International Journal of Advanced Trends in Engineering and Management
(ISSN 2583-7052)

Vol. IV, No. 04, pp. 20-24 , April

Mobile App Success Rate Prediction

Mrs. Parrem Rajini[1], Erram Reddy Aravind [2], K. Muralidhar Goud[3*], Kothuri Rama Krishna[4]

[1]Department of Computer Science, Chaitanya Deemed to be University
[2]Department of Computer Science, Chaitanya Deemed to be University
[3]Department of Computer Science, School of Engineering, Malla Reddy University
[4]Department of Electrical and Electronics Engineering, B. V. Raju Institute of Technology, Narsapur, Medak, Telangana.

***Corresponding Author E-mail: muralidhar.3n1@gmail.com**

ABSTRACT: With the exponential rise of mobile applications in recent years, predicting the success rate of an app before or shortly after launch has become a crucial challenge for developers, investors, and marketing strategists. This study explores machine learning techniques to build predictive models capable of estimating a mobile app's success based on features such as app category, user ratings, number of installs, review sentiment, update frequency, and monetization strategy. We analyze a publicly available dataset and apply supervised learning methods including Random Forest, Logistic Regression, and Gradient Boosting. The model is evaluated based on precision, recall, and F1 score, with SHAP values used for interpretability. Our results demonstrate that user reviews and app ratings have the strongest predictive power. By providing a predictive and explainable model, our research offers valuable insights for stakeholders in the mobile app ecosystem aiming to optimize release strategies and investment decisions.

Keywords: Mobile Applications, Success Prediction, Machine Learning, SHAP, User Analytics.

1. Introduction

The mobile application industry has witnessed explosive growth over the past decade, with millions of apps available on platforms like Google Play Store and Apple App Store. This surge is fueled by increasing smartphone adoption, low development barriers, and demand for digital services ranging from entertainment and education to finance and health. As of 2024, mobile apps contribute significantly to global digital economy revenues, accounting for billions of downloads and trillions in user engagement hours. Despite this scale, only a small percentage of apps achieve substantial success, while the vast majority struggle to gain traction or face early abandonment.

Predicting the success of a mobile app early in its lifecycle is of immense value to developers, investors, product managers, and marketing teams. Success is typically associated with a combination of factors such as high user ratings, sustained download growth, positive review sentiments, frequent updates, and effective monetization strategies. However, predicting such outcomes is inherently challenging due to market dynamics, rapidly changing user behavior, and evolving platform policies. Traditional analytical tools are limited in their ability to capture the complex, non-linear interactions among app features and performance metrics.

Recent advances in machine learning (ML) provide an opportunity to model these complexities by learning patterns from historical app data. ML algorithms can assess both structured data (e.g., rating scores, installs) and unstructured data (e.g., user reviews) to classify or predict an app's potential performance.

However, while predictive accuracy is valuable, it is equally important to ensure that predictions are explainable and actionable. Stakeholders need more than just a binary success/failure output—they require a clear understanding of *why* the model made a specific prediction and *how* to improve outcomes.

This manuscript proposes an explainable AI-based approach for predicting mobile app success rates. Unlike black-box models that prioritize accuracy alone, our methodology combines powerful machine learning models (e.g., Gradient Boosting) with interpretability frameworks such as SHAP (SHapley Additive exPlanations) and counterfactual reasoning. These tools not only reveal which features most influence predictions but also suggest tangible changes that could turn a failing app into a successful one. By integrating these elements into a unified framework, we aim to offer a scalable, interpretable, and effective tool for early success prediction in mobile app development.

Furthermore, this research contributes to the growing demand for ethical and transparent AI in commercial applications. Regulators and consumers alike are demanding explainability, especially in decision-making systems that influence business investments or consumer experiences. Our work aligns with these trends by emphasizing both predictive performance and human-centered insights—making the solution not just technically sound but also practically deployable across real-world app ecosystems.

2. Literature Survey

Research in the domain of mobile app success prediction has evolved rapidly, leveraging various analytical and machine learning techniques. Early studies primarily relied on statistical models and regression analysis to correlate simple attributes such as app rating and number of installs with success metrics. While these approaches were effective in capturing linear relationships, they often failed to model the complex, non-linear interactions between user engagement metrics, temporal trends, and qualitative aspects like review sentiments.

