Enhancing Dublin Rental Price Prediction by Integrating Policy Insights via Large Language Models and Explainable AI

MSc Research Project
Data Analytics

Emmanuella Anita Odiaka
Student ID: X23282819

School of Computing
National College of Ireland

Supervisor: Dr. Abid Yaqoob

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name:	Emmanuella Anita Odiaka		
Student ID:	X23282819		
Programme:	MSc Data Analytics	**Year:**	2024/2025
Module:	Msc Research Practicum 2		
Supervisor:	Dr. Abid Yaqoob		
Submission Due Date:	11/08/2025		
Project Title:	Enhancing Dublin Rental Price Prediction by Integrating Policy Insights via Large Language Models and Explainable AI		
Word Count:	6879	**Page Count:**	28

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: Emmanuella Anita Odiaka

Date: 10/08/2025

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

AI Acknowledgement Supplement

Masters Research Project

Enhancing Dublin Rental Price Prediction by Integrating Policy Insights via Large Language Models and Explainable AI

Your Name/Student Number	Course	Date
X23282819	MSc Data Analytics	09/08/2025

This section is a supplement to the main assignment, to be used if AI was used in any capacity in the creation of your assignment; if you have queries about how to do this, please contact your lecturer. For an example of how to fill these sections out, please click here.

AI Acknowledgment

This section acknowledges the AI tools that were utilized in the process of completing this assignment.

Tool Name	Brief Description	Link to tool
Open AI's GPT-3.5 Turbo	This is a language model that optimizes textual and conversational tasks.	https://platform.openai.com/docs/overview

Description of AI Usage

This section provides a more detailed description of how the AI tools were used in the assignment. It includes information about the prompts given to the AI tool, the responses received, and how these responses were utilized or modified in the assignment. **One table should be used for each tool used**.

Open AI's GPT-3.5 Turbo	
I used this tool as my main LLM to extract policy insights from government files in line with my project's core objectives.	
Sample Prompt	Sample Response
prompt = f""" From the following Irish housing policy excerpt, list specific rules or measures that could influence rental prices in Dublin, and output them as bullet points with short descriptions. Policy excerpt: {policy_text} """	- Rent Pressure Zones (RPZs): Capping annual rent increases at 2% can help prevent drastic spikes in rental prices, providing more stability for tenants. - Affordable Housing Scheme: Delivering 10,000 units per year can increase the supply of housing, potentially leading to lower rental prices due to increased availability.

Evidence of AI Usage

This section includes evidence of significant prompts and responses used or generated through the AI tool. It should provide a clear understanding of the extent to which the AI tool was used in the assignment. Evidence may be attached via screenshots or text.

Additional Evidence:

Prompt template used:

```
# Define the prompt for the OpenAI LLM
prompt_template = """
Please extract the following information from the provided text about Irish housing and rental policies:
- Dates of policy implementation (e.g., "Effective from YYYY-MM-DD")
- Maximum allowed rent increase percentages (e.g., "Rent increases capped at X% per year")
- Specific criteria for rent reviews (e.g., "Rent review permitted every X months under conditions...")
- Geographical areas mentioned in relation to specific policies (e.g., "Rent Pressure Zones include...")
- Details about new housing supply targets or initiatives (e.g., "Target of X new homes by YYYY", "Initiatives include...")

Format the output as a JSON object with keys:
"source_document",
"implementation_date",
"rent_increase_cap",
"rent_review_criteria",
"geographical_areas",
"housing_supply_details". If information is not found, use "N/A".

Text to analyze:
{}
"""
```

Context and role assignment for the LLM:

```
        # Send the text chunk and the prompt to the OpenAI API
        try:
            response = openai.chat.completions.create(
                model="gpt-3.5-turbo", # Or another suitable model
                messages=[
                    {"role": "system", "content": "You are a helpful assistant that extracts specific information from policy documents and provides it in JSON format."},
                    {"role": "user", "content": current_prompt}
                ],
                response_format={"type": "json_object"}
            )
            llm_output = response.choices[0].message.content
            try:
                extracted_info = json.loads(llm_output)
                # Add the source document name to the extracted information
                extracted_info['source_document'] = os.path.basename(file_path)
                extracted_data.append(extracted_info)
            except json.JSONDecodeError:
                print(f"Warning: Could not parse JSON from LLM response for chunk {i+1} in {os.path.basename(file_path)}")
                print(f"LLM Output: {llm_output[:200]}...") # Print a snippet of the problematic output
                # Append a record with "Error" values for this chunk if JSON parsing fails
                extracted_data.append({"source_document": os.path.basename(file_path), "implementation_date": "Error", "rent_increase_cap": "Error", "rent_review_criteria": "Er


        except Exception as api_error:
            print(f"Error calling OpenAI API for chunk {i+1} in {os.path.basename(file_path)}: {api_error}")
            # Append a record with "Error" values for this chunk if an API error occurs
            extracted_data.append({"source_document": os.path.basename(file_path), "implementation_date": "Error", "rent_increase_cap": "Error", "rent_review_criteria": "Error"
```

Enhancing Dublin Rental Price Prediction by Integrating Policy Insights via Large Language Models and Explainable AI

Emmanuella Anita Odiaka
X23282819

Abstract

This body of work proposes the use of AI techniques to determine whether incorporating policy insights obtained with Large Language Models (LLMs), improve the accuracy of rental price forecast models for the Dublin housing market. It will bring together structured rental data with unstructured government policy texts for the construction of a hybrid model. It will employ machine learning, natural language processing components (NLP), and Explainable AI (XAI) software for making precise predictions and deriving valuable insights. Additionally, language models like GPT-4 will be used for extracting key themes, sentiments, and proposals from government texts. It will verify the performance of the model with benchmark regression metrics, fairness checks, and thorough analysis of insights generated. It will build an evidence-based system for enabling decision makers to make clearer, fairer, and realistic housing policy plans.

Keywords: Rental Price Forecasting, Large Language Models, Dublin Housing Market, Explainable AI, Policy Insight Extraction.

1 Introduction

1.1 Background

Affordability of housing and rental market instability continue to be a major headache especially for the younger generation in most urban cities, including Dublin. As the city's population continues to expand and economic change at a breakneck pace, rental prices have varied extensively, often influenced by reasons other than locational or property-specific considerations. Whereas traditional prediction models have been largely based on structured inputs such as square footage, locale, or bed count, they tend to be lacking the complicated, policy-shaped interactions that play a determining role in shaping the housing picture. In contrast, interpretability has been a very significant problem where AI models have been used to determine public policy or monetary concerns. Latest breakthroughs in Explainable AI (XAI) and Large Language Models (LLMs) offer new opportunities to overcome these shortcomings.

New developments in Artificial Intelligence (AI), particularly in machine learning (ML) and natural language processing (NLP) fields, offer potential mediums to overcome this problem. Particularly, the advent of Large Language Models (LLMs) like OpenAI's GPT-4 enabled the mining of useful information from vast amounts of unstructured text data such as government reports, white papers, and housing policy reports. These studies often contain deep but unused contextual information e.g., rent regulation, social housing initiatives, or tax incentives, that can contribute significantly to rental prices and supply-demand dynamics.

By marrying structured data with unstructured policy data, we are able to construct a more comprehensive and robust forecasting system. The crux of this study is to leverage a hybrid AI-based approach - a method that marries traditional property data with policy knowledge extracted by LLMs, to enhance the accuracy of prediction as well as model explainability in Dublin's rental market.

1.2 Research Question and Objectives

The fundamental question that this research sets out to address is:

To what extent does the inclusion of LLM-derived policy insights enhance the accuracy and interpretability of ML models used to predict rental price trends in Dublin?

To overcome this, this study elaborates the following objectives:

- Create a simple model that can predict for rental prices using traditional structured features like size of the property, number of rooms, location etc.
- Create a hybrid model combining these traditional features with policy insights (such as indicators of rent regulation) obtained from textual government documents using an LLM.
- Evaluate both models for predictive accuracy and explainability by using standard evaluation metrics (RMSE, MAE, R^2).
- Use Explainable AI (XAI) methods such as SHAP and LIME to interpret model outputs and assess the contribution of policy features.
- Assess bias and fairness in predictions and possible implications towards housing equity and public policy.

The motivation for this research lies in the ever-changing game of urban housing markets, where political intervention is both decisive and complex. Traditional data-driven approaches do not take these dimensions into consideration, which yield incomplete forecasts. This research aims to enhance predictive performance not only by incorporating semantically rich policy data but also to provide policymakers, housing advocates, and urban planners with actionable insights based on this data.

1.3 Potential Contribution

This study aims to enhance the field of real estate analytics and policy informatics in academia and practice through:

1. A new hybrid model that integrates ML, LLM-based policy extraction, and XAI.

2. Applying policy analysis to AI rental prediction, essentially bridging the gap between data science and governmental administration.
3. Rendering housing price prediction models more transparent, so that policymakers, urban planners, and real estate analysts can utilize them more effectively.
4. Providing a replicable methodology for other urban markets experiencing similar regulatory flux and data complexity.

