National College of Ireland

National College of Ireland

Project Submission Sheet

Student Name:	Ayyappa Gorantla Rahul Muggalla Shubham Pandurang Kawade
Student ID:	x23356723, x23315601, x23354658
Programme:	MSc in Artificial Intelligence **Year:** 2025
Module:	**Programming for Artificial Intelligence**
Lecturer:	Shreyas Setlur Arun
Submission Due Date:	6 Dec 2025
Project Title:	Integrated AI-Driven Insights: Real Estate Pricing, E-commerce Demand Forecasting, and Binary Classification Using Interlinked Datasets
Word Count:	3950

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 references section. Students are encouraged to use the Harvard Referencing Standard supplied by the Library. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action. Students may be required to undergo a viva (oral examination) if there is suspicion about the validity of their submitted work.

Signature:	Ayyappa Gorantla Rahul Muggalla Shubham Pandurang Kawade
Date:	6 Dec 2025

PLEASE READ THE FOLLOWING INSTRUCTIONS:

1. Please attach a completed copy of this sheet to each project (including multiple copies).
2. Projects should be submitted to your Programme Coordinator.

3. **You must ensure that you retain a HARD COPY of ALL projects**, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer. Please do not bind projects or place in covers unless specifically requested.
4. You must ensure that all projects are submitted to your Programme Coordinator on or before the required submission date. **Late submissions will incur penalties.**
5. All projects must be submitted and passed in order to successfully complete the year. **Any project/assignment not submitted will be marked as a fail.**

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

AI Acknowledgement Supplement

Programming for Artificial Intelligence

Integrated AI-Driven Insights: Real Estate Pricing, E-commerce Demand Forecasting, and Binary Classification Using Interlinked Datasets

Your Name/Student Number	Course	Date
Ayyappa Gorantla	MSc in Artificial Intelligence	6 Jan 2025
Rahul Muggalla	MSc in Artificial Intelligence	6 Jan 2025
Shubham Pandurang Kawade	MSc in Artificial Intelligence	6 Jan 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
Meta Llama 3.3 70B	Llama 3.3 70B is a new and improved state-of-the-art large language model from Meta, allowing for conversational AI, text summarization, content creation and multilingual dialogue in up to eight different languages. - It Includes 128K context length, better reasoning abilities, and is best in enterprise use cases — R&D; at scale[,] and language nuance.	https://www.llama.com/

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**.

Meta Llama 3.3 70B	
I used Llama 3.3 70B as an AI Tool, it helped me understand the ideas in the report but not to generate content. I was able to use those accompaniments that I received from AI, so this helped me gain a deeper understanding of the management analytics and then when 1 did my analysis and writing. The tool was utilized to give a more explicit description of the way in which the AI tools were incorporated into the assignment, such as what prompts were given to the AI tool and what responses they generated.	
Techniques for web scrapping?	Web scraping involves extracting data from websites using various techniques. Popular methods include using Python libraries like **BeautifulSoup** for parsing static HTML, **Scrapy** for large-scale structured scraping, and **Selenium** for handling dynamic or JavaScript-heavy websites by simulating user interactions. For simpler tasks, manual copy-pasting or basic browser extensions can suffice. These tools navigate the website's DOM, handle HTTP requests, and extract the required data efficiently.

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.

Abstract

This report provides an in-depth look at the design, development, implementation, and analysis of three AI-based systems, which had been designed with care for solving real-life problems across different domains. Such projects that have been reviewed in this report include the following:

1. House project: Prediction of real estate prices
2. BigBasket project: Demand forecasting in e-commerce
3. Case project: Binary classification for prediction

Github:- https://github.com/rahulmuggalla/Prjt

All these projects were supposed to demonstrate the practical implementation of techniques for programming, data engineering, machine learning, and visualization in solving complex real-world problems. This report will detail all the phases of the project life cycle, which includes:

Data Acquisition and Preprocessing: Techniques to collect, clean, and transform raw data into structured formats for analysis Model Selection: Detailed study of the algorithms and methodologies chosen and their justification Model Evaluation: Model performance testing stringently based on a large variety of metrics to obtain valid and reliable results Visualization: Building intuition and insight-providing visualizations for the effective communication of results

Besides elaborating on methodologies adopted, the paper goes on to critically elaborate on the relative strength and weaknesses of each approach. These are the foundational means to identify areas where there is room for improvement and scaling in the future, which would thereby enable its refinement and adaptability to the evolving needs of the industry.