The increasing availability of rich app store metadata has encouraged more sophisticated modeling techniques. Supervised learning models such as Decision Trees, Support Vector Machines (SVM), and Logistic Regression have been employed to classify apps based on binary success labels or multiclass tiers (e.g., low, medium, high performance). For example, Zhao et al. (2021) used SVMs trained on historical app data to forecast download growth, incorporating both numeric and textual review features. Their results highlighted the importance of incorporating Natural Language Processing (NLP) techniques to better understand user perception, which often precedes tangible metrics like ratings or uninstall rates.

More recently, ensemble methods such as Random Forest, XGBoost, and Gradient Boosting Machines (GBMs) have become prominent due to their superior accuracy and robustness in handling noisy, multi-modal datasets. These models have been shown to outperform traditional algorithms, especially when dealing with heterogeneous data combining structured metadata and unstructured text. Studies like those by Arora et al. (2022) and Wang et al. (2023) applied ensemble models on large-scale datasets from the Google Play Store and App Annie, demonstrating strong predictive power for app popularity and revenue classification.

Parallel to these advancements in predictive modeling, explainable AI (XAI) has become a critical area of focus. The deployment of black-box models in real-world decision-making has raised concerns about transparency and fairness, especially in commercial and consumer-focused domains like app marketplaces. Techniques such as LIME (Local Interpretable Model-Agnostic Explanations) and SHAP (SHapley Additive exPlanations) provide model-agnostic interpretations by quantifying the impact of each feature on a specific prediction. While initially used in healthcare and finance, these techniques are increasingly being adopted in app analytics to understand the drivers behind predictions and guide feature or content optimization.

In addition, studies have explored the integration of sentiment analysis into predictive pipelines. Sentiment analysis allows for the quantification of user opinions and satisfaction, extracted from app reviews and ratings. Tools like VADER and Text Blob have been widely used to assign polarity scores to user reviews, which can then be fed into machine learning models as auxiliary features. Work by Mishra and Chatterjee (2022) showed that combining sentiment scores with quantitative features significantly improved classification accuracy for app performance prediction.

Despite these advancements, several gaps remain. Few studies focus on early-stage prediction, where limited usage data is available. Also, while many papers emphasize accuracy, few offer interpretability that can guide real-world decision-making. Moreover, studies often use static datasets, overlooking the evolving nature of app performance due to updates, seasonal usage trends, and market competition. These limitations highlight the need for comprehensive, explainable, and adaptive prediction frameworks—such as the one proposed in this work—which can bridge predictive performance with actionable insight.

3. Proposed Work Explanation

The proposed work introduces a robust and interpretable framework to predict mobile app success using a combination of machine learning algorithms and explainability techniques. The pipeline consists of four major components: data preprocessing, model training, prediction explainability, and actionable feedback generation through counterfactual reasoning. The objective is to create a system that is not only highly accurate but also capable of providing transparent, understandable, and useful insights to developers, marketers, and decision-makers.

3.1 Data Acquisition and Preprocessing

We start with a real-world dataset extracted from the Google Play Store, enriched by web-scraping and natural language processing to gather app descriptions, user reviews, install counts, ratings, categories, and permissions. Additional metadata includes update frequency, in-app purchases, price, and developer reputation metrics. Review data undergoes sentiment analysis using pre-trained models (e.g., VADER or BERT) to extract polarity and subjectivity scores. These features are essential in capturing user perception and early feedback signals.

The dataset is cleaned to handle missing values (using median or mode imputation), transformed using label encoding for categorical variables, and normalized to ensure numerical stability in model training. Feature correlation analysis and recursive feature elimination (RFE) are applied to reduce dimensionality and remove redundant attributes that do not contribute significantly to model performance.

3.2 Model Training and Evaluation

We employ several supervised learning models, including Logistic Regression (baseline), Random Forest (RF), and Gradient Boosting Machine (GBM). Among these, GBM is chosen as the final model due to its ability to capture non-linear interactions and high performance on tabular data. The models are evaluated using k-fold cross-validation, with key metrics including accuracy, precision, recall, and F1-score. Performance tuning is conducted using grid search for hyper parameter optimization.

The success label is binary—apps are categorized as “successful” or “unsuccessful” based on a composite score derived from the number of installs, average user rating, and review sentiment polarity. Apps with scores above a dynamic threshold are marked as successful, allowing the model to learn patterns associated with high-performing applications across categories and regions.

3.3 Explainability and SHAP Integration

To address the black-box nature of high-performing models like GBM, we integrate SHAP (SHapley Additive exPlanations) to provide global and local interpretability. SHAP values are calculated for each feature in the model to assess their contribution to the prediction. These values help identify key drivers of success such as “User Rating”, “Sentiment Polarity”, “Install Count”, and “Update Frequency”.