1.4 Structure of the Document

This paper is divided into seven principal sections other than the Introduction. Section 2 provides a critical review of relevant work in rental price prediction, interpretability in ML, regulatory impact in rental markets and the use of LLMs in policy insights extraction, highligting the gaps addressed in this study. Section 3 outlines the research methodology used to guide data collection, preprocessing, model development and evaluation. Section 4 covers the dydtems design including the data flow and distinction between the baseline and hybrid models. Section 5 digs into the practical implementation of the outliened methodology, covering feature engineering, GPT-4 based feature extraction and model training. Section 6 dives into the model evaluation, results and interpretability analysis with SHAP and LIME. Section 7 closes out the report by reflecting on the contributions of the study, recommendations and further improvements.

2 Related Work

This review discusses the changing landscape of research where artificial intelligence, real estate economics, policy analysis, and explainability come together. The review is structured around four related themes that address the core question of this research directly: Is the interpretability and predictability of AI predictive models for Dublin rental price patterns materially improved by the inclusion of policy insights gleaned from Large Language Model (LLM) analysis of government reports? By combining findings from recent academic studies, this review highlights both the strengths of current methods and the areas that this research aims to address.

2.1 AI in Rental Price Prediction: Techniques and Limitations

Machine learning (ML) and Artificial Intelligence (AI) have been a major player in real estate price forecasting over the last few years with the rise in urban rental markets. These trends make the use of advanced computational methods for modeling spatial, temporal and contextual features useful in enhancing predictions. Other recent progress in this direction includes the use of ensemble learning techniques that combine more than one model to achieve something more robust. Furthermore, scientists started using open data to improve model results. It has helped to make rental price prediction more accurate and reliable, giving more valued results to researchers and SMEs.

On a deeper level, Hu et al. (2023) built a geographically weighted ensemble model (GERPM) considering spatial heterogeneity for Chinese cities Nanjing and Shanghai with improved performance over traditional models. On the other hand, Vivekananda et al. (2024) design a regression meta-model ensemble (Random Forest, Ridge, SVR), where performance is improved by feature selection and parameter optimization. Although both studies demonstrate that ensemble models outperform their constituent individual models, they do not include external policy-related variables, a shortcoming that will apply to the aims of the current study to embed information on regulations in forecasting models.

Random Forest was used by JS et al. (2024) for non-linear rental data analysis with Kriging being optimized for spatial interpolation by Seya et al. (2019)and utilized to assist deep neural networks (DNNs) with large datasets. While these methods can identify patterns, they do not link to the drivers of real-world policy, which this research seeks to examine further. As a whole, these studies confirm the efficacy of AI methods in price prediction; however, they have a relatively narrow focus on internal market data. The current study proposes to scale this method with the introduction of policy signals gleaned through the application of large language models (LLMs).

2.2 Improving Interpretability with Explainable AI

The majority of machine learning (ML) models used in rental price prediction are not explainable. They are complex heavy and do not offer transparent explanations of decisions being made that could generate distrust and ignorance. This is especially important for real estate markets, where rational justification for price forecasts is needed by the players. It is still a great challenge to balance the accuracy of models and their interpretability as in general cases, accurate models have low interpretability whereas models which are less complex in nature have poor predictive capability.

In answering these challenges, Lenaers et al. (2023) employed six explainable AI techniques to Belgian rental predictions, showing how multiple interpretability instruments can yield a concert of insights into that model behavior. On a similar note, Seo et al. (2024) proposed SEE-Net with a shallow neural network to direct deep neural networks' predictions in a manner that renders them explainable. The two approaches equally point to the paramount importance of balancing being technically accurate while still aiding transparency, which is a key stakeholder trust and regulatory concern.

Kästner et al. (2024) take a different route, driven by the life sciences, with a mechanistic interpretability call. Kucklick (2024) proposes a holistic evaluation framework (HIEF) for assessing quality of explanations. Although these models respond to transparency issues, they tend to be missing contextual inputs like policy variables. This study seeks to extend this work by employing policy insights from large language models to generate interpretable and context aware policies that can be beneficial for real world applications like Dublin's.

2.3 Government Policy as a Market Shaper

Government policies and regulations play a key role in determining the supply, demand, and stability of rental markets; as a result, they are an important aspect of the rental market landscape. To learn more about how government actions affect what's happening on the ground in rental markets, researchers could study the impacts of financial regulations, land use laws and short-term rental rules. One can use this profound knowledge to add predictive features in predicting market trends or can be used to support policy decisions.

A prime instance of this is the work by Hur (2024) and Sun et al. (2023) whose findings show the extent to which Korean tax and financial regulations influence the housing market, and their often widely unpredictable results. These findings also illustrate how uncertainty in policy can inject volatility into markets that many of the existing AI models will fail to factor in. Lönnroth et al. (2022) proactive land use planning in Finland, including land banking and flexible developer requirements, also serve to stabilize prices. This research indicates that integrating policy awareness into predictive algorithms is required. This study indicates the need to embed policy awareness into prediction algorithms.

At European level, Wessel et al. (2024) also investigate European short-term rental regulations, showing mixed results in terms of effectiveness, owing to the differences in enforcement. These studies highlight the fragmented and localized nature of policy impacts, which this research will seek to model with Large Language Models trained on a variety of governmental texts. Existing models largely ignore these sorts of interconnections, preventing them from achieving their full forecasting potential, particularly in politically volatile, urban environments.

2.4 LLMs for Policy Insight Extraction

Large Language Models (LLMs), for example, GPT-4 and BERT, are transforming policy analysis as well as economic forecasting through automated extraction and categorization of insight in government reports. Their ability to work on vast amounts of text quickly, accurately, as well as efficiently reduces dependence on traditional methods. In applications in finance, governance, as well as law studies, LLMs provide decision-makers with improved decision-making information. LLMs have enhanced financial projections through identifying market trends as well as policy trends, making them increasingly significant tools for policymakers, analysts, as well as experts in economics across different sectors.

Fatouros et al. (2024) demonstrate this capability in MarketSenseAI using GPT-4 towards improving financial projections through policy-driven data creation. Gudiño-Rosero et al. (2024) demonstrate how LLMs are capable of extracting subtle patterns of citizens' preferences, the essence of regulations' effectiveness, while Meshkin et al. (2024) depict LLM use in automated compliance through data extraction in filings at high accuracy rates.

The study conducted by Choudhary (2025) raised doubts around bias in LLMs outputs and law text imterpretations, suggesting transparency in training data and symbolic reasoning methodologies as ways of addressing performance improvement. Such findings are the foundation for this research in relying on LLMs not just for the summarizing of documents, but as a bridge for the dynamic integration of policy context into rental forecasting models; a gap previously not addressed in AI forecasting.

Chen et al. (2024) takes a different route in that they used LLMs like ChatGPT directly as a prediction engine through few-shot prompting to estimate rental prices of Shanghai Lane houses. Their research shows the feasibility of LLMs in rental price prediction tasks but treats them as black box regressors. In contrast, my work will use these LLMs for extraction and reasoning, not prediction, using them to parse these policy texts and then use existing ML techniques in my prediction. It is also worth mentioning that for more explainability, I will be incorporating the use of SHAP and LIME for the model's explainability to overcome this black box.

2.5 Research Niche and Contribution

For predicting rental prices, most existing known models focus on predictive power by use of market variables such as location, proximity to amenities etc. But they take insufficient account of how government rules can also affect prices. Some researchers, including Trindade Neves et al. (2024) and Ghosh et al. (2023), which found that predictions could be made more accurate by utilizing public information or clustering similar texts; But even these methods neglect how government decisions matter for prices. And, sure, we've become more adept at explaining how AI works, but we haven't taken a good look at how this could aid us in the face of changing policies.

This is where this research identifies a pertinent gap: merging insights from policies, processed with huge language models, and with AI forecasting tools, while also ensuring that all of it is easy to digest. This method helps boost accuracy and clarity in a particularly policy-heavy space. By doing this, we can not only improve our forecasts for rental markets, such as the one in Dublin, but we can also help those in power understand how their decisions impact the market.

3 Research Methodology

The research methodology lays out the order of steps taken to realize the study objectives presented in Section 1. As opposed to following a generic framework like CRISP-DM, this approach was adapted specifically to meet the needs of the project, incorporating traditional rental data with LLM-extracted housing policy insights.

Figure 1 presents the overall workflow, from data collection to model evaluation. The methodology is original in its design, particularly in the way unstructured policy documents were processed and combined with structured features. The use of SHAP and LIME further supplements model transparency, making this approach different from previous work on similar datasets.

