



Análises Imobiliárias com Inteligência Artificial para Investidores

RÚBEN ALEXANDRE ALBUQUERQUE SANTOS
(Licenciado)

Trabalho de Projeto para obtenção do Grau de Mestre em
Engenharia Informática e Multimédia

Orientador: Doutor Artur Jorge Ferreira

Júri:

Presidente: Doutor Pedro Mendes Jorge

Vogais: Doutor Paulo Trigo
Doutor Artur Jorge Ferreira

Dezembro de 2024



Análises Imobiliárias com Inteligência Artificial para Investidores

RÚBEN ALEXANDRE ALBUQUERQUE SANTOS
(Licenciado)

Trabalho de Projeto para obtenção do Grau de Mestre em
Engenharia Informática e Multimédia

Orientador: Doutor Artur Jorge Ferreira, DEETC/ISEL

Júri:

Presidente: Doutor Pedro Mendes Jorge, DEETC/ISEL

Vogais: Doutor Paulo Trigo, DEETC/ISEL

Doutor Artur Jorge Ferreira, DEETC/ISEL

Dezembro de 2024

Agradecimentos

A elaboração deste trabalho não teria sido possível sem a colaboração, estímulo e muito empenho de diversas pessoas. Gostaria assim, de expressar toda a minha gratidão e apreço a todos aqueles que, direta ou indiretamente, contribuíram para que esta tarefa se tornasse uma realidade. A todos quero manifestar os meus sinceros agradecimentos.

Em primeiro lugar, gostaria de agradecer ao meu orientador, Professor Artur Ferreira, por sua orientação precisa e apoio ao longo deste processo. Sua paciência, conhecimento e incentivo foram fundamentais para a realização deste trabalho. Agradeço por todas as reuniões, conselhos e por sempre me guiar nas direções certas, mesmo nos momentos mais desafiadores.

Aos meus familiares, especialmente à minha mãe Florbela, agradeço pelo apoio constante durante toda essa jornada. Sua presença e incentivo foram fundamentais para que eu pudesse concluir este trabalho.

Agradeço à minha namorada Inês, pelo apoio e pela parceria ao longo deste percurso. Sua paciência, além de suas valiosas contribuições nas ideias e nas mudanças de perspectiva sobre diversos pontos, foram fundamentais em várias etapas do processo.

Agradeço aos meus amigos pelo apoio, pelos conselhos e pela disponibilidade para discutir ideias e esclarecer dúvidas ao longo desta jornada. Em especial, agradeço a Gonçalo, Inês, João Gegaloto, Diogo e Vanessa, que estiveram sempre prontos para ajudar e proporcionar um grande apoio nos momentos importantes.

A todos, um muito obrigado por permitirem que este projeto final de mestrado se realizasse.

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, 27 de Dezembro 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, o valor por metro quadrado, entre outras, 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 utilizadores no processo de tomada de decisão.

Como resultado da implementação desta solução, os utilizadores 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 utilizadores na área do imobiliário, fornecendo previsões personalizáveis para o mercado português.

Palavras-chave: Aprendizagem Automática, Aplicação Web, Indicadores de Mercado, Inteligência Artificial, Mercado Imobiliário, Modelos de Previsão, Modelos de Regressão

Abstract

In this project, a solution is proposed for the problem of predicting real estate prices in the Portuguese market, which is volatile and influenced by multiple macroeconomic variables. This issue is relevant because property prices depend on factors such as interest rates, inflation, and price per square meter, among others, making investment decision-making a complex and risky task.

The proposed solution involves developing a web application that utilizes machine learning models like Linear Regression, Random Forest, Neural Networks, and LSTM to predict real estate prices in different municipalities of Portugal. The application allows for interactive visualization of data and forecasts, facilitating the analysis of market trends and supporting investors in the decision-making process.

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

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

Keywords: Machine Learning, Web Application, Market Indicators, Artificial Intelligence, Real Estate Market, Forecasting Models, Regression Models

Índice

Índice de Figuras	xv
Índice de Tabelas	xvii
Siglas	xix
1 Introdução	1
1.1 Enquadramento	1
1.2 Objetivos	3
1.3 Abordagem	4
1.4 Organização do Documento	6
2 Revisão da Literatura	7
2.1 Caracterização do Setor Imobiliário	7
2.2 Abordagens Existentes	9
2.2.1 Fatores Determinantes	9
2.2.2 Análise de Séries Temporais	12
2.2.3 Análise de Regressão	13
2.2.4 Modelo de Preços Hedónico	13
2.2.5 Algoritmos de Inteligência Artificial	14
2.2.6 Métodos de Previsão Econométricos	15
2.2.7 Métodos de Previsão da Análise Espacial	15
2.3 Avaliação dos Métodos	16
2.4 Sumário	17
3 Dados, Metodologias e Métricas	19
3.1 Fontes de Dados	20
3.1.1 Instituto Nacional de Estatística (INE)	20
3.1.2 Casafari	21
3.1.3 Doutor Finanças	22
3.1.4 Banco de Portugal	22
3.1.5 Qualidade dos Dados e Limitações	23
3.2 Processamento de Dados	23
3.2.1 Etapa de Pré-processamento	24
3.2.2 Etapa de Transformação	25
3.2.3 Etapa de Integração	25

3.3	Análises dos Dados	26
3.3.1	Exploração de Dados (EDA)	26
3.3.2	Correlação entre as Variáveis	28
3.3.3	Tendências encontradas nos Dados	29
3.4	Desenvolvimento de Modelos de Aprendizagem	30
3.4.1	Principais Métricas de Avaliação	31
3.4.2	Ajuste de Dados para o Treino dos Modelos	32
3.4.3	Regressão Linear	36
3.4.4	Random Forest	38
3.4.5	Redes Neurais Artificiais: MLP (Multi-Layer Perceptron)	41
3.4.6	Redes Neurais Recorrentes: LSTM (Long Short-Term Memory)	44
3.5	Análise Comparativa e Seleção do Modelo Preditivo	47
3.5.1	Comparação de Modelos Preditivos	47
3.5.2	Modelo Selecionado	48
4	Aplicação Web	51
4.1	Plataformas Similares	51
4.2	Descrição da Aplicação	53
4.3	Design da Aplicação	54
4.3.1	Esboços em Papel (Paper Sketch)	54
4.3.2	<i>Wireframes</i> e <i>Mockups</i>	55
4.3.3	Comparação com Princípios de Design	57
4.4	Implementação	58
4.4.1	Tecnologias	58
4.4.2	Arquitetura	58
4.4.3	Implantação da Aplicação	59
5	Avaliação Experimental	63
5.1	Análise e Comparação do Desempenho dos Modelos	63
5.2	Avaliação da Aplicação Web	64
5.2.1	Participantes	64
5.2.2	Resultados da Avaliação	65
5.2.3	Sugestões Futuras para a Aplicação	66
6	Discussão de Resultados	67
7	Conclusão	71
7.1	Trabalho Futuro	72
	Referências	73
	Anexos	
I	Anexo 1	77

Índice de Figuras

1.1	Evolução projetada do setor imobiliário pelo Statista (2024-2028) [5].	2
1.2	Etapas de Aquisição de Dados e Desenvolvimento de Modelos	4
1.3	Processo de Desenvolvimento da Aplicação	5
2.1	Caraterização do bem imobiliário [19].	8
2.2	Métodos de previsão e avaliação de bens imóveis [5].	17
3.1	Distribuição dos preços dos imóveis por região	27
3.2	Distribuições de preços médios de venda por concelho	27
3.3	A Matriz de Correlação: Mapa de Relações entre Indicadores Económicos e Financeiros	28
3.4	Relação Preço/Venda e Arrendamento por Região	29
3.5	Volatilidade dos Preços de Venda e arrendamento nas Regiões de Oeiras, Odiveelas, Amadora, Sintra, Loures e Cascais	30
3.6	Preço médio de venda após a normalização	33
3.7	Localização da divisão dos Dados	34
3.8	Divisão Temporal com TimeSeriesSplit (5 Folds)	35
3.9	Previsão do modelo de Regressão Linear	37
3.10	As 20 variáveis mais importantes do modelo Regressão Linear	38
3.11	Performance do modelo Random florest	40
3.12	As 20 variáveis mais importantes do modelo Random Forest	41
3.13	Performance do modelo Redes Neurais Artificiais (MLP)	43
3.14	Performance do modelo Redes Neurais Artificiais	44
3.15	Performance do modelo LSTM ao longo do treino	46
3.16	Performance do modelo LSTM	47
3.17	Previsão do Modelo LSTM: Comparação entre Valores Reais e Previstos	49
4.1	Zestimate - Zillow’s estimate of a home’s market value	52
4.2	Housing Market Data From Redfin	52
4.3	Esboço em papel da aplicação	55
4.4	<i>Wireframes</i> da aplicação	56
4.5	<i>Mockups</i> da aplicação	57
4.6	Arquitetura da aplicação	59
4.7	Primeira Página da aplicação implementada	60
4.8	Segunda Página da aplicação implementada	61
4.9	Terceira Página da aplicação implementada	62

5.1	Resultados do formulário: Distribuição dos géneros	65
5.2	Resultados do formulário: Distribuição das Idades	65
6.1	Principais KPIs considerados pelos investidores no mercado imobiliário [34] .	68
I.1	Resultados do Questionário sobre a Usabilidade	80

Índice de Tabelas

3.1	Resultados do Modelo de Regressão Linear	36
3.2	Resultados da Regressão Linear em Diferentes Divisões Temporais com Time-SeriesSplit	37
3.3	Desempenho do Modelo de Regressão Linear no Conjunto de Treino e Teste	38
3.4	Resultados do modelo Random Forest	39
3.5	Resultados do <i>TimeSeriesSplit</i> com Random Forest	40
3.6	Análise de sobre-aprendizagem do modelo Random Florest	41
3.7	Resumo da Arquitetura do Modelo	42
3.8	Resultados e análise da sobre-aprendizagem do modelo MLP	43
3.9	Resumo da Arquitetura do Modelo LSTM	45
3.10	Desempenho do modelo LSTM nos conjuntos de treino e teste	46
3.11	Comparação de Modelos: Resultados Gerais	48
3.12	Resultados de Desempenho do Modelo LSTM	49
5.1	Comparação de Modelos	63
I.1	Dados de Preços de Venda e Arrendamento	77
I.2	Endividamento dos Particulares	78
I.3	Desemprego e Indicadores Econômicos	79
I.4	Dados Temporais e Euribor	79

Siglas

R^2	Coeficiente de Determinação	46
API	Application Programming Interface	59
CI	Continuous Integration (Integração Contínua)	59
EDA	Análise exploratória de dados	26
INE	Instituto Nacional de Estatística	19
LSTM	Long Short-Term Memory	44
MAE	Mean Absolute Error	46
MLP	Multi-Layer Perceptron	41
MSE	Mean Squared Error	43
REST	Representational State Transfer	59
SPA	Single Page Application	59
UX	User Experience (Experiência do Utilizador)	54



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, entre outros., 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 [30], atingirá 637,80 triliões 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.

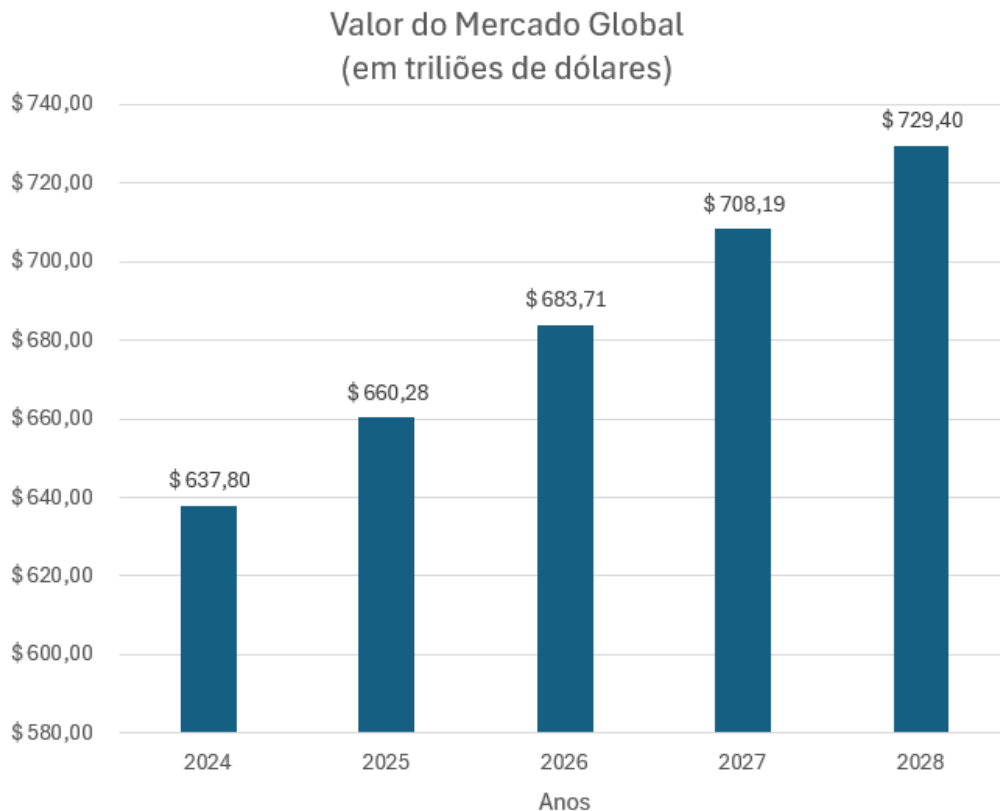


Figura 1.1: *Evolução projetada do setor imobiliário pelo Statista (2024-2028) [5].*

Conforme Statista [32], com um volume projetado de 518,90 trilhões 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 trilhões de dólares.

A China destaca-se como o principal contribuinte para o valor total do setor imobiliário no mundo [29]. Em 2024, atingiu 117,40 trilhões de dólares e, com uma previsão de crescimento anual (CAGR 2024–2028), atingirá um volume de mercado de 153,10 trilhões 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 Doutor Finanças [9] partir do segundo trimestre de 2022, as transações de propriedade diminuíram 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 perspectiva 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 IA 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 [11] [37].

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.

O desenvolvimento desta pesquisa, pretende das resposta a quatro questões fundamentais serão tratadas com o intuito de alcançar os objetivos delineados, como:

Questão 1: Quais foram as limitações enfrentadas ao integrar dados de diferentes fontes, como APIs e bases de dados externas?

Questão 2: Quais os melhores algoritmos para realizar previsão do mercado?

Questão 3: Qual é a melhor maneira de apresentar as análises e entendimentos aos utilizadores da aplicação de forma clara e intuitiva?

Questão 4: Quais os fatores que os investidores consideram quando pretendem investir?

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 indicações 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 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 tratamento de dados, remoção de valores atípicos, 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 Exploratory Data Analysis (EDA) será conduzida para identificar padrões iniciais, relações entre variáveis e preparar o conjunto de dados para uso nos modelos de IA.

A Figura 1.2 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.

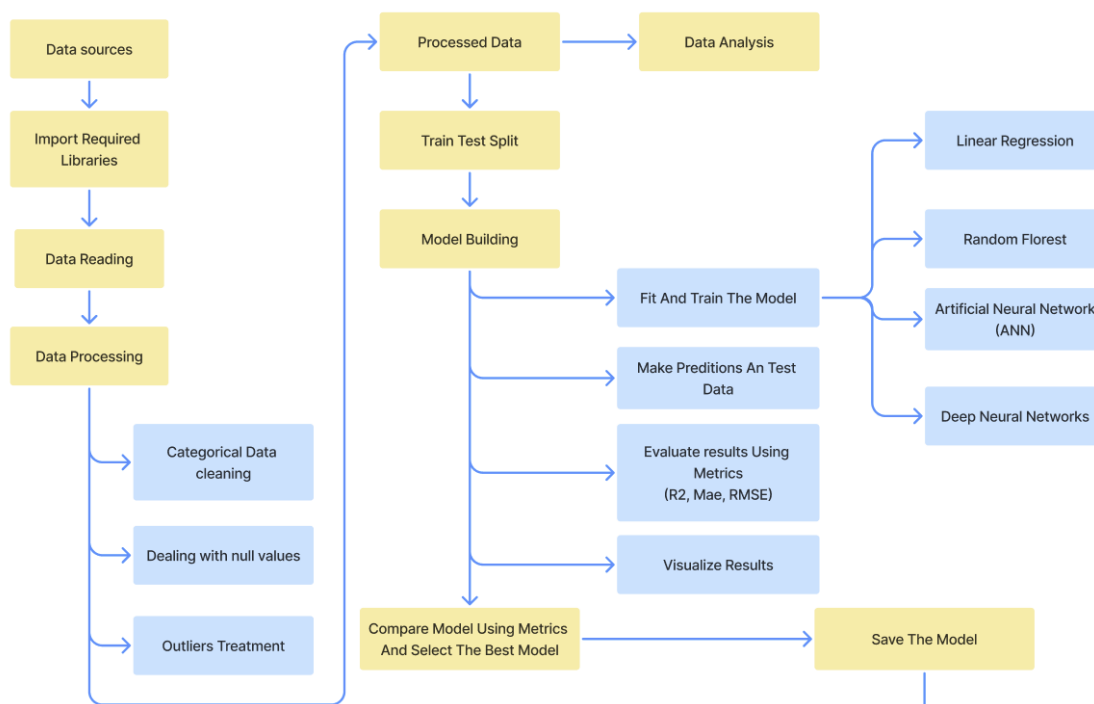


Figura 1.2: *Etapas de Aquisição de Dados e Desenvolvimento de Modelos*

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 aprendizagem automática, 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 (R^2).

