Project And Data Management Plan

Comparison of Machine Learning vs Deep Learning Algorithms for House Price Prediction

Student Name: Shivakumar Pole

Student Number: 21066186

Course: MSc Data Science with Advanced Research

Supervised by: Dr.William cooper

# Overview:

Machine Learning and Deep Learning are two sides of the same coin. These statistical and mathematical techniques help the user to study past data and create an algorithm trained on the past data to predict future (*Deep learning vs. machine learning*, 2020). In this research these techniques will be tested on “House Price Prediction” problem ([link](https://www.kaggle.com/datasets/shree1992/housedata)). This research compares multiple Machine Learning and Deep Learning techniques like SVM, XGBoost, Multi-Layer Perceptron etc., (Ray, 2017) By evaluating multiple approaches, this research aims to understand which algorithm provides the best and most reliable predictions in the context of house prices prediction problem.

# Research Questions:

1. How do traditional ML algorithms perform compared to DL algorithms in the context of house price predictions?
2. Which features are most significant in influencing the predictions?
3. What impact does the volume of data have in performance of ML vs DL algorithms?

# Aim:

The main aim of this research is to compare and contrast the performance of ML and DL algorithms to determine which of them is more suitable for predicting house prices based on various housing features like location, size, amenities, etc.,

# Objectives:

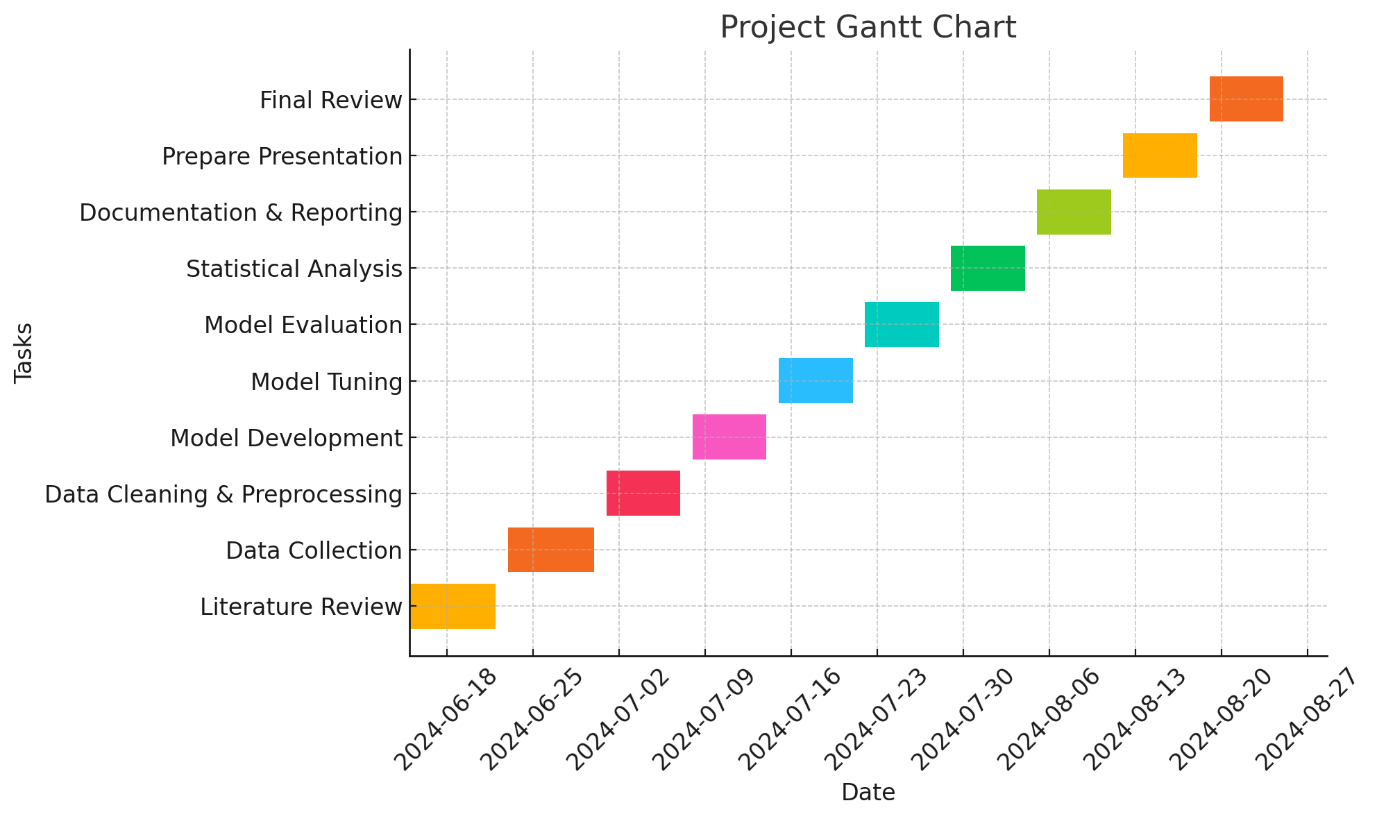
* To review existing literature on the ML applications of real estate prices
* To analyse and process data obtained from open-source website without any ethical issues.
* To train multiple ML and DL models
* To design and tune DL models to increase its performance.
* To compare the results from ML and DL models using performance metrics like Mean Squared Error and R squared (Goyal, 2021).
* To document the findings and provide a critical discussion about the results.

# Methodology:

1. Literature Review: This phase typically is used to understand the current methodologies used in this domain.
2. Data Collection: Using open-source websites that have data to understand and explore the challenges that involve data cleaning, feature engineering, and normalization (Zhang, 2019). These steps are required to get the most value from the dataset.
3. Model Development: This phase is the critical phase as it holds the core of the research. It includes training multiple ML and DL models to fine tuning the best models to increase their performance.
4. Model Evaluation: Using techniques like splitting dataset into multiple parts, cross validation along with comparing the performance metrics to understand the strengths and weaknesses of the algorithms.
5. Statistical Analysis: Apply statistical tests to compare the performance of the algorithms used in this research (ML and DL)

# Conclusion:

This project will encompass the results of the comparison between ML and DL algorithms. Along with the comparison of the algorithms the other part is to understand how the data works, which feature has the most impact, how the volume of the data affects the performance of these algorithms. To highlight the unique strengths and limitations of each approach while fine tuning these algorithms. This research will also include potential for future research and practical applications in real estate area.



# References:

*Deep learning vs. machine learning* (2020) *Zendesk*. Available at: https://www.zendesk.com/in/blog/machine-learning-and-deep-learning/ (Accessed: June 8, 2024).

Goyal, S. (2021) *Evaluation metrics for regression models - analytics Vidhya - medium*, *Analytics Vidhya*. Available at: https://medium.com/analytics-vidhya/evaluation-metrics-for-regression-models-c91c65d73af (Accessed: June 8, 2024).

Ray, S. (2017) *Top 10 machine learning algorithms to use in 2024*, *Analytics Vidhya*. Available at: https://www.analyticsvidhya.com/blog/2017/09/common-machine-learning-algorithms/ (Accessed: June 8, 2024).

Zhang, Z. (2019) *Understand data normalization in machine learning*, *Towards Data Science*. Available at: https://towardsdatascience.com/understand-data-normalization-in-machine-learning-8ff3062101f0 (Accessed: June 8, 2024).