica e fácil de navegar.

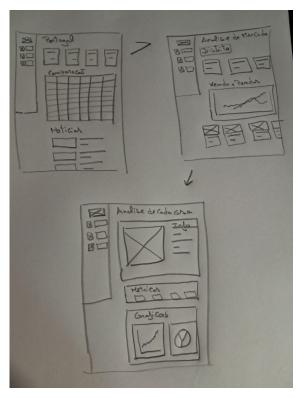


Figura 4.1: Esboço em papel da aplicação

Após a criação dos esboços em papel, a próxima etapa envolve a transição para ferramentas de design digital, como o Figma, que facilitam a criação de *wireframes*, *mockups* e protótipos de alta fidelidade.

4.3.2 Wireframes e Mockups

Após os esboços em papel, o design evoluiu para a criação de wireframes, que são representações estruturais da interface, mapeando a disposição dos principais elementos como botões, menus e áreas de conteúdo. No caso desta aplicação, os ajudaram a definir o layout das páginas principais, como o *Dashboard*, a Análise de Mercado e a Análise de Investimento, garantindo uma navegação lógica e eficiente para os utilizadores.



Figura 4.2: da aplicação

Após a validação dos Wireframes, o processo evoluiu para a criação de Mockups. Enquanto os focam na estrutura, os mockups trazem um nível mais elevado de detalhe visual, incorporando elementos gráficos como cores, tipografia, ícones e imagens. Os mockups fornecem uma visão mais próxima do design final da aplicação, permitindo visualizar como a aplicação será percebida pelo utilizador. Na aplicação desenvolvida, os mockups ajudaram a visualizar como os elementos visuais e gráficos se integram, permitindo a comparação de variações de cor e estilo, e como estas afetam a experiência do utilizador.

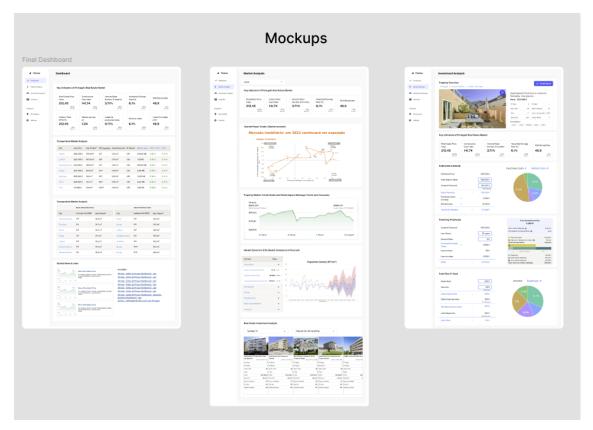


Figura 4.3: Mockups da aplicação

Com os wireframes validados, o processo avançou para a criação de mockups, que trazem um nível mais elevado de detalhe visual. Enquanto os wireframes focam na estrutura, os mockups incorporam elementos gráficos como cores, tipografia, ícones e imagens. Essa fase permitiu que a interface se aproximasse mais do design final, ajudando os stakeholders a visualizar como a aplicação seria percebida pelos utilizadores. Na Figura 4.3, estão representados os mockups finais da aplicação, mostrando como os elementos gráficos foram organizados para otimizar a experiência do utilizador.

Este processo contínuo de iteração e refinamento de ideias, desde os esboços iniciais até os protótipos interativos, garantiu que o design da aplicação não apenas fosse visualmente atrativo, mas também funcional e eficiente no que se refere à experiência do utilizador.

4.3.3 Comparação com Princípios de Design

O design da aplicação foi desenvolvido seguindo princípios de usabilidade, com forte influência das heurísticas de Nielsen, que são diretrizes amplamente aceitas para garantir uma experiência de utilizador otimizada. Entre as heurísticas aplicadas, destacam-se:

 Consistência e padrões: A aplicação mantém uma interface coerente em todas as suas páginas, utilizando padrões familiares para navegação e interação. A consistência é fundamental para garantir que os utilizadores saibam o que esperar ao interagir com diferentes elementos da interface, reduzindo a necessidade de aprender novos comportamentos e minimizando a confusão. • Controle e Liberdade do Utilizador: A aplicação foi desenhada para dar ao utilizador o máximo de controle sobre suas ações, permitindo a reversão fácil de comandos ou o retorno a páginas anteriores sem perder o progresso. Além disso, a interface possibilita que o utilizador navegue livremente entre as diferentes secções, ajustando filtros ou desfazendo ações, promovendo uma experiência sem frustração e mais confiável.

Essas heurísticas foram integradas ao design para garantir que a interface fosse intuitiva, clara e eficiente, proporcionando uma experiência de utilizador otimizada e sem barreiras para a navegação e interação.

4.4 Implementação

A implementação da aplicação focou-se no desenvolvimento de um front-end robusto e interativo, utilizando tecnologias modernas para garantir uma experiência de utilizador fluida e eficiente. A ausência de um back-end complexo não impediu a criação de uma aplicação funcional e bem estruturada, com a capacidade de integrar dados de fontes externas e fornecer visualizações interativas úteis para os investidores imobiliários. Esta secção descreve as principais tecnologias utilizadas, a arquitetura do front-end, e a gestão de dados.