Já a Figura 1.3 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:

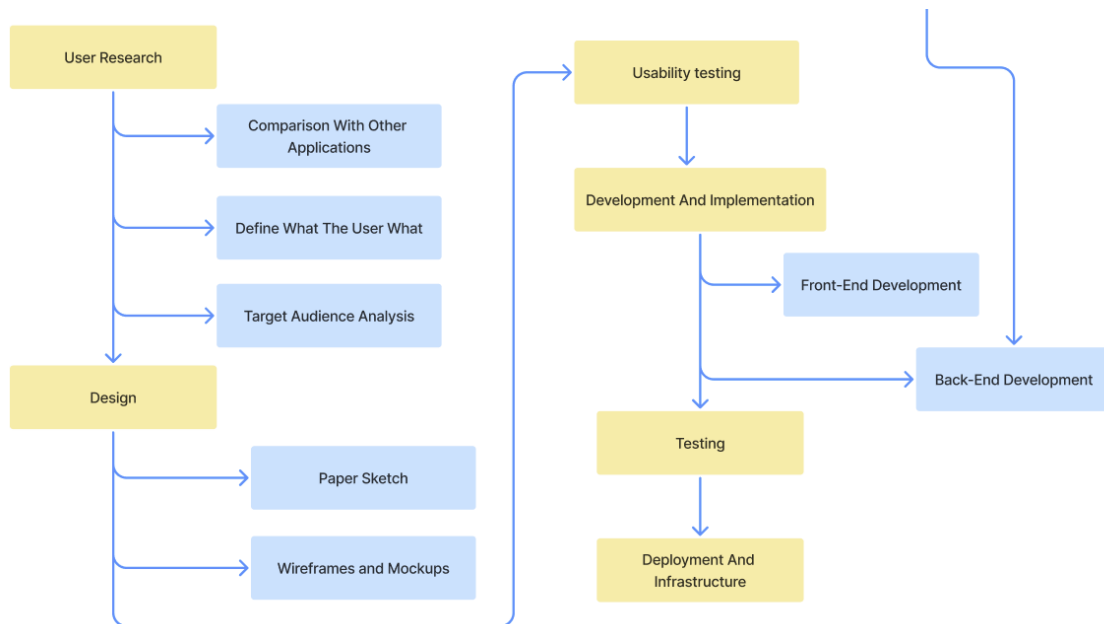


Figura 1.3: *Processo de Desenvolvimento da Aplicação*

Após o desenvolvimento dos modelos de IA, a aplicação web oferecerá uma interface interativa, com acesso às 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.

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

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.

O Capítulo 3 apresenta as fontes de dados utilizadas no projeto, explicando o processo de tratamento e preparação dessas informações para análise. Também são descritos os modelos de aprendizagem automática desenvolvidos, incluindo as etapas de desenvolvimento e avaliação. O capítulo termina com uma comparação dos resultados obtidos pelos diferentes modelos para perceber qual o melhor modelo para utilizar na aplicação web.

O Capítulo 4 descreve o desenvolvimento da aplicação web, desde a análise de plataformas semelhantes até o design e implementação, detalhando as tecnologias utilizadas. O capítulo também aborda a avaliação técnica, assegurando que a solução cumpra os requisitos funcionais e proporcione uma experiência otimizada ao utilizador.

O Capítulo 5 concentra-se na avaliação experimental do projeto, comparando o desempenho dos modelos de aprendizagem automática e da aplicação web. Essa comparação é baseada em testes específicos e no *feedback* dos utilizadores, permitindo identificar pontos fortes e áreas de melhoria.

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.

Para aceder ao código-fonte e à documentação adicional sobre a aplicação desenvolvida, consulte o repositório GitHub associado a este projeto, disponível em: <https://github.com/rubenAlbuquerque/AI-Powered-Real-Estate-Analytics-for-Investors>.



2

Revisão da Literatura

Neste capítulo, será realizada uma revisão da literatura sobre os trabalhos relacionados ao tema da pesquisa, abordando diferentes aspectos do mercado imobiliário, análise de dados e inteligência artificial. Começamos pela caracterização do setor imobiliário, explorando sua complexidade e os fatores macroeconômicos que o influenciam, conforme apresentado na Seção 2.1. Em seguida, examinaremos as abordagens existentes 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 e técnicas de Inteligência Artificial na Seção 2.2. Na Seção 2.3, avaliamos os métodos utilizados, destacando uma variedade de métricas, desde as convencionais, até abordagens mais sofisticadas, como análise de erro e validação cruzada. Finalmente, na Seção 2.4, apresentamos um sumário das principais conclusões, contextualizando o estudo e justificando a metodologia adotada.

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 econômicas 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 procura. Ao reconhecer a importância desses elementos e sua influência no mercado, os interessados podem avaliar melhor as oportunidades, mitigar riscos e promover um desenvolvimento sustentável no setor imobiliário [3].

A interação entre esses atributos é exemplificada na Figura 2.1, que ilustra a relação entre as diversas características do mercado imobiliário:

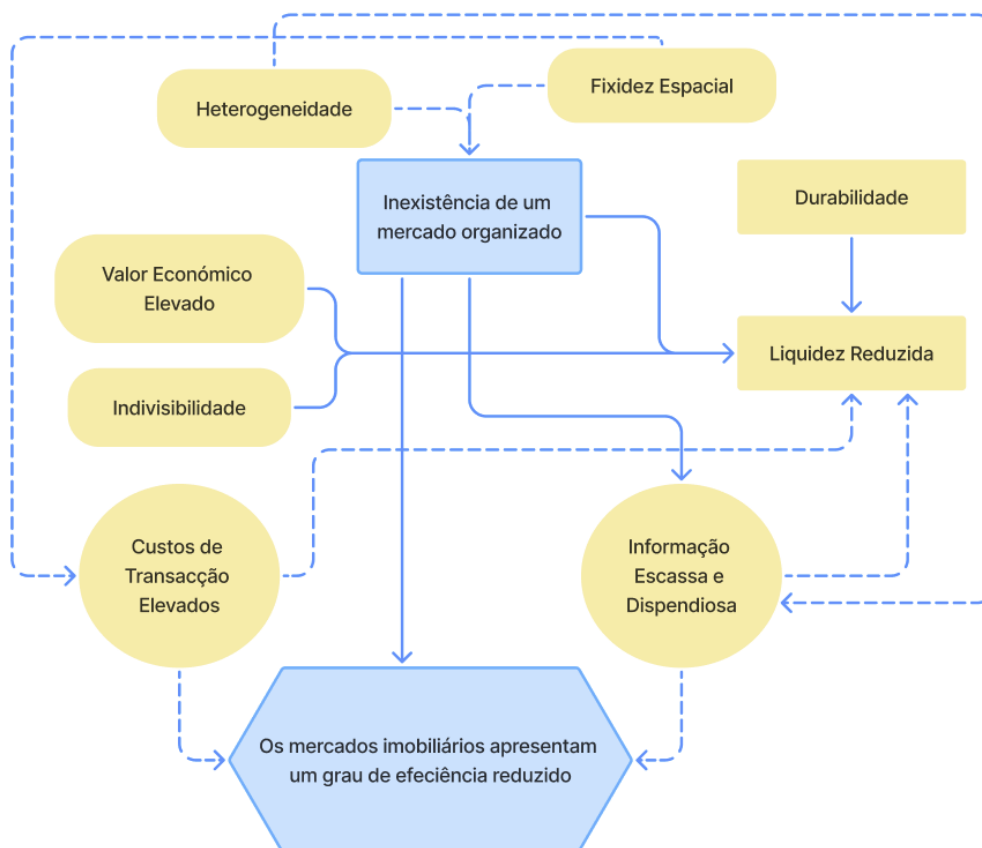


Figura 2.1: Caracterização do bem imobiliário [19].

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 oferta 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 Abordagens Existentes

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.

2.2.1 Fatores Determinantes

Segundo Pai e Wang (2020) [27], os fatores que afetam o mercado imobiliário podem ser categorizados como qualitativos ou quantitativos. Os fatores qualitativos referem-se 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. oferta 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 ausência 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 percepções úteis para profissionais do mercado imobiliário e contribui para um mercado mais transparente e eficaz.

O artigo de Pashardes et al. (2009) [28] 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 estudo de 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 et al. da Escola de Economia da Universidade de Finanças e Economia de Shanxi, China [23], 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 macroeconômicos 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 (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] 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, adquirindo 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) [25] sobre o impacto da pandemia de COVID-19 nos preços das casas em Alicante, Espanha, oferece uma compreensão das dinâmicas do mercado imobiliário perante 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, 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 efetuado por Sjöling (2012) [33] 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 econômica global, tal como a crise de 2008.

Além dos fatores já mencionados, como atributos dos imóveis, fatores macroeconômicos e eventos singulares como crises, outros estudos abordam a previsão e avaliação de imóveis que podem ser igualmente interessantes numa perspectiva mais abrangente. O artigo de Ernawati Kamal et al. (2016) [18] aborda a perspectiva 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 perspectiva dos agentes era o fator mais importante, especialmente quando combinada com infraestrutura adequada e design de alta qualidade, 2) fatores macroeconômicos, que no estudo se destaca a influência da taxa de inflação e das taxas 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 tais como 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-seções seguintes abordamos métodos como Análise de Séries Temporais, Análise de Regressão, Modelos de Preços Hedónico e Inteligência Artificial, destacando as 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 (Auto Regressive 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 usado 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 [21].

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 das suas relações com os preços imobiliários. De acordo com o artigo de Liu [20], 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 apresenta sensibilidade a valores atípicos, o que significa que os valores extremos, ou ausências nos dados, podem afetar significativamente os resultados do modelo [31].

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) [22] 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 hedônico. Estes algoritmos de aprendizagem automática são capazes de aprender artificialmente a partir dos dados subjacentes e melhorar continuamente a sua performance preditiva. No entanto, a principal crítica aos modelos de aprendizagem automática é a 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 Chan, [35], destacou limitações significativas do modelo hedônico 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 hedônico não foi capaz de produzir estimativas precisas de valoração. Essas descobertas destacam a importância de reconhecer as limitações do modelo hedônico 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 tem-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 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 quem 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 valor de mercado. Por outro lado, métodos mais antigos, como o modelo hedônico, são mais fáceis de entender e transparentes, pois permitem uma análise mais direta das variáveis, sendo que avaliam cada característica do imóvel de forma individual, o que proporciona uma visão clara e intuitiva de como cada característica contribui para o valor final do imóvel [22].

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 contribuindo para 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 modelo VAR foi a sua utilização na resposta do setor de construção a choques macroeconómicos, prever a 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.

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 percepções 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 [26].

A Previsão Geoespacial, que utiliza Sistemas de Informação Geográfica (SIG) [7], é um exemplo desses métodos. Os sistemas SIG permitem-nos compreender, organizar e analisar 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)[24] também usa 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.

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 autores procuraram categorizar esses erros.

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 raiz quadrada da 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) [22] 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 relevantes sobre sua aplicabilidade prática e potencial de generalização. Na Seção 3.5, apresentam-se em mais detalhe essas métricas e a importância de sua aplicação para a avaliação dos modelos.

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 macroeconômicas 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 2.2, destacando as diferenças entre os métodos de previsão e avaliação imobiliária.

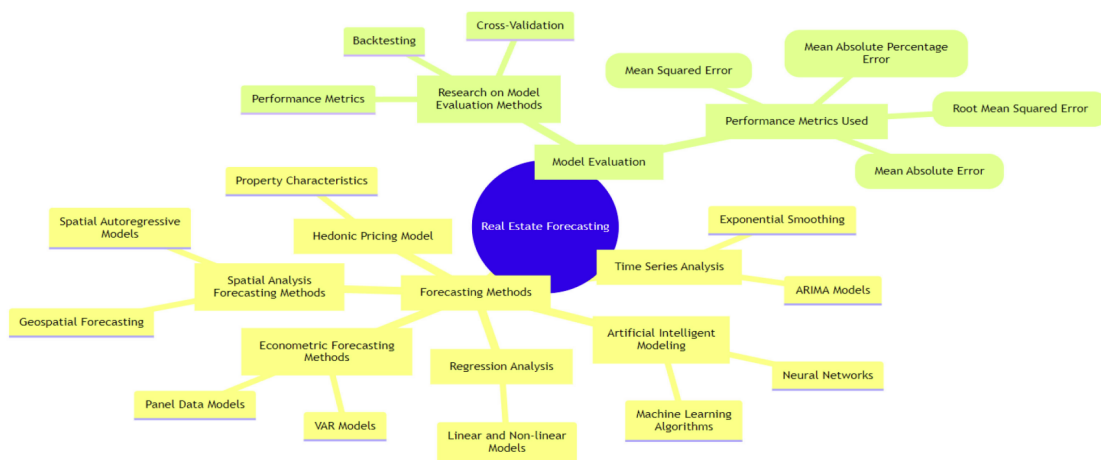


Figura 2.2: *Métodos de previsão e avaliação de bens imóveis [5].*

A Figura 2.2 apresentada organiza métodos e métricas de avaliação na previsão de preços imobiliários, centralizando o tema "*Real Estate Forecasting*" e destacando várias abordagens fundamentais. Dentre os métodos de previsão, incluem-se a Análise de Séries Temporais (com ARIMA e Suavização Exponencial) e a Modelagem com Inteligência Artificial (com algoritmos de aprendizado de máquina e redes neurais), que capturam padrões complexos. A Análise Espacial (incluindo modelos auto regressivos espaciais e previsão geoespacial) e os Métodos Econométricos (como modelos de dados em painel e VAR) abordam aspectos de localização e relações econômicas. O Modelo Hedônico de Preço utiliza características de propriedades para determinar valores específicos. A avaliação dos modelos é feita com métricas como MSE, RMSE, MAE e MAPE, medindo a precisão, eficiência e complexidade das previsões.

Técnicas como backtesting e validação cruzada são usadas para garantir a robustez dos modelos. Essa estrutura combinada oferece uma abordagem ampla e multidimensional, essencial para capturar a complexidade e a variabilidade do mercado imobiliário, resultando em previsões mais precisas e confiáveis.

3 Dados, Metodologias e Métricas

Neste capítulo, é apresentada a metodologia adotada para a análise do mercado imobiliário, bem como os dados utilizados e os resultados obtidos. O estudo integra 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.

Primeiramente na Secção 3.1, abordam-se as fontes de dados utilizadas neste estudo. Diversas instituições, como o Instituto Nacional de Estatística (INE), Casafari, Doutor Finanças e o Banco de Portugal, foram consultadas para reunir uma combinação abrangente de informações sobre o mercado imobiliário e indicadores macroeconómicos. A integração dessas fontes garante uma análise completa das dinâmicas do setor.

A Secção 3.2 aborda o processamento de dados, discutindo as etapas de pré-processamento para tratamento de valores omissos e remoção de duplicados. Também são exploradas as transformações necessárias para gerar variáveis mais informativas, além da integração dos dados provenientes de diferentes fontes, resultando em um conjunto único e coerente, pronto para as análises e modelagens.

Na Secção 3.3, são realizadas as análises exploratórias de dados (EDA) com o intuito de identificar padrões e características importantes. Também é explorada a correlação entre variáveis, ajudando a entender as inter-relações entre os diferentes fatores que influenciam o mercado imobiliário. Por fim, são apresentados *insights* iniciais com base nessas análises, que guiam as etapas subsequentes.

O desenvolvimento dos modelos de aprendizagem é discutido na Secção 3.4, onde são aplicados diversos algoritmos preditivos, como regressão linear, Random Forest e redes neurais. Além disso, são descritas as métricas de avaliação utilizadas para medir o desempenho de cada modelo.

Por fim, a Secção 3.5 apresenta a comparação dos resultados dos modelos, destacando os principais desempenhos e identificando o melhor modelo com base nas métricas de avaliação, oferecendo uma análise criteriosa sobre o modelo mais adequado para prever os preços no mercado imobiliário.

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 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. Os dados 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 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 programa 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 arrendamento. Para este estudo, a plataforma foi utilizada para extrair dados relevantes ao mercado imobiliário nos concelhos da área da grande Lisboa, nomeadamente 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 ficheiros *Comma Separated Values* (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, designadamente:

- **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 inter-regionais.
- **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** - 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 Doutor Finanças

[Doutor Finanças](#) 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, conseqüentemente, 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, com os seguintes significados:

- **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 na concessão de 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 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. Esses 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, indica 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 procura 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 pela qualidade e precisão, alguns desafios foram encontrados durante a aquisição e processamento dos dados. Nalguns casos, como na Casafari, a granularidade excessiva dos dados resultou em dificuldades para padronizar informações entre diferentes períodos. Além disso, algumas fontes apresentaram valores omissos 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 adquiridos, 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 omissos até à criação de novas variáveis, essenciais para enriquecer as análises.

3.2.1 Etapa de Pré-processamento

Os dados utilizados neste estudo foram adquiridos a partir de várias fontes, com diferentes arquivos, cada um com o seu próprio formato e características. Para permitir a integração dessas informações, foi necessário aplicar 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. 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, na qual havia ocorrências de múltiplas linhas com a mesma informação. Os dados duplicados foram removidos para garantir unicidade e que as análises não fossem distorcidas por entradas repetidas.

Outro aspeto importante no pré-processamento foi o tratamento de valores omissos. 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 intermédios com base numa função linear. Este 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 reter apenas informações referentes ao período de 2018 a 2024 fossem consideradas. Isso foi feito para focar a análise num intervalo de tempo mais recente e relevante, conforme o objetivo do estudo.

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 Etapa de Transformação

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 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 analisar a evolução do mercado imobiliário, foram criadas 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 arrendamento, 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 perspectiva 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 Etapa de Integração

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 juro 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 juro 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 arrendamento 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 (I.1), Endividamento dos Particulares (I.2), Desemprego e Indicadores Económicos (I.3), e Dados Temporais e Euribor (I.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 omissos, uma vez que o processo de limpeza de dados garantiu que o conjunto estivesse completo e pronto para análise.