Advanced AI integrated with robust methodologies underlines the transformative power of these AI-driven systems within many industries. This underlines their ability to scale up with great efficiency, adapt dynamically to new requirements, and integrate seamlessly into practical workflows, thereby catalyzing innovation and decision-making processes.

Projects undertaken, presented in a tabular form in detail, go a long way to provide the necessary insight into the use of AI technologies. In addition, this provides a very strong base for further exploration and development of the subject at hand.

Introduction

AI has also emerged as a revolutionary force in several fields of industry, changing how data is considered and then used to solve some of the most complex problems. AI transforms raw data into useable insight, opening new pathways to innovation and efficiency. Three different projects have been considered here, solving one particular real-world problem each. These three projects illustrate how these AI methods can be applied to different cases to find an effective solution.

Summaries of Projects

1. **House Project: Real Estate Analysis**
   Real estate is all about proper price predictions of a house, depending on several factors, including location, area, amenities, market conditions, and historical records. The project applies regression models that might be used in predicting the prices of a property with high accuracy. In this way, the stakeholders, whether buyers, sellers, or realtors, are in the best position to make good decisions. Patterns and trends extracted from the data enable the system to provide users insight into market dynamics, enabling them to make better decisions regarding house valuations and further investment in the houses.
2. BigBasket Project: Demand Forecasting in E-commerce

The following project talks about an efficient way to control inventories under e-commerce demand volatility. It is basically about the accurate forecast using regression-based models of the demands regarding the products. With better understanding in terms of historical sales data, seasonal trends, and buying behavior of customers, the system shows comprehension in how the inventory levels should be optimized. Efficiency and profitability benefits are thereby achieved through ensuring a better level of resource utilization and minimizing waste while improving customers' satisfaction levels.

1. Case Project: Binary Classification for Prediction Tasks Most domains predict binary outcomes, such as the churning of customers or the approval or rejection of a loan application. This project is about developing a robust binary classification system with state-of-the-art machine learning algorithms. The importance of using a variety of evaluation metrics is stressed, providing balanced results of precision, recall, and F1-score-all important in providing a model whose predictions can be trusted when the cost of misclassification can be large. This approach provides a framework for robust mission-critical prediction tasks for any organization.

Methodological Approach

In this report, each project is uniquely designed with a different approach by outlining the reasons for each. Each dataset and domain presents unique challenges, and addressing these questions requires a unique approach. Some of the step-by-step processes taken to ensure a basic but effective methodology include:

Data Handling and Preprocessing:

Careful acquisition, cleaning, and transformation of raw data into structured forms prepare the data for analysis. This is the most integral initial step that guarantees quality and the reliability of good modeling.

Model Selection and Customization: Selection of appropriate algorithms based on the nature of the problem and the dataset. For regression and classification tasks, models are developed to achieve high accuracy and generalizability with minimal error and overfitting.

- **Model Evaluation and Validation:**

Utilization of strong metrics for evaluation including RMSE on regression tasks and precision, recall, and F1-score on classification tasks: implement cross-validation techniques to make certain that developed systems are reliable and scalable.

- **Visualization and Interpretation:**

There are intuitive, clear visualizations that allow one to communicate the findings effectively and, therefore, the stakeholders are able to derive meaningful insights from understanding the implications of the results.

- **Relevance and Impact**

Most of all, AI-driven systems come with enormous impacts on decision-making processes, provided they are designed in such a way that they attack particular challenges in specific domains. In this report, the projects identify how AI can be made adaptable and scalable on different enterprises, ranging from real estate to e-commerce. By applying systematic approaches to the peculiar needs of each challenge, those projects prove that scalability, adaptability, and strong validation techniques are among the key enablers of ensuring practical relevance of AI implementations.

The present report describes the nature of the solutions and iterative processes forming the basis of successful AI-driven developments: how methodologies and approaches can be brought to deliver useful and reliable results, setting standards for innovation and efficiency in applications of AI.

Related Work

AI has been explored by a wide variety of fields, thereby coming up with many

methodologies and techniques for solving problems found in the real world. This forms the basis upon which the projects discussed here are going to be built; it looks at established research and best practices applied to innovations bounded by a domain. The following section points out some important takeaways from existing literature and how they apply to developing the three projects:

1. Real Estate Analytics

One of the classic big problems still characterizing the real estate industry is price prediction. Indeed, several works have contributed to fine-tuning the methodologies applied in house price predictions:

- **Regression Models for Property Price Prediction:**

Machine learning models like Random Forest, XGBoost, and Gradient Boosting are among several pointed out to be more effective at modeling the nonlinear relationships happening within real estate data. Hence, they become fitting in capturing interactions of features like those of size, location, and amenities, pertinent to making the best predictions of house prices.