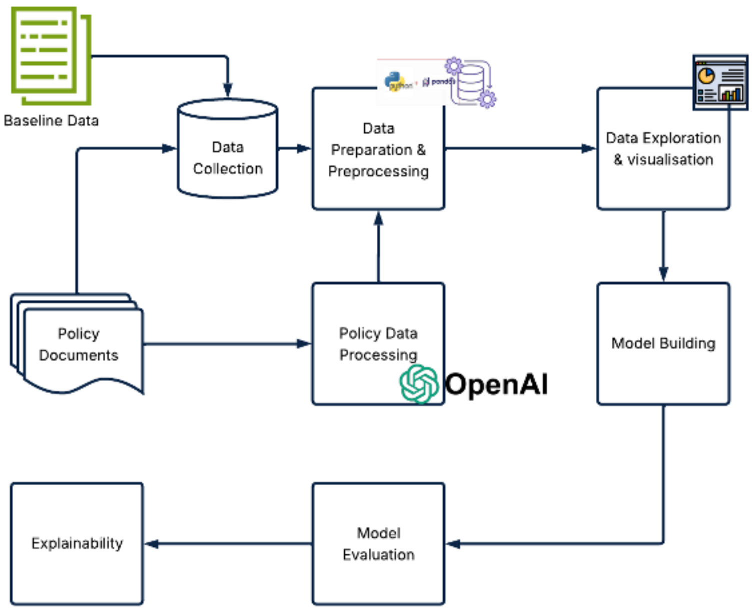

Figure 1: Proposed methodology

3.1 Data Collection and Understanding

Two categories of data were collected:

- **Structured Data:** A Dublin rental data set, which had been previously acquired from the now-defunct Daft.ie API and hosted on Kaggle, was used. Daft.ie is a well known real estate website in Ireland. The data set has ~ 2,718 rows and includes features such as price (target), bedrooms, bathrooms, furnishing status, property type, geographic coordinates, and address[1].
- **Unstructured Policy Data:** A total of 14 housing policy documents (PDFs) were sourced from authorized Irish government and institutional websites. These included the *Affordable Housing Act*, *Residential Tenancies Acts*, *Cost Rental Housing Policy*, and *Review of Rent Pressure Zones*. These documents formed the basis for LLM-based extraction of structured policy insights (e.g., rent caps, affordability provisions, regional zoning)[2].

[1] https://www.kaggle.com/code/shxluo/dublin-rental-prediction
[2] See Appendix for Policy Documents source links

The rental dataset was loaded as a CSV and transformed into a Pandas Dataframe. Policy documents were manually curated and stored locally. These two data streams formed the foundation for the baseline and hybrid models.

Table 1 summarizes the attributes in the structured rental dataset, while Table 2 lists the policy documents and their sources, while source links will be provided in the appendix.

Table 1 Summary of the Dataset's Attributes:

Feature	Data Type	Description
price	String	Rental price of the property.
address	String	Full address of the property including street name, area, and region.
bathroom	Integer	Number of bathrooms in the property.
bedroom	Integer	Number of bedrooms in the property.
furnish	Categorical	Furnishing status of the property.
description	String	Full textual description of the property including features, amenities, surroundings, and agent notes.
property type	Categorical	Type of property. Weather it is a House, Apartment or others.
ID	Integer	Unique identifier for each property listing.
longitude	Float	Geographic coordinate (east-west) of the property location.
latitude	Float	Geographic coordinate (north-south) of the property location.

Table 2 Summary of Policy Data and their sources:

Document name	Source Name	Policy Type	Policy Goal
Affordable Housing Act 2021.pdf	Irish Statue Book	Housing	To establish frameworks for affordable housing purchase and cost rental schemes, promoting long-term affordability in urban areas like Dublin.
Cost rental housing.pdf	Citizens Information	Housing	To introduce and manage cost rental housing targeted at middle-income households with long-term secure rentals priced at least 25% below market value.
Final Review RPZ and Policy Options for Rent Controls in PRS.pdf	The Housing Agency	Rental	To assess and review the effectiveness of Rent Pressure Zones (RPZs) and explore future rent control policy options to balance affordability with housing supply.
Housing Policy: Actions to Deliver Change.pdf	National Economic and Social Council	Housing	To restructure Ireland's housing system with a focus on sustainability, affordability, and equity through coordinated institutional reforms and long-term planning.
housing-for-all-a-new-housing-plan-for-ireland-executive-summary.pdf	Assets.gov.ie	Housing	To provide a roadmap for delivering 300,000 homes by 2030 across all tenure types, including cost rental, affordable, and social housing.

Planning and Development (Housing) and Residential Tenancies Act 2016.pdf	Irish Statue Book	Housing	To facilitate strategic housing developments and introduce provisions for efficient planning and development processes, especially concerning environmental reviews.
Proposed Changes to Irish Residential Rent Controls.pdf	Matheson law firm	Rental	To propose amendments to rent control laws aimed at protecting tenants while maintaining investment incentives in the residential rental market.
Residential Policy Covid.pdf	ESRI Research Series	Rental	To safeguard tenants during the COVID-19 crisis through eviction bans, rent freeze policies, and financial supports aimed at preventing homelessness.
Residential Tenancies (No. 2) Act 2021.pdf	Irish Statue Book	Rental	To implement further amendments to residential tenancy laws, particularly concerning eviction protections and rent review procedures during emergency periods.
RESIDENTIAL TENANCIES ACT 2004 REVISED Updated to 15 May 2025.pdf	Irish Statue Book	Rental	To define legal relationships between tenants and landlords, govern rent increases, and enforce compliance within the private rental sector.
Residential Tenancies Act 2021.pdf	Irish Statue Book	Rental	To enhance protections in tenancy law, extend rent caps, and provide further safeguards during public health and economic emergencies.
Residential Tenancies Bill 2021.pdf	House of the Oireachtas	Rental	To further modernize tenancy law, particularly around notices of termination, landlord obligations, and clarity in rent setting and review.
Revised Residential Tenancies and Valuation Act 2020.pdf	Irish Statue Book	Rental	To revise tenancy regulations, reinforce tenant rights, and adjust rent control frameworks within updated residential valuation and tenancy law.
The Impact of Cost Rental Housing-5.pdf	The Housing Agency	Housing	To evaluate the impact of cost rental housing on affordability, stability, and residents' well-being.

3.2 Data Preparation and Processing

An exhaustive data preparation was done:

- **Structured Data**: Missing values were handled (e.g., default imputation for furnishing), and outliers (e.g., outlier values in price per bedroom) were removed by the IQR technique. Address parsing was employed to obtain 'city' and 'county' features, and one-hot encoding was performed on categorical variables.
- **Policy Data**: Content pulled by LLM, using a defined prompt was returned in JSON format. Error handling for poor responses was included in parsing. Location names were also normalized, and Dublin-related entries were matched using keyword tagging. Full preprocessing is described in more detail in Section 5.

3.3 Data Exploration and Visualisation

Exploratory Data Analysis (EDA) was conducted on the cleaned rental dataset. Visualizations included histograms, boxplots, heatmaps, and count plots to reveal key patterns:

- Distribution of rental prices
- Frequency listing by number of bedrooms and bathrooms
- Price and location/type feature correlations

These informed feature engineering and model decisions, as discussed in Section 5.

3.4 Policy Data Processing and Feature Engineering

After feature extraction process, feature engineering was the essential next step needed to create the hybrid model:

- Dictionaries containing geographical details were standardized to standard Dublin locations and used to identify related policies
- Boolean features were created to flag policy entries matching the standardized locations.
- Policy features containing rent pressure zones within these locations was processed

The policy outputs from LLM were summed up and converted into numerical features, and these features were appended to the structured rental data. The structured and policy-derived features were combined to build the hybrid model (see Section 5).

3.5 Data Splitting and Scaling

Following preprocessing, the two datasets were divided into **training and test sets (80:20 split)**. A **StandardScaler** was used to normalize feature values in both models so that model evaluation was consistent and comparable.

3.6 Model Building

Two predictive models were developed:

- The **Baseline Model** utilized only structured rental data.
- The **Hybrid Model** included policy-derived features learned by GPT-3.5.

A **Random Forest Regressor** was selected due to its high resilience and strong performance in regression tasks as indicated by V. Mirg (2022). Hyperparameters were further adjusted using **GridSearchCV** for optimal accuracy. This two-stage modelling approach allowed for a comparative evaluation of baseline and policy-augmented models.

3.7 Evaluation and Explainability

Model performance was assessed using standard regression metrics: **Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and R-squared (R^2)**. Other, explainability tools like SHAP and **LIME** were used to analyze the impact of every feature on predictions. it was particularly critical in the case of evaluating the impact of policy variables in the hybrid model. The results and the plot of feature importance are included in Section 6.