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

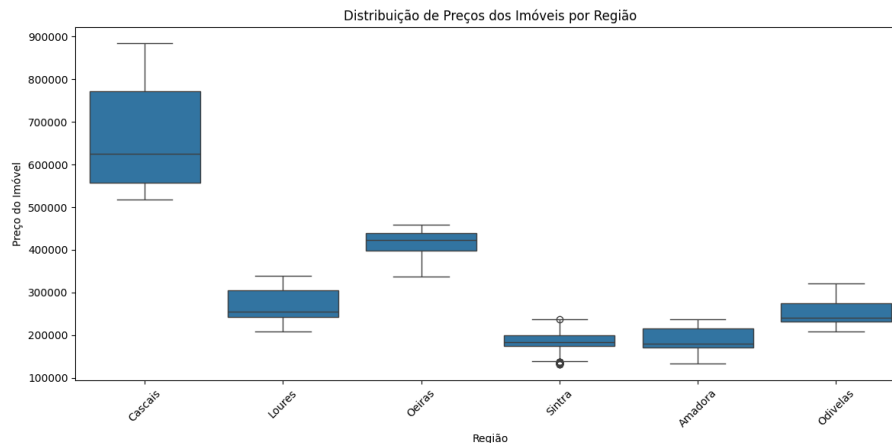


Figura 3.1: *Distribuição dos preços dos imóveis por região*

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 3.2. Esses gráficos permitem uma análise mais detalhada da concentração de preços em cada região.

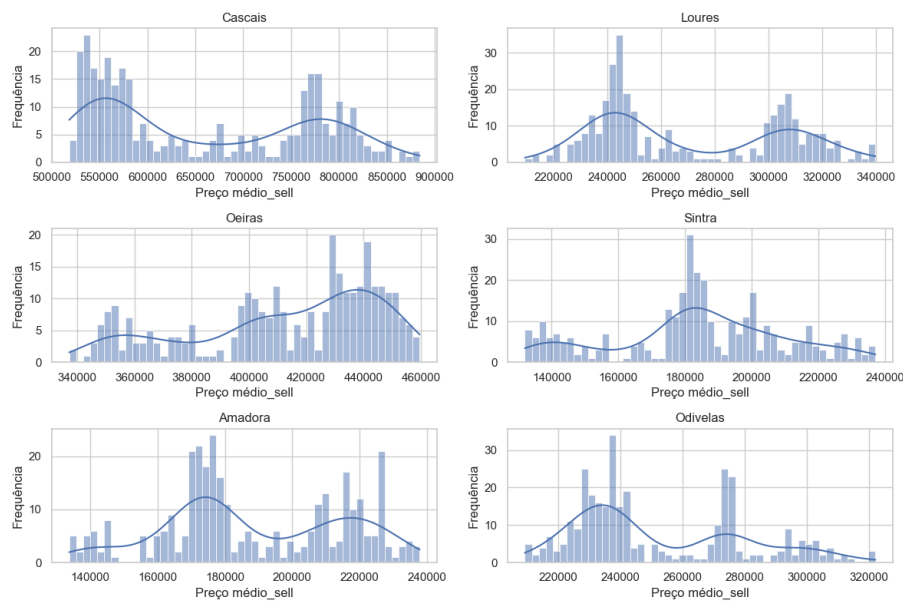


Figura 3.2: *Distribuições de preços médios de venda por concelho*

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€. Esta 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 3.3 foi gerado para identificar as relações entre diferentes variáveis, destacando as conexões entre indicadores económicos e o mercado imobiliário.

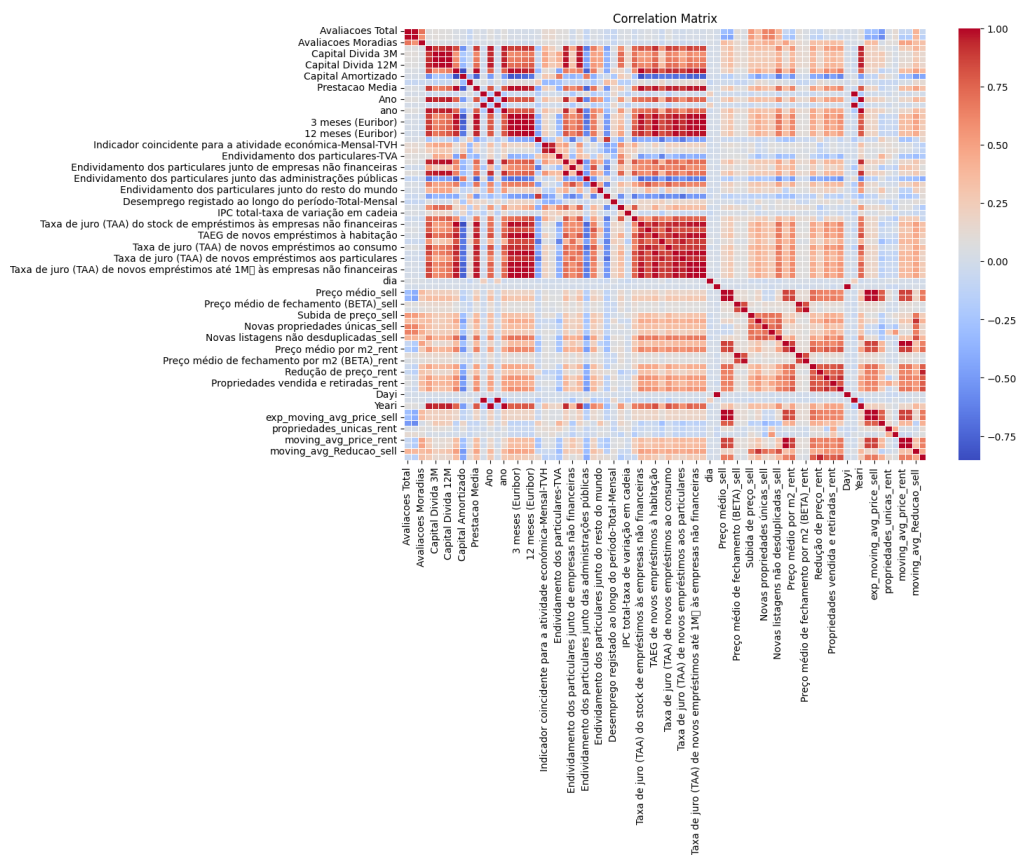


Figura 3.3: A Matriz de Correlação: Mapa de Relações entre Indicadores Económicos e Financeiros

A matriz de correlação mostra, uma forte correlação entre a TAEAG e os preços de venda de imóveis, indicando que as taxas de juro afetam diretamente o comportamento dos preços. Além disso, a análise com desfasagem temporal (lags) revelou que o impacto das taxas de juro 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 Tendências encontradas nos 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 3.4), evidenciando uma forte valorização imobiliária na região. Esse comportamento reflete o alto interesse do mercado e a procura crescente por imóveis nesta área, consolidando-a como uma das mais caras do país.

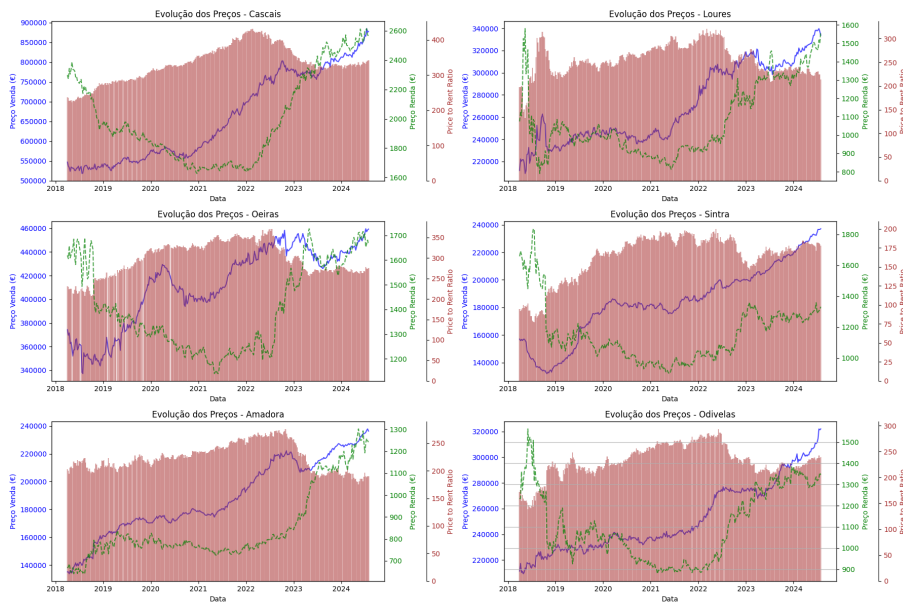


Figura 3.4: *Relação Preço/Venda e Arrendamento por Região*

O *price-to-rent ratio*, conforme mostrado na Figura 3.4, forneceu informações valiosas sobre a atratividade relativa entre comprar ou Arrendar imóveis. 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.

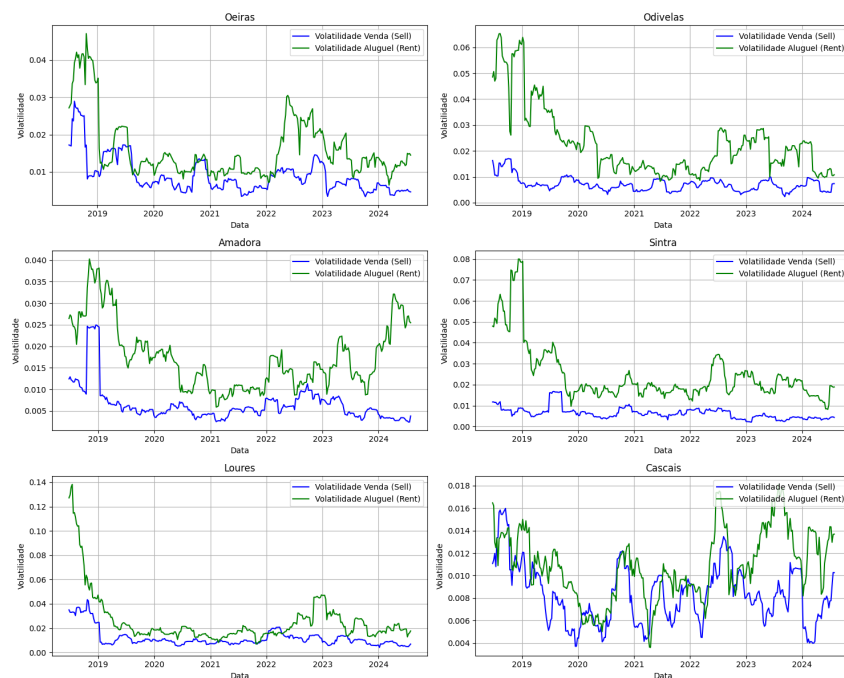


Figura 3.5: *Volatilidade dos Preços de Venda e arrendamento nas Regiões de Oeiras, Odivelas, Amadora, Sintra, Loures e Cascais*

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 procura.

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 procura 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 no concelho de Odivelas. Foram aplicadas tanto abordagens simples, como a Regressão Linear, quanto técnicas mais avançadas, como o *Random Forest* e as Redes neuronais. 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 detetar as nuances dos dados. Os resultados dos modelos apresentados nesta

secção são exclusivamente referentes ao concelho de Odivelas.

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 o modelo que oferece os melhores resultados em termos de capacidade de detetar as variações dos dados reais. Cada uma delas oferece uma perspectiva diferente sobre o comportamento dos modelos, desde a medição do erro médio até à capacidade de explicar a variabilidade nos preços dos imóveis. As principais métricas utilizadas são:

- **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. A fórmula do MAE é dada por:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

onde:

- n é o número total de observações,
 - y_i é o valor real da observação i ,
 - \hat{y}_i é o valor previsto da observação i ,
 - $|y_i - \hat{y}_i|$ é o erro absoluto para a observação i .
- **Erro Quadrático Médio (MSE)** - Penaliza os erros grandes com mais intensidade, uma vez que os erros são elevados ao quadrado antes de serem somados. É adequado quando se deseja reduzir previsões com grandes desvios. A fórmula do MSE é dada por:

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

onde:

- n é o número total de observações,
 - y_i é o valor real da observação i ,
 - \hat{y}_i é o valor previsto da observação i ,
 - $(y_i - \hat{y}_i)^2$ é o erro quadrático para a observação i .
- **Coefficiente de Determinação (R²)** - Mede a proporção de variação explicada pelas variáveis independentes. Varia de $-\infty$ a 1, sendo que valores próximos de 1 indicam um ajuste mais preciso aos dados reais. Valores negativos indicam que o

modelo tem pior desempenho do que uma simples média dos dados. A fórmula do R^2 é dada por:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

onde:

- y_i é o valor real da observação i ,
 - \hat{y}_i é o valor previsto da observação i ,
 - \bar{y} é a média dos valores reais y_i ,
 - $\sum_{i=1}^n (y_i - \hat{y}_i)^2$ é a soma dos erros quadráticos,
 - $\sum_{i=1}^n (y_i - \bar{y})^2$ é a soma da variância total dos valores reais.
- **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. Neste caso, a acurácia pode ser calculada como:

$$\text{Acurácia} = \left(1 - \frac{\text{MAE}}{\text{Média}(y_{\text{real}})}\right) \times 100$$

onde:

- MAE é o Erro Absoluto Médio,
- $\text{Média}(y_{\text{real}})$ é a média dos valores reais.

3.4.2 Ajuste de Dados para o 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, descrevem-se 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 modelo 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 K-nearest neighbors (KNN).

Frequentemente, variáveis com valores significativamente maiores, como o preço médio de venda dos imóveis, podem dominar outras com valores menores, como taxas de juro ou áreas dos imóveis. Isso pode comprometer o desempenho do modelo, dificultando a aprendizagem das relações entre as variáveis, 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, também ajusta os dados com base na 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 3.6 apresenta o gráfico do preço médio de venda dos imóveis por localização, após a aplicação da normalização com o *StandardScaler*.

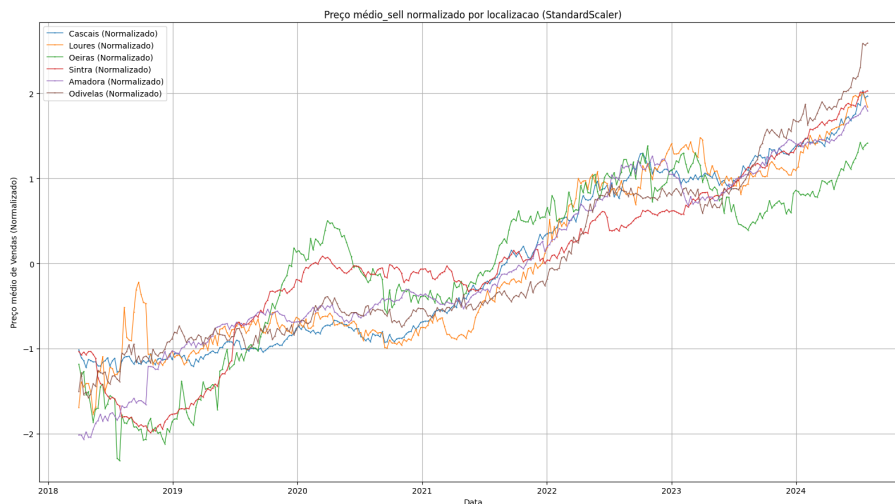


Figura 3.6: Preço médio de venda após a normalização

3.4.2.2 Estrutura da Divisão dos Dados

Antes do treino dos modelos foi necessário fazer uma divisão adequada do conjunto de dados para 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, este poderia ajustar-se 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 seguir 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 é 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. Tal 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 3.7 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.

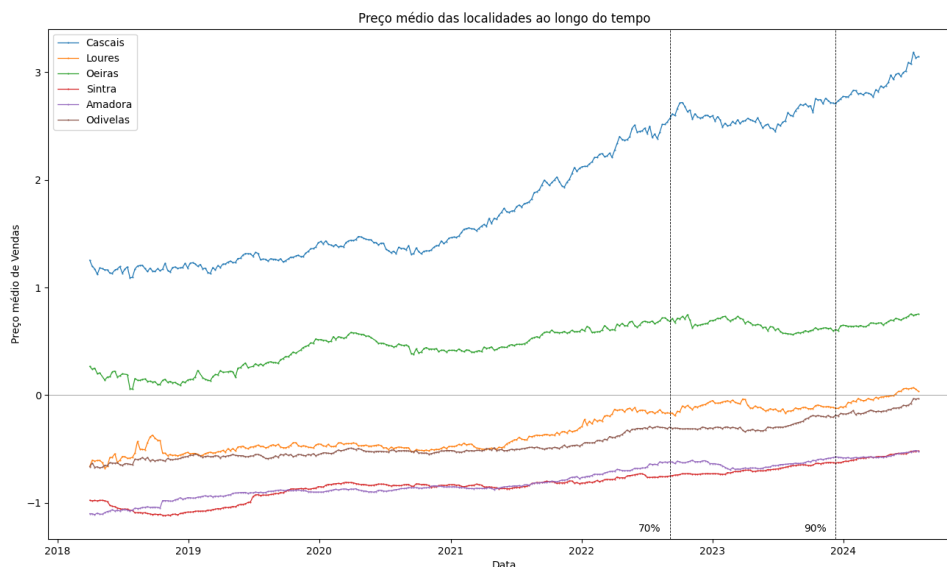


Figura 3.7: Localização da divisão dos Dados

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, respeitando a ordem cronológica dos mesmos.

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.

