A logo for college computing

Description automatically generated

**Assessment Cover Page**

|  |  |
| --- | --- |
| *Student Full Name* | Miqueias S. dos Santos |
| *Student Number 2* | 2016287 |
| *Module Title* | Strategic Thinking (HDip in Data Analytics) |
| *Assessment Title* | CA 1 – Capstone Project Proposal |
| *Assessment Due Date* | 29th March 2024 |
| *Date of Submission* |  |

**Declaration**

By submitting this assessment, I confirm that I have read the CCT policy on academic misconduct and understand the implications of submitting work that is not my own or does not appropriately reference material take from a third party or other source.

I declare it to be my own work and that all material from third parties has been appropriately referenced.

I further confirm that this work has not previously been submitted for assessment by myself or someone else in CCT College Dublin or any other higher education institution.

Introduction

The House Price Prediction project aims to create a machine learning model to predict house prices. It holds significance for the real estate industry, public sectors, and serves as an informative resource for property transactions. The primary goal is to offer timely market insights by combining current and historical data for future predictions. The project intends to contribute to strategic initiatives, addressing issues such as rising prices, homelessness, and legislation for fair buyer opportunities. It plays a pivotal role in enhancing transparency and efficiency in the real estate market, empowering stakeholders, aiding government decisions, and providing valuable insights for comprehensive concerns.

1. **Objectives**
   1. **Development of the Model**: Create a machine learning model to predict house prices based on key features like location, room count, and construction year.
   2. **Enhance Market Transparency**: Utilize the model to provide clear insights into house prices, empowering both buyers and sellers to make well-informed decisions.
   3. **Price Negotiations:** Improve efficiency and fairness in negotiations by using the model's predictions to establish a clearer understanding of property values.
   4. **Inform Governmental Decisions:** Explore using predictive insights to assist government decisions, particularly in identifying areas for social housing construction to address housing challenges.
   5. **Contribute to Strategic Policies:** Explore the model's potential role in shaping strategic initiatives and governmental policy formulation, tackling issues like rising house prices, homelessness, and ensuring fair opportunities for buyers. These objectives align with the broader goal of developing a house price prediction model, showcasing its practical applications and impacts in the real estate industry and beyond.
2. **Problem Definition**

The House Price Prediction project seeks to tackle a pressing challenge within the Irish real estate market— the persistent escalation of housing prices. According to Statista (published on Jan 8, 2024), between 2015 and 2023, the ratio in Ireland increased by nearly 17 index points. This substantial rise is a barrier to homeownership as property values rose in ratio to earnings, limiting affordability.

This imbalance not only affects individuals but also disrupts community stability and impedes fair real estate transactions. The consequent affordability drop presents a complex issue for governmental sectors in developing effective solutions. In response, the project aims to develop a machine learning model for house price predictions. Through comprehensive analyses, the project strives to confront the challenges associated with rising housing prices and propose solutions for a future characterized by accessible and equitable real estate.

1. **Scope**

The project encompasses a comprehensive exploration of machine learning methodologies to accurately forecast house prices, with a focus on the Irish real estate market. The project aims to include:

* 1. **Data Collection and Preprocessing:** In-depth gathering of relevant real estate data, including property details, location information, and historical pricing. Cleaning and preprocessing of the data to ensure its quality and suitability for machine learning model training.
  2. **Data Sources**: The data intended for use in the project is publicly available and can be accessed at The Property Services Regulatory Authority - Property Price Register. This information is open for download and use by the public.
  3. **Model Development:** Creation and fine-tuning of a robust machine learning model, incorporating features such as location, number of rooms, and year built. Implementation of advanced techniques to enhance the model's predictive accuracy.
  4. **Transparency and Efficiency Enhancement**: Utilization of the developed model to provide transparent insights into house prices, empowering buyers, and sellers with valuable information. Integration of the model into the real estate market to optimize efficiency in price negotiations.
  5. **Government Decision Support**: Exploration of the potential application of predictive insights to aid public government decisions, particularly in identifying areas for strategic social housing construction.
  6. **Strategic Initiatives and Policies Contribution**: Investigation of the model's predictions' applicability in contributing to strategic initiatives and policy formulation, addressing challenges like rising house prices and homelessness.
  7. **Exclusions**: The project will not delve into predictive analysis beyond the scope of house prices in the Irish real estate market. Implementation of policies; the focus is on providing insights for policy considerations.
  8. **Boundaries**: The project is limited to the capabilities of machine learning algorithms and the quality of available data. Ethical considerations and data privacy will be strictly adhered to.
  9. **Planned Methods, Techniques, and Approaches**: Utilization of supervised learning algorithms for model development. Feature engineering to enhance model performance. Evaluation of various regression techniques for predicting house prices. Ethical handling of data, ensuring privacy and compliance.

1. **Timeline**
   1. **Semester One**

Weeks 1-4: Data Collection and Preprocessing.

Weeks 5-8: Model Development and Fine-tuning.

Weeks 9-12: Initial Transparency and Efficiency Enhancement.

* 1. **Semester Two**

Weeks 1-4: Government Decision Support Exploration.

Weeks 5-8: Contribution to Strategic Initiatives and Policies.

Weeks 9-12: Final Model Optimization, Documentation, and Presentation Preparation.

* 1. **Expected Deliverables by End of Semester Two**: A well-trained machine learning model capable of predicting house prices accurately. Transparent insights into real estate market dynamics for informed decision-making. Integration of the model into a user-friendly interface for market participants. Comprehensive documentation of methodologies, results, and recommendations.

1. **Ethical Considerations**

The project is committed to upholding the highest ethical standards throughout its lifecycle. Key ethical considerations include:

* 1. **Data Privacy and Confidentiality:** Ensuring that all collected data adheres to privacy regulations and is anonymized to prevent the identification of individuals.
  2. **Informed Consent**: If needed, obtaining consent from data sources, acknowledging the intended use of their information, and ensuring transparency about data handling practices.
  3. **Bias and Fairness:** Regularly assessing and mitigating bias within the machine learning model to prevent unfair treatment of specific demographic groups. Implementing fairness-aware algorithms and techniques to ensure equitable outcomes.
  4. Transparency: Providing clear and understandable explanations of the model's predictions to users, avoiding black-box scenarios. Disclosing the limitations and potential biases of the model to users and stakeholders.
  5. Data Permissions: Acquiring explicit permissions for data usage from relevant authorities or organizations that own or manage the datasets. Adhering to any restrictions or guidelines stipulated by data sources to ensure ethical data utilization.
  6. Societal Impact: Evaluating potential societal impacts of the project, including its influence on property pricing, market dynamics, and housing accessibility. Ensuring that the project's outcomes contribute positively to the welfare of individuals and communities.
  7. No Medical Capstone Projects: Abiding by ethical guidelines and avoiding the inclusion of any medical-related data or analyses in the capstone project. Ensuring that the project's focus remains on addressing challenges within the real estate domain without inadvertently impacting medical privacy or ethics. During the project development, I am committed to continuous monitoring of ethical considerations, adjusting as needed, and researching when necessary to uphold the highest standards of integrity and responsibility in all project activities.

# References

Statista. (2024). Annual house price change in Ireland. Retrieved from <https://www.statista.com/statistics/1155332/annual-house-price-change-in-ireland/>

The Property Services Regulatory Authority - Property Price Register. (Accessed on [insert date]). Retrieved from <https://www.propertypriceregister.ie/>