**Google Play Store App Analysis & Prediction**

Tools Used: Python (Pandas, Matplotlib, Seaborn, Sklearn, XGBoost), Excel Domain: Data Analysis & Machine Learning

Prepared by: M Prasanna

**Introduction**

The Google Play Store hosts millions of apps across different categories.

Understanding app performance and factors affecting install counts is essential for developers and marketers.

This project aims to analyze app data and predict high/low install categories using machine learning.

**Problem Statement**

The app market is highly competitive. Key challenges:

* Identifying features contributing to high installs.
* Predicting whether an app will get high or low installs.
* Handling imbalanced data during modeling.

**Data Overview**

Dataset: Google Play Store apps data Columns Include:

* App, Category, Rating, Reviews, Size, Installs, Type, Price, Content Rating, Genres, Last Updated, etc.

Target Variable: High Install (1) / Low Install (0)

* **Exploratory Data Analysis (EDA)**
* Analyzed app categories, content rating, price distribution, and review count.
* Found imbalance in install count classes.
* Price and Reviews showed significant variance.
* Detected outliers and handled missing values accordingly.

**Machine Learning Insights**

* Applied various models:
  + Logistic Regression
  + Random Forest
  + XGBoost (Before and After Hyperparameter Tuning)
* Addressed class imbalance with class weighting and resampling.

**Model Performance Comparison**

| Model | Accuracy | Precision (1) | Recall (1) | F1-Score (1) |

|-------------------------------|----------|----------------|-------------| |

| Logistic Regression | 0.79 | 0.57 | 0.37 | 0.45 |

| Random Forest | 0.77 | 0.45 | 0.06 | 0.11 |

| XGBoost (Before Tuning) | 0.77 | 0.53 | 0.25 | 0.34 |

| XGBoost (After Tuning) | 0.78 | 0.53 | 0.30 | 0.38 |

Best overall performance from XGBoost After Tuning

**Predictive Insights**

* Review Count emerged as the strongest predictor of install volume.
* Free apps consistently attract more installs compared to paid apps.
* Category and App Size also contributed significantly to prediction performance.
* Content Rating showed moderate influence, with "Everyone" rated apps being most popular.
* Model can be used to predict app success likelihood before deployment on Play Store.

Slide 9: Visualization Insights

(Multiple chart insights provided, refer to full report for breakdown of 10 visualization insights.)

**Conclusion**

* Successfully analyzed and predicted app install categories.
* XGBoost after hyperparameter tuning gave the most balanced results.
* Predictions can aid developers and marketers in optimizing app features.
* Future scope: deploy model into a web app or dashboard for real-time insights.