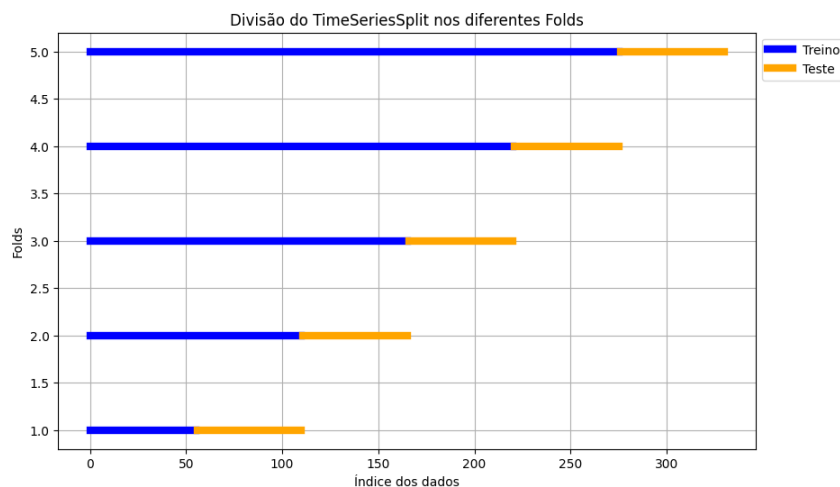


Figura 3.8: *Divisão Temporal com TimeSeriesSplit (5 Folds)*

A Figura 3.8 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 *TimeSeriesSplit* ajuda 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, a 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 que resultou nos parâmetros `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 apresentam-se na Tabela 3.1.

Tabela 3.1: Resultados do Modelo de Regressão Linear

Métrica	Valor
MAE Linear	22948
R ² Linear	-1.96
Acurácia	97.07%

Esses resultados indicaram desempenho insatisfatório, com um valor de R² negativo, sugerindo que o modelo não foi capaz de explicar a variação nos dados de teste.

Os resultados para as diferentes divisões temporais com TimeseriesSplit estão apresentados na Tabela 3.2.

Tabela 3.2: Resultados da Regressão Linear em Diferentes Divisões Temporais com Time-SeriesSplit

Dim. Treino	Dim. Teste	MAE Linear	R ² Linear	Acurácia (%)
31	30	14424.92	-0.48	97.43
61	30	273097307.34	-245352047.57	-46860.98
91	30	32463.29	0.19	95.30
121	30	75334.57	-52.13	90.25
151	30	8309.28	0.85	98.98

Os resultados variaram significativamente, com valores de R² negativos em algumas divisões, indicando que o modelo estava ajustando pior do que uma previsão média simples. Contudo, à medida que a quantidade de dados de treino aumentou, o modelo apresentou melhorias, especialmente nas últimas divisões, onde o R² se tornou positivo e o MAE foi reduzido, demonstrando uma maior precisão.

Essa tendência pode ser observada na Figura 3.9, que apresenta a previsão do modelo de Regressão Linear.

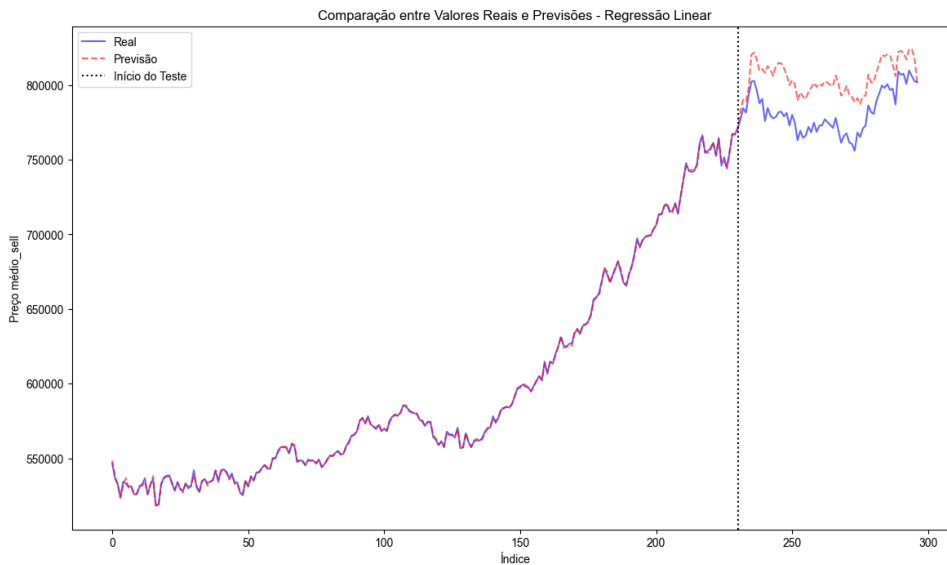


Figura 3.9: *Previsão do modelo de Regressão Linear*

Os resultados indicam que o modelo de Regressão Linear conseguiu detetar 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 ocorrência de sobre-aprendizagem como podemos ver na Tabela 3.3.

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

Métrica	Valor
MAE (Treino)	894
R ² (Treino)	1.00
MAE (Teste)	8309
R ² (Teste)	0.85

O Figura 3.10 mostra as 20 variáveis mais influentes no modelo de Regressão Linear, de acordo com seus coeficientes.

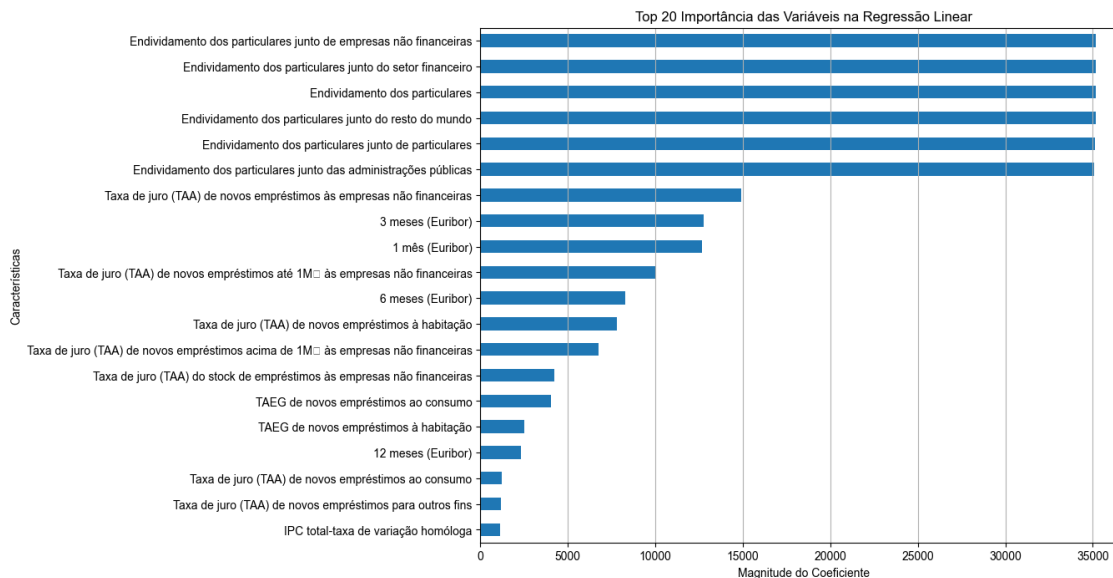


Figura 3.10: As 20 variáveis mais importantes do modelo Regressão Linear

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, resultando num 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 sua capacidade de lidar com interações complexas entre as variáveis imobiliárias e macroeconómicas. Além disso, o modelo possui uma alta resistência a sobre-aprendizagem, característica especialmente útil ao trabalhar com dados temporais, onde as relações entre variáveis podem mudar ao longo do tempo.

A flexibilidade desse modelo permite detetar padrões não lineares que outros modelos, como a Regressão Linear, podem 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, com valores entre 400 e 800.
- **max_features** - número máximo de variáveis consideradas ao dividir cada nó (teste realizado com os parâmetros None, sqrt e log2, 0.5, 0.75).
- **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 (3, 5, 10, 15 ou 20)
- **min_samples_leaf** - número mínimo de amostras numa folha (1, 2, 3, 4, e 5).

Após testar várias combinações de hiperparâmetros, o melhor modelo encontrado foi o RandomForestRegressor com os seguintes parâmetros: max_depth=None, max_features=0.75, min_samples_leaf=2, min_samples_split=5, n_estimators=400 e random_state=42. Essa configuração mostrou-se eficaz para modelar os dados, proporcionando um bom equilíbrio entre a complexidade do modelo e a capacidade de 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 Tabela 3.4 apresenta as métricas de avaliação obtidas.

Tabela 3.4: Resultados do modelo Random Forest

Métrica	Valor
MAE Linear	5974
R ² Linear	-0.49
Acurácia	97.85%

Esses resultados indicam que o modelo foi capaz de detetar padrões complexos, embora tenha apresentado um valor negativo de R² em algumas divisões, o que pode sugerir um sub-ajuste em determinadas partes dos dados.

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.

Na Tabela 3.5 apresenta os resultados das três divisões temporais.

Tabela 3.5: Resultados do *TimeSeriesSplit* com Random Forest

Dim. Treino	Dim. Teste	MAE Linear	R ² Linear	Acurácia (%)
56	55	6151.55	-1.76	97.35
111	55	1296.25	0.27	99.45
166	55	9990.18	-0.79	96.01
221	55	4825.35	-5.84	98.24
276	55	22749.44	-4.60	92.36

Esses resultados indicam que o modelo Random Forest captou parte da variação nos preços médios de venda dos imóveis, mas ainda apresentou limitações significativas. Com um erro absoluto médio (MAE) de 22.749,44 na última divisão de treino, um R² negativo de -4,60 e uma acurácia de 92,36%, o modelo teve dificuldades em ajustar corretamente as nuances dos dados de teste. Embora o modelo mantenha uma precisão relativamente alta em algumas divisões, os valores de R² negativos sugerem que ele não consegue generalizar adequadamente, especialmente em divisões onde a variação nos dados é mais complexa.

Para ilustrar o desempenho do modelo, foi gerado um gráfico (Figura 3.11) 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 vertical pontilhada preta indica o ponto onde o conjunto de treino termina e o conjunto de teste começa.

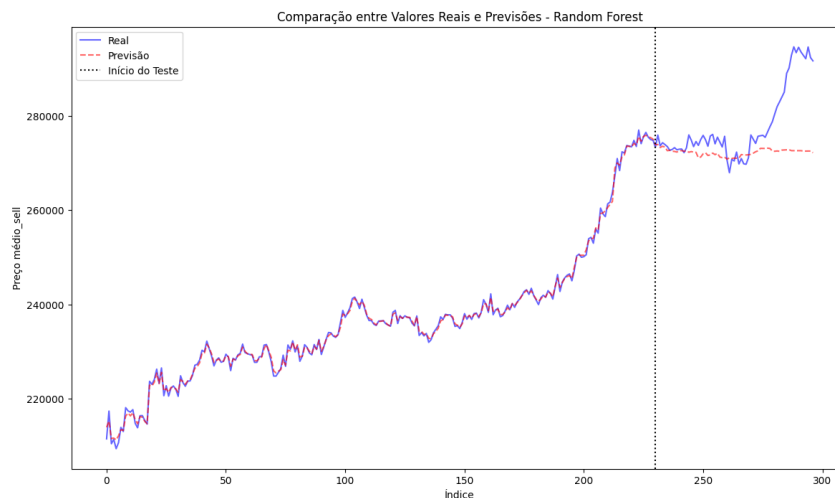


Figura 3.11: *Performance do modelo Random florest*

O gráfico da Figura 3.11 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 mantêm-se bastante próximas dos valores reais, validando a eficácia do Random Forest na previsão dos preços do mercado imobiliário.

Os resultados apresentados na Tabela 3.6 ilustram a avaliação do modelo Random Forest em relação ao possível sobre-aprendizagem. Nessa tabela, observa-se o desempenho do modelo tanto no conjunto de treino quanto no conjunto de teste, permitindo uma análise comparativa da sua capacidade de generalização.

Tabela 3.6: Análise de sobre-aprendizagem do modelo Random Florest

Métrica	Valor
MAE (Treino)	429
R^2 (Treino)	0.99
Accurary (Treino)	99.81%
MAE (Teste)	5974
R^2 (Teste)	-0.48
Accurary (Teste)	97.85%

Esses resultados sugerem que o modelo Random Forest se ajustou muito bem aos dados de treino, com um MAE baixo de 429, um R^2 de 0,99 e uma acurácia de 99,81%. No entanto, o desempenho no conjunto de teste foi consideravelmente inferior, com um MAE elevado de 5.974, um R^2 negativo de -0,48 e uma acurácia de 97,85%. Esses resultados indicam uma possível sobre-aprendizagem, sugerindo que o modelo captou excessivamente os padrões específicos do conjunto de treino, mas teve dificuldades em generalizar para novos dados.

O gráfico da Figura 3.12 mostra as 20 variáveis mais influentes no modelo Random Forest, de acordo com suas importâncias. Essas variáveis influenciaram significativamente o desempenho do modelo, sendo fatores-chave para a previsão de preços imobiliários.

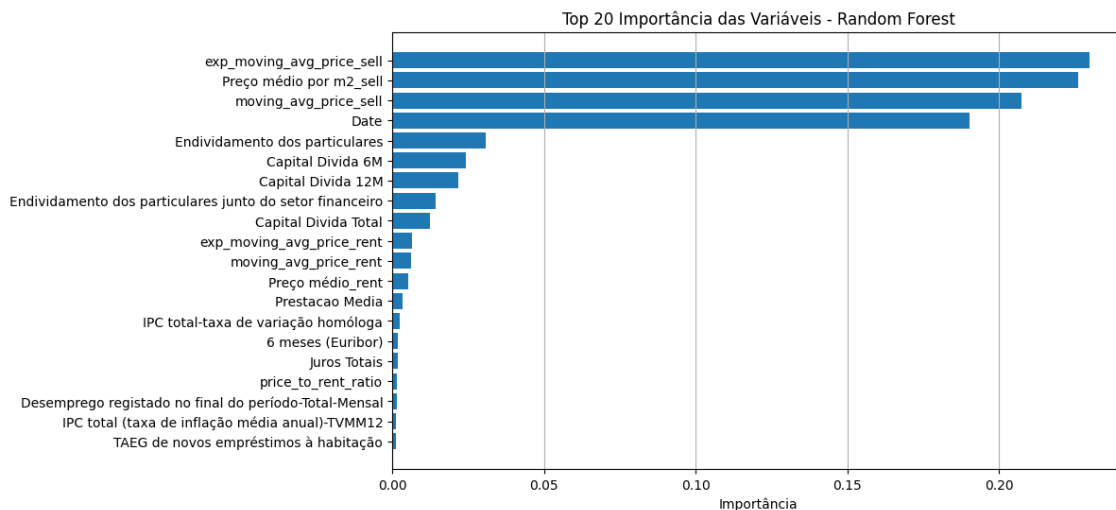


Figura 3.12: As 20 variáveis mais importantes do modelo Random Forest

3.4.5 Redes Neurais Artificiais: MLP (Multi-Layer Perceptron)

As Redes Neurais Artificiais, mais especificamente a **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 detetar 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 (preço médio de venda dos imóveis).

O modelo foi treinado 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 3.7 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.

Tabela 3.7: Resumo da Arquitetura do Modelo

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

Treino com 70% 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 do conjunto de teste e restaurando os melhores pesos após 150 épocas sem melhoria significativa. Os gráficos resultantes do treino e validação do modelo estão ilustrados na Figura 3.13.

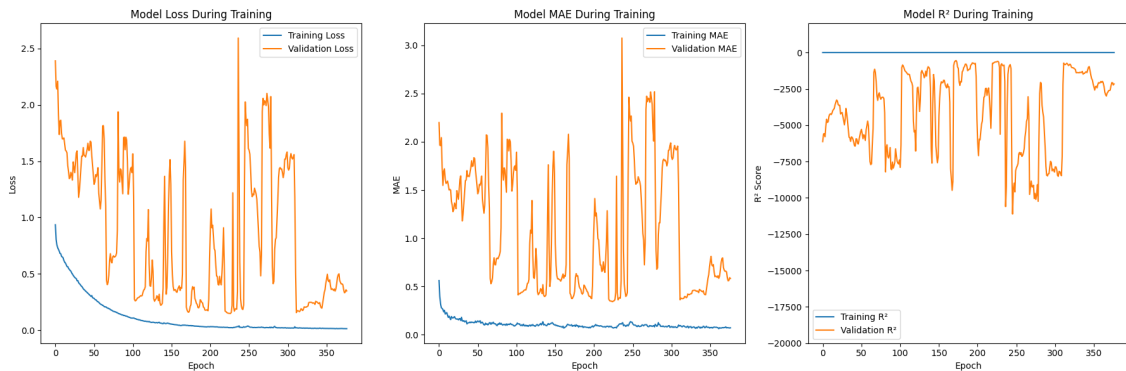


Figura 3.13: Performance do modelo Redes Neurais Artificiais (MLP)

Esses gráficos indicam que a perda e o Erro Absoluto Médio (MAE) do conjunto de treino diminuíram de forma consistente ao longo das épocas, sugerindo que o modelo aprendeu a capturar os padrões dos dados de treino. No entanto, a perda e o MAE no conjunto de validação mostram uma grande variação e permanecem significativamente mais altos e instáveis, o que é um forte indicativo de sobreajuste. O gráfico do R^2 também evidencia essa diferença: enquanto o R^2 do treino se mantém próximo de 1 (indicando um ajuste quase perfeito), o R^2 do conjunto de validação apresenta valores muito baixos e altamente variáveis, inclusive com valores negativos em várias épocas. Esses resultados sugerem que o modelo está ajustado demais aos dados de treino, mas não consegue generalizar bem para o conjunto de validação.

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 3.8.