4 Design Specification

In this section, the sequential process followed in order to accomplish the research objectives is explained. Figure 2 shows the system architecture with the data collection pipeline to evaluation implemented in Python. The architecture is divided into four major layers: **Data Ingestion**, **Data Preparation**, **Modelling & Evaluation**, and **Output & Visualisation**.

Data Ingestion Layer: Rental data was imported from a CSV file, while housing policy documents were obtained from official sources. The data ingestion involved web scraping and download of relevant data, followed by cleaning to eliminate inconsistencies, handling missing values, and creating initial features for modelling.

Data Preparation Layer: There are two concurrent tracks forming this layer. The **Baseline Model** focused on processing the structured rental listing attributes through exploratory data analysis, feature engineering, and categorical data encoding. The **Policy Model** applied GPT-3.5 to derive structured policy features (rent caps and Rent Pressure Zone (RPZ) indicators) from unstructured documents. The policy features were then engineered and merged with the rental dataset to form the hybrid dataset.

Modelling & Evaluation Layer: Each stream of data was split into test and training sets as earlier mentioned. Both models were trained using the same regression algorithm, after which their performance were compared on the same metrics for the sake of ensuring consistency and comparability of models.

Output & Visualisation Layer: The obtained results were visualised to show prediction accuracy, the impact of policy features, and model interpretability. SHAP and LIME were applied to highlight how different rental and policy features influenced the predictions, to address transparency in the decision-making process.

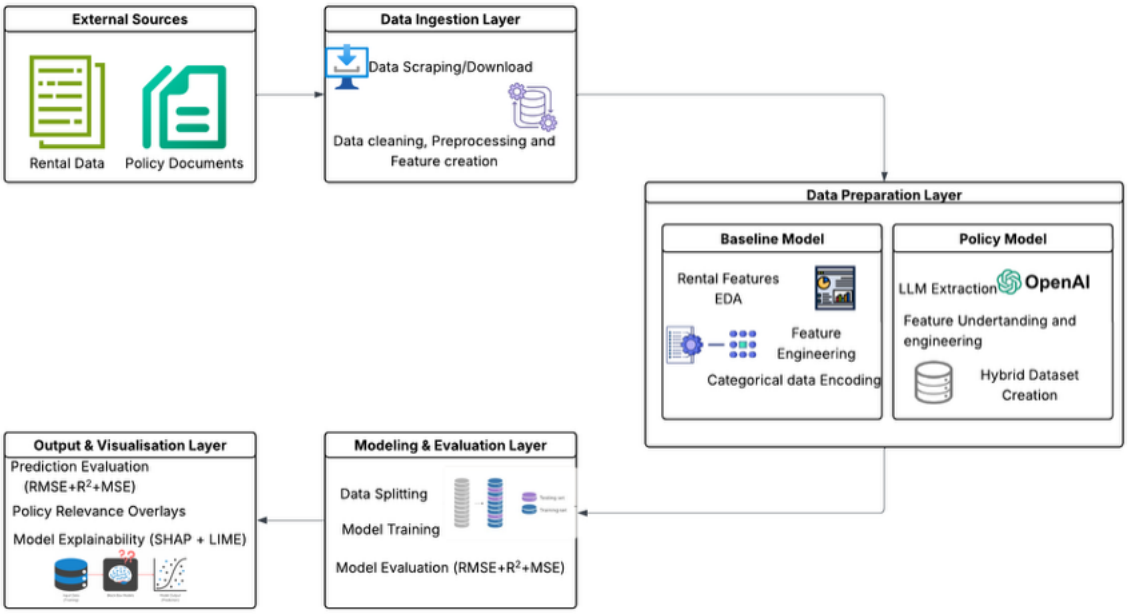

Figure 2: Design Specification

5 Implementation

This section takes a close-up view of the technical rollout of the project as done in the Jupyter notebook environment. It lays down the processing, transformation, and modelling stages for the baseline and hybrid approaches based on the methodology detailed in Section 3. The baseline model is constructed based on traditional rental listing features, with the hybrid model incorporating additional features gleaned from policy documents as per a large language model (LLM). Each phase from data ingestion to model testing is detailed below with a focus on how the fusion of policy data impacted the modelling pipeline.

5.1 Data Understanding

Rental data Loading and Initial Exploration:

The rental dataset (daft_v_2.csv) was loaded into a Pandas Dataframe after mounting and accessing the Google Drive file. Preliminary checks were performed to examine shape and content of data, first few rows were inspected, checking data types, numeric attribute summary, and missing value. The 'description' and 'furnish' columns were identified with high missing data, handled later in the preprocessing stage. Finally, summary statistics was printed to assess statistical distributions of the features, which formed the basis for further cleaning.

Policy Documents Exploration:

The folder housing the gathered fourteen housing policy PDFs was accessed from the Google Drive . To guarantee access and compatibility with subsequent text chunking and LLM query, two sample documents were selected and deconstructed using PyPDF2.PdfReader. A loop was used to read every page, and the first 1000 characters were printed to obtain information about document formatting.

5.2 Data Preparation

Rental Data Preprocessing (Baseline):

- **Data Cleaning***:* The whole rental data was passed through a full data cleaning operation. The 'description' column was entirely discarded since it was sparse and unstructured. Just the 'furnish' column was retained and imputed with a default string. Text prices (containing symbols and time units) were converted to monthly numeric format by a special cleaning function. Rows for which the price formats failed to convert were eliminated.
- **EDA and Summary***:* An EDA confirmed data patterns and distribution were as anticipated. Further EDA details are presented in Section 5.3.
- **Outlier Detection***:* A new feature, 'price_per_bedroom' was created for outlier detection to normalise the dataset for property size. This was done in an effort to prevent incorrectly flagging expensive properties as outliers. The Identified outliers were removed using the IQR technique for model robustness.
- **Feature Engineering***:* This entailed extracting geographic data from 'address' column using a custom function to obtain 'county' and 'city' features. Next, categorical features (‘furnish', 'property type', 'county', and 'city'.) were One-hot encoded, generating the new cleaned data frame with 2,717 rows and 72 features in the dataset. This dataset served as the input data for the baseline model, as can be seen from Figure 4.

Policy Data Preparation (for Hybrid Model):

Section 5.4 elaborates further on the processing that has been carried out to achieve the hybrid dataset obtained after policy information extraction using Open AI-s GPT-3.5 Turbo as the LLM. This section elaborates on steps taken in preparation:

- **OpenAI API Setup:** As OpenAI has a strong natural language processing capability, OpenAI was utilized as the LLM in this project. The OpenAI API key was securely acquired by saving and setting the pass key as an environment variable.
- **Text Extraction and Querying***:* Policy documents were scanned using PyPDF2, which read all the pages of each file and combined these outputs to representative strings. Following this, the string was then segmented into chunks of 10000 maximum characters, which was then sent along with a well-crafted prompt template (see Figures 9 and 10) to the GPT-3.5-turbo model.
- **JSON Parsing and Cleanup:** Data received was returned in JSON format, with error checking in place to bypass or correct malformed entries. Outputs were then written into a new policy refined dataframe.
- **Data Saving and Reloading:** This refined Policy dataset was saved into CSV and loaded back to verify data integrity.

- **Feature Engineering from Policy Data:** As the resulting data stored relevant details in dictionaries, further standardisation (format normalising, whitespace stripping, etc) was carried out. Policies speaking strictly on Dublin were filtered out using keyword-based filtering, yielding the creation of five major policy indicators, aggregates of which were merged into the rental dataset, yielding the df_hybrid_refined dataset. A description of the resulting dataframe is displayed in figure 4**.**

```
cleaned_df_encoded.info()

<class 'pandas.core.frame.DataFrame'>
Index: 2519 entries, 1 to 2717
Data columns (total 75 columns):
 #   Column                            Non-Null Count  Dtype
---  ------                            --------------  -----
 0   price                             2519 non-null   object
 1   address                           2519 non-null   object
 2   bathroom                          2519 non-null   int64
 3   bedroom                           2519 non-null   int64
 4   ID                                2519 non-null   int64
 5   longitude                         2519 non-null   float64
 6   latitude                          2519 non-null   float64
 7   price_per_bedroom                 2519 non-null   object
 8   furnish_Furnished                 2519 non-null   bool
 9   furnish_Furnished or unfurnished  2519 non-null   bool
 10  furnish_Unfurnished               2519 non-null   bool
 11  property type_Apartment           2519 non-null   bool
 12  property type_Flat                2519 non-null   bool
 13  property type_House               2519 non-null   bool
 14  property type_Studio              2519 non-null   bool
 15  county_Adamstown                  2519 non-null   bool
 16  county_Balbriggan                 2519 non-null   bool
 17  county_Ballsbridge                2519 non-null   bool
 18  county_Ballyboughal               2519 non-null   bool
 19  county_Blackrock                  2519 non-null   bool
 20  county_Booterstown                2519 non-null   bool
 21  county_Citywest                   2519 non-null   bool
 22  county_Dalkey                     2519 non-null   bool
 23  county_Deansgrange                2519 non-null   bool
 24  county_Donabate                   2519 non-null   bool
 25  county_Dublin 1                   2519 non-null   bool
 26  county_Dublin 10                  2519 non-null   bool
 27  county_Dublin 11                  2519 non-null   bool
 28  county_Dublin 12                  2519 non-null   bool
 29  county_Dublin 13                  2519 non-null   bool
 30  county_Dublin 14                  2519 non-null   bool
 31  county_Dublin 15                  2519 non-null   bool
 32  county_Dublin 16                  2519 non-null   bool
 33  county_Dublin 17                  2519 non-null   bool
 34  county_Dublin 18                  2519 non-null   bool
 35  county_Dublin 2                   2519 non-null   bool
 36  county_Dublin 20                  2519 non-null   bool
 37  county_Dublin 22                  2519 non-null   bool
 38  county_Dublin 24                  2519 non-null   bool
 39  county_Dublin 3                   2519 non-null   bool
```