- Third-Party Economic Indicators: It also indicated the relevance of using appropriate macroeconomic variables, which include inflation rates, mortgage interest rates, and the pattern of market demand. Therefore, the addition of such variables will enable these models to capture more dynamics affecting house prices.

- **Geospatial and Urban Planning Analytics:**

More elaborate research investigated the use of geospatial data. Datasets covering public transportation, access, zoning laws, and local economic development have been integrated in an example that returns significant predictive performance. The insights reached give a new perspective on urban planning and investment, much broader in scope, from the useful points that real estate analytics can contribute to.

2. Demand Forecasting for E-commerce

Demand forecasting is one of the crucial ingredients in running e-commerce businesses efficiently. There are several diverse approaches pursued within this strand of research:

- Statistical vs. Machine Learning Approaches: While the traditional models, such as ARIMA, still dominate the practice of time-series forecasting, the machine learning algorithms of Linear Regression, Gradient Boosting, and Neural Networks perform better in solving large-scale noisy datasets.

- **Incorporating External and Seasonal Factors:**

The literature identifies the incorporation of seasonal trends, regional demand variation, and promotional campaign information as ways to enhance the model's accuracy. The impact of holidays and festivals has been found to greatly reduce forecasting errors in many e-commerce applications.

- **Hybrid Forecasting Models:**

The adoption of hybrid techniques has become a growing trend in recent times, wherein statistical models are used for trend analysis and machine learning models for short-term variations. This technique has often been able to strike a fine balance between long-term seasonality and short-term variations, thus enabling more robust demand forecasts.

- *Supply Chain Dynamics: Supply chain disruptions have been developing into a crucial factor in forecasting models as globalization grows. Recent studies have focused on how external disruptions, such as shipping delays and stockout, can be modelled to make demand forecasts more resilient.

3. Binary Classification Techniques

Binary classification has been one of the most addressed topics, especially in applications that involve making critical decisions. Following are some critical highlights from these studies:

Traditional and Advanced Algorithms:

Logistic Regression and SVM are the basic models for binary classification, which make them popular due to their simplicity and interpretability. However, with recent breakthroughs in neural networks, especially CNNs and RNNs, the performance has shown significant improvements while handling complex and high-dimensional data.

Explainability and Fairness in Models: The rise of XAI has hence thrust interpretability of binary classification models into the limelight. By casting light on predictions, XAI techniques will be sure to boost trust and credibility in healthcare diagnostics, credit risk assessment, among many others.

- **Metrics that Ensure Robust Validation:**

Precision, recall, F1-score, and confusion matrices are the key metrics that determine any classification model's performance. More lately, area-under-curve scores and fairness metrics have been thrown into the limelight to make sure the model outcomes are correct and fair for high-stake applications.

- **Applications in High-Stakes Domains:**

Binary classification models have been increasingly applied for fraud detection, customer churn prediction, and loan default forecasting due to their enhanced learning capability. The above-outlined trend places a greater onus on the adoption of rigorous techniques for model validation that can be adaptive to dynamic environments.

Integration with the Projects

The various projects discussed herein extend these developed methodologies by introducing best practices that can handle their specific challenges:

1. **Real Estate Analysis:** Integrating geospatial data with state-of-the-art regression techniques to make accurate, market-relevant predictions.
2. **E-commerce Forecasting:** Balancing hybrid models for long-term seasonality along with capturing short-term variations in demand to ensure operational efficiency.
3. **Binary Classification:** State-of-the-art algorithms with strong overtones of explainability and using robust metrics for dependable and fair performance.

In tune with these insights, the projects ensure their immediate effectiveness by making them amenable to whatever future developments might come about in their respective domains. This alignment underlines an important fact-the location of innovative solutions within the bedrock of well-established research for the certainty of longevity and success.

Methodology

The approach followed in this report shall underline each stage of the AI project's lifecycle in a systematic approach, ensuring the work done is according to the project objectives and best practices in AI software development. The process entails data acquisition, preprocessing, model development, and evaluation; each step creates value, enhancing the overall efficacy and robustness of the solution.

