Module D10: Introduction to Artificial Intelligence

Adam Darmanin

January 15, 2025

MASTERS OF BUSINESS ADMINISTRATION (MBA) in Leadership Excellence
Assignment: Application of Machine Learning Techniques to Housing Dataset

Contents

1   Introduction

This assignment will have you go through the process of predictive modeling a Housing dataset. You will engage in data cleaning and preprocessing, followed by exploratory data analysis (EDA) to uncover insights. You will then implement machine learning algorithms, including a deep learning model, to predict housing prices. Finally, you will discuss your findings with the models and their performances, and interpret the results to provide actionable recommendations for potential homebuyers. Students are encouraged to integrate insights from academic literature and papers provided, referencing them appropriately (1).

2   Assignment Objectives

The primary objectives of this assignment are:

- Perform data cleaning and preprocessing: This includes handling missing values, addressing outliers, and applying feature engineering techniques covered in the course.

- Conduct exploratory data analysis (EDA): Identify patterns, trends, and relationships within the dataset to uncover key insights and assess its usefulness for predictive modeling.

- Methodology and Discussion:

  - Build and evaluate machine learning models: Implement two classical models from supervised and unsupervised learning families, and design and train a neural network as part of a deep learning approach.

  - Analyze how models perform on key metrics, understand feature importance, and identify influential factors affecting housing prices. This model and results discussion section has the most score, and you should be clear and thorough, providing detailed analysis and well-reasoned comparisons supported by visualizations and quantitative metrics.

  - Use feature importance, explainable AI (XAI), and other tools to interpret the models and highlight recommendations for potential homebuyers or investors.

- Integrate academic references: Critically engage with the provided literature and any papers of your choice. Use these to support your methodology, interpretations, and conclusions.

3   Dataset Description

The Ames dataset (1) is a more advanced version of the California dataset we used in our lessons. It captures numerous property characteristics, such as lot size, year built, living area, and neighborhood.

Continuous features like lot area and living area provide dimensional data, while discrete variables count items like bedrooms and bathrooms. Categorical variables describe qualitative attributes such as house style and neighborhood. The target variable is the **sale price**, and is continuous. More details and access to the dataset are available at `https://cran.r-project.org/web/packages/AmesHousing/index.html`.

4   Submission Guidelines

Your submission should include:

- A detailed report delivered as a PDF and formatted in ACM style (2) (maximum length of 10 pages).

- References to academic literature: Ensure citations are included for all referenced papers and sources.

- Supporting code provided in a Jupyter Notebook and any other files zipped. Ensure the zip archive includes:

  - A README file that explains how to run the code.

  - A requirements.txt file listing all necessary dependencies.

- Ensure that the code is functional, well-documented, and reproducible.

5   Evaluation Criteria

Your assignment will be evaluated as follows:

- **Data Engineering (20%)**: Quality and completeness of your feature engineering and the depth and clarity of your EDA.

- **Modeling (30%)**: The accuracy, evaluations and robustness of the selected models and the network. You have to utilize the machine learning techniques you have learned and researched.

- **Writeup and Analysis (45%)**: A discussion on the insights derived from your experiments, the EDA, and the comparison of the models selected including the metrics you selected to vet and compare them.

- **Functional Code (5%)**: The code should run seamlessly, be clean, and well-documented.

6   Plagiarism and Quality Expectations

Students are reminded that plagiarism will not be tolerated. All submissions must represent original work and include proper citations for any references used. Assignments will be checked using plagiarism detection tools, and any instance of academic dishonesty will result in a failing grade.

Deliverables are expected to meet masters-level standards of academic quality. Reports should be well-structured, clearly written, and free of errors. Code should be clean, efficient, and thoroughly documented. Visualizations and analyses should be presented in a manner that is both informative and visually clear.

References

[1] C. R. Project, "Ameshousing: Alternative to the boston housing data," 2025. [Online]. Available: https://cran.r-project.org/web/packages/AmesHousing/index.html

[2] "Acm sig proceedings templates," 2025. [Online]. Available: https://www.overleaf.com/latex/templates/association-for-computing-machinery-acm-sig-proceedings-template/bmvfhcdnxfty