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

Métrica	Valor
MAE (Treino)	758
R^2 (Treino)	1.00
MSE (Treino)	969 851
MAE (Teste)	5 072
R^2 (Teste)	0.06
MSE (Teste)	54 771 786

Os valores de MAE, MSE e R^2 foram calculados para avaliar o desempenho do modelo no conjunto de teste. Com um MAE de 5.072, um MSE de 54.771.786 e um R^2 de 0,06, observa-se que o modelo apresentou um desempenho razoável, mas ainda teve dificuldades em captar toda a variabilidade dos dados de teste. Embora o R^2 positivo indique que o modelo conseguiu prever algumas tendências dos dados, o valor baixo sugere uma capacidade de generalização limitada.

Em comparação com o ajuste aos dados de treino (MAE de 758, MSE de 969.851 e R^2 de 1,00), esses resultados apontam para uma possível sobre-aprendizagem, onde o modelo ajustou-se muito bem aos dados de treino, mas teve dificuldade em generalizar para novos dados.

Demonstração dos Dados Previstos em Gráfico de Série Temporal

Na Figura 3.14 temos 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.

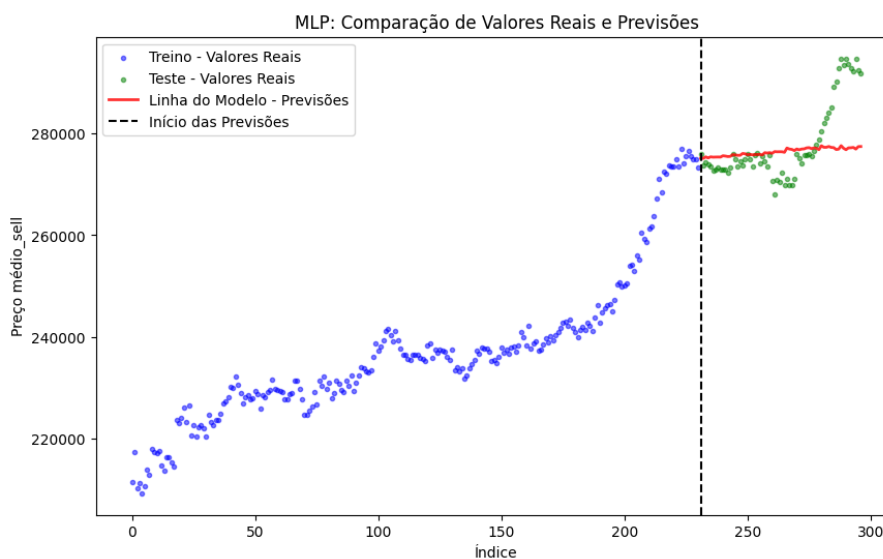


Figura 3.14: *Performance do modelo Redes Neurais Artificiais*

Apesar de uma pequena divergência ao longo do conjunto de teste, as previsões do modelo MLP acompanham razoavelmente bem os valores reais, especialmente no conjunto de treino. A linha preta tracejada indica o ponto de início do conjunto de teste, onde é visível uma perda de precisão nas previsões. No entanto, o modelo consegue capturar a tendência geral dos preços médios de venda dos imóveis, embora falhe em acompanhar as oscilações mais acentuadas no conjunto de teste. Esses resultados sugerem uma boa capacidade do modelo em aprender o padrão dos dados de treino, mas uma limitação em generalizar para novos dados.

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

O modelo *Long Short-Term Memory* (LSTM) foi desenvolvido para detectar 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 detectar padrões e tendências no mercado imobiliário.

A otimização dos hiper-parâmetros 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 sub-ajuste.

Arquitetura do Modelo LSTM

Após diversas iterações de ajustes manuais, a arquitetura final do modelo LSTM foi selecionada com base na performance otimizada para a previsão de preços de imóveis. A configuração final consistiu nas seguintes camadas:

- **Primeira Camada LSTM** - Com 1280 unidades e ativação *tanh*, configurada para retornar sequências, permitindo que a próxima camada processe as informações de maneira eficiente. Essa camada utiliza regularização L2 com peso de 0.01 para reduzir o risco de sobreajuste. Além disso, foram aplicadas normalização por lote e um *Dropout* de 0,2 para regularizar o modelo.
- **Segunda Camada LSTM** - Esta camada possui 128 unidades, ativação *tanh*, e também usa regularização L2 (peso de 0.004), mas não retorna sequências. Assim, fornece uma saída condensada que é encaminhada para as camadas densas subsequentes. Normalização por lote e *Dropout* (0,2) também são aplicados nesta camada.
- **Camada Densa** - Com 32 unidades e ativação *tanh*, utilizada para combinar as saídas da camada anterior e prepará-las para a camada de saída.
- **Camada de Saída** - Um único neurónio sem função de ativação, responsável pela previsão do valor contínuo dos preço médio de venda dos imóveis.

A Tabela 3.9 contem o resumo da arquitetura do modelo.

Tabela 3.9: Resumo da Arquitetura do Modelo LSTM

Layer (type)	Output Shape	Param #
lstm_3 (LSTM)	(None, 1, 1280)	6,952,960
batch_normalization_3 (BatchNormalization)	(None, 1, 1280)	5,120
dropout_4 (Dropout)	(None, 1, 1280)	0
lstm_4 (LSTM)	(None, 128)	721,408
batch_normalization_4 (BatchNormalization)	(None, 128)	512
dropout_5 (Dropout)	(None, 128)	0
dense_5 (Dense)	(None, 32)	4,128
dense_6 (Dense)	(None, 1)	33

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

Treino e Avaliação

O treino foi configurado com um tamanho de lote (`batch_size`) de 32 e um total de 650 épocas. Para otimização, foi utilizado o otimizador Adam com uma taxa de aprendizagem inicial de 0,0008, ajustada por um decaimento exponencial. O decaimento da taxa de aprendizagem ocorre a cada 10 épocas, reduzindo a taxa em 4,4% (decay rate de 0,956) para estabilizar o processo de treino. Uma função de perda Huber foi escolhida para lidar melhor com *outliers*, e o treino foi monitorado com paciência de 150 épocas para parada antecipada (early stopping), restaurando os melhores pesos ao final. As principais métricas utilizadas para avaliar o desempenho do modelo foram o **Mean Absolute Error (MAE)** e o **Coefficiente de Determinação (R^2)**.

Após o treino, o desempenho do modelo foi avaliado nos conjuntos de treino e teste. A Tabela 3.10 resume os resultados obtidos.

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

Métrica	Valor
MSE (Treino)	3 250 438
MAE (Treino)	1 413
R^2 (Treino)	0.98
MSE (Teste)	49 516 827
MAE (Teste)	4 550
R^2 (Teste)	0.15

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

Para compreender melhor como o modelo está a aprender, podemos visualizar, na Figura 3.15 a evolução da perda (*Loss*), do MAE e do R^2 durante o treino, tanto para o conjunto de treino quanto para o conjunto de teste.

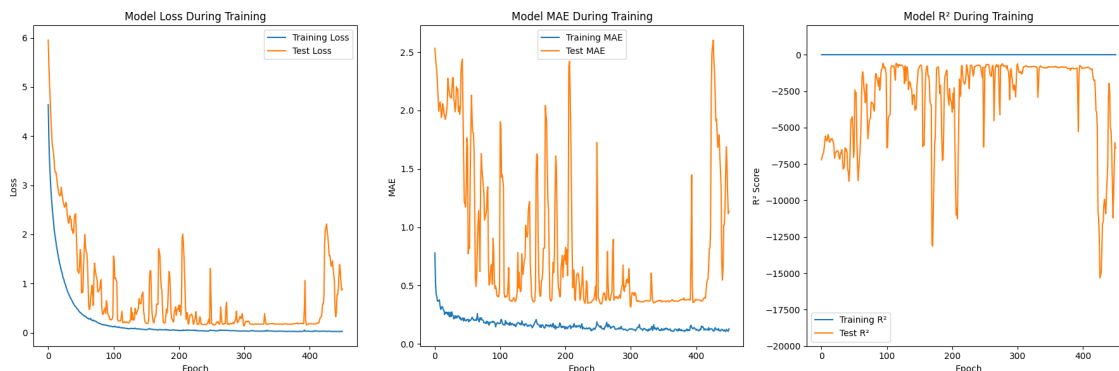


Figura 3.15: Performance do modelo LSTM ao longo do treino

A Figura 3.15 mostra que o modelo LSTM ajusta-se bem aos dados de treino, com perda e MAE reduzidos e R^2 próximo de 1. No entanto, no conjunto de teste, as métricas são instáveis, com perda e MAE variando muito e R^2 frequentemente negativo, indicando sobreajuste e dificuldade de generalização.

O gráfico da Figura 3.16 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:

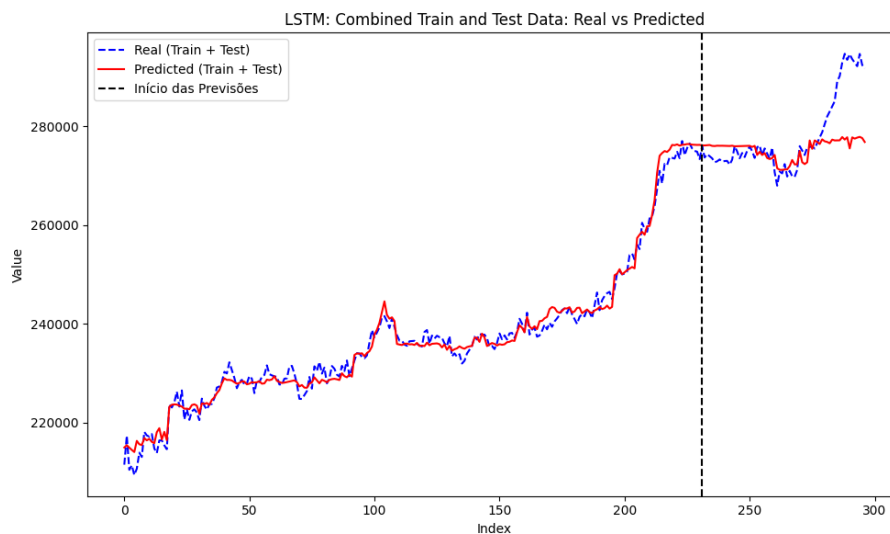


Figura 3.16: *Performance do modelo LSTM*

Esses gráficos mostram que o modelo LSTM foi capaz de detetar 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 Análise Comparativa e Seleção do Modelo Preditivo

Nesta secção, comparamos o desempenho de diferentes modelos de aprendizagem automática treinados e avaliados com base num conjunto de dados de treino e teste.

3.5.1 Comparação de Modelos Preditivos

Após o treino e avaliação, os resultados das principais métricas foram observados e comparados entre os modelos testados. A Tabela 3.11 a seguir resume os resultados de cada modelo.

Tabela 3.11: Comparação de Modelos: Resultados Gerais

Modelo	MAE	R ²	Acurácia (%)	MSE
Regressão Linear	22 948	-1.96	97.07	582 980 257
Regressão Linear TimeSeriesSplit	8 309	0.85	98.98	104 289 812
Random Forest	5 974	-0.49	97.85	86 825 203
Random Forest TimeSeriesSplit	22 749	-4.60	92.36	644 031 910
MLP	5 072	0.06	98.17	54 771 786
LSTM	4 550	0.15	98.36	49 516 827

A comparação das métricas 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 apresentou o menor MAE no conjunto de teste (4.550), indicando que ele foi o mais preciso em prever os valores reais dos preços dos imóveis. O MLP também teve um desempenho próximo, com um MAE de 5.072. A Regressão Linear com *TimeSeriesSplit* também mostrou um bom desempenho, com um MAE de 8.309. Em contrapartida, o modelo Random Forest com *TimeSeriesSplit* apresentou o maior MAE (22.749), sugerindo maior dificuldade em capturar a precisão nas previsões.
- **Erro Quadrático Médio (MSE)** - O LSTM obteve o menor MSE (49.516.827), seguido de perto pelo MLP (54.771.786), indicando que ambos os modelos de redes neurais conseguiram minimizar erros mais extremos, o que é importante em contextos sensíveis como o mercado imobiliário. O Random Forest com *TimeSeriesSplit* teve o maior MSE (644.031.910), sugerindo que ele foi menos eficaz em evitar grandes desvios nas previsões.
- **Coefficiente de Determinação (R²)** - O LSTM apresentou o maior R² no conjunto de teste (0,15), seguido pelo MLP com um valor de 0,06. Esses valores sugerem que esses modelos foram capazes de explicar uma fração da variabilidade dos dados de teste, embora ainda apresentem limitações. A Regressão Linear com TimeSeriesSplit também teve um desempenho positivo, com R² de 0,85, indicando boa capacidade explicativa, mas em uma abordagem específica de divisão temporal. O Random Forest com *TimeSeriesSplit* e a Regressão Linear tradicional apresentaram valores negativos de R², indicando que suas previsões foram menos eficazes que uma simples média.
- **Acurácia (%)** -A acurácia foi calculada para alguns modelos, com o LSTM obtendo a maior acurácia (98,36%), seguido pelo MLP (98,17%) e pela Regressão Linear com TimeSeriesSplit (98,98%). O Random Forest apresentou acurácias ligeiramente inferiores, especialmente na versão com TimeSeriesSplit (92,36%), sugerindo que ele foi menos preciso em capturar as nuances dos dados.

3.5.2 Modelo Selecionado

O modelo LSTM foi selecionado como a melhor opção para prever os preços médios de venda dos imóveis, devido à sua capacidade de captar padrões temporais e tendências

gerais no conjunto de dados. Abaixo estão os resultados de desempenho do LSTM em diferentes conjuntos (treino, teste e validação), resumidos na Tabela 3.12.

Tabela 3.12: Resultados de Desempenho do Modelo LSTM

Conjunto	MAE	MSE	R ²
Treino	1 413	3 250 438	0.98
Teste	4 550	49 516 827	0.15
Validação	27 312.50	792 523 586.45	-13.51

A Tabela 3.12 evidencia o desempenho do LSTM em cada conjunto. Observa-se que o modelo ajustou-se muito bem aos dados de treino, com um MAE e MSE baixos e um R² próximo de 1. No conjunto de teste, o LSTM também manteve uma precisão razoável, embora o R² tenha caído para 0,15, indicando limitações na capacidade de generalização para novos dados. No conjunto de validação, o desempenho foi consideravelmente inferior, com um MAE elevado e um R² negativo, o que sugere dificuldades em captar variações mais complexas e rápidas nos preços.

Além disso, a Figura 3.17 ilustra visualmente as previsões do modelo em comparação com os valores reais, permitindo observar a precisão e as limitações do LSTM ao longo do tempo.

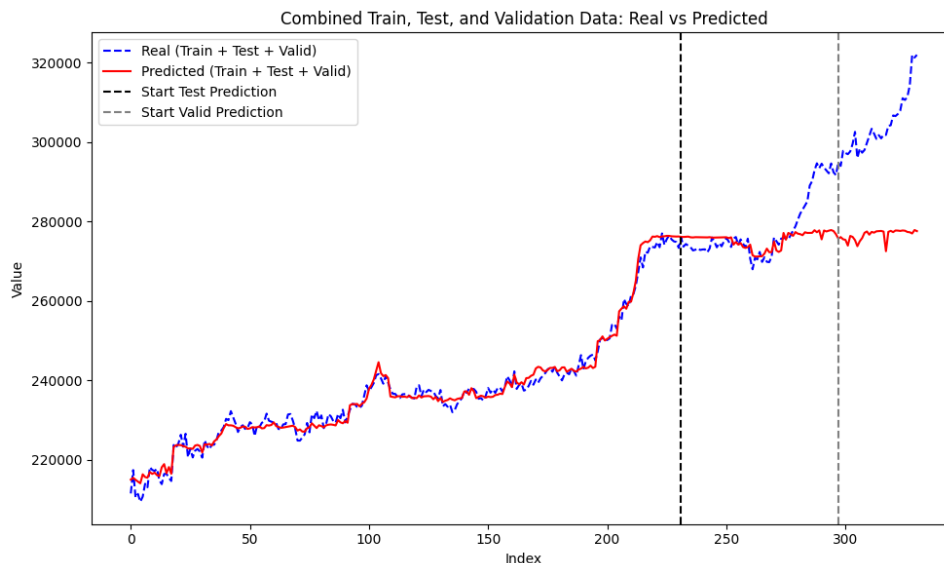


Figura 3.17: *Previsão do Modelo LSTM: Comparação entre Valores Reais e Previstos*

Na Figura 3.17, a linha azul tracejada representa os valores reais dos preços médios de venda dos imóveis, enquanto a linha vermelha sólida representa as previsões do modelo LSTM. A linha preta tracejada marca o início das previsões nos conjuntos de teste e validação. Observa-se que o LSTM acompanha bem a tendência geral dos preços no conjunto de treino, mas no conjunto de validação (após a segunda linha tracejada), o modelo subestima significativamente as variações acentuadas. Esses resultados sugerem que, embora o LSTM capture a tendência principal, ele apresenta dificuldades ao prever flutuações extremas e eventos atípicos, especialmente no conjunto de validação.



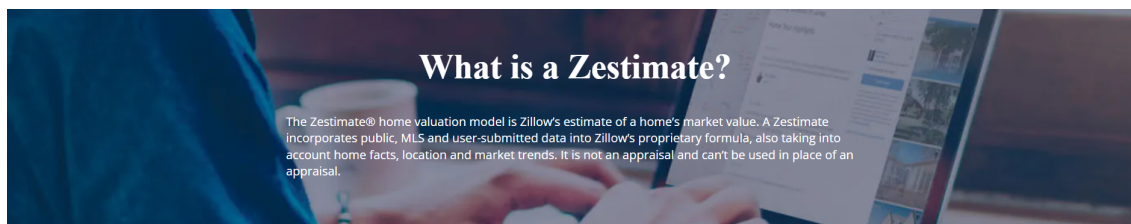
4 Aplicação Web

Pretende-se desenvolver uma aplicação web que integre modelos de aprendizagem automática para prever preços 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 com personalização de filtros e o acesso a informações detalhadas sobre os mercados imobiliários de diferentes regiões. Apresenta-se 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 secção 4.1 trata de plataformas similares que inspiraram funcionalidades da aplicação. Na secção 4.2, temos uma descrição detalhada das principais funcionalidades da aplicação, incluindo análise de mercado e avaliação de investimento. A secção 4.3 aborda o design da interface, desde os esboços iniciais até os mockups e wireframes. Na secção 4.4, temos a implementação, descrevendo as tecnologias utilizadas, a arquitetura da aplicação e o processo de implantação.

