Software Requirements Specification

Pharmaceutical Success Prediction Tool

1. Functional Requirements

Data Input:

Load an Excel file (pharmaceutical_industry_sample.xlsx) containing drug-related data:

Columns: Drug Name, Company, Category, Approval Year, Revenue (USD millions), Region.

Generate a new binary target column named Success based on Revenue threshold.

Modeling & Prediction:

Train a Random Forest Classifier to predict Success.

Evaluate the model using:

- Accuracy score
- Confusion matrix
- Classification report

Visualization:

Generate a bar chart displaying the feature importance of predictors.

Save the visualization as a PDF file.

2. Non-Functional Requirements

Software Requirements Specification

Pharmaceutical Success Prediction Tool

Platform:

- OS: Windows, macOS, or Linux
- Environment: Python 3.7+

Performance:

- Should efficiently handle small datasets (up to ~1000 rows)

Usability:

- Run via a Python script or Jupyter Notebook
- Command-line friendly

3. Software Requirements

Programming Language:

- Python 3.7 or higher

Libraries/Dependencies (installable via pip):

- pandas - for data manipulation
- scikit-learn - for machine learning modeling
- seaborn - for visualization
- matplotlib - for plotting

Installation Command:

Software Requirements Specification

Pharmaceutical Success Prediction Tool

pip install pandas scikit-learn seaborn matplotlib

4. Input File Format

Excel file named pharmaceutical_industry_sample.xlsx located at:

- ./pharmaceutical_industry_sample.xlsx or /mnt/data/pharmaceutical_industry_sample.xlsx

Required Columns:

- Drug Name, Company, Category, Approval Year, Revenue (USD millions), Region