Figure 3: Cleaned Baseline Dataframe

```
 67  county_Swords                     10 non-null     bool
 68  city_Co. Dublin                   10 non-null     bool
 69  city_Dublin City Centre           10 non-null     bool
 70  city_North Co. Dublin             10 non-null     bool
 71  city_North Dublin City            10 non-null     bool
 72  city_South Co. Dublin             10 non-null     bool
 73  city_South Dublin City            10 non-null     bool
 74  city_West Co. Dublin              10 non-null     bool
 75  policy_rent_increase_mentions     10 non-null     int64
 76  policy_rent_review_mentions       10 non-null     int64
 77  policy_housing_supply_mentions    10 non-null     int64
 78  policy_dublin_related_mentions    10 non-null     int64
 79  policy_rpz_mentions               10 non-null     int64
dtypes: bool(67), float64(2), int64(8), object(3)
memory usage: 1.7+ KB

==================================================

Info for the last 20 rows:
<class 'pandas.core.frame.DataFrame'>
Index: 20 entries, 2663 to 2717
Data columns (total 80 columns):
 #   Column                            Non-Null Count  Dtype
---  ------                            --------------  -----
```

Figure 4: Summary of Hybrid Dataframe showing integrated policy features

5.3 Exploratory Data Analysis

EDA was conducted on the cleaned rent dataframe to observe trends and identify feature relationships:

- **Visualisation 1:** Figure 5 aggregates the prices of the properties and finds that most of the properties are being rented out for a price of €1000 – €3500, with most of the properties valued around €2,000. The right-skewed distribution finds that there are properties over €30,000. These can be attributed to high-end properties in high-end locations like Glenageary or Killiney, which are causing the upper levels of this attribute.
- **Visualisation 2:** The mean monthly rent based on property type in Dublin can be identified from Figure 6. From the graph, it can be understood that houses are the most costly with. Flats and studios are less costly, where average prices range from €1,380 - €1,130 respectively. This is a tendency that shows how property size and facilities influence the rentals, where larger properties like houses with more rooms tend to charge higher tenants like family units and smaller units aim for lower-budget tenants like students.

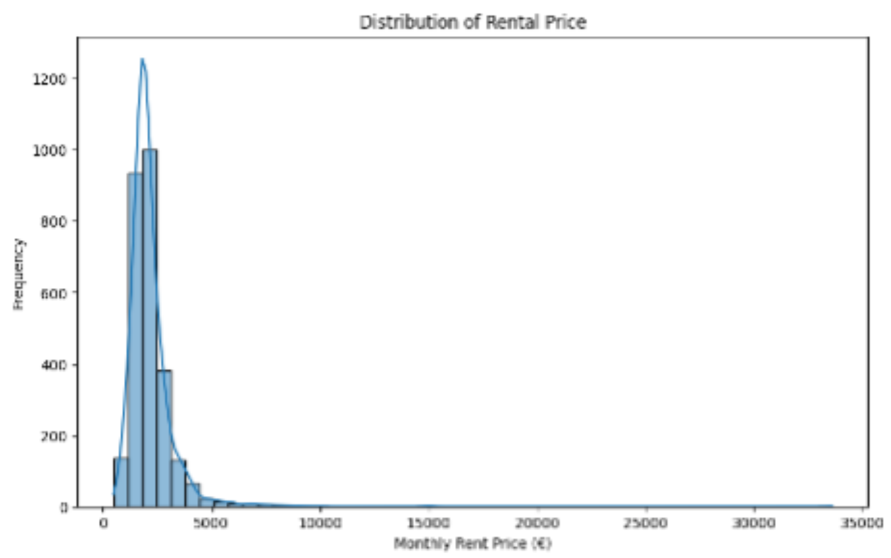

Figure 5: Distribution plot of Rental Price

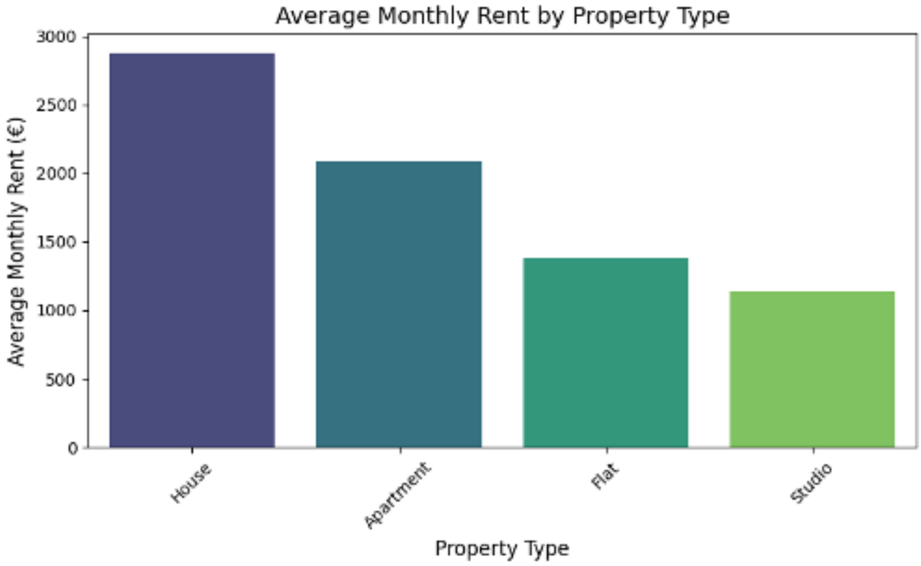

Figure 6: Average rent per Property Type

- **Visualisation 3:** Figure 7 presents the top 10 most expensive counties in Dublin, calculated on the average monthly rent at the time of the original data collection. Glenageary and Killiney are obviously the highest priced, with rentals of more than 5,300 a month. Upscale suburbs such as Kilmacud and Dalkey also have relatively high average rentals between 3,500 and 4,000. Lastly, beachfront suburbs such as Monkstown and Ballsbridge have above-average prices, as shown in the chart. What this implies is that Southern Dublin suburbs maintain the high-prices rental market due to their control of the seaside suburbs and established upscale suburbs, which drives prices upward.

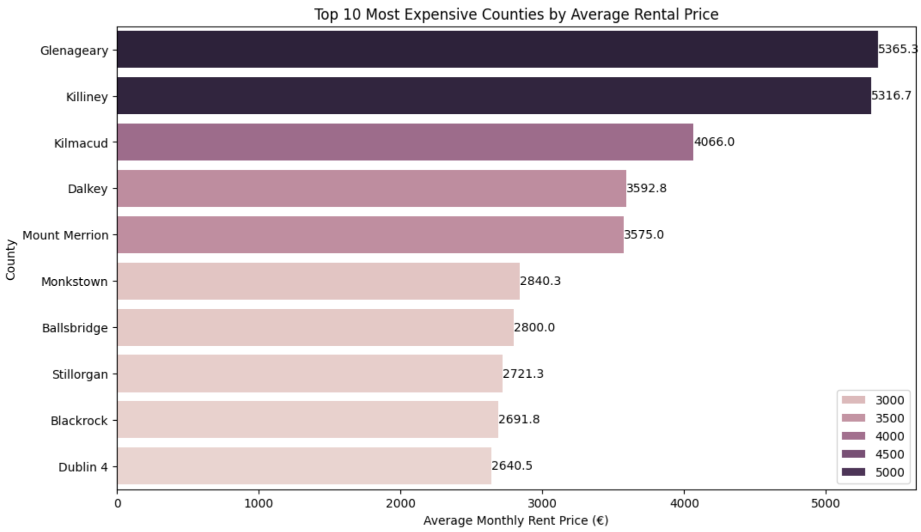

Figure 7: Top Counties by Average Rent Price

- **Visualisation 4:** Figure 8 shows a geocoded map, which displays all the regions of Dublin where the property listings appear. The folium library was used to create the map in order to efficiently map address coordinates to geo coordinates.

