MINI PROJECT: PROGRESS REPORT - I

Post-Listing Performance and Risk Assessment using Machine Learning and Power BI

1 Project Definition

The project “Post-IPO Listing Performance Prediction using Machine Learning and Power BI” focuses on predicting the post-listing performance of Initial Public Offerings (IPOs) using machine learning techniques. The project uses two independent prediction approaches, namely fundamental-based prediction and GMP-based prediction, to estimate the listing gain percentage and classify IPOs into different risk categories. Machine learning models such as Random Forest and Support Vector Machines are employed for prediction, while Power BI is used to visualize analytical results and predictions through interactive dashboards. This project integrates predictive analytics with business intelligence to assist in understanding IPO performance trends.

2 Project Objective

The primary objective of this project is to design and develop a data-driven system that predicts the post-listing performance of Initial Public Offerings (IPOs) and assesses associated investment risks using machine learning techniques and business intelligence tools. By leveraging historical IPO data, the project aims to provide accurate performance predictions and meaningful risk classification to support analytical understanding of IPO behavior.

The project integrates two independent prediction approaches—fundamental-based analysis and Grey Market Premium (GMP)-based analysis—to evaluate IPO listing gains and compare their predictive effectiveness. Additionally, interactive Power BI dashboards are used to visually represent trends, predictions, and risk insights, enabling easier interpretation and informed decision-making.

- To analyze IPO fundamental parameters for performance prediction
- To analyze GMP-based parameters for short-term IPO prediction
- To classify IPOs into different risk categories
- To compare fundamental-based and GMP-based prediction approaches
- To visualize predictions and trends using Power BI dashboards

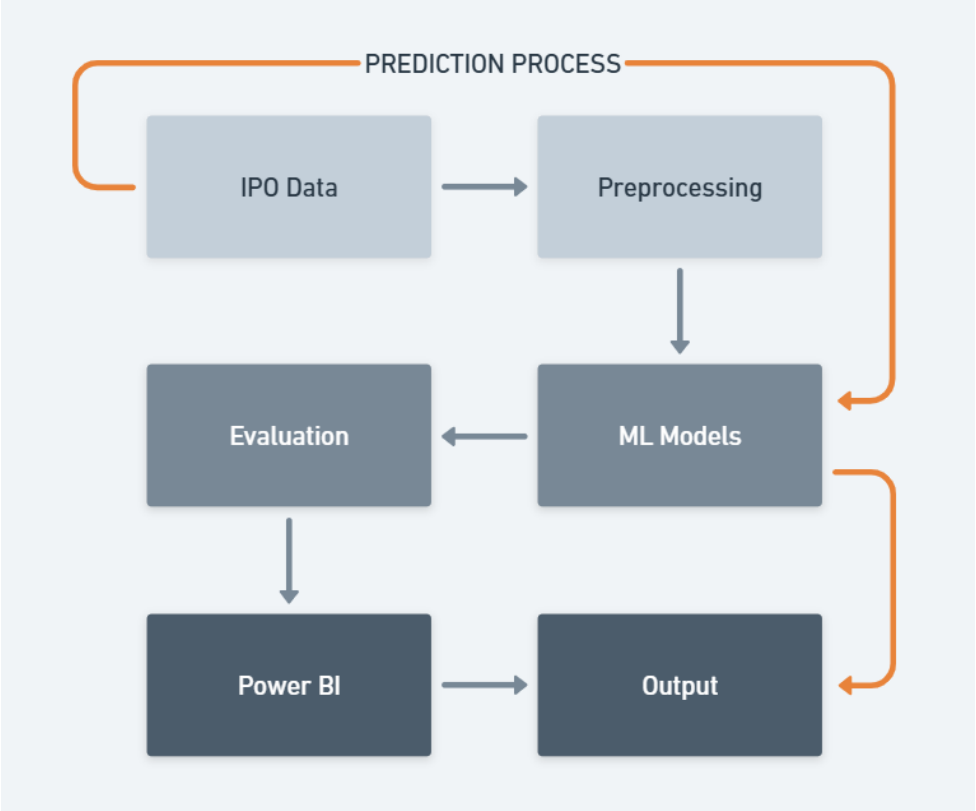

Fig. 1 Process flow diagram

3 Project Scope

The scope of this project is limited to mainline IPOs and focuses only on post-listing performance prediction. The project uses historical IPO datasets and applies supervised machine learning models to generate predictions. Power BI is utilized for presenting visual insights such as performance trends, predicted gains, and risk classification. The project does not include real-time stock trading or live market forecasting.