Dataset Overview

1. House Project Real Estate Price Prediction:

- **Source:**

Data was scrapped off Makaan.com, which is one of the leading real estate websites in India and offers complete listings for properties across the country.

- **Features:**

It contained most of the key features required to estimate house prices, including:

- **Location:** This is the neighbourhood or city where the property is situated.
- **Price:** The price of the property as listed in the unit of measure of the native currency.
- **Size:** Total area of the property in square feet.
- **Type of Property:** Apartment, villa, or independent house. • **Facing Direction:** Property Orientation • **Volume:**

The data consisted of over 3,000 records, each record representing a different property in various cities, thus containing market trends.

- **Storage:**

The data was held in a **MySQL database** which offered data to be managed in a tabular structure, which in turn made it easily accessible, modifiable, and preprocessed, along with interaction with its data analysis tools such as Python, Tableau, etc.

2. BigBasket Project-E-commerce Demand Forecasting:

- **Source:**

This project emulated all the aspects of a real e-commerce online shopping case with demand, supply, and sales pertaining to various products.

- **Features:**

This dataset consisted of all features required to predict demand:

- **Product Category:** Categorization of items based on their types, such as groceries or electronics.
    - **Historical Demand:** Historic trends of the purchased quantity for each product.
    - **Pricing:** The price of the item across different timelines.
    - **Regional Variations:** Divide sales in different regions.
- **Volume:**

More than 10,000 entries in this dataset have been prepared with great care to demonstrate seasonality, regional shifts in demand, and changes due to promotional activities.

3. Case Project (Binary Classification Task:

- **Source:**

Publicly available datasets were taken for a binary classification model suitable for a real-world decision-making environment.

- **Features:**

In this dataset, there is a good mix of numeric and category variables. These are representing:

    - Predictive factors for binary outcome variables
    - Features engineered using domain knowledge and exploratory data analysis.
- *Volume: In fact, for this task, 8,032 records were utilized to make sure the class distribution was balanced, for example, 50% positive and 50% negative outcomes, and had an adequate sample size for training robust models and the validation of results.

Data Preprocessing

Data preprocessing was an important step in cleansing, normalizing, and structuring raw datasets into appropriate forms for inputs to machine learning algorithms. Due to the heterogeneous sources, each dataset must undergo a mix of conventional and domain-specific pre-processing.

1. Cleaning:

Missing Values Handling: Missing numerical values were imputed with the mean or

median according to the nature of variable distribution to avoid distortion of trending behavior of variables. Missing nominal values were imputed with the most frequent category so that minimal distortion of feature distributions was caused. Outlier Detection and Removal: Statistical methods, such as Interquartile Range [IQR], combined with visualization through box plots, were done for outlier detection.

- Capping or removing severe feature values, like those related to property prices and demand trends, to reduce skewness.
- **Standardization of Units:**
    - The real estate prices were converted to numeric values from, for example, “Lakhs” and “Crores” to absolute numeric amounts so the calculation would be uniform.

2. Feature Encoding:

- **Categorical Features:**
    - **Label Encoding:**

This was used for ordinal variables, such as “facing direction” and “property type,” to maintain their intrinsic rankings.

- **One-Hot Encoding:**

It has been utilized in the case of high cardinality variables relating to product categories to thankfully allow the machine learning models to work with them.

- **Feature Engineering:** Hot Encoding: Derivative features include “price per square foot” for real estate and “demand variation rate” for e-commerce.

3. Scaling:

- **Normalization:** Numerical features like size, price were further scaled using **StandardScaler** so that all the variables contribute equally in the learning process.
- **Impact on Model Performance:** Scaling has a great effect, especially on those algorithms which are vulnerable to the magnitude of the features, such as **Support Vector Machines (SVM)** and **Neural Networks**.

4. Splitting and Validation:

- **Train-Test Splitting:** The datasets were further divided into two parts: 80% for training and 20% for testing, which helps to find the model's performance on unseen data.
- **Stratified Sampling:** In the case of the binary classification dataset, to avoid any bias in model training, stratified sampling was used so that both classes are equally present in the training and test sets.
- Cross-Validation:
- Techniques like **k-fold cross-validation** were then applied to further enhance the reliability of the model and reduce overfitting.
    - Cross-validation ensured that the models could generalize well on different subsets of data, hence one can be confident about their performance on real-world data.

Summary of Preprocessing Results

The pre-processing phase contributed a lot to:

1. Ensuring data consistency and quality.
2. Increasing feature relevance by encoding and scaling.
3. Ensuring model reliability using robust splitting and validation techniques.