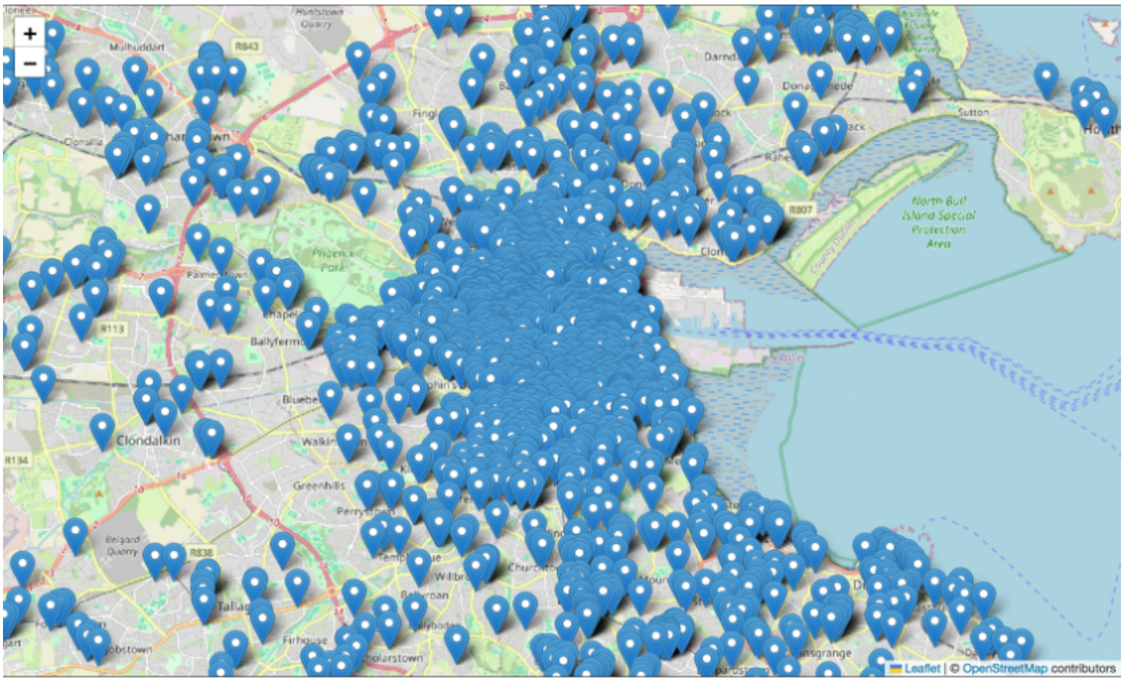

Figure 8: Geographical Distribution of Rental Listings

5.4 Policy Data Processing and Feature Engineering

There were two primary methods of obtaining policy information.

Initial Prompt Strategy: The first strategy used a general prompt (template in Figure 9) to get overall policy information like start dates, rent levels, review rules, areas, and housing goals. The GPT-3.5 model was prompted to return this data in a structured JSON format. While this approach was partially successful and yielded a DataFrame (policy_df), the geographical_areas field was either too ambiguous or not formatted uniformly. Outputs varied from generic zone headers, incomplete place names, and inconsistent output types (lists, dictionaries), making it impossible to consistently link policies to specific locations such as Dublin.

```
# Define the prompt for the OpenAI LLM
prompt_template = """
Please extract the following information from the provided text about Irish housing and rental policies:
- Dates of policy implementation (e.g., "Effective from YYYY-MM-DD")
- Maximum allowed rent increase percentages (e.g., "Rent increases capped at X% per year")
- Specific criteria for rent reviews (e.g., "Rent review permitted every X months under conditions...")
- Geographical areas mentioned in relation to specific policies (e.g., "Rent Pressure Zones include...")
- Details about new housing supply targets or initiatives (e.g., "Target of X new homes by YYYY", "Initiatives include...")

Format the output as a JSON object with keys:
"source_document",
"implementation_date",
"rent_increase_cap",
"rent_review_criteria",
"geographical_areas",
"housing_supply_details". If information is not found, use "N/A".

Text to analyze:
{}
"""
```

Figure 9: Initial LLM Prompt

Refining by Geographical Context: Recognising the central role of location in rental pricing, a revised prompt (see Figure 10) was created to target geographical information specifically "relevant to Dublin and its sub-areas." This new prompt requested more precise structured data, including each area's name and type. Responses were much clearer, and outputs were collected in a new DataFrame (policy_df_refined. Structure of dataframe is shown in Figure 11).

```
# Refined prompt for the OpenAI LLM
refined_prompt_template = """
Please extract the following specific and location-aware information from the provided Irish housing
 and rental policy text, focusing on details relevant to Dublin and its sub-areas:
- Dates of policy implementation (e.g., "Effective from YYYY-MM-DD")
- Maximum allowed rent increase percentages (e.g., "Rent increases capped at X% per year")
- Specific criteria for rent reviews (e.g., "Rent review permitted every X months under conditions...")
- **Detailed Geographical Areas mentioned:** List specific locations, focusing on Dublin postal codes,
sub-areas (like Ballsbridge, Dalkey, city council areas),
or broader regions if specific areas are not mentioned.
For each area, specify the *type* of area if mentioned
(e.g., "Rent Pressure Zone", "County", "City Council Area").
Explicitly state if a policy is noted as applying *specifically* to Dublin or parts of Dublin.
- Details about new housing supply targets or initiatives, noting if they are specific to Dublin.

Format the output as a JSON object with keys:
"source_document",
"implementation_date",
"rent_increase_cap",
"rent_review_criteria",
"geographical_areas_detailed",
"housing_supply_details".
For "geographical_areas_detailed", provide a list of objects, each with keys "area_name" and "area_type"
(use "N/A" if type is not specified). If information is not found or not relevant to Dublin,
use "N/A" for the respective key or an empty list for "geographical_areas_detailed".

Text to analyze:
{}
"""
```

Figure 10: Refined LLM Prompt

```
Structure of the Aggregated Refined Policy Information DataFrame:
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 180 entries, 0 to 179
Data columns (total 6 columns):
 #   Column                       Non-Null Count  Dtype
---  ------                       --------------  -----
 0   source_document              180 non-null    object
 1   implementation_date          180 non-null    object
 2   rent_increase_cap            180 non-null    object
 3   rent_review_criteria         180 non-null    object
 4   geographical_areas_detailed  180 non-null    object
 5   housing_supply_details       180 non-null    object
dtypes: object(6)
memory usage: 8.6+ KB
```

Figure 11: Structure of Extracted Policies Dataframe

Standardisation and Linking:

- Geographical references field from the LLM output was processed (i.e trimmed, deduplicated and lowercased) to aid accurate matching with geographical entities present in the cleaned rental dataset.
- Clean lists of locations were compared against unique county and city names that existed in the rent data to determine whether any of the policy records were associated with Dublin.
- A simple binary tag was created to flag these matches.
- For each document, policy mentions were counted and summarised under five key indicators and aggregated into numeric features:
  - policy_rent_increase_mentions
  - policy_rent_review_mentions

- policy_housing_supply_mentions
- policy_dublin_related_mentions
- policy_rpz_mentions

These features were then merged into the cleaned rental dataset, producing the final enriched version (df_hybrid_refined). The merge ensured that every property record was paired with contextual policy information relevant to its area, forming the basis of the hybrid model.

5.5 Data Splitting and Scaling

Baseline Model: The baseline dataset was separated into training and test sets, ensuring a 80:20 ratio was kept Only structured features such as bedroom, bathroom, property type, and location-based variables were included. A StandardScaler was applied to normalize features before training.

Hybrid Model: The same data splitting and scaling procedure was applied to the hybrid dataset. However, the feature set was expanded to include the five policy-related features. Before splitting, columns like address, ID, and other non-informative identifiers were dropped. The data was then passed to the same StandardScaler for consistency.

5.6 Model Training and Implementation

In section 2.3, multiple authors discussed traditional and hybrid ensemble learning techniques, such as Random Forest. Random Forest is a technique that uses a strategy called bagging. It consists of generating multiple decision trees using repeated random samples of the data to make predictions and averages the outcomes, for accuracy.

Two key parameters: n_estimators (i.e., the number of trees) and max_depth were optimized through GridSearchCV on a five-fold cross-validation basis. The above procedure examined all possible combinations of parameters and selected those having the fewest errors when making rent predictions in Dublin.

To allow for a controlled comparison between the effect of traditional features and the added policy signals, the scaled training data subset of baseline and hybrid models were fitted on the regressor, along with the optimal hyperparameters and trained.

