Software Requirements Specification

Delayed Defect Detection Tool

1. Functional Requirements

Data Input:

- Load a CSV file (`defect_data.csv`) with columns:

  - Defect_ID, Product_ID, Production_Date, Detection_Date, Defect_Type, Severity, Root_Cause

Feature Engineering:

- Compute Detection_Delay_Days
- Label Delayed_Detection (1 if Delay > 30 days, else 0)

Machine Learning Model:

- Train Random Forest Classifier
- Evaluate performance with classification report
- Save model as defect_delay_model.pkl

User Output:

- Display classification report
- Predict delayed detection for new data
- Export results to Excel and PDF

2. Non-Functional Requirements

Platform:

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

Performance:

- Handle up to 100,000 records efficiently
- Prediction latency < 5 seconds

Usability:

- Run via Python scripts (Jupyter, VS Code, PyCharm)
- Streamlit dashboard optional

3. Software Requirements

Programming Language:

- Python 3.7 or higher

Dependencies (pip install):

- pandas, numpy, scikit-learn, joblib, streamlit, matplotlib, seaborn, fpdf, openpyxl

Library	Purpose
pandas	Data loading & manipulation
numpy	Numerical operations
scikit-learn	ML model training & evaluation
joblib	Model serialization
streamlit	Web dashboard (optional)
matplotlib	Visualizations
seaborn	Enhanced plots
fpdf	PDF generation
openpyxl	Excel file operations

4. Input File Format

- CSV or Excel file (`defect_data.csv` / `defect_data.xlsx`)
- Required columns:
  - Production_Date, Detection_Date, Defect_Type, Severity, Root_Cause

5. Output Files

- Model: defect_delay_model.pkl
- Predictions: Predicted_Delayed_Defect_Detection.xlsx
- PDF Report: Predicted_Delayed_Defect_Detection.pdf