4.4.1 Tecnologias

A aplicação foi dividida em duas camadas principais: back-end e front-end. O back-end foi implementado utilizando Flask, um microframework web em Python, escolhido por sua leveza e flexibilidade na criação de APIs. O Flask foi utilizado para servir os dados armazenados em ficheiros no servidor e processar as requisições do front-end. Esses ficheiros contêm informações sobre o mercado imobiliário, e o Flask é responsável por preparar os dados, transformando-os em respostas JSON que são consumidas pelo front-end.

O front-end da aplicação foi desenvolvido utilizando ReactJS, uma biblioteca JavaScript amplamente utilizada para a construção de interfaces de utilizador dinâmicas. O ReactJS foi escolhido por sua modularidade e eficiência na criação de aplicações dinâmicas através do uso de componentes reutilizáveis. Essa estrutura facilitou o desenvolvimento e garantiu a escalabilidade do projeto, permitindo a criação de uma interface interativa que se ajusta dinamicamente à entrada do utilizador.

Além de ReactJS, foi utilizado o Material UI como biblioteca para facilitar a criação de estilos da pagina. O Material UI oferece um conjunto de componentes pré-construídos, que permitiram acelerar o processo de desenvolvimento e garantir uma aparência moderna e profissional para a aplicação. Entre os componentes utilizados estão botões, menus de navegação, tabelas e painéis interativos, todos consistentes com as práticas de design material.

4.4.2 Arquitetura

A arquitetura da aplicação foi desenhada para ser modular e escalável, com uma clara divisão entre o front-end e o back-end, como mostrado na Figura 4.3.

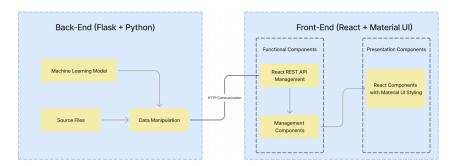


Figura 4.4: Arquitetura da aplicação

O back-end, implementado com Flask (Python), é responsável por processar os dados e modelos de aprendizagem automática. Manipula os source files (arquivos de origem) e envia os dados pre-processados para o front-end via uma Application Programming Interface (API) baseada em Representational State Transfer (REST). Além disso, o Flask executa o modelo de aprendizagem automática para gerar previsões, que também são enviadas ao front-end.

O front-end, desenvolvido com ReactJS e estilizado com Material UI, consome os dados enviados pelo back-end e apresenta visualizações interativas. A comunicação entre o front-end e o back-end é feita através de requisições HTTP (API REST). O front-end faz as requisições ao Flask, que lê os dados, os processa e retorna as respostas no formato JSON. Esses dados são então renderizados pelo React na interface, exibindo gráficos, tabelas e painéis interativos.

A navegação entre as diferentes páginas da aplicação, como o *Dashboard*, *Market Analysis* e *Investment Analysis*, é gerida utilizando o React Router. Isso permite que a aplicação funcione como uma Single Page Application (SPA), onde a navegação ocorre sem recarregar a página inteira, proporcionando uma experiência de utilizador mais fluida.

Além disso, todos os componentes React foram desenvolvidos utilizando o Material UI, garantindo uma estética moderna e consistente com os princípios do Material Design da Google. Isso otimizou o processo de desenvolvimento e assegurou uma interface visualmente coesa em todas as secções da aplicação.

4.4.3 Deploy da Aplicação

A aplicação foi implantada utilizando a plataforma Vercel, amplamente reconhecida por sua integração otimizada com projetos React. O Vercel oferece uma experiência de Continuous Integration (Integração Contínua) (CI)/Continuous Deployment (Deploy Contínuo) (CD), onde qualquer modificação no código é automaticamente implementada e refletida na versão de produção. Essa abordagem garante uma atualização rápida e sem complicações, tornando o processo de manutenção ágil e eficiente.

O Vercel foi escolhido como plataforma de *deploy* por suas inúmeras vantagens, especialmente em projetos baseados em React. Além de integrar-se perfeitamente com repositórios GitHub e GitLab, o Vercel oferece um processo de *deployment* contínuo (CI/CD), o que facilita a implementação automática de novas alterações no código.

A aplicação pode ser acedida através do seguinte link: Aplicação no Vercel.

Nas figuras abaixo, é possível ver capturas de ecrã do website implementado:

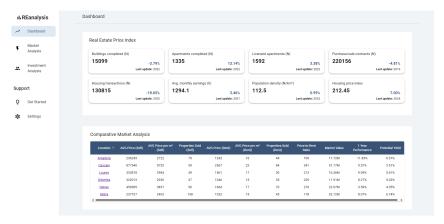


Figura 4.5: Primeira Pagina da aplicação implementada

4.5 Avaliação

Após a implementação da aplicação, foi realizada uma avaliação para analisar a experiência do utilizador (UX) e garantir que a plataforma atendesse às necessidades dos investidores imobiliários de forma eficiente e intuitiva. Para conduzir essa avaliação, utilizei um questionário criado no Google Forms, disponível **aqui**, no qual os participantes puderam fornecer feedback detalhado sobre diversos aspetos da aplicação, como a usabilidade, clareza das informações preditivas e o design da interface. O objetivo era identificar áreas de melhoria e confirmar a eficácia da aplicação em apresentar dados de mercado de maneira acessível.