6 Evaluation

This section evaluates the two models developed in this study: the Baseline Model and the Hybrid Model. The evaluation is against already established measures; Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and R-squared (R^2) that provide information on the accuracy and reliability of the predictions of the models.

6.1 Model Evaluation

Baseline Model: The baseline model was trained on the structured rental dataset with traditional property features such as number of bedrooms, number of bathrooms, furnishing status, property type, and geographic location. As established in the implementation section, the trained Random Forest model was used to evaluate the test set.

- **Mean Absolute Error (MAE):** 25.93
- **Mean Squared Error (MSE):** 20278.33
- **Root Mean Squared Error (RMSE):** 142.40
- **R-squared (R^2):** 0.98

These results show that the model effectively predict rental prices with high accuracy. The low RMSE value and high R^2 value show that the traditional features alone were sufficient to explain a large portion of the observed difference in rental prices.

Hybrid Model: The hybrid model extended the baseline by adding five new features extracted from rent-related policy documents such as rent cap mentions, RPZ mentions and housing supply provision mentions, as they affect Dublin. The hybrid model was trained using the same Regressor and hyperparameters as the baseline for a comparison purposes.

- **Mean Absolute Error (MAE):** 25.53
- **Mean Squared Error (MSE):** 21102.70
- **Root Mean Squared Error (RMSE):** 142.40
- **R-squared (R^2):** 0.98

Despite the addition of new features, both models performed almost identically in terms of metrics. The R^2 remained 0.98, and the variations in MAE and RMSE were not significant. These results show that the added policy features, in their current aggregated state, had a near insignificant influence on increasing prediction accuracy.

6.2 Feature Importance

Feature importance analysis supports the evaluation's results by revealing that traditional property features dominated the model's predictions. The top three predictors were:

- **price_per_bedroom** (0.476)
- **bedroom** (0.308)
- **bathroom** (0.176)

Figure 12 shows a side by side comparison of the feature importance analysis carried out on both models, further supporting the findings.

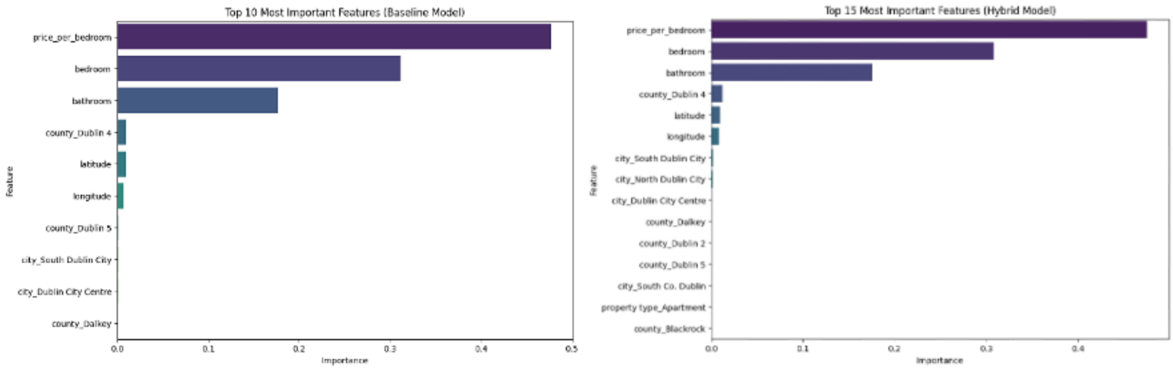

Figure 12: Feature Importance

Location features such as county_Dublin 4 and city_South Dublin City also contributed to the model, but to much smaller degrees. Interestingly, all five policy-related features had zero importance (see Figure 13) on the Random Forest model, which means they were not used at all in making predictions.

```
Importance of Policy Features:
- policy_rent_increase_mentions: 0.000000
- policy_rent_review_mentions: 0.000000
- policy_housing_supply_mentions: 0.000000
- policy_dublin_related_mentions: 0.000000
- policy_rpz_mentions: 0.000000
```

Figure 13: Policy Feature Importance

6.3 Explainability Results

SHAP Analysis: The SHAP summary plots confirmed earlier results by reasserting the global importance of traditional features and showed that policy-related features had no influence on the output.

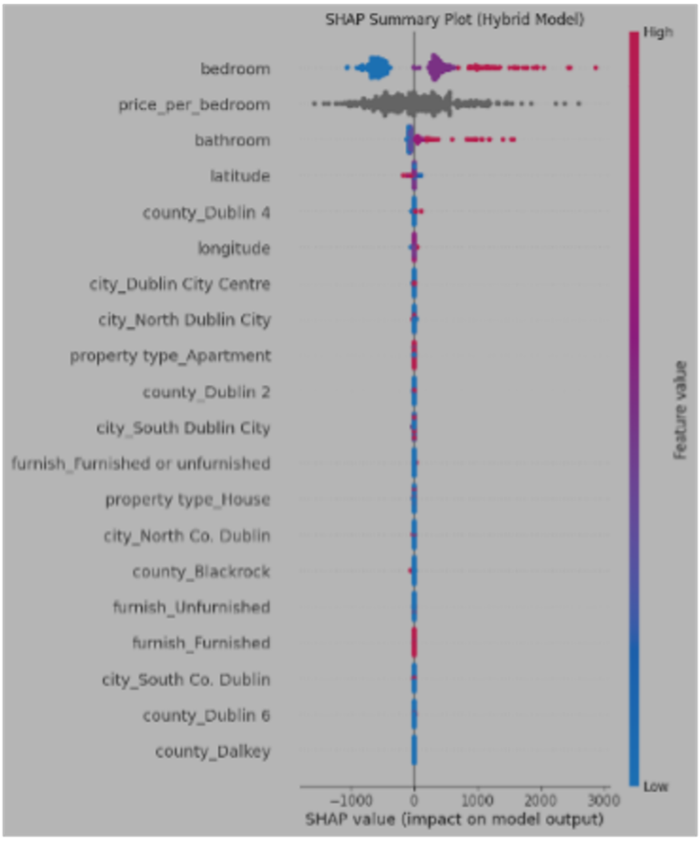

Figure 14: SHAP Summary Plot

LIME Analysis: LIME provides insights into why a model made a specific prediction for a single instance. The explanation lists the most significant features which had the most positive or negative impact for that singular example. From the sample given to explainer, we see a similar trend following previous results in that the policy features made no contribution to the influence of the predicted value.

Figure 15: LIME Explanation Plot

7 Conclusion and Future Work

This study set out to determine whether the inclusion of policy derived insights from governmental housing information could improve the predictive capabilities of rental price forecasting in Dublin. To achieve this five core objectives were derived: developing a baseline rental price prediction model, constructing a hybrid version augmented with policy insights obtained from LLM extraction, assessing both models using common metrics, result interpretation with Explainable AI, and examining fairness assumptions in the prediction made by the models..

7.1 Achievement of Research Objectives

- **Objective 1 – Baseline Model Using Structured Features:** Traditional features like number of bedrooms, bathrooms, and location features, presented in a structured format formed the basis for constructing the baseline model. This model performed strongly, with an R^2 score of 0.98 and low MAE and RMSE values, validating that core property characteristics are excellent predictors of rental prices in Dublin.

- **Objective 2 – Hybrid Model with Policy Insights via LLM:** Policy related features as they affect the Dublin rental market like Rent Pressure Zones or rent control were engineered out of the LLM extracted Policy documents and merged with the structured data to form the backbone of the hybrid model. While technically successful, these features did not provide any significant impact on the model's performance, uncovering flaws with how the policy data was represented.
- **Objective 3 – Model Evaluation Using Standard Metrics:** Both models underwent thorough evaluation through the established metrices. As earlier observed, the performance of both models showed nearly the same results, reinforcing that the added policy features, by aggregates of keyword mentions, did not improve predictive power.
- **Objective 4 – Explainability with SHAP and LIME:** SHAP and LIME analyses were conducted to interpret both global and individual feature contributions. The results of which consistently pinpointed features like price per bedroom, bedroom count, and location as the most significant. In contrast, the policy indicators displayed zero SHAP values and minimal presence in LIME explanations, thus confirming their insignificant influence.
- **Objective 5 – Fairness and Policy Implications:** Even though there were no indications of bias in the predictions, the fact that the policy features contributed no value raises concern in terms of how well the current pipeline capture real-world policy impact. Failure to capture policy effects in model predictions would be problematic if such models are to be applied to policy-critical domains such as rent control or housing equity.

7.2 Why Policies Fell Short

In spite of using meticulously crafted prompting and policy documents processing, the reason why key-word aggregation was used to curate policy-rich features lost their impact. The reasons are mainly:

- **Low variation:** Policy mentions count is more or less the same for most of the properties, so it does not add any value towards price prediction.
- **Superficial detailing:** Deciding the merging based on frequency of mentions ignores important details that can actually influence rental prices.
- **Prevalence of traditional property features:** Presence of features directly linked to rental price like number of bedrooms can overshadow policy features in general.
- **Lack of direct linkage:** The way the policy details were linked with individual rentals was not well informed and thus influenced the actual impact.

