Model Training (Predictive and Descriptive Analysis)

Goal	Model Type	Algorithm	Inputs (Features)	Output
Predict app popularity (installs)	Regression	Linear Regression, Random Forest	Rating, Reviews, Price, Category, Year	Predicted installs
Predict user satisfaction (rating)	Regression	XGBoost	Type, Installs, Reviews, Price	Predicted rating
Cluster apps by market behaviour	Clustering	K-Means	Installs, Ratings, Reviews	Market segments
Identify review topics	NLP	LDA Topic Modelling	Text reviews	Top concern clusters
Classify app success (High vs Low)	Classificatio n	Logistic Regression, Decision Tree	Category, Reviews, Type	Success label

Tools and Technologies

Category	Tool	Use
Data Handling	Python (Pandas, NumPy)	Cleaning and feature engineering
Visualization	Seaborn, Plotly, Dash	Interactive graphs
Modeling	scikit-learn, XGBoost, spaCy	Regression, classification, NLP
Collaboration	GitHub, Google Colab	Code versioning and team tracking
Documentation	Google Docs, Notion	Project reports and literature review