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. A Figura 4.1 apresenta uma imagem ilustrativa da página do Zillow, destacando essas funcionalidades.



What is a Zestimate?

The Zestimate® home valuation model is Zillow's estimate of a home's market value. A Zestimate incorporates public, MLS and user-submitted data into Zillow's proprietary formula, also taking into account home facts, location and market trends. It is not an appraisal and can't be used in place of an appraisal.

How accurate is the Zestimate?

The nationwide median error rate for the Zestimate for on-market homes is 2.4%, while the Zestimate for off-market homes has a median error rate of 7.49%. The Zestimate's accuracy depends on the availability of data in a home's area. Some areas have more detailed home information available — such as square footage and number of bedrooms or bathrooms — and others do not. The more data available, the more accurate the Zestimate value will be.

These tables break down the accuracy of Zestimates for both active listings and off-market listings.

Active listings accuracy

	Top Metro Areas		States	National	
Metropolitan Areas	Median Error	Homes With Zestimates	Within 5% of Sales Price	Within 10% of Sales Price	Within 20% of Sales Price
Atlanta	2.02%	25.2 K	82.43%	94.14%	98.12%
Austin	2.34%	14.2 K	80.78%	94.29%	98.62%
Baltimore	2.04%	7.4 K	77.03%	90.51%	95.38%

Figura 4.1: Zestimate - Zillow's estimate of a home's market value

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. Na Figura 4.2, é apresentado um exemplo visual da página do Redfin, destacando essas ferramentas e funcionalidades.

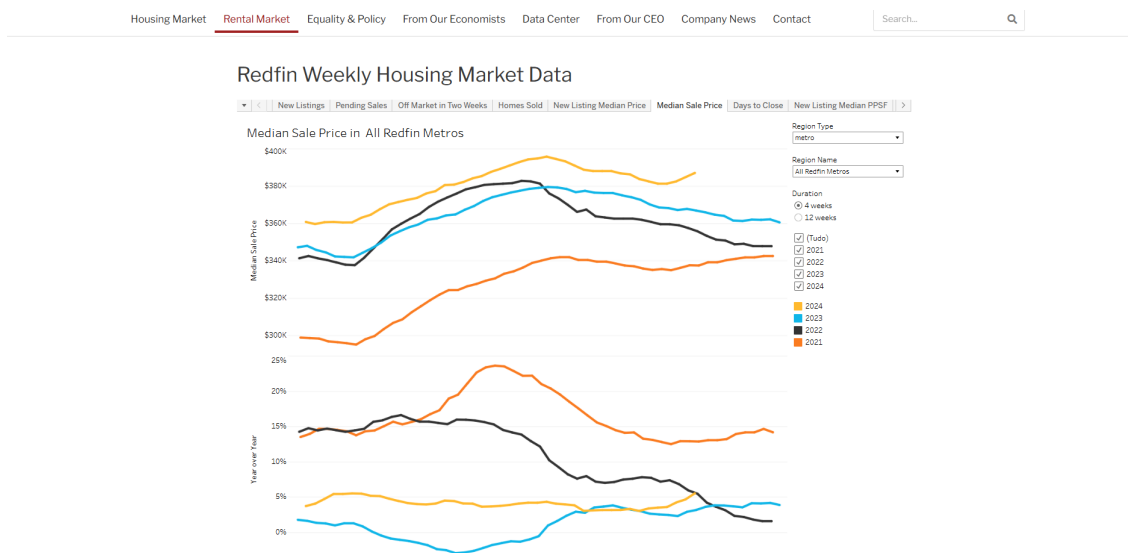


Figura 4.2: Housing Market Data From Redfin

Embora essas plataformas sejam eficazes nos seus mercados, focadas principalmente nos Estados Unidos, não contêm dados nem previsões do mercado português. Por isso, a aplicação desenvolvida foi projetada para explorar modelos preditivos ao contexto local, oferecendo funcionalidades ajustadas às particularidades do mercado de Portugal.

4.2 Descrição da Aplicação

A aplicação 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.

As principais funcionalidades a aplicação são:

- **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 ter 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 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 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 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: [Link do 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 esboço em papel criado durante essa fase. O esboço organizou as principais funcionalidades da aplicação de maneira lógica e fácil de navegar.

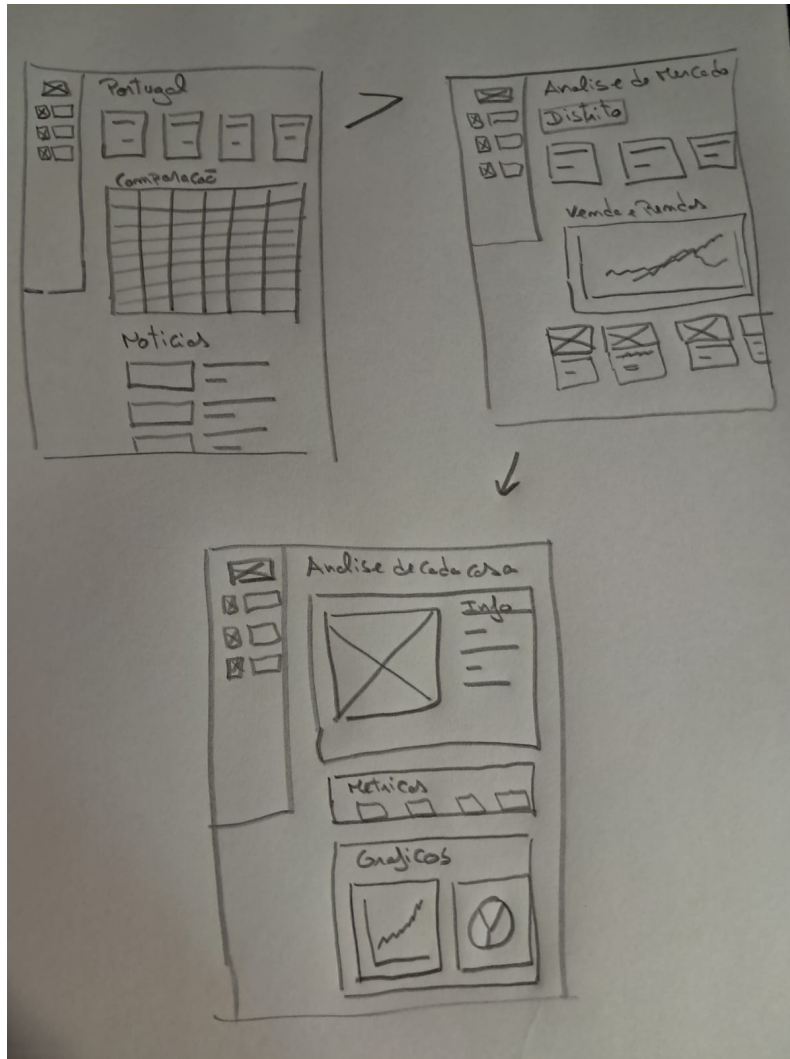


Figura 4.3: Esboço em papel da aplicação

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*, como podemos ver na Figura 4.4, que são representações estruturais da interface, mapeando a disposição dos principais elementos como botões, menus e áreas de conteúdo. No caso desta aplicação, so *wireframes* 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.

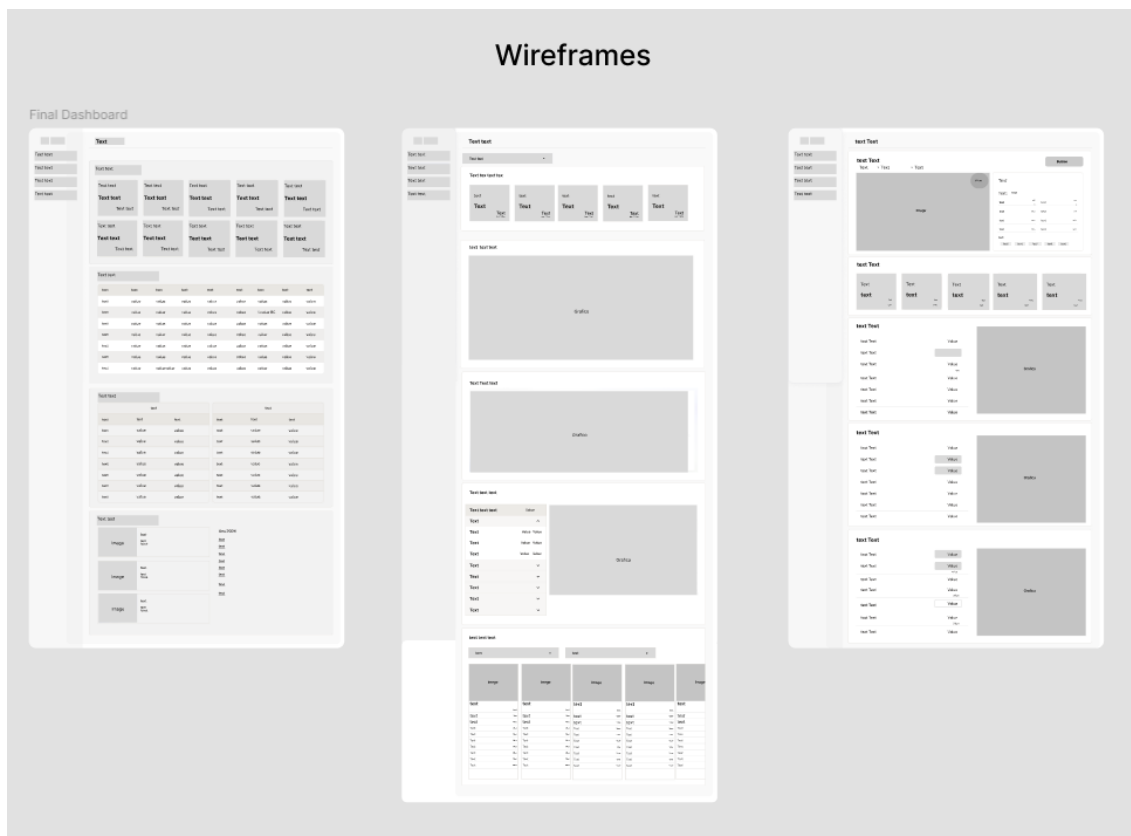


Figura 4.4: Wireframes da aplicação

Após a validação dos *Wireframes*, o processo evoluiu para a criação de *Mockups*, como podemos ver na Figura 4.5. Enquanto os *Mockups* 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.

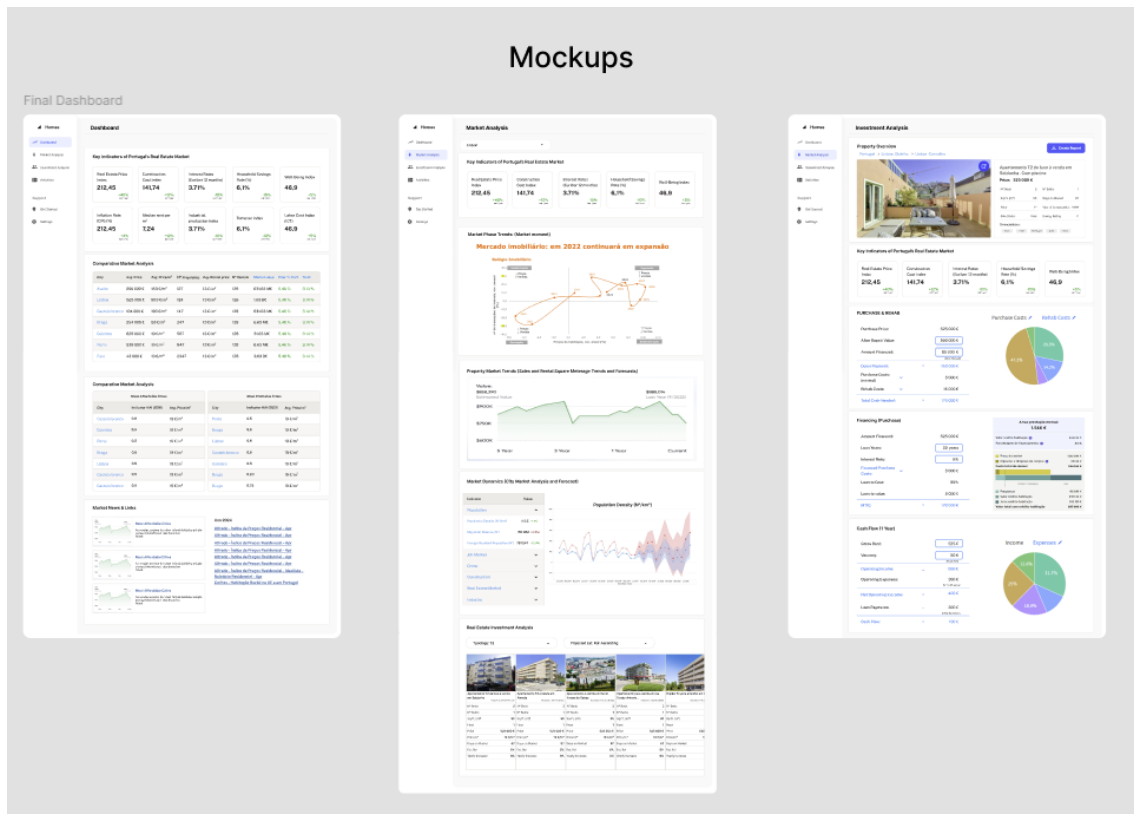


Figura 4.5: *Mockups da aplicação*

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 a visualizar como a aplicação seria percebida pelos utilizadores. Na Figura 4.5, 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 aos protótipos interativos, garantiu que o design da aplicação é visualmente atrativo e funcional sendo 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 aceites 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 no 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. Esta secção descreve as principais tecnologias utilizadas, a arquitetura do *front-end*, e a gestão de dados.

4.4.1 Tecnologias

A aplicação foi organizada em duas camadas principais: *back-end* e *front-end*. O *back-end* foi implementado utilizando [Flask](#), um *microframework web* em [Python](#), escolhido pela leveza e flexibilidade na criação de APIs. O Flask foi utilizado para servir os dados armazenados em ficheiros no servidor e processar os pedidos 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 [React](#), uma biblioteca JavaScript amplamente utilizada para a construção de interfaces de utilizador dinâmicas. O React foi escolhido pela 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 React, foi utilizado o [Material UI](#) como biblioteca para facilitar a criação de estilos da página. O Material UI oferece um conjunto de componentes pré-construídos, para 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 4.6](#).

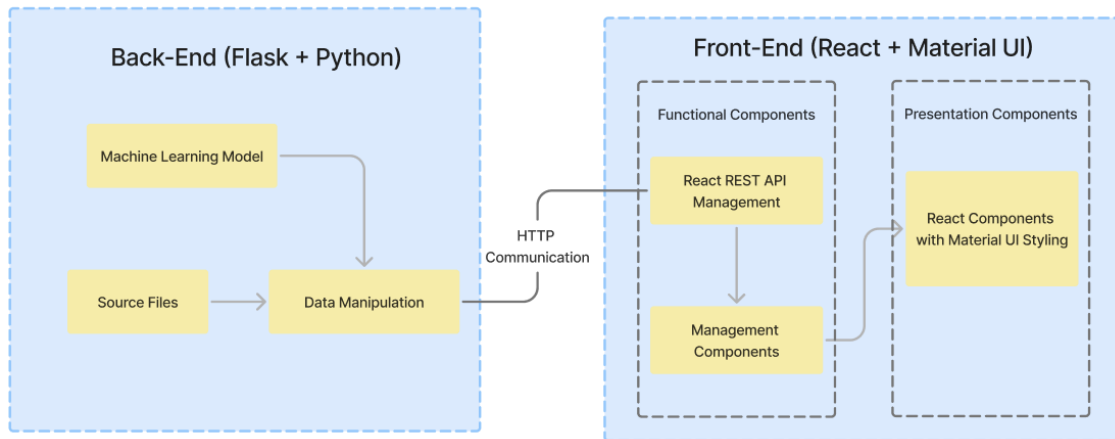


Figura 4.6: Arquitetura da aplicação

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 pré-processados para o *front-end* via [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 React 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 pedidos 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. A aplicação funciona 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 Implantação 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 \(Integração Contínua\) \(CI\)](#), 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 implantação dado ter vantagens, especialmente em projetos baseados em React. Além de integrar-se 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 acessada através do link: [Aplicação no Vercel](#).

Na Figura 4.7, é apresentada a primeira página, o *Dashboard*, que oferece uma visão geral dos principais indicadores do mercado imobiliário em Portugal.