4 Project Modules

The proposed system is divided into multiple well-defined modules to ensure modular development, scalability, and ease of maintenance. Each module is responsible for a specific functionality in the overall IPO prediction and risk analysis system. The modular approach helps in systematic data processing, accurate prediction, and effective visualization of results for end users.

- Data Collection Module
- Data Preprocessing Module
- Fundamental-Based Prediction Module
- GMP-Based Prediction Module
- Evaluation Module
- Visualization Module

4.1 Data Collection Module

The Data Collection Module is responsible for gathering historical IPO-related data from reliable financial sources. This module collects datasets containing both fundamental parameters and Grey Market Premium (GMP) information of IPOs. Fundamental parameters include issue size, price band, lot size, company financial ratios, promoter holding, and subscription details. GMP data reflects market sentiment prior to listing and is collected from trusted market tracking platforms.

The collected data is stored in a structured format to ensure consistency and ease of access for further processing. This module plays a crucial role as the quality and reliability of the prediction models depend heavily on the accuracy and completeness of the collected data.

4.2 Data Preprocessing Module

The Data Preprocessing Module focuses on transforming raw data into a clean and usable format suitable for machine learning models. This module handles missing values using appropriate statistical techniques, removes duplicate and inconsistent records, and corrects data anomalies. Feature selection is performed to identify the most influential attributes affecting IPO performance.

Additionally, normalization and scaling techniques are applied to ensure uniformity across numerical features. Labels for classification tasks, such as IPO risk categories (Low, Medium, High), are generated based on predefined criteria. This module enhances model efficiency, reduces noise, and improves prediction accuracy.

4.3 Fundamental-Based Prediction Module

The Fundamental-Based Prediction Module utilizes company financial and issue-related attributes to predict IPO performance. A Random Forest Regressor is implemented to predict

the expected listing gain percentage based on fundamental indicators. Random Forest is chosen due to its robustness, ability to handle non-linearity, and resistance to overfitting.

In addition, a Random Forest Classifier is used to classify IPOs into different risk categories based on fundamental data. This classification assists investors in understanding the risk profile of an IPO before making investment decisions. The module provides both numerical predictions and categorical risk assessments.

4.4 GMP-Based Prediction Module

The GMP-Based Prediction Module focuses on analyzing market sentiment through Grey Market Premium data. A Support Vector Regression (SVR) model is employed to predict the expected listing gain percentage using GMP values and related features. SVR is selected due to its effectiveness in handling small datasets and non-linear relationships.

Furthermore, a Support Vector Classifier (SVC) is used to classify IPO risk levels based on GMP trends. This module helps capture short-term market behavior and investor sentiment, complementing the fundamental-based analysis and providing a more comprehensive prediction framework.

4.5 Evaluation Module

The Evaluation Module assesses the performance of all prediction and classification models developed in the system. Various evaluation metrics are used to measure model effectiveness. For classification tasks, metrics such as accuracy, precision, recall, and F1-score are employed. For regression tasks, R^2 score, Mean Absolute Error (MAE), and Mean Squared Error (MSE) are used.

This module enables comparison between fundamental-based and GMP-based models, helping to identify the most reliable prediction approach. The evaluation results guide further model tuning and optimization.

4.6 Visualization Module

The Visualization Module provides an interactive and user-friendly representation of prediction results using Power BI. Dynamic dashboards are designed to display predicted listing gains, IPO risk levels, and comparative performance between fundamental-based and GMP-based models. Visual elements such as charts, graphs, filters, and slicers enhance data interpretation.

Power BI enables real-time insights and helps stakeholders, investors, and analysts make informed decisions based on clear visual analytics. This module bridges the gap between complex machine learning outputs and practical decision-making.

5 Project Basic Requirements

5.1 Software Requirements

- Operating System: Windows 10
- Programming Language: Python
- Python Libraries: Pandas, NumPy, Scikit-learn, Matplotlib, Seaborn
- Visualization Tool: Power BI
- Reliable Data sources for Model training and testing
- IDE: VS Code / Jupyter Notebook / Google Colab Notebook

6 Literature Review

Several studies related to IPO performance prediction have been reviewed to understand the current state of research and identify gaps that this project aims to address.

6.1 Recent Advances in IPO Prediction

Banerjee and Ghosh [1] employed multinomial logistic regression to predict IPO performance from prospectus data, demonstrating the effectiveness of statistical methods in IPO analysis. Their work in 2025 highlights the importance of textual information extracted from IPO prospectuses in predicting post-listing performance.