O questionário foi dividido em diferentes categorias de perguntas para facilitar a obtenção de feedback específico:

- Compreensão e Clareza: Os participantes foram questionados sobre a facilidade de entender as informações preditivas apresentadas e se as previsões de mercado eram claras e acessíveis.
- Facilidade de Navegação: Avaliou-se a experiência de navegação entre as várias secções da aplicação, focando na fluidez e simplicidade da interface.
- Utilidade das Funcionalidades: Os utilizadores deram sua opinião sobre a utilidade das ferramentas de análise de mercado e investimento para a tomada de decisões, refletindo sobre a eficácia da aplicação em fornecer suporte prático.
- Estética e Apelo Visual: Foram feitas perguntas sobre o design visual e a estética da aplicação, analisando se a interface era visualmente atraente e contribuía para uma boa experiência do utilizador.

 Sugestões de Melhoria: Esta categoria recolheu sugestões sobre aspetos da aplicação que poderiam ser melhorados, possibilitando ajustes focados nas necessidades dos utilizadores.

Além disso, o questionário incluiu uma secção de caracterização dos utilizadores, com perguntas sobre género, faixa etária, frequência de uso de aplicações web e experiência com tecnologias, garantindo que os participantes representassem uma amostra diversificada e relevante do público-alvo.

Esta avaliação permitiu recolher um *feedback* valioso, tanto quantitativo quanto qualitativo, proporcionando uma análise abrangente do desempenho da aplicação e ajudando a direcionar futuras melhorias.

4.5.1 Resultados da Avaliação

Os resultados da avaliação forneceram *insights* valiosos sobre a eficácia da aplicação e a experiência do utilizador. De forma geral, a aplicação foi bem recebida, com *feedback* positivo nas seguintes áreas:

- Compreensão e Clareza: A maioria dos participantes considerou as informações preditivas claras e fáceis de entender. Os gráficos e as tabelas foram elogiados pela sua organização e visualização clara dos dados de mercado.
- Facilidade de Navegação: A navegação entre as secções da aplicação, como o Dashboard e as páginas de Análise de Mercado, foi considerada fluida e intuitiva. Os participantes mencionaram que as transições entre páginas eram rápidas e a interface era fácil de explorar, mesmo para novos utilizadores.
- Utilidade das Funcionalidades: Muitos utilizadores destacaram a utilidade das ferramentas de análise de mercado, especialmente na personalização de filtros para obter *insights* detalhados. A possibilidade de analisar diferentes concelhos e regiões foi considerada uma das funcionalidades mais úteis para investidores.
- Estética e Apelo Visual: O design moderno da interface, baseado no Material Design, recebeu elogios pela sua aparência profissional e por criar uma experiência de utilizador visualmente agradável. A consistência visual e o uso de cores e ícones adequados contribuíram para essa avaliação positiva.

No entanto, algumas áreas de melhoria foram identificadas com base nas sugestões de melhoria fornecidas pelos participantes:

Sugestões de Melhoria: Algumas sugestões comuns incluíram a adição de mais opções de personalização nos gráficos e uma maior variedade de dados apresentados, como indicadores socioeconómicos adicionais. Além disso, alguns utilizadores sugeriram a inclusão de tutoriais ou dicas em contexto para ajudar novos utilizadores a compreenderem melhor as funcionalidades mais complexas da aplicação.

5

Avaliação experimental

Neste capítulo, serão apresentados os resultados obtidos a partir dos diferentes modelos de aprendizagem automática aplicados à previsão dos preços imobiliários. Em 5.1 Análise e Comparação dos Resultados, será feita uma análise detalhada e comparativa do desempenho de cada modelo, utilizando métricas de avaliação como o Erro Absoluto Médio (MAE), Erro Quadrático Médio (MSE) e o Coeficiente de Determinação (R²). Em seguida, na secção 5.2 Avaliação da Aplicação Web, será realizada uma avaliação qualitativa da aplicação desenvolvida, discutindo a sua usabilidade, funcionalidade e valor prático para o mercado imobiliário. Por fim, em 5.3 Discussão de Resultados, serão discutidos os *insights* obtidos a partir dos resultados, destacando as implicações para o mercado imobiliário e as possíveis melhorias nos modelos e na aplicação.

5.1 Análise e comparação do desempenho dos Modelos

Nesta secção, será realizada uma avaliação quantitativa dos diferentes modelos de aprendizagem automática utilizados para prever os preços imobiliários. O objetivo é identificar qual modelo apresentou melhor desempenho ao prever os preços médios de venda, considerando diferentes variáveis económicas e do mercado imobiliário. A comparação será feita com base em métricas como MAE, MSE e R², que fornecem uma visão ampla sobre a precisão e a eficácia dos modelos em capturar as variações dos dados reais.

Tabela 5.1: Comparação de Modelos

Modelo		\mathbb{R}^2	Acurácia (%)	MSE
Regressão Linear		-156.46	72.59	4 435 166 683
Regressão Linear TimeSeriesSplit	3 117	0.600	98.61	13 695 927
Random Forest	9 742	-2.89	95.68	133 518 709
Random Forest TimeSeriesSplit	9 742	-2.89	95.68	133 518 709
Dense Neural Network (MLP)	6 775	0.636	-	71 710 290
Redes Neurais Recorrentes (LSTM)	6 257	0.706	_	57 885 163

Com base na análise comparativa apresentada na tabela, podemos observar diferenças significativas no desempenho dos modelos aplicados à previsão dos preços imobiliários. A

Regressão Linear apresentou o pior desempenho, com um MAE elevado e um valor de R² negativo, o que indica uma baixa capacidade de explicação da variação dos dados. Mesmo após a aplicação do *TimeSeriesSplit*, embora tenha mostrado uma ligeira melhoria no R² e MAE, seu desempenho foi inferior em relação a outros modelos mais avançados.