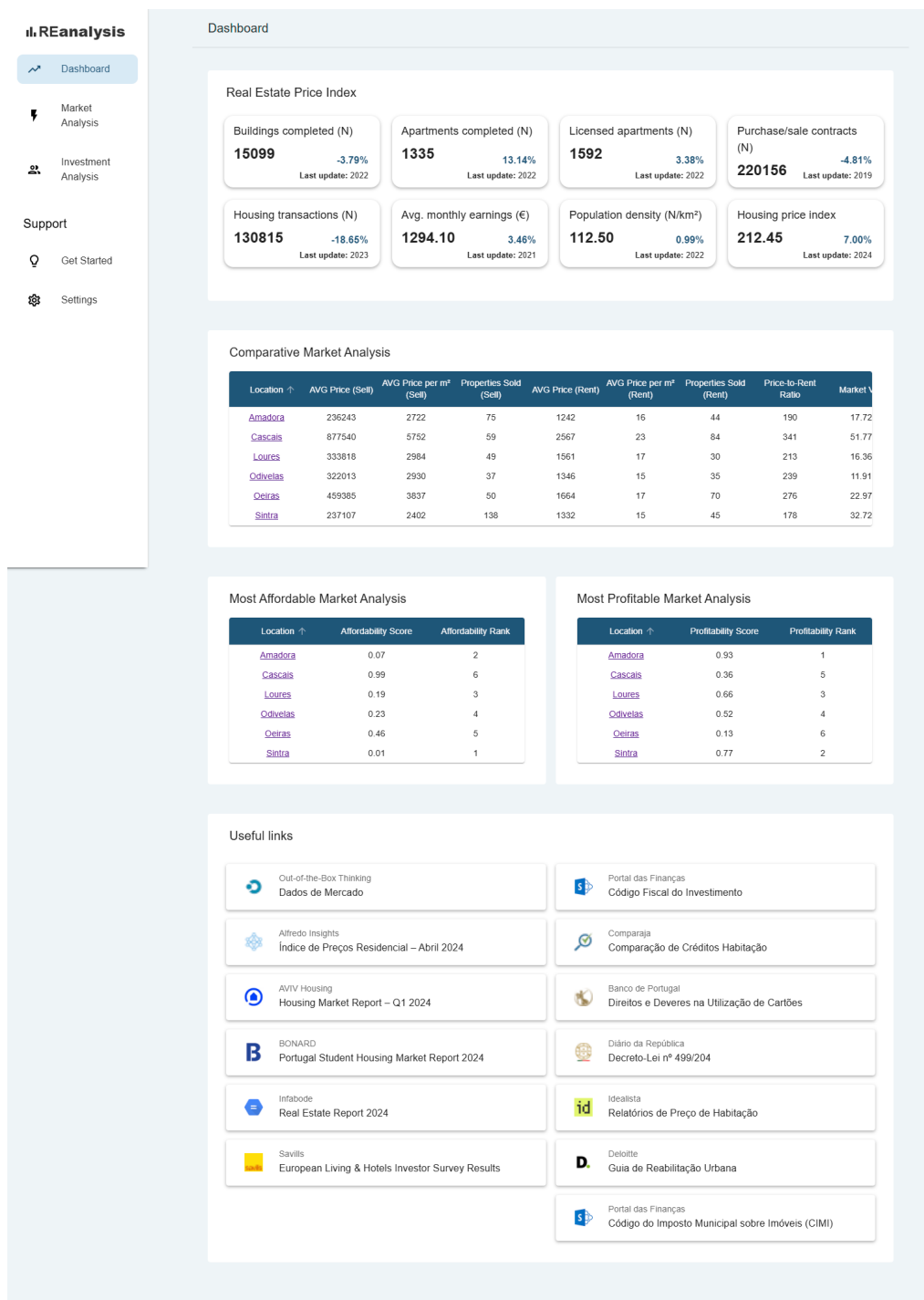


Figura 4.7: Primeira Página da aplicação implementada

Na Figura 4.8, a segunda página, *Market Analysis*, foca-se numa análise detalhada do mercado de um concelho específico.

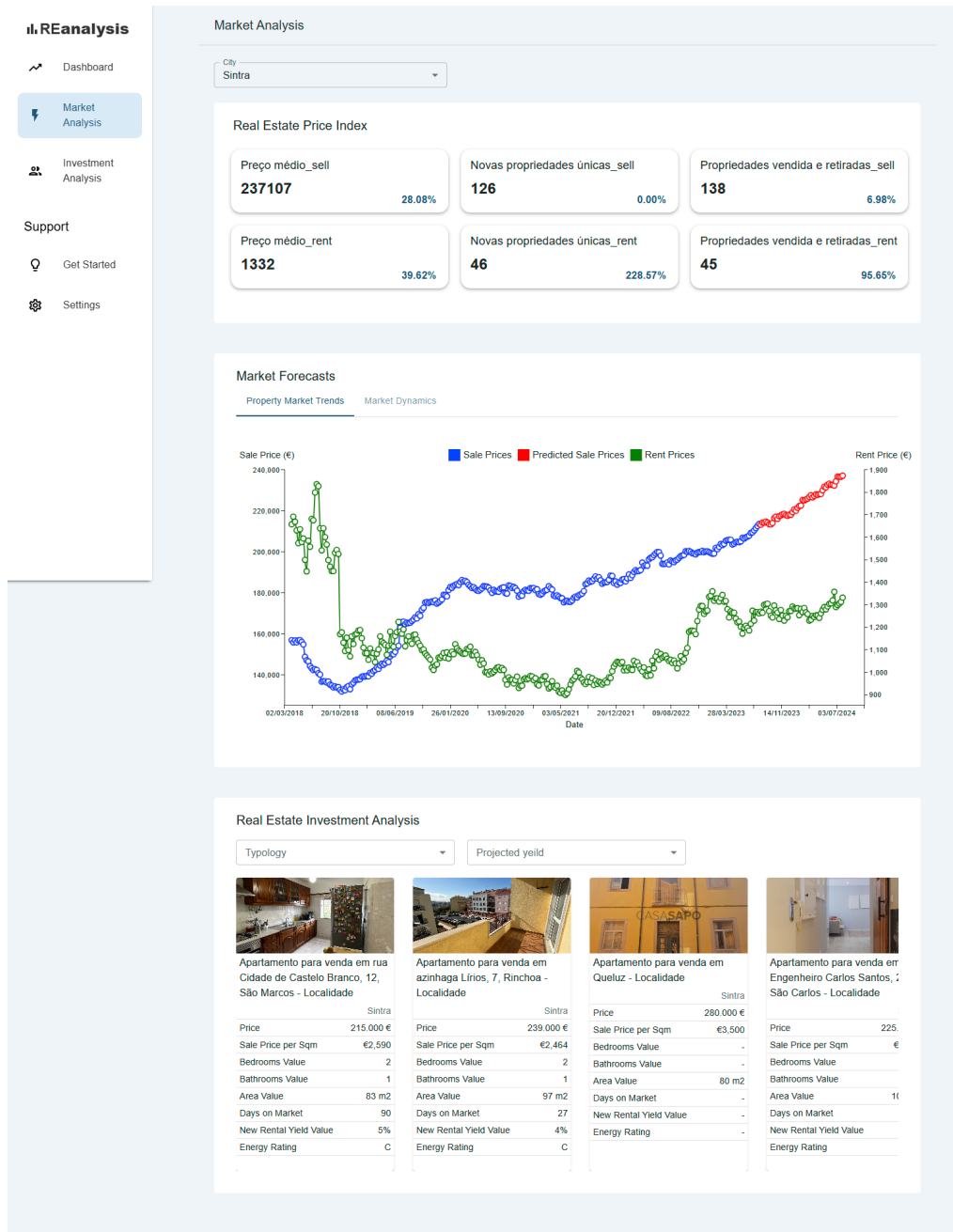


Figura 4.8: Segunda Página da aplicação implementada

A terceira página, *Investment Analysis*, ilustrada na Figura 4.9, fornece uma análise financeira detalhada de propriedades específicas.

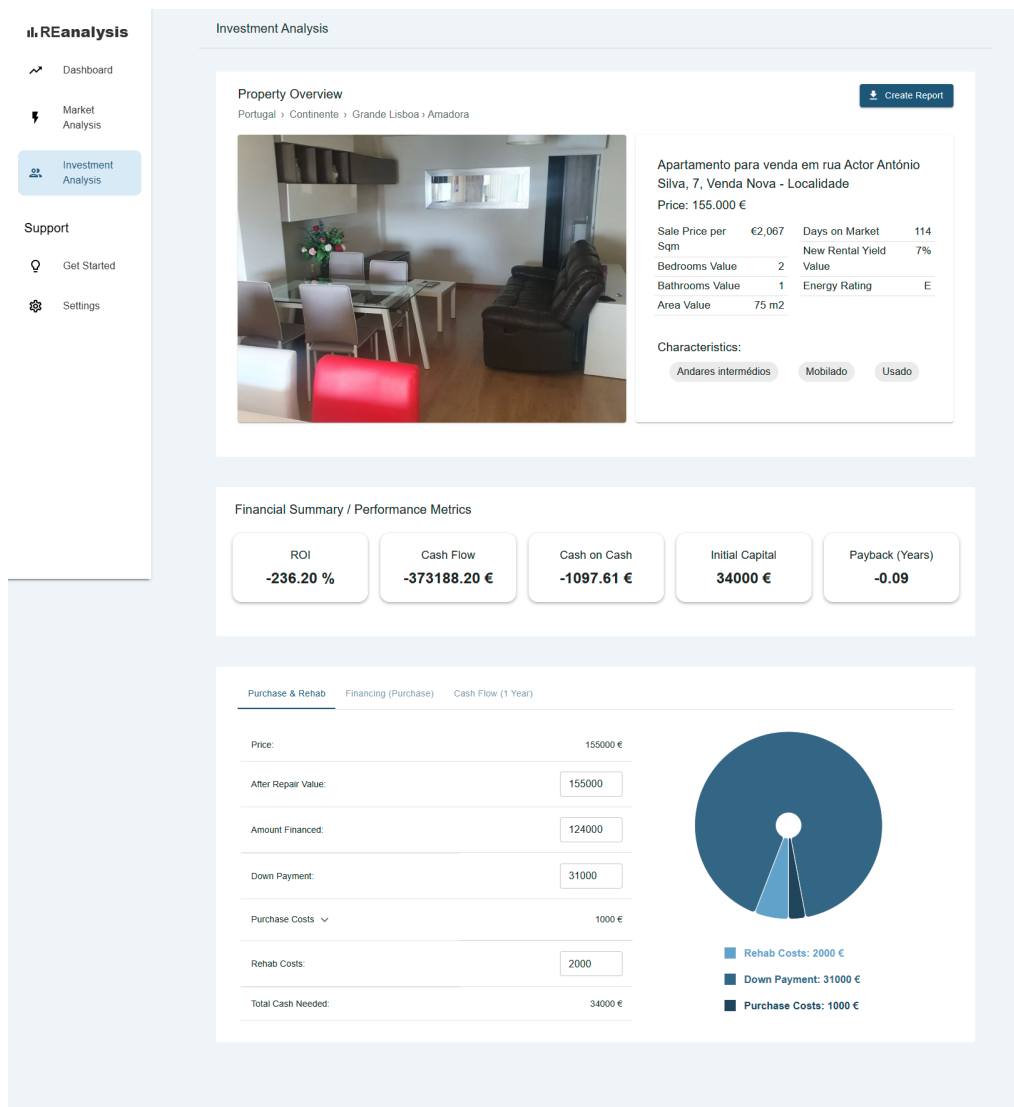


Figura 4.9: Terceira Página da aplicação implementada

5

Avaliação Experimental

Neste capítulo, são apresentados os resultados obtidos a partir dos diferentes modelos de aprendizagem automática aplicados à previsão dos preços imobiliários. A secção 5.1 realiza 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 (R^2). Em seguida, na secção 5.2 é realizada uma avaliação qualitativa da aplicação desenvolvida, discutindo a sua usabilidade, funcionalidade e valor prático para o mercado imobiliário. Por fim, na secção 5.3 são apresentadas as indicações obtidas 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. O objetivo é identificar o modelo que apresenta melhor desempenho ao prever os preços médios de venda, considerando diferentes variáveis económicas e do mercado imobiliário. Na Tabela 5.1 podemos ver a comparação feita com base em métricas como MAE, MSE e R^2 , 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	MAE	R^2	Acurácia (%)	MSE
Regressão Linear	22 948	-1.96	97.07	582 980 257
Regressão Linear TimeSeriesSplit	8 309	0.85	98.98	104 289 812
Random Forest	5 974	-0.49	97.85	86 825 203
Random Forest TimeSeriesSplit	22 749	-4.60	92.36	644 031 910
MLP	5 072	0.06	98.17	54 771 786
LSTM	4 550	0.15	98.36	49 516 827

Após uma análise comparativa dos modelos testados, o modelo LSTM destacou-se como a melhor opção para prever os preços médios de venda dos imóveis. Apesar de apresentar um desempenho inferior no conjunto de validação, o LSTM demonstrou uma maior capacidade de captar a tendência geral dos dados no conjunto de treino e de teste, além de ser mais

robusto na modelagem de dados sequenciais, como séries temporais.

No conjunto de validação, o modelo LSTM obteve um MAE de 27.312,50, um R^2 de -13,51 e um MSE de 792.523.586,45, com uma acurácia de 91,01%. Embora o R^2 negativo indique que o modelo ainda enfrenta desafios em capturar a variabilidade completa dos dados, o LSTM manteve uma boa precisão e foi capaz de seguir a tendência geral dos preços, conforme ilustrado na figura. No entanto, o modelo mostra limitações para prever oscilações extremas e padrões mais complexos, especialmente nas fases de validação e teste, onde subestima o crescimento acentuado dos preços.

A escolha do LSTM foi justificada principalmente pela sua capacidade de "memorizar" padrões sequenciais, o que é essencial para dados temporais. Além disso, o LSTM apresentou os menores erros de previsão no conjunto de treino e testou favoravelmente com relação aos outros modelos, mesmo com o desafio do conjunto de validação.

Para garantir a continuidade e a aplicação do modelo, o LSTM, juntamente com os parâmetros e pesos ajustados, foi salvo e armazenado para ser utilizado na próxima etapa: a aplicação web. Este modelo será integrado no *backend* da aplicação, permitindo previsões em tempo real para os utilizadores, que poderão obter estimativas dos preços médios de venda de imóveis. Essa aplicação prática do modelo permite que seja utilizado em cenários do mundo real, ampliando o valor prático do projeto.

5.2 Avaliação da Aplicação Web

Relativamente à avaliação das funcionalidades e da usabilidade do sistema, analisamos o *feedback* recolhido dos utilizadores através de questionários e testes de utilização. 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 [online](#) e os resultados em anexo I.1, 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.

5.2.1 Participantes

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 constituíssem uma amostra diversificada e relevante do público-alvo.

Para garantir uma avaliação representativa, foram selecionados seis participantes com perfis variados, refletindo a diversidade de utilizadores. Entre as características principais dos participantes, destacam-se: a predominância do género masculino, idades que variam entre 18 e 65 anos, e uma frequência diária de uso de aplicações Web como podemos ver nas seguintes Figuras 5.1 e 5.2:

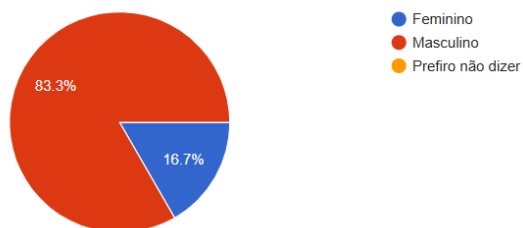


Figura 5.1: Resultados do formulário: Distribuição dos gêneros

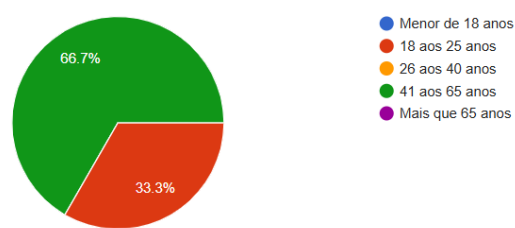


Figura 5.2: Resultados do formulário: Distribuição das Idades

A maioria declarou-se como experiente ou com alguma familiaridade com ferramentas digitais, além de possuir experiência com aplicações similares, como Idealista, Remax, e Imovirtual. Quanto à atualização sobre o mercado imobiliário, todos os participantes consideram essa prática importante em diferentes graus. A diversidade nos níveis de experiência e o uso frequente de tecnologias asseguram uma análise abrangente e relevante sobre a aplicação Web em avaliação.

5.2.2 Resultados da Avaliação

O questionário foi estruturado em categorias distintas de perguntas, visando recolher *feedback* específico sobre vários aspetos da aplicação Web:

- **Compreensão e Clareza** - Os participantes foram questionados sobre a facilidade em entender as informações preditivas e se as previsões de mercado eram claras e acessíveis. De forma geral, a maioria dos utilizadores considerou as informações claras, com uma média de 4,17 nas respostas para facilidade de uso e clareza. Gráficos e tabelas foram bem avaliados, sendo elogiados pela organização e visualização eficiente dos dados de mercado.
- **Utilidade das Funcionalidades** - Os participantes compartilharam suas impressões sobre a eficácia das ferramentas de análise de mercado e investimento oferecidas. Houve um destaque para a utilidade de funcionalidades específicas, como *Market Forecasts* e Análise Demográfica Detalhada, bem como a possibilidade de personalizar filtros para uma análise detalhada de diferentes zonas e faixas de preço.
- **Estética e Apelo Visual** - Questões sobre o design e estética da aplicação foram incluídas para avaliar o apelo visual e sua contribuição para uma boa experiência de utilizador. Embora o design tenha sido elogiado por ser visualmente atraente e profissional, alguns participantes sugeriram melhorias, como um design mais apelativo ou simplificação de algumas áreas visuais, para otimizar a experiência de navegação e interação.

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

5.2.3 Sugestões Futuras para a Aplicação

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

- **Variedade de Dados** - Foi recomendado incluir uma maior diversidade de dados, especialmente indicadores adicionais, como indicadores socioeconômicos, para enriquecer a análise e oferecer uma visão mais abrangente do mercado.
- **Filtro Inicial para Segmentação** - Sugestões frequentes destacaram a necessidade de um filtro inicial para facilitar a segmentação por zonas geográficas, tipologia de propriedades ou faixa de preços, agilizando a busca por dados relevantes.
- **Design Mais Apelativo** - Alguns participantes sugeriram melhorias estéticas, como um design mais moderno e visualmente atrativo, para tornar a interface mais agradável e intuitiva.
- **Tutoriais e Dicas Contextuais** - Foi recomendada a inclusão de tutoriais ou dicas em contexto para orientar novos utilizadores, especialmente em funcionalidades mais complexas, facilitando o aprendizado e aumentando a acessibilidade da aplicação.