For example, SHAP plots show that apps with recent updates and positive sentiment tend to have higher predicted success probabilities. This insight enables developers to understand which aspects of their app most influence performance and where improvements are needed. Global SHAP summary plots give an overall picture of feature importance, while local force plots provide detailed analysis at the individual app level.

3.4 Counterfactual and What-If Analysis

An innovative aspect of our work is the incorporation of counterfactual explanations. These are generated using optimization techniques that seek the minimal set of changes to an input instance X such that the model output changes from “unsuccessful” to “successful.” For example, if an app is predicted to fail, the system might suggest that increasing its average user rating from 3.5 to 4.2 and reducing the app size from 50 MB to 30 MB would change the outcome.

This kind of “what-if” analysis is critical for early-stage app developers, as it provides concrete, data-driven feedback on how to improve their chances of success. Unlike post-hoc suggestions based on guesswork, these counterfactuals are model-aware and scenario-specific, making them highly actionable.

3.5 System Architecture and Deployment Considerations

The complete framework is designed for modular deployment. The preprocessing and training modules are implemented in Python using libraries such as Scikit-learn, XGBoost, and SHAP. A dashboard frontend can be built using tools like Streamlit or Dash to allow interactive exploration of predictions and explanations. This setup enables integration into app development environments or analytics dashboards, providing real-time decision support.

Security and ethical considerations are addressed through anonymization of personally identifiable information (PII) and adherence to data protection standards such as GDPR. The system also includes logging of all decisions and explanations for audit ability**.**

4. Results and Discussion

We evaluated our models using a dataset of over 10,000 mobile applications sourced from Kaggle and Google Play data aggregators. Features included numerical values (e.g., rating, installs, size), categorical variables (e.g., category, content rating), and derived text features (e.g., sentiment polarity of reviews).

The Gradient Boosting classifier outperformed other models with:

- Accuracy: **89.5%**
- Precision: **0.87**
- Recall: **0.90**
- F1-score: **0.885**

SHAP analysis identified the following top features:

1. **User Rating** – high ratings strongly indicate success.
2. **Number of Installs** – a key indicator of popularity.
3. **Review Sentiment** – more positive reviews lead to higher success probability.
4. **Update Frequency** – apps updated regularly tend to perform better.

5. **App Category** – certain categories (e.g., Productivity, Tools) show higher success rates.

The interpretability framework enabled clear visualization of these influences, as shown in Figure 1 (SHAP bar chart). Moreover, counterfactual tests revealed that small changes like reducing app size or improving review responses could change the predicted outcome for borderline apps.

Table 1. *Model Comparison Metrics*

Model	Accuracy	Precision	Re-Call	F1-Score
Logistic Regression	81.2%	0.78	0.79	0.786
Random forest	86.7%	0.85	0.87	0.86
Gradient Boosting	89.5%	0.87	0.90	0.885

5. Conclusion

This research presents an end-to-end pipeline for predicting the success rate of mobile applications using machine learning with integrated explainability. The model leverages both structured and unstructured data (reviews and metadata) and provides not only highly accurate predictions but also actionable insights via SHAP and counterfactual methods. This makes the tool valuable for stakeholders—developers, product managers, investors—who need to make informed decisions prior to or soon after app release.

Our work bridges the gap between opaque AI systems and transparent decision support. By making predictions interpretable, we contribute toward building trust in automated evaluation systems. Future directions include real-time prediction dashboards, integration with app stores for live monitoring, and the incorporation of user engagement time as a dynamic feature for temporal analysis.

References

1. J. Lee; K. Park; S. Kim, Year: 2022, “A Predictive Framework for Mobile App Success Using Multivariate Analysis”, Mobile Computing Journal, Vol: 19, No: 3, pp. 102–112.

2. Marco Tulio Ribeiro; Sameer Singh; Carlos Guestrin, Year: 2016, “Why Should I Trust You?: Explaining the Predictions of Any Classifier”, ACM KDD, pp. 1135–1144.

3. Scott M Lundberg; Su-In Lee, Year: 2017, “A Unified Approach to Interpreting Model Predictions”, NeurIPS, pp. 4765–4774.

4. H. Nguyen; T. Ngo; V. Huynh, Year: 2023, “Sentiment-Driven Success Modeling for Mobile Applications”, Applied Intelligence, Vol: 58, No: 1, pp. 55–68.