O Random Forest demonstrou ser mais eficaz do que a Regressão Linear, especialmente na versão com *TimeSeriesSplit*, conseguindo capturar melhor as variações nos dados. No entanto, apresentou um R² negativo em algumas situações, sugerindo que o modelo pode ter dificuldades em explicar as flutuações do mercado imobiliário, embora ainda tenha alcançado uma acurácia considerável.

Entre os modelos de redes neurais, o MLP (Rede Neural Multicamadas) teve um desempenho razoável, com um MAE relativamente baixo, embora seu R² não tenha sido tão alto quanto o esperado. O modelo LSTM (Redes Neurais Recorrentes) foi o que apresentou os melhores resultados gerais, com o menor MAE e o maior R², demonstrando uma excelente capacidade de prever os preços ao capturar padrões temporais de longo prazo no mercado imobiliário.

Em conclusão, o LSTM foi o modelo mais adequado para a previsão dos preços imobiliários, devido à sua capacidade de lidar com séries temporais e captar melhor as flutuações e tendências de longo prazo. No entanto, o Random Forest também mostrou ser uma boa alternativa para cenários em que é necessário lidar com variáveis macroeconómicas e imobiliárias complexas.

5.2 Avaliação da Aplicação Web

A secção Avaliação da Aplicação Web será dedicada à análise da experiência de utilização da aplicação, bem como à avaliação das funcionalidades e da usabilidade do sistema. Baseia-se no feedback recolhido dos utilizadores através de questionários e testes de utilização.

5.2.1 Participantes

Para garantir uma avaliação representativa, foram selecionados 10 participantes com perfis diversificados dentro do setor imobiliário. Estes incluíram:

5.2.2 Feedback de Utilizadores

Os primeiros testes com utilizadores revelaram uma aceitação positiva, especialmente por parte de investidores e agentes imobiliários, que consideraram a funcionalidade de previsão de preços e a simulação de cenários como altamente relevantes para suas atividades. A interface simples e intuitiva permitiu que os utilizadores interagissem facilmente com as previsões e realizassem ajustes nos parâmetros de entrada para explorar diferentes cenários económicos.

Durante as sessões de teste, os participantes relataram que a aplicação facilitou a visualização de tendências de mercado e a comparação de preços entre concelhos, ajudando-os a identificar regiões com potencial de valorização. A ferramenta também se destacou pela

capacidade de personalizar os filtros e ajustar variáveis, como as taxas de juro e os rendimentos por metro quadrado, tornando as previsões mais adaptadas às suas necessidades específicas.

5.2.3 Participantes

Para garantir uma avaliação representativa, foram selecionados 10 participantes com perfis diversificados dentro do setor imobiliário. Estes incluíram investidores, agentes imobiliários e analistas de mercado, que utilizaram a aplicação para realizar simulações de cenários de investimento e explorar previsões de preços em diferentes regiões.

5.2.4 Feedback de Utilizadores

Os primeiros testes com utilizadores revelaram uma aceitação positiva, especialmente por parte de investidores e agentes imobiliários, que consideraram a funcionalidade de previsão de preços e a simulação de cenários como altamente relevantes para suas atividades. A interface simples e intuitiva permitiu que os utilizadores interagissem facilmente com as previsões e realizassem ajustes nos parâmetros de entrada para explorar diferentes cenários económicos.

Durante as sessões de teste, os participantes relataram que a aplicação facilitou a visualização de tendências de mercado e a comparação de preços entre concelhos, ajudando-os a identificar regiões com potencial de valorização. A ferramenta também se destacou pela capacidade de personalizar os filtros e ajustar variáveis, como as taxas de juro e os rendimentos por metro quadrado, tornando as previsões mais adaptadas às suas necessidades específicas.

5.2.5 Melhorias Futuras da Aplicação

Com base no *feedback* dos utilizadores, algumas melhorias foram identificadas para versões futuras da aplicação, incluindo:

- Integração de novos dados: Incorporar dados em tempo real de fontes públicas como o INE e o Banco de Portugal, permitindo que as previsões reflitam as mudanças mais recentes no mercado.
- Otimização de performance: Melhorar o tempo de resposta, especialmente em modelos mais complexos, como o LSTM, que demandam maior poder computacional.
- Novas Funcionalidades de Simulação: Expandir a simulação de cenários para incluir múltiplas variáveis simultaneamente e permitir a comparação de resultados lado a lado.
- Melhorias na experiência de utilizador: Adicionar tutoriais em contexto e dicas sobre como interpretar os resultados preditivos, especialmente para novos utilizadores.

5.3 Discussão de Resultados

A análise dos resultados experimentais indica que os modelos de aprendizagem automática são uma ferramenta poderosa para prever preços imobiliários, mas é necessário um ajuste cuidadoso de cada modelo para otimizar o desempenho. Os resultados sugerem que o LSTM é a melhor escolha para capturar padrões temporais de longo prazo, enquanto o Random Forest se mostrou mais eficaz na captura de interações complexas entre variáveis económicas e imobiliárias.

No entanto, há espaço para melhorias, especialmente na integração de dados em tempo real e na otimização do tempo de processamento dos modelos mais complexos. Além disso, o desenvolvimento de funcionalidades adicionais na aplicação web, como previsões personalizadas e simulações de múltiplos cenários, pode aumentar a utilidade da ferramenta para investidores e agentes do setor imobiliário.