7.3 Future Work and Recommendations:

To address these limitations and move towards capturing the true real-world effect of policy on rental markets, the following directions are recommended:

- **Extract Specific Policy Details:** Attempt to extract concrete values (e.g percentage rent increases, rent review periods) instead of mere number of keyword mentions.
- **Create Location-Based Features:** Develop features that bind particular information (e.g., having been in a specific zone with a particular rent limit) to specific properties as a function of address.
- **Analyze Time-based Variables:** Include the time variable if available to allow the analysis of the impact of policies after they have been put into effect.
- **Run Other Models:** Pass the refined hybrid set through additional machine learning models to determine whether they will be more sensitive to the policy data.

Conclusion:

This study fully shows that traditional property attributes reign supreme in terms of reliability of predictors for rental prices in Dublin. While no improvement was picked up from the hybrid model, this study demonstrates a successful process of incorporating unstructured policy context with structured rent data via LLM, presenting a promising area of future research.

Making improvement on how policy information is extracted, described and connected to rental property data in future iterations could better equip real estate regulators and decision makers with a fair, clear and policy-aware prediction tool.

Appendix

Table 3: Policy Document Names and Source Links

Document name	Source Link
Final Review RPZ and Policy Options for Rent Controls in PRS.pdf	https://www.housingagency.ie/publications/review-rent-pressure-zones-report
Proposed Changes to Irish Residential Rent Controls.pdf	https://www.matheson.com/insights/detail/proposed-changes-to-irish-residential-rent-controls
Residential Tenancies (No. 2) Act 2021.pdf	https://www.irishstatutebook.ie/eli/2021/act/17/enacted/en/html
Planning and Development (Housing) and Residential Tenancies Act 2016.pdf	https://www.irishstatutebook.ie/eli/2016/act/17/enacted/en/html
housing-for-all-a-new-housing-plan-for-ireland-executive-summary.pdf	https://www.gov.ie/en/department-of-housing-local-government-and-heritage/publications/housing-for-all-a-new-housing-plan-for-ireland/
Revised Residential Tenancies and Valuation Act 2020.pdf	https://www.irishstatutebook.ie/eli/2020/act/7/enacted/en/print.html
Cost rental housing.pdf	https://www.citizensinformation.ie/en/housing/renting-a-home/help-with-renting/cost-rental-housing/
Residential Tenancies Act 2021.pdf	https://www.irishstatutebook.ie/eli/2021/act/39/enacted/en/print.html
The Impact of Cost Rental Housing-5.pdf	https://www.housingagency.ie/sites/default/files/The%20Impact%20of%20Cost%20Rental%20Housing.pdf
RESIDENTIAL TENANCIES ACT 2004 REVISED Updated to 15 May 2025.pdf	https://www.irishstatutebook.ie/eli/2025/act/5/enacted/en/print.html
Affordable Housing Act 2021.pdf	https://www.irishstatutebook.ie/eli/2021/act/25/enacted/en/print.html
Residential Tenancies Bill 2021.pdf	https://data.oireachtas.ie/ie/oireachtas/libraryResearch/2021/2021-04-06_bill-digest-residential-tenancies-bill-2021_en.pdf
Housing Policy: Actions to Deliver Change.pdf	https://www.nesc.ie/publications/housing-policy-actions-to-deliver-change/
Residential Policy Covid.pdf	https://www.esri.ie/publications/exploring-the-short-run-implications-of-the-covid-19-pandemic-on-affordability-in-the

References

Hur, S.-K. (2024). Assessing Policy Effects on the Interacting Housing, Chonsei, and Wolse Markets. (n.d.). Scilit. Retrieved August 10, 2025, from https://www.scilit.com/publications/f709bfe7c21b992f9e54e2e384c6a64a

Chen, T., & Si, S. (2024). *Predicting Rental Price of Lane Houses in Shanghai with Machine Learning Methods and Large Language Models* (No. arXiv:2405.17505). arXiv. https://doi.org/10.48550/arXiv.2405.17505

Choudhary, T. (2025). Political Bias in Large Language Models: A Comparative Analysis of ChatGPT-4, Perplexity, Google Gemini, and Claude. *IEEE Access*, *13*, 11341–11379. https://doi.org/10.1109/ACCESS.2024.3523764

Fatouros, G., Metaxas, K., Soldatos, J., & Kyriazis, D. (2024). Can Large Language Models beat wall street? Evaluating GPT-4's impact on financial decision-making with MarketSenseAI. *Neural Computing and Applications*. https://doi.org/10.1007/s00521-024-10613-4

Ghosh, I., Sanyal, M. K., & Pamucar, D. (2023). Modelling Predictability of Airbnb Rental Prices in Post COVID-19 Regime: An Integrated Framework of Transfer Learning, PSO-Based Ensemble Machine Learning and Explainable AI. *International Journal of Information Technology & Decision Making*, *22*(03), 917–955. https://doi.org/10.1142/S0219622022500602

Gudiño-Rosero, J., Grandi, U., & Hidalgo, C. A. (n.d.). Large Language Models (LLMs) as Agents for Augmented Democracy. *Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences*. https://doi.org/10.1098/rsta

Hu, G., & Tang, Y. (2023). GERPM: A Geographically Weighted Stacking Ensemble Learning-Based Urban Residential Rents Prediction Model. *Mathematics*, *11*(14), 3160. https://doi.org/10.3390/math11143160

JS, N., Sravya, S. G., & Tharun, K. (2024). Housing Market Intelligence: Data Science for Rental Price Forecasting. *2024 3rd International Conference for Innovation in Technology (INOCON)*, 1–9. https://doi.org/10.1109/INOCON60754.2024.10511646

Kästner, L., & Crook, B. (2024). Explaining AI through mechanistic interpretability. *European Journal for Philosophy of Science*, *14*(4), 52. https://doi.org/10.1007/s13194-024-00614-4

Kucklick, J.-P. (2024). HIEF: A holistic interpretability and explainability framework. *Journal of Decision Systems*, *33*(3), 335–375. https://doi.org/10.1080/12460125.2023.2207268

Lenaers, I., & De Moor, L. (2023). Exploring XAI techniques for enhancing model transparency and interpretability in real estate rent prediction: A comparative study. *Finance Research Letters*, *58*, 104306. https://doi.org/10.1016/j.frl.2023.104306

Lönnroth, T., Krigsholm, P., Falkenbach, H., & Oikarinen, E. (2022). The impact of land use regulation on regional variations in housing supply elasticity. *ERES*, Article 2022_53. https://ideas.repec.org//p/arz/wpaper/2022_53.html

Meshkin, H., Zirkle, J., Arabidarrehdor, G., Chaturbedi, A., Chakravartula, S., Mann, J., Thrasher, B., & Li, Z. (2024). Harnessing large language models' zero-shot and few-shot learning capabilities for regulatory research. *Briefings in Bioinformatics*, *25*(5), bbae354. https://doi.org/10.1093/bib/bbae354

Seo, B., & Li, J. (2024). Explainable machine learning by SEE-Net: Closing the gap between interpretable models and DNNs. *Scientific Reports*, *14*(1), 26302. https://doi.org/10.1038/s41598-024-77507-2

Seya, H., & Shiroi, D. (2019). *A comparison of apartment rent price prediction using a large dataset: Kriging versus DNN* (No. arXiv:1906.11099). arXiv. https://doi.org/10.48550/arXiv.1906.11099

Sun, S.-B., & Choi, M.-S. (2023). A Study on the Effect of Government Housing Policy on the Monthly Rental of Apartment Housing in the Seoul Metropolitan Area Focused on Lessors. *Residential Environment Institute Of Korea*, *21*, 121–135. https://doi.org/10.22313/reik.2023.21.2.121

Trindade Neves, F., Aparicio, M., & de Castro Neto, M. (2024). The Impacts of Open Data and eXplainable AI on Real Estate Price Predictions in Smart Cities. *Applied Sciences*, *14*(5), 2209. https://doi.org/10.3390/app14052209

Vivekananda, M. N., & Shidlyali, P. A. (2024). Enhancing Housing Price Prediction Using AI and Machine Learning: A Stacked Regression Meta-Modeling Approach. *2024 8th International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)*, 1954–1960. https://doi.org/10.1109/I-SMAC61858.2024.10714734

Wessel, M., Schmidt-Kessen, M. J., & Hukal, P. (2024). Regulating short-term rental platforms: The effects of local regulatory responses on Airbnb's operations in Europe. *Industrial and Corporate Change*, *33*(5), 1158–1179. https://doi.org/10.1093/icc/dtad075