
Manufacturing Defects Prediction

DA24 C4
27 October 2024

By Wan Mohd Azzid bin Wan Hossen





Self Introduction

Education

- Diploma in Management at Open University Malaysia (2010)

Work Experience

- Ordnance Officer at Malaysian Army (2010 - 2021)

About Me

- Interest in gaining new experience, knowledge & skills
- 10 years' experience as an officer of organization & management
- My motto: 'Knowledgeable maketh success!'

Course Timeline

1

Month 1

- Self Resilience Mastery
- Communication Mastery

2

Month 2

- Introduction to Data Science
- Introduction to Python Programming

3

Month 3

- Introduction to Machine Learning
- Introduction to Object Detection with AI

4

Month 4

- Advanced Deep Learning Techniques
- Career Launchpad Mastery

5

Month 5

- Final Review



Project Overview

Insights into Factors Influencing Defect Rates and Production Efficiency

Project Problem Statement

This dataset provides insights into factors influencing defect rates in a manufacturing environment.



Objective

- Identify primary factors contributing to manufacturing defects.
- Balance the dataset effectively.
- Predict defect rates with accuracy.

Technology Stack

1

Exploratory Data Analysis

- Pandas, NumPy, Matplotlib & Seaborn
- Random Forest (Feature Importances)

2

Scaled & Balanced Values

- MinMaxScaler
- SMOTE

3

Model Development

- Scikit-Learn
- XGBoost & LightGBM