Verma et al. [2] explored multi-modal information fusion for predicting the success of Indian IPOs in 2024. Their study combined textual, numerical, and sentiment-based features to improve prediction accuracy, indicating that diverse data sources can enhance IPO performance forecasting.

6.2 Machine Learning Approaches

Patel and Mehta [3] conducted a systematic review of machine learning methodologies for IPO forecasting in 2023. Their comprehensive analysis identified Random Forest, Support Vector Machines, and Neural Networks as the most promising algorithms for IPO prediction tasks.

Singh and Sharma [4] performed a comparative analysis of machine learning algorithms for predicting IPO underperformance in 2022. Their findings suggested that ensemble methods, particularly Random Forest, outperformed traditional statistical approaches in capturing non-linear relationships in IPO data.

Aggarwal et al. [5] applied machine learning analyses to predict IPO first-day returns in 2022. Their research demonstrated that machine learning models could achieve superior predictive accuracy compared to conventional financial models, especially when incorporating market sentiment indicators.

6.3 Sentiment Analysis and Textual Information

Liu et al. [6] investigated the use of prospectus sentiment for predicting IPO performance using machine learning in 2021. Their study revealed that sentiment extracted from IPO documents significantly influences first-day returns and can be effectively incorporated into predictive models.

Jegadeesh and Wu [7] adopted a machine learning approach to analyze textual information and its relationship with IPO underpricing in 2020. They found that natural language processing techniques combined with machine learning algorithms could extract valuable predictive signals from IPO-related documents.

6.4 Random Forest Applications

Cortez et al. [8] utilized Random Forest algorithms to predict IPO initial returns in 2019. Their work established Random Forest as a robust and interpretable method for IPO prediction, capable of handling high-dimensional feature spaces and non-linear relationships.

6.5 Research Gap and Project Contribution

While existing research has explored various aspects of IPO prediction, most studies focus on either fundamental analysis or market sentiment indicators in isolation. This project addresses this gap by developing a comprehensive system that integrates both fundamental-based and GMP-based prediction approaches. Furthermore, the incorporation of Power BI for interactive visualization distinguishes this project from existing IPO analysis platforms that primarily provide descriptive statistics. The comparative analysis of two independent prediction methodologies provides investors with multiple perspectives for informed decision-making.

Research indicates that IPO subscription parameters such as QIB, HNI, and RII have a strong influence on listing performance. The Grey Market Premium (GMP) serves as a short-term market sentiment indicator that captures investor enthusiasm before listing. Traditional statistical methods are limited in capturing non-linear IPO behavior, whereas machine learning models such as Random Forest and Support Vector Machines provide better predictive accuracy through their ability to model complex relationships in data.

7 Project Feasibility Study

The feasibility study assesses whether the proposed IPO prediction and risk analysis system can be developed and implemented effectively. The study evaluates the project from technical, economic, and operational perspectives to ensure its practicality.

7.1 Technical Feasibility

The project is technically feasible due to the availability of reliable and well-supported machine learning libraries such as Scikit-learn, Pandas, and NumPy. Algorithms like Random Forest and Support Vector Machines can be efficiently implemented using these tools. Additionally, Power BI provides robust capabilities for data visualization and dashboard creation, enabling effective presentation of prediction results. The system can be developed and executed on standard computer systems without the need for specialized hardware.

7.2 Economic Feasibility

The project is economically feasible as it utilizes open-source software and free tools, eliminating any additional cost. Python libraries used for data analysis and machine learning are freely available, and Power BI Desktop can be used without licensing fees. Therefore, the project does not require any financial investment, making it suitable for academic and educational purposes.

7.3 Operational Feasibility

The system is easy to operate and understand, as the core analysis is performed in the background while users interact with visual outputs through Power BI dashboards. The graphical representation of predictions and risk levels simplifies interpretation, requiring minimal technical knowledge. This ensures smooth operation and effective usage of the system.