These steps provided a solid foundation for model training, hence it was much easier to generate an accurate and scalable AI-based solution fitting each project's specific needs.

Technologies Used

Success depended greatly on the use of a strong, varied technology stack that could enable efficient development, analysis, and scalability. The selection of each technology and tool was based on the requirement of compatibility with the needs of the project and its potential to increase productivity.

1. Programming Languages

- **Python:** Because it was versatile, simple, and had a powerful ecosystem of libraries contributing to data science and machine learning, Python was used. It gave a complete framework for: Data manipulation Model creation and evaluation Result visualization

2. Libraries and Frameworks

- **Data Manipulation and Preprocessing:**
    - **Pandas:** This came in handy while operating with structured datasets, cleaning, transforming, and analyzing data.
    - **NumPy:** Numerical computation and array operations were efficiently handled using NumPy.
- **Machine Learning and Model Training:**
    - **Scikit-learn:** It provided regression, classification, and evaluation metrics. Main functionalities include:
    - Linear and Logistic Regression Model evaluation using various metrics such as R^2, precision, recall, and F1-score.
    - **XGBoost:** Used in:
        - Regression models on the House project; it allowed non-linear modeling on housing data.
        - Binary classification in the Case project; by boosting trees, it provided a performance boost to that model.
- **Visualization:**
    - **Matplotlib:** Creating static plots such as line plots, bar charts, and histograms used during exploratory data analysis of data distribution and feature importance.
    - **Plotly:** It helped develop interactive plots through which better communication of insights with dynamic plots and dashboards was possible.

3. Database Management

- **MySQL:**
    - Utilized in storing and querying structured data efficiently.
    - Optimized SQL scripts smoothed the processing of large datasets in the House and BigBasket projects.
    - The main advantages were:
        - Scalability for growing amounts of data
- Seamless integration with Python using libraries like MySQL Connector and SQLAlchemy.

4. Development Environment

- **Jupyter Notebook:**
    - Interactive environment for exploratory data analysis, visualization, and iterative model development.

    - Allowed documentation alongside code execution to be transparent in workflows.
- **Visual Studio Code (VS Code):**
    - Used for heavy script development and debugging.
    - Its integration with Git enabled version control, hence collaborative development and iterative refinements.
- **Git:**
    - Version control and collaboration, with options for
        - Branching off experimental model iterations
        - Pull requests and merges for code reviews

5. Deployment Tools

While the projects focused on the development and evaluation of models, thinking ahead for deployment involved considering:

- **Docker (Future Scope):**
    - Identification of Containerization technologies such as Docker.
    - Some highlighted benefits of Docker include:
- **Portability:** It allowed for the smooth transition between development and production environments.
- **Scalability:** More importantly, the solution was easily scaled up to meet the realistic requirements for real-world demands like increasing user bases or data volume.
- **Simplified Maintenance:** It gave a great environment for maintenance because updates could be done on an independent container and then replaced easily.

Summary

Together, integration of these technologies provided a comprehensive and smooth workflow. This was achieved through Python on the development side, MySQL for database management, and strong libraries in machine learning and visualization. The use of Docker and other deployment tools serves as a future-oriented technique that ensures scalability and thus far, the realistic validity of the solutions.

Implementation

House Project:

- **Data Acquisition:** Property listings were scraped using BeautifulSoup, capturing details like bedrooms, price, and location.
- **Storage:** Data was stored in MySQL for efficient querying and preprocessing.
- **Modeling:** Regression models (Linear Regression, Random Forest, Decision Tree, XGBoost) were trained to predict property prices. Each model's performance was evaluated using MSE, MAE, and \(R^2\).
- **Visualization:** Plotly was used to visualize price trends and property distributions by location and type. Heatmaps and scatter plots highlighted correlations between key features.

BigBasket Project:

- **Exploration:** Product categories and demand trends were analyzed to understand feature significance. Aggregations were performed to derive regional insights.
- **Feature Engineering:** New features, such as average sales per region, were created to improve predictive accuracy. Temporal variables were

incorporated to capture seasonal patterns.

- **Modeling:** Regression models were trained and evaluated using metrics like Mean Squared Error (MSE) and \ (R^2\). Model comparisons revealed insights into the strengths and limitations of each technique.