Este trabalho demonstra o potencial da combinação de técnicas de aprendizagem automática e uma plataforma interativa para fornecer previsões precisas e insights práticos sobre o mercado imobiliário. No entanto, a continuação da pesquisa e a expansão das funcionalidades da aplicação serão fundamentais para maximizar o seu impacto no setor.

6 Conclusão e Trabalhos Futuros

Neste trabalho, foi realizada uma análise abrangente do mercado imobiliário em Portugal, utilizando diversas técnicas de aprendizagem automática para prever os preços médios de venda de imóveis. Foram exploradas diferentes fontes de dados, incluindo variáveis macroeconómicas, taxas de juro, indicadores específicos do mercado imobiliário e dados históricos de preços, com o objetivo de criar modelos preditivos robustos e avaliar a influência de múltiplos fatores sobre o comportamento dos preços de imóveis.

6.1 Conclusão

Os resultados deste estudo demonstraram que modelos de aprendizagem automática podem ser ferramentas poderosas para prever preços no mercado imobiliário, especialmente quando múltiplas fontes de dados são combinadas. O Random Forest destacou-se como o modelo com melhor desempenho, capturando de forma eficaz as interações complexas entre variáveis económicas e de mercado. O LSTM, embora tenha mostrado bom desempenho em séries temporais, teve tempos de processamento mais longos e menor precisão em comparação com o Random Forest. A Regressão Linear, por sua vez, mostrou-se limitada para capturar as relações não lineares presentes no mercado imobiliário.

A análise das principais variáveis influentes confirmou que fatores macroeconómicos, como as taxas de juro e o índice de preços ao consumidor (IPC), têm um impacto significativo nos preços dos imóveis, juntamente com variáveis específicas do mercado, como o preço médio por metro quadrado e a relação preço/arrendamento. Esses resultados ressaltam a natureza multifacetada do mercado imobiliário, onde condições económicas mais amplas e características locais do mercado interagem para definir o comportamento dos preços.

Além disso, a implementação da aplicação web provou ser uma ferramenta útil e prática para visualização de dados e previsão de preços imobiliários, permitindo que os utilizadores façam simulações de cenários e explorem diferentes variáveis que influenciam o mercado. O feedback positivo dos primeiros testes com utilizadores reforça a relevância de ferramentas interativas que oferecem previsões e análises baseadas em dados.

6.2 Trabalho Futuro

Embora os resultados obtidos sejam promissores, existem várias oportunidades para aprimorar o trabalho realizado e expandir suas aplicações. A seguir, estão algumas sugestões para trabalhos futuros:

- Otimização de Modelos Avançados: Embora o Random Forest tenha se destacado, outros modelos avançados, como LSTM, podem ser otimizados para melhorar o desempenho, tanto em termos de precisão quanto de tempo de processamento. Ajustes de hiperparâmetros e a incorporação de técnicas como grid search podem ajudar a encontrar configurações mais eficientes.
- Integração de Dados em Tempo Real: Um aprimoramento significativo seria a integração de dados em tempo real na aplicação web, possibilitando que as previsões sejam sempre atualizadas com base nas mais recentes variações macroeconómicas, como alterações nas taxas de juro e indicadores de inflação. Isso aumentaria a utilidade da ferramenta em ambientes voláteis e dinâmicos.
- Incorporação de Novas Variáveis: Embora as variáveis utilizadas no estudo tenham mostrado forte correlação com os preços dos imóveis, futuras pesquisas podem explorar a incorporação de novas variáveis, como características demográficas e sociais, fatores ambientais e dados sobre infraestrutura e desenvolvimento urbano.
- Expansão Geográfica: O estudo foi focado no mercado imobiliário de Portugal, mas o modelo pode ser adaptado para outros países ou regiões. Comparar o desempenho dos modelos em diferentes mercados internacionais poderia fornecer insights valiosos sobre os padrões globais de precos de imóveis.
- Exploração de Modelos de Explainable AI (XAI): Para melhorar a interpretabilidade dos modelos, técnicas de Explainable AI podem ser utilizadas. Isso permitiria que as previsões fossem mais facilmente compreendidas pelos utilizadores, aumentando a confiança na tomada de decisões e ajudando a explicar como cada variável influencia os preços.
- Desenvolvimento de Funcionalidades Adicionais para a Aplicação Web: A aplicação pode ser expandida com funcionalidades como previsões personalizadas para imóveis específicos, avaliação do impacto de eventos económicos, e melhorias na simulação de múltiplos cenários. Além disso, a integração de ferramentas de visualização de dados mais avançadas pode melhorar a experiência do utilizador.
- Colaboração com Setores Públicos e Privados: Parcerias com órgãos públicos
 e empresas do setor imobiliário podem fornecer previsões customizadas e insights
 práticos. A aplicação pode ser usada como uma ferramenta de suporte à tomada de
 decisão, tanto em nível governamental quanto corporativo.