8 Project Requirement Gathering

8.1 Requirement Gathering Techniques

- Study of IPO datasets
- Review of existing IPO analysis tools
- Literature study
- Discussion with faculty guide

8.2 Comparison of Existing Applications with Proposed Application

Feature	Existing Applications	Proposed Application
IPO Analysis	Manual / Descriptive	ML-Based Prediction
Risk Assessment	Not Available	Available
GMP vs Fundamentals	Separate	Comparative
Visualization	Static Charts	Power BI Dashboards

Table 1: Comparison of Existing Applications with Proposed Application

9 Project Timeline Chart

Phase	Duration
Requirement Analysis	First Week of January
Dataset Collection	Second Week of January
Data Preprocessing	Third Week of January
Model Development	February
Testing and Evaluation	Early March
Power BI Dashboard Creation	Mid March
Documentation and Presentation	Last Week of March

Table 2: Project Timeline Chart

ID	Task Name	Start	Finish	Duration
1	Requirement Analysis & System Study	01-01-2026	08-01-2026	1w
2	Dataset Collection	08-01-2026	15-01-2026	1w
3	Data Preprocessing	15-01-2026	22-01-2026	1w
4	Model Development	22-01-2026	26-02-2026	4w
5	Testing and Evaluation	26-02-2026	12-03-2026	2w
6	Power BI Dashboard Creation	12-03-2026	26-03-2026	2w
7	Documentation and Presentation	26-03-2026	09-04-2026	2w

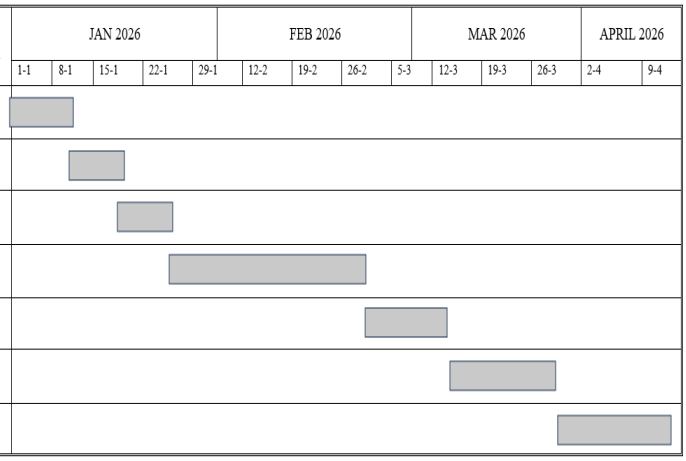

Fig. 2 Gantt chart of the system

10 Project Work Distribution Among Team Members

10.1 Team Member 1: Desai Karansinh

Responsible for dataset collection, preprocessing, exploratory data analysis and development of Model for Fundamental-based Prediction.

10.2 Team Member 2: Aryan Changela

Responsible for dataset collection, preprocessing, exploratory data analysis and development of Model for GMP-based Prediction.

10.3 Team Member 3: Hriday Prajapati

Responsible for Power BI dashboard creation, documentation, and presentation preparation.

11 Reference

1. T. Banerjee and S. Ghosh, ”Prediction of IPO Performance from Prospectus UsingMultinomial Logistic Regression,” AIMS Mathematics, vol. 10, no. 2, pp. 2301–2315, 2025.

2. A. Verma, K. Shah, and R. Iyer, ”Experimenting with Multi-modal Information toPredict Success of Indian IPOs,” arXiv preprint arXiv:2412.16174, 2024.

3. S. Patel and N. Mehta, ”IPO Forecasting Using Machine Learning Methodologies: A Systematic Review,” International Journal of Management and Applied Science, vol. 9, no. 1, pp. 12–20, 2023.

4. A. K. Singh and R. Sharma, ”A Comparative Analysis of Machine Learning Algorithms for IPO Underperformance Prediction,” Journal of Advances and Applications in Statistics and Research, vol. 5, no. 2, pp. 45–53, 2022.

5. R. K. Aggarwal, S. Prabhala, and P. Puri, ”Predicting IPO First-Day Returns: Evidence from Machine Learning Analyses,” Journal of Banking & Finance, vol. 134, pp. 1–18, 2022.

6. J. Liu, Y. Wang, and H. Zhang, ”Predicting IPO Performance from ProspectusSentiment Using Machine Learning,” Business Management Review, vol. 14, no. 3, pp. 101–110, 2021.

7. M. Jegadeesh and J. Wu, ”Textual Information and IPO Underpricing: A MachineLearning Approach,” Journal of Financial Markets, vol. 48, pp. 1–15, 2020.

8. J. L. Cortez, P. Silva, and A. Gomes, ”Predicting IPO Initial Returns Using RandomForest,” Procedia Computer Science, vol. 164, pp. 593–600, 20

Guide Remark/Suggestions

Guide Signature with Date