

Phase 2 Project Submission

Predicting IMDb Scores

Name:	Tharun E
Reg.No:	410121104058
NM ID:	au410121104058
Department:	CSE-III
Domain&Project:	Applied Data Science & Predicting IMDb Scores
College:	4101-Adhi College of engineering and Technology.

Introduction:-

Provide an overview of the project, emphasizing the importance of predicting IMDb scores for movies and TV shows.

Explain that the document explores advanced regression techniques like Gradient Boosting and Neural Networks to enhance prediction accuracy.

Data Collection and Preparation:-

Describe where the IMDb data was sourced and the format it came in.

Highlight the crucial data preprocessing steps, including cleaning, handling missing values, and encoding categorical variables.

Mention the feature selection and engineering strategies used to improve model performance.

Exploratory Data Analysis (EDA):-

Present basic statistics about the dataset, such as mean IMDb scores, standard deviations, and data distribution.

Showcase data visualization techniques like histograms, scatter plots, and correlation matrices to gain insights into the data.

Discuss correlations between IMDb scores and various features like genre, director, and budget.

Regression Techniques:-

Define what regression analysis is and its role in predicting continuous target variables.

Contrast traditional linear regression with advanced techniques like Gradient Boosting and Neural Networks, justifying the choice of these methods.

Gradient Boosting :-

Explain the concept of Gradient Boosting and its iterative nature in building predictive models.

Describe the implementation of Gradient Boosting using popular libraries like XGBoost or LightGBM.

Discuss the importance of hyperparameter tuning and the metrics used to evaluate model performance.

Neural Networks:-

Provide an introduction to Neural Networks, highlighting their flexibility and ability to capture complex patterns.

Walk through the process of building a simple feedforward neural network for IMDb score prediction.

Explore deep learning architectures suitable for this regression task, discussing key components like optimizers and activation functions.

Model Training and Evaluation:-

Explain the importance of splitting the data into training, validation, and test sets.

Detail the steps involved in training both the Gradient Boosting and Neural Network models.

Introduce cross-validation techniques to ensure robust model evaluation.

Comparative Analysis:-

Conduct a thorough comparison of the performance of Gradient Boosting and Neural Networks.

Highlight the strengths and weaknesses of each technique in the context of IMDb score prediction.

Utilize visualizations and metrics to support the comparative analysis.

Model Interpretability:-

Discuss methods for interpreting Gradient Boosting models, such as feature importance analysis.

Address the challenges of interpreting Neural Networks due to their black-box nature.

Conclusion:-

Summarize the key findings of the project, including which technique yielded superior prediction accuracy.

Mention potential real-world applications of accurate IMDb score prediction.