Bibliografia

- [1] [Online; accessed 19. Mar. 2024]. Set. de 2017. URL: https://pdfs.semanticscholar.org/0e5f/ab193cf788e6bc76305f9e2d5664b6351496.pdf (ver p. 17).
- [2] [Online; accessed 13. Mar. 2024]. Out. de 2018. URL: https://ceejme.eu/wp-content/uploads/2018/10/ceejme_4_7_art_10.pdf (ver pp. 10, 11).
- [3] [Online; accessed 13. Mar. 2024]. Out. de 2021. URL: https://www.superiorschoolnc.com/wp-content/uploads/2019/10/NC-8e-Superior-Chs-1-3.pdf (ver p. 9).
- [4] [Online; accessed 18. Mar. 2024]. Mar. de 2024. URL: https://www.ijcsma.com/articles/literature-review-on-real-estate-value-prediction-using-machine-learning.pdf (ver p. 16).
- [5] E. A. Alburshaid e A. I. Al-Alawi. "Forecasting Methods for the Real Estate Market: A Review". Em: ResearchGate (nov. de 2023). URL: https://www.researchgate.net/publication/376033071_Forecasting_Methods_for_the_Real_Estate_Market_A_Review (ver pp. 2, 14, 19).
- [6] P. P. Angelov, E. A. Soares, R. Jiang, N. I. Arnold e P. M. Atkinson. "Explainable artificial intelligence: an analytical review". Em: WIREs Data Min. Knowl. Discovery 11.5 (set. de 2021), e1424. ISSN: 1942-4787. DOI: 10.1002/widm.1424 (ver p. 16).
- [7] J. Barros. "O que são os Sistemas de Informação Geográfica (SIG) e quais suas aplicações Geoaplicada". Em: *Geoaplicada* (mar. de 2023). URL: https://www.geoaplicada.com/sig-e-suas-aplicacoes/#0_que_e_um_Sistema_de_Informação_Geográfica (ver p. 17).
- [8] E. Circle. "REAL ESTATE MARKET AND ITS CHARACTERISTICS". Em: VA-LUER WORLD (jan. de 2022). URL: https://www.valuerworld.com/2022/01/20/real-estate-market-and-its-characteristics (ver pp. 10, 11).
- [9] Doutor Finanças Cuidamos da sua saúde financeira. "A evolução dos preços das casas em Portugal". Em: Doutor Finanças Cuidamos da sua saúde financeira (set. de 2023). URL: https://www.doutorfinancas.pt/imobiliario/a-evolucao-dos-precos-das-casas-em-portugal (ver p. 2).
- [10] M. Elsaraiti e A. Merabet. "A Comparative Analysis of the ARIMA and LSTM Predictive Models and Their Effectiveness for Predicting Wind Speed". Em: *Energies* 14.20 (out. de 2021), p. 6782. ISSN: 1996-1073. DOI: 10.3390/en14206782 (ver p. 14).

- [11] S. Enterprises. "5 Ways AI May Impact Real Estate Investing | Stark Enterprises". Em: Stark Enterprises | (fev. de 2024). URL: https://starkenterprises.com/5-ways-ai-may-impact-real-estate-investing (ver p. 3).
- [12] European Central Bank. "Projeções macroeconómicas para a área do euro elaboradas por especialistas do BCE, setembro de 2023". Em: *European Central Bank* (set. de 2023). ISSN: 2529-4466. DOI: 10.2866/746241 (ver p. 2).
- [13] N. Ferlan, M. Bastic e I. Psunder. "Influential Factors on the Market Value of Residential Properties". Em: EE 28.2 (abr. de 2017), pp. 135–144. ISSN: 2029-5839. DOI: 10.5755/j01.ee.28.2.13777 (ver p. 11).
- [14] Forecasting and Econometric Models Econlib. [Online; accessed 19. Mar. 2024].

 Out. de 2018. URL: https://www.econlib.org/library/Enc/ForecastingandEconometricModels.

 html (ver p. 17).
- [15] B. Grum e D. K. Govekar. "Influence of Macroeconomic Factors on Prices of Real Estate in Various Cultural Environments: Case of Slovenia, Greece, France, Poland and Norway". Em: *Procedia Economics and Finance* 39 (jan. de 2016), pp. 597–604. ISSN: 2212-5671. DOI: 10.1016/S2212-5671(16)30304-5 (ver p. 12).
- [16] A. Grybauskas, V. Pilinkienė e A. Stundžienė. "Predictive analytics using Big Data for the real estate market during the COVID-19 pandemic". Em: *Journal of Big Data* 8.1 (2021). DOI: 10.1186/s40537-021-00476-0 (ver p. 13).
- [17] Hedonic Pricing: Definition, How the Model Is Used, and Example. [Online; accessed 18. Mar. 2024]. Mar. de 2024. URL: https://www.investopedia.com/terms/h/hedonicpricing.asp (ver p. 15).
- [18] E. M. Kamal, H. Hassan e A. Osmadi. "Factors influencing the housing price: developers' perspective". Em: *International Journal of Humanities and Social Sciences* 10.5 (2016), pp. 1676–1682 (ver p. 14).
- [19] Y. Kong, J. L. Glascock e R. Lu-Andrews. "An Investigation into Real Estate Investment and Economic Growth in China: A Dynamic Panel Data Approach". Em: Sustainability 8.1 (jan. de 2016), p. 66. ISSN: 2071-1050. DOI: 10.3390/su8010066 (ver p. 17).
- [20] A. N. Laia, J. Montezuma e J. C. d. Neves. Análise de Investimentos Imobiliários. Texto Editores, 2009, pp. 1–410. ISBN: 978-972-47-3967-6 (ver p. 10).
- [21] G. Liu. "Research on Prediction and Analysis of Real Estate Market Based on the Multiple Linear Regression Model". Em: Sci. Program. 2022 (mai. de 2022). ISSN: 1058-9244. DOI: 10.1155/2022/5750354 (ver p. 15).
- [22] L. Liu e L. Wu. "Predicting housing prices in China based on modified Holt's exponential smoothing incorporating whale optimization algorithm". Em: Socioecon. Plann. Sci. 72 (dez. de 2020), p. 100916. ISSN: 0038-0121. DOI: 10.1016/j.seps. 2020.100916 (ver p. 14).

