Research Paper Metadata Document

Publication Information

Paper Title: Machine Learning-Based IPO Price Prediction: A Data-Driven Analysis

Format Standard: IEEE Conference Paper Format

Publication Year: 2025

Publication Date: April 2025

Author Information

Primary Author

- **Name:** Pratyush Rawat
- **Student ID:** 23FE10CDS00288
- **Email:** pratyush.rawat@example.com
- **Department:** Data Science and Engineering
- **School:** School of Computer Science and Engineering-2
- **Institution:** Manipal University Jaipur
- **Location:** Rajasthan, India

Co-Author

- **Name:** Dudhat Ayush
- **Student ID:** 23FE10CDS00293
- **Email:** dudhat.ayush@example.com
- **Department:** Data Science and Engineering
- **School:** School of Computer Science and Engineering-2
- **Institution:** Manipal University Jaipur
- **Location:** Rajasthan, India

Academic Context

Course Code: DSE2220

Course Title: Machine Learning

Project Name: IPOVista

Team Name: Proto

Institution: Manipal University Jaipur

Department: Data Science and Engineering

Academic Level: Undergraduate

Research Classification

Primary Domain: Financial Technology (FinTech)

Sub-domains:

- Machine Learning
- Predictive Analytics
- Financial Forecasting
- Investment Analysis
- Computational Finance

Research Type: Applied Research / Empirical Study

Methodology: Quantitative Analysis with Machine Learning

Keywords

1. Initial Public Offering
2. Machine Learning
3. IPO Prediction
4. Ensemble Methods
5. Neural Networks
6. Feature Engineering
7. Investment Analysis
8. Financial Forecasting

Dataset Specifications

Dataset Name: Indian IPO Market Data

Format: CSV (Comma-Separated Values)

Size: 326 unique IPO records

Features: 13 attributes per record

Market: Indian Stock Market

Time Period: Multiple years (historical data)

Data Type: Structured financial and subscription metrics

Technical Specifications

Machine Learning Models Evaluated

1. Logistic Regression
2. Random Forest
3. Gradient Boosting
4. Support Vector Machine (SVM)
5. Voting Ensemble
6. Neural Network (Feedforward)

Programming Languages & Libraries

- **Primary Language:** Python
- **ML Frameworks:**
    - scikit-learn
    - TensorFlow
    - Keras
- **Data Processing:**
    - pandas
    - NumPy
- **Visualization:**
    - matplotlib
    - seaborn (implied)

Optimization Techniques

- Bayesian Optimization (200 trials)
- Hyperparameter Tuning
- SMOTETomek (Class Imbalance Handling)
- PowerTransformer (Yeo-Johnson method)

Performance Metrics

Best Test AUC: 0.7025 (Logistic Regression)

Highest Accuracy: 68.8% (Random Forest)

Neural Network Training Accuracy: 85%

Neural Network Validation Accuracy: 74%

Neural Network Training AUC: ~0.90

Optimal Decision Threshold: 0.5

Repository Information

GitHub Repository: FLACK277/INITIAL-IPO-PREDICTION

Repository URL: https://github.com/FLACK277/INITIAL-IPO-PREDICTION

Access Date: October 15, 2025

Repository Type: Public (implied)

Document Structure

Total Sections: 5 major sections

Section Breakdown:

1. Introduction
2. Related Work
3. Methodology

4. Results and Discussion
5. Conclusion and Future Scope

Figures: 5 figures

- Fig. 1: ROC Curve comparison
- Fig. 2: Confusion Matrix
- Fig. 3: Metrics vs Threshold analysis
- Fig. 4: Model performance comparison
- Fig. 5: IPO prediction output example

Tables: 2 tables

- Table I: Dataset Structure and Feature Summary
- Table II: Comprehensive Model Performance Comparison

References

Total References: 15

Reference Style: IEEE Citation Format

Key Reference Topics:

- Deep Learning (LSTM networks)
- IPO Underpricing Research
- Sentiment Analysis
- Ensemble Methods
- SMOTE Technique
- Machine Learning Libraries

Date Range: 1986-2018 (primary literature)

Supplementary Materials

Case Study: MRF IPO Prediction

Predicted Listing Gain: 31.24%

Predicted Opening Price: ₹59.06

Neural Network Confidence: 64%

Ensemble Confidence: 72%

Recommendation: Moderate BUY

Ethical & Research Considerations

Data Source: Publicly available Indian IPO market data

Transparency: Complete methodology disclosure

Reproducibility: GitHub repository provided for replication

Limitations: Clearly documented in Section V.B

Future Work: Comprehensive roadmap provided in Section V.C

Document Specifications

Format: IEEE Conference Paper

Page Layout: Two-column format (standard IEEE)

Font: Times New Roman (IEEE standard)

Citation Style: IEEE numerical citation [1], [2], etc.

Copyright: IEEE Copyright 2025

Language: English (US)

Contact Information

For Correspondence:

- Primary Contact: Pratyush Rawat (pratyush.rawat@example.com)
- Secondary Contact: Dudhat Ayush (dudhat.ayush@example.com)

Institutional Contact:

- Manipal University Jaipur
- Department of Data Science and Engineering
- Rajasthan, India

Version Information

Document Version: Final Publication Version

Submission Date: April 2025

Review Status: Academic Project Submission

Peer Review: Course Evaluation (DSE2220)

Additional Metadata

Computational Requirements:

- Training Environment: Standard ML development environment
- Computational Intensity: Moderate (Neural Network training)
- Hardware: Not specified (standard GPU/CPU configuration)

Data Availability: Indian IPO market data (CSV format)

Code Availability: GitHub repository (FLACK277/INITIAL-IPO-PREDICTION)

Conflict of Interest: None declared (academic project)