6

Discussão de Resultados

Neste capítulo, analisamos e discutimos os principais resultados obtidos a partir da aplicação dos diversos modelos de aprendizagem automática para a previsão de preços imobiliários. A discussão é organizada em torno de quatro questões que guiaram esta pesquisa, permitindo uma análise sobre os desafios enfrentados e as descobertas realizadas em cada uma delas.

Questão 1: Quais foram as limitações enfrentadas ao integrar dados de diferentes fontes, como APIs e bases de dados externas?

A integração de dados de múltiplas fontes, como APIs e bases de dados externas, trouxe tanto benefícios quanto desafios ao estudo. A diversidade de fontes ampliou a análise do mercado imobiliário, mas também apresentou dificuldades relacionadas à inconsistência nos formatos e estruturas dos dados, que exigiram esforços significativos de normalização. A sincronização temporal entre fontes com diferentes frequências e períodos de atualização foi outro obstáculo, assim como a presença de valores ausentes e discrepantes, que aumentaram a incerteza dos resultados. Apesar desses obstáculos, a diversidade dos dados permitiram uma análise mais completa e detalhada do mercado, contribuindo significativamente para os objetivos da pesquisa.

Algumas variáveis relevantes, como dados sobre condições de crédito específicas (quantidades de créditos ou valor monetário), dados demográficos, como os saldos populacionais, ou até mesmo rendimentos médios, não puderam ser extraídas, pois não continham o mesmo nível de detalhe temporal que as demais fontes. Essa limitação impediu uma análise mais aprofundada sobre certos fatores que poderiam influenciar o mercado imobiliário, restringindo o escopo das variáveis consideradas no estudo.

Questão 2: Quais os melhores algoritmos para realizar previsão do mercado?

Com base nos resultados obtidos, o LSTM mostrou-se o melhor algoritmo para a previsão de preços imobiliários, destacando-se com o menor Erro Absoluto Médio (MAE) e o maior coeficiente de determinação (R^2). Sua arquitetura permite armazenar e processar

informações sequenciais, o que justifica seu desempenho superior. Modelos mais avançados, como XGBoost ou Transformers, poderiam aprimorar ainda mais as previsões, graças à sua habilidade em capturar padrões complexos e lidar com grandes volumes de dados. No entanto, esses modelos apresentam desafios, como a necessidade de maior capacidade computacional e um ajuste mais rigoroso de hiper-parâmetros para otimizar o desempenho.

Questão 3: Qual é a melhor maneira de apresentar as análises e entendimentos aos utilizadores da aplicação de forma clara e intuitiva?

A melhor forma de apresentar as análises na aplicação é por meio de uma interface visualmente apelativa e interativa, utilizando gráficos claros e intuitivos, como os apresentados nas páginas de *Market Analysis* e *Investment Analysis*. A inclusão de comparações entre concelhos, gráficos de tendências e diagramas interativos facilita a interpretação dos dados. Além disso, o uso de filtros e uma navegação simples permite que os utilizadores adaptem as informações às suas necessidades específicas, tornando a análise mais eficiente e acessível para diferentes perfis de investidores.

Questão 4: Quais os fatores que os investidores consideram quando pretendem investir?

Antes do desenvolvimento do Website, foi realizada uma análise para identificar os fatores mais importantes considerados pelos investidores no mercado imobiliário. Esses fatores incluíram o preço de compra, potencial de valorização, condições de financiamento, entre outros. O artigo do Databox [34] apresenta uma hierarquia dessas prioridades, conforme ilustrado no gráfico da Figura 6.1. Com base nesses dados, as funcionalidades da aplicação foram projetadas para atender a essas necessidades.



Figura 6.1: Principais KPIs considerados pelos investidores no mercado imobiliário [34]

Após o desenvolvimento da aplicação, o feedback dos utilizadores indicou a necessidade de novas funcionalidades para aprimorar a análise e facilitar a tomada de decisões. As

sugestões para o Dashboard incluíram Mapas Interativos, um Calendário de eventos, filtros para segmentação inicial por zona e dados mais detalhados.

Na Market Analysis, os utilizadores solicitaram Análise Demográfica, alertas, filtros avançados e informações adicionais. Já na Investment Analysis, destacaram a importância de uma Simulação de Créditos, Benchmarking de Investimentos, Ferramentas de Gestão de Risco e dados mais detalhados. Essas melhorias visam tornar a aplicação mais personalizada e útil para investidores imobiliários.



7

Conclusão

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.

Os resultados deste estudo demonstraram que modelos de aprendizagem automática são ferramentas promissoras para prever preços no mercado imobiliário, especialmente quando várias variáveis económicas e de mercado são integradas. Entre os modelos testados, o LSTM destacou-se com o melhor desempenho geral, apresentando o menor MAE (4.550) e MSE (49.516.827) no conjunto de teste, e uma acurácia de 98,36%. Esses resultados indicam que o LSTM conseguiu capturar de forma eficiente os padrões temporais nos dados, embora seu R^2 relativamente baixo (0,15) e o alto MAE no conjunto de validação indiquem limitações em cenários de validação mais complexos.

Entre os modelos mais explicativos, como a Regressão Linear e o Random Forest, a Regressão Linear apresentou o melhor desempenho e o mais interpretável. O gráfico de importância das variáveis para este modelo revela que o modelo atribui maior peso ao "Endividamento dos Particulares" em diversas categorias (junto a empresas não financeiras, setor financeiro e resto do mundo), além de considerar as taxas de juro (Euribor e TAEG) como fatores relevantes. Isso reflete a dependência do modelo linear de variáveis macroeconómicas diretamente relacionadas ao crédito e ao nível de endividamento, destacando sua capacidade de captar relações económicas estruturais no mercado imobiliário.

Além disso, a 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.

7.1 Trabalho Futuro

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

- **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.
- **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.
- **Novas Funcionalidades de Simulação** - Expandir a simulação de cenários para incluir múltiplas variáveis simultaneamente e permitir a comparação de resultados lado a lado.
- **Exploração de Outros Métodos de Machine Learning e Modelos Ensemble** - Para aprimorar a precisão e robustez das previsões, métodos adicionais de aprendizagem automática, como Gradient Boosting Machines (GBM), XGBoost ou até redes neurais convolucionais (CNNs) para séries temporais, poderiam ser explorados. Modelos ensemble, que combinam previsões de múltiplos modelos, também podem ser investigados para aproveitar os pontos fortes de cada abordagem e melhorar a precisão final das previsões.

Referências

- [1] [Online; accessed 03. Nov. 2024]. Set. de 2017. URL: <https://pdfs.semanticscholar.org/0e5f/ab193cf788e6bc76305f9e2d5664b6351496.pdf> (ver p. 15).
- [2] [Online; accessed 03. Nov. 2024]. Out. de 2018. URL: https://ceejme.eu/wp-content/uploads/2018/10/ceejme_4_7_art_10.pdf (ver pp. 8, 9).
- [3] [Online; accessed 03. Nov. 2024]. Out. de 2021. URL: <https://www.superiorschoolnc.com/wp-content/uploads/2019/10/NC-8e-Superior-Chs-1-3.pdf> (ver p. 7).
- [4] [Online; accessed 03. Nov. 2024]. Mar. de 2024. URL: <https://www.ijcsma.com/articles/literature-review-on-real-estate-value-prediction-using-machine-learning.pdf> (ver p. 14).
- [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, 12, 17).
- [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. 14).
- [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_Informacao_Geografica (ver p. 15).
- [8] E. Circle. “REAL ESTATE MARKET AND ITS CHARACTERISTICS”. Em: *VALUER WORLD* (jan. de 2022). URL: <https://www.valuerworld.com/2022/01/20/real-estate-market-and-its-characteristics> (ver pp. 8, 9).
- [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. 12).

- [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](https://doi.org/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](https://doi.org/10.5755/j01.ee.28.2.13777) (ver p. 9).
- [14] *Forecasting and Econometric Models - Econlib*. [Online; accessed 03. Nov. 2024]. Out. de 2018. URL: <https://www.econlib.org/library/Enc/ForecastingandEconometricModels.html> (ver p. 15).
- [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](https://doi.org/10.1016/S2212-5671(16)30304-5) (ver p. 10).
- [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](https://doi.org/10.1186/s40537-021-00476-0) (ver p. 11).
- [17] *Hedonic Pricing: Definition, How the Model Is Used, and Example*. [Online; accessed 03. Nov. 2024]. Mar. de 2024. URL: <https://www.investopedia.com/terms/h/hedonicpricing.asp> (ver p. 13).
- [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. 12).
- [19] 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. 8).
- [20] 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](https://doi.org/10.1155/2022/5750354) (ver p. 13).
- [21] 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](https://doi.org/10.1016/j.seps.2020.100916) (ver p. 13).
- [22] 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](https://doi.org/10.1111/1540-6229.12397) (ver pp. 13, 14, 16).
- [23] 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. 10).

- [24] 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](https://doi.org/10.3390/app12168123) (ver p. 16).
- [25] 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](https://doi.org/10.3390/land11112100) (ver p. 11).
- [26] *O que são SIG? | Sistemas de Informação Geográfica*. [Online; accessed 03. Nov. 2024]. Mar. de 2024. URL: <https://www.esri-portugal.pt/pt-pt/o-que-sao-os-sig/overview> (ver p. 15).
- [27] 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](https://doi.org/10.3390/app10175832) (ver p. 9).
- [28] 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. 10).
- [29] *Real Estate - China | Statista Market Forecast*. [Online; accessed 03. Nov. 2024]. Abr. de 2023. URL: <https://www.statista.com/outlook/fmo/real-estate/china> (ver p. 2).
- [30] *Real Estate - Worldwide | Statista Market Forecast*. [Online; accessed 03. Nov. 2024]. Fev. de 2024. URL: <https://www.statista.com/outlook/fmo/real-estate/worldwide#methodology> (ver p. 1).
- [31] *Regressão Linear: Principais conceitos, tipos, limitações, aplicações e exemplo – MakerZine*. [Online; accessed 03. Nov. 2024]. Mar. de 2024. URL: <https://www.makerzine.com.br/cozinha-de-dados/regressao-linear-principais-conceitos-tipos-limitacoes-aplicacoes-e-exemplo> (ver p. 13).
- [32] *Residential Real Estate - Global | Statista Market Forecast*. [Online; accessed 03. Nov. 2024]. Abr. de 2023. URL: <https://fr.statista.com/outlook/fmo/real-estate/residential-real-estate/worldwide> (ver p. 2).
- [33] B.-O. Sjöling. *Indicators for bubble formation in housing markets*. 2012 (ver p. 11).
- [34] F. Stojanovic. “17 KPIs and Metrics You Should Include in a Real Estate KPI Dashboard | Databox”. Em: *Databox* (ago. de 2024). URL: <https://databox.com/real-estate-metrics-dashboard> (ver p. 68).
- [35] *View of Hedonic Valuation of Real Estate Properties in Nigeria*. [Online; accessed 03. Nov. 2024]. Mar. de 2024. URL: <https://journals.uct.ac.za/index.php/JARER/article/view/452/537> (ver pp. 14, 16).
- [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](https://doi.org/10.1155/2022/1695265) (ver p. 16).

- [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).



I Anexo 1

Tabela I.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 (BETA)_sell	0.0	0.0	0.0	0.0	0.0
Preço médio de fechamento por m ² (BETA)_sell	0.0	0.0	0.0	0.0	0.0
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 únicas_sell	35.0	30.0	25.0	58.0	34.0
Propriedades vendidas e retiradas_sell	30.0	21.0	27.0	103.0	37.0
Novas listagens não desduplicadas_sell	68.0	50.0	49.0	89.0	60.0
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 (BETA)_rent	0.0	0.0	0.0	0.0	0.0
Preço médio de fechamento por m ² (BETA)_rent	0.0	0.0	0.0	0.0	0.0
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 únicas_rent	16.0	2.0	9.0	7.0	3.0
Propriedades vendidas e retiradas_rent	20.0	4.0	12.0	15.0	14.0
Novas listagens não desduplicadas_rent	24.0	3.0	10.0	7.0	5.0
price_to_rent_ratio	356.482664	267.311819	299.476985	157.493312	256.921615

Tabela I.2: Endividamento dos Particulares

Métrica	248	579	903	1234	1565
Endividamento dos particulares	151348.35	151348.35	151348.35	151348.35	151348.35
Endividamento dos particulares junto do setor financeiro	136796.08	136796.08	136796.08	136796.08	136796.08
Endividamento dos particulares junto de empresas não financeiras	10067.38	10067.38	10067.38	10067.38	10067.38
Endividamento dos particulares junto das administrações públicas	96.07	96.07	96.07	96.07	96.07
Endividamento dos particulares junto de particulares	953.64	953.64	953.64	953.64	953.64
Endividamento dos particulares junto do resto do mundo	3435.19	3435.19	3435.19	3435.19	3435.19
TAEG de novos empréstimos à habitação	4.82	4.82	4.82	4.82	4.82
TAEG de novos empréstimos ao consumo	10.48	10.48	10.48	10.48	10.48
Taxa de juro (TAA) de novos empréstimos à habitação	3.30	3.30	3.30	3.30	3.30
Taxa de juro (TAA) de novos empréstimos ao consumo	8.45	8.45	8.45	8.45	8.45
Taxa de juro (TAA) de novos empréstimos às empresas não financeiras	4.69	4.69	4.69	4.69	4.69
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 I.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 subsídio de desemprego (milhares)-mensal	183.9	183.9	183.9	183.9	183.9
Desemprego registado no final do período-Total-Mensal	322086.0	322086.0	322086.0	322086.0	322086.0
Desemprego registado ao longo do período-Total-Mensal	55841.0	55841.0	55841.0	55841.0	55841.0
Indicador coincidente para a atividade económica-Mensal-TVH	3.5	3.5	3.5	3.5	3.5
Indicador coincidente para o consumo privado-Mensal-TVH	1.7	1.7	1.7	1.7	1.7
IPC total-taxa de variação homóloga	8.4	8.4	8.4	8.4	8.4
IPC total-taxa de variação em cadeia	-0.8	-0.8	-0.8	-0.8	-0.8
IPC total (taxa de inflação média anual)-TVMM12	8.2	8.2	8.2	8.2	8.2

Tabela I.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_Reducão_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_Reducão_rent	6.66667	0.33333	0.00000	0.66667	0.00000

ANEXO I. ANEXO 1

Pergunta	Resposta 1	Resposta 2	Resposta 3	Resposta 4	Resposta 5	Resposta 6
Indique o seu gênero	Masculino	Feminino	Masculino	Masculino	Masculino	Masculino
Em que intervalo de idade se encontra?	18 aos 25 anos	41 aos 65 anos	41 aos 65 anos	41 aos 65 anos	18 aos 25 anos	18 aos 25 anos
Com que frequência usa aplicações Web?	Diariamente	Diariamente	Diariamente	Diariamente	Diariamente	Diariamente
Considera-se experiente no uso de tecnologias?	Sim, sou muito experiente	Sim, tenho alguma	Sim, tenho alguma	Sim, tenho alguma	Sim, sou experiente	Sim, sou experiente
Quais outros aplicativos similares já usou para analisar dados no mercado imobiliário no seu dia-a-dia?	Idealista	Idealista, CasaFati, Google Sheets, Supercasa, Remax	Imovirtual	Bolsa	Anúncio	Sites de Remax, Imovirtual, etc.
Qual é a importância para você estar atualizado sobre o mercado imobiliário no dia a dia?	Importante	Muito importante	Importante	Muito importante	Não é tão importante	Importante
A aplicação Web é fácil de utilizar?	4	5	3	4	5	5
A aplicação Web é bem projetada e intuitiva?	5	5	3	3	4	5
Gostaria de ver mais recursos adicionais na página "Dashboard" da aplicação Web? Se sim	Mapas Interativos e Filtros	Calendário com a visualização de eventos importantes e preços	-	Mais dados e Mais informações	-	Poder logo no início filtrar por zona
Gostaria de ver mais recursos adicionais na página "Market Analysis" da aplicação Web? Se sim	Análise Demográfica Detalhada e Alertas	Não	-	Mais informações	-	Poder logo no início filtrar por zona (tipologia, range de preços)
Gostaria de ver mais recursos adicionais na página "Investment Analysis" da aplicação Web? Se sim	Simulação de créditos	Benchmarking e Comparação de Investimentos	Ferramentas de Gestão de Risco	-	Sim, precisa de mais informações	Não
As páginas disponíveis na aplicação web atendem às suas necessidades?	Sim	Sim	Sim	Não	Sim	Sim
Com base na sua experiência geral	5	5	4	3	3	4
A aplicação forneceu as informações que esperava de forma clara e acessível?	Sim	Sim	Sim	Não	Sim	Sim
Você considera que a aplicação atende às necessidades do público em geral?	Talvez	Sim	Sim	Talvez	Sim	Sim
O que mais gostou na aplicação?	Market Forecasts	Velocidade	personalização e simplicidade	-	Estrutura	Simplicidade e rapidez
O que acha que poderia ser melhorado na aplicação?	Página Investment Analysis	Fundo branco	-	Ter mais informações e conteúdo	-	So um design mais apelativo
Qual dos seguintes recursos você considera mais útil na aplicação?	Market Analysis	Investment analysis	Página Dashboard	Investment analysis	Market Analysis	Market Analysis
Você recomendaria esta aplicação para outras pessoas?	Sim	Sim	Sim	Talvez	Sim	Sim

Figura I.1: Resultados do Questionário sobre a Usabilidade

@is@figure