- [23] F. Lorenz, J. Willwersch, M. Cajias e F. Fuerst. "Interpretable machine learning for real estate market analysis". Em: Real Estate Econ. 51.5 (set. de 2023), pp. 1178– 1208. ISSN: 1080-8620. DOI: 10.1111/1540-6229.12397 (ver pp. 15, 16, 18).
- [24] R. Ma et al. "Research on the Main Factors Affecting House Prices". Em: Academic Journal of Business & Management 4.3 (2022), pp. 103–105 (ver p. 12).
- [25] C. V. Marian e M. Iacob. "The ArchTerr Project—A GIS-Based Integrated System for Cultural and Archaeological Heritage Protection (Pilot Phase Tested in Romania)". Em: Appl. Sci. 12.16 (ago. de 2022), p. 8123. ISSN: 2076-3417. DOI: 10.3390/app12168123 (ver p. 18).
- [26] R.-T. Mora-Garcia, M.-F. Cespedes-Lopez e V. R. Perez-Sanchez. "Housing Price Prediction Using Machine Learning Algorithms in COVID-19 Times". Em: Land 11.11 (nov. de 2022), p. 2100. ISSN: 2073-445X. DOI: 10.3390/land11112100 (ver p. 13).
- [27] O que são SIG? | Sistemas de Informação Geográfica. [Online; accessed 22. Mar. 2024]. Mar. de 2024. URL: https://www.esri-portugal.pt/pt-pt/o-que-sao-os-sig/overview (ver p. 17).
- [28] P.-F. Pai e W.-C. Wang. "Using Machine Learning Models and Actual Transaction Data for Predicting Real Estate Prices". Em: *Appl. Sci.* 10.17 (ago. de 2020), p. 5832. ISSN: 2076-3417. DOI: 10.3390/app10175832 (ver p. 11).
- [29] P. Pashardes, C. S. Savva et al. "Factors affecting house prices in Cyprus: 1988-2008".
 Em: Cyprus Economic Policy Review 3.1 (2009), pp. 3-25 (ver p. 12).
- [30] Real Estate China | Statista Market Forecast. [Online; accessed 30. Apr. 2024]. Abr. de 2023. URL: https://www.statista.com/outlook/fmo/real-estate/china (ver p. 2).
- [31] Real Estate Worldwide | Statista Market Forecast. [Online; accessed 29. Feb. 2024]. Fev. de 2024. URL: https://www.statista.com/outlook/fmo/real-estate/worldwide#methodology (ver p. 1).
- [32] Regressão Linear: Principais conceitos, tipos, limitações, aplicações e exemplo MakerZine. [Online; accessed 18. Mar. 2024]. Mar. de 2024. URL: https://www.makerzine.com.br/cozinha-de-dados/regressao-linear-principais-conceitostipos-limitacoes-aplicacoes-e-exemplo (ver p. 15).
- [33] Residential Real Estate Global | Statista Market Forecast. [Online; accessed 30. Apr. 2024]. Abr. de 2023. URL: https://fr.statista.com/outlook/fmo/real-estate/residential-real-estate/worldwide (ver p. 2).
- [34] B.-O. Sjöling. Indicators for bubble formation in housing markets. 2012 (ver p. 13).
- [35] View of Hedonic Valuation of Real Estate Properties in Nigeria. [Online; accessed 18. Mar. 2024]. Mar. de 2024. URL: https://journals.uct.ac.za/index.php/ JARER/article/view/452/537 (ver pp. 16, 18).

- [36] L. Wu, Y. Zhang, Y. Wei e F. Chen. "A BP Neural Network-Based GIS-Data-Driven Automated Valuation Framework for Benchmark Land Price". Em: Complexity 2022 (abr. de 2022). ISSN: 1076-2787. DOI: 10.1155/2022/1695265 (ver pp. 17, 18).
- [37] Y. Yarotska. "AI Real Estate Investing: Top 7 Cases of Using AI for Real Estate Investment". Em: *Ascendix Tech* (fev. de 2024). URL: https://ascendixtech.com/ai-real-estate-investment (ver p. 3).