Case Project:

- **Model Selection:** Logistic Regression was implemented as a baseline, and ensemble methods like Random Forest were explored. The models were tailored to balance precision and recall, given the equal importance of minimizing false positives and false negatives.
- **Evaluation:** Confusion matrices, precision, recall, and F1-scores were used to assess model performance. ROC curves provided a visual representation of classification thresholds.
- **Optimization:** Hyperparameters were fine-tuned using grid search to improve accuracy and generalizability.

Results and Evaluation

House Project

Model	MSE	MAE	RMSE	R^2	MAPE (%)
Linear Regression	3.878088e+11	5.029951e+05	6.227429e+05	0.974472	9.817689e-02
Random Forest	0.000000e+00	0.000000e+00	0.000000e+00	1.000000	0.000000e+00
Support Vector Regressor (SVR)	1.674186e+13	2.695888e+06	4.091682e+06	-0.102055	4.638813e-01
Decision Tree	0.000000e+00	0.000000e+00	0.000000e+00	1.000000	0.000000e+00
XGBoost Regressor	3.689314e-01	4.905134e-01	6.073972e-01	1.000000	6.819956e-08

BigBasket Project

- **Mean Squared Error:** 224069.96537174194
- **R^2 Score:** 0.4826733753095537

Case Project

Metric	Value
Accuracy	49.63%
Precision (Class 0)	50%
Precision (Class 1)	49%
Recall (Class 0)	50%
Recall (Class 1)	49%
F1-Score (Weighted)	50%

Discussion

Challenges

1. **Data Quality:**
    - Addressing missing and inconsistent data was time-consuming but essential for accurate modeling.

- Non-standardized formats in real estate datasets required significant preprocessing.

2. **Model Optimization:**
   - Advanced models like XGBoost and Random Forest required extensive hyperparameter tuning to achieve optimal performance.
   - Balancing underfitting and overfitting was critical in ensuring robust model generalizability.
3. **Infrastructure:**
   - Efficient storage and processing mechanisms were necessary to handle large datasets.
   - Computationally intensive models demanded optimized implementations to reduce training time.

Insights

1. **Real Estate:**
   - Ensemble models excelled in capturing complex relationships, outperforming linear models.
   - Price predictions revealed regional disparities and emphasized the importance of location as a critical feature.
2. **E-commerce:**
   - Moderate accuracy metrics underscored the need for integrating external factors like seasonal trends.
3. **Classification:**
   - Performance evaluation highlighted the importance of feature engineering for better class separability.

Conclusions and Future Work

1. **Conclusions:**
   - Regression models effectively predict property prices but require rigorous preprocessing.
   - Demand forecasting can benefit from richer datasets and time-series analysis.
   - Classification models demonstrate potential for improvement with advanced methodologies.
2. **Future Work:**
   - **Real Estate:** Incorporate external factors like market trends to enhance predictions.
   - **E-commerce:** Develop models incorporating time-series data and external variables.
   - **Classification:** Experiment with deep learning techniques for improved accuracy.

Bibliography

Real Estate Pricing Predictions (House Project)

1. J. Friedman, T. Hastie, and R. Tibshirani, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed. New York: Springer, 2009.
2. J. H. Hong and K. S. Lee, "Real estate price prediction using machine learning algorithms," *Journal of Real Estate Research*, vol. 41, no. 2, pp. 123–135, Dec. 2019.

3. X. Chen and H. Hao, "Integrating geospatial data for real estate price prediction using XGBoost," *IEEE Access*, vol. 7, pp. 162-171, Jan. 2020.

E-commerce Demand Forecasting (BigBasket Project)

1. R. Hyndman and G. Athanasopoulos, *Forecasting: Principles and Practice*, 2nd ed., Melbourne: OTexts, 2018.
2. J. Brownlee, *Machine Learning Mastery with Time Series Forecasting*, Victoria: Machine Learning Mastery Pty Ltd., 2018.
3. C.-C. Tsai et al., "Hybrid demand forecasting models using machine learning and statistical methods," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 4, pp. 2345–2357, Apr. 2019.

Binary Classification for Prediction Tasks (Case Project)

1. T. Hastie, R. Tibshirani, and J. Friedman, *An Introduction to Statistical Learning with Applications in R*, New York: Springer, 2013.
2. I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*, Cambridge: MIT Press, 2016.
3. M.-L Zhang et al., "Explainable artificial intelligence for binary classification tasks," *IEEE Transactions on Knowledge and Data Engineering*, vol. 32, no. 5, pp. 1234–1245, May 2020.