Tabela .1: Dados de Preços de Venda e Arrendamento

ID	248	579	903	1234	1565
Preço médio_sell	781410.0	318903.0	448916.0	200174.0	216328.0
Preço médio por m ² _sell	5340.0	2784.0	3564.0	2153.0	2484.0
Preço médio de fechamento	0.0	0.0	0.0	0.0	0.0
(BETA)_sell					
Preço médio de fechamento	0.0	0.0	0.0	0.0	0.0
por m ² (BETA)_sell		10.0		10.0	
Subida de preço_sell	5.0	10.0	5.0	10.0	5.0
Redução de preço_sell	19.0	15.0	12.0	22.0	11.0
Novas propriedades	35.0	30.0	25.0	58.0	34.0
únicas_sell					
Propriedades vendidas e	30.0	21.0	27.0	103.0	37.0
retiradas_sell					
Novas listagens não	68.0	50.0	49.0	89.0	60.0
desduplicadas_sell					
Preço médio_rent	2192.0	1193.0	1499.0	1271.0	842.0
Preço médio por m ² _rent	18.0	12.0	14.0	13.0	12.0
Preço médio de fechamento	0.0	0.0	0.0	0.0	0.0
(BETA)_rent					
Preço médio de fechamento	0.0	0.0	0.0	0.0	0.0
por m ² (BETA)_rent					
Subida de preço_rent	4.0	0.0	2.0	0.0	0.0
Redução de preço_rent	5.0	1.0	0.0	1.0	0.0
Novas propriedades	16.0	2.0	9.0	7.0	3.0
únicas_rent					
Propriedades vendidas e	20.0	4.0	12.0	15.0	14.0
retiradas_rent					
Novas listagens não	24.0	3.0	10.0	7.0	5.0
desduplicadas_rent					
price_to_rent_ratio	356.482664	267.311819	299.476985	157.493312	256.921615

@is@a@figure

Tabela .2: Endividamento dos Particulares

Métrica	248	579	903	1234	1565
Endividamento dos	151348.35	151348.35	151348.35	151348.35	151348.35
particulares					
Endividamento dos	136796.08	136796.08	136796.08	136796.08	136796.08
particulares junto do setor					
financeiro					
Endividamento dos	10067.38	10067.38	10067.38	10067.38	10067.38
particulares junto de					
empresas não financeiras					
Endividamento dos	96.07	96.07	96.07	96.07	96.07
particulares junto das					
administrações públicas					
Endividamento dos	953.64	953.64	953.64	953.64	953.64
particulares junto de					
particulares					
Endividamento dos	3435.19	3435.19	3435.19	3435.19	3435.19
particulares junto do resto do					
mundo					
TAEG de novos empréstimos	4.82	4.82	4.82	4.82	4.82
à habitação					
TAEG de novos empréstimos	10.48	10.48	10.48	10.48	10.48
ao consumo					
Taxa de juro (TAA) de novos	3.30	3.30	3.30	3.30	3.30
empréstimos à habitação					
Taxa de juro (TAA) de novos	8.45	8.45	8.45	8.45	8.45
empréstimos ao consumo					
Taxa de juro (TAA) de novos	4.69	4.69	4.69	4.69	4.69
empréstimos às empresas não					
financeiras					
Capital Divida Total	62356.00	62356.00	62356.00	62356.00	62356.00
Capital Divida 3M	126262.00	126262.00	126262.00	126262.00	126262.00
Capital Divida 6M	127634.00	127634.00	127634.00	127634.00	127634.00
Capital Divida 12M	125771.00	125771.00	125771.00	125771.00	125771.00
Juros Totais	114.00	114.00	114.00	114.00	114.00
Capital Amortizado	201.00	201.00	201.00	201.00	201.00

Tabela .3: Desemprego e Indicadores Econômicos

Métrica	248	579	903	1234	1565
Unemployed population	391.2	391.2	391.2	391.2	391.2
Número de beneficiários de	183.9	183.9	183.9	183.9	183.9
subsídio de desemprego					
(milhares)-mensal					
Desemprego registado no final	322086.0	322086.0	322086.0	322086.0	322086.0
do período-Total-Mensal					
Desemprego registado ao	55841.0	55841.0	55841.0	55841.0	55841.0
longo do					
período-Total-Mensal					
Indicador coincidente para a	3.5	3.5	3.5	3.5	3.5
atividade					
económica-Mensal-TVH					
Indicador coincidente para o	1.7	1.7	1.7	1.7	1.7
consumo					
privado-Mensal-TVH					
IPC total-taxa de variação	8.4	8.4	8.4	8.4	8.4
homóloga					
IPC total-taxa de variação	-0.8	-0.8	-0.8	-0.8	-0.8
em cadeia					
IPC total (taxa de inflação	8.2	8.2	8.2	8.2	8.2
média anual)-TVMM12					

Tabela .4: Dados Temporais e Euribor

Métrica	248	579	903	1234	1565
Data	1.012023e+06	1.012023e+06	1.012023e+06	1.012023e+06	1.012023e+06
Ano	2023	2023	2023	2023	2023
Mês	1.00	1.00	1.00	1.00	1.00
dia	31.00	31.00	31.00	31.00	31.00
1 mês (Euribor)	1.981	1.981	1.981	1.981	1.981
3 meses (Euribor)	2.345	2.345	2.345	2.345	2.345
6 meses (Euribor)	2.858	2.858	2.858	2.858	2.858
12 meses (Euribor)	3.337	3.337	3.337	3.337	3.337
Dayi	1.00	1.00	1.00	1.00	1.00
Monthi	1.00	1.00	1.00	1.00	1.00
Yeari	2023	2023	2023	2023	2023
moving_avg_price_sell	780915.3	316897.3	447138.3	200168.0	217517.3
exp_moving_avg_price_sell	780833.0	316940.0	447408.9	200084.9	217336.9
moving_avg_Reducao_sell	15.66667	11.00000	14.66667	22.33333	13.00000
moving_avg_price_rent	2159.333	1192.000	1522.000	1275.333	843.333
exp_moving_avg_price_rent	2161.802	1192.356	1516.559	1270.812	843.431
moving_avg_Reducao_rent	6.66667	0.33333	0.00000	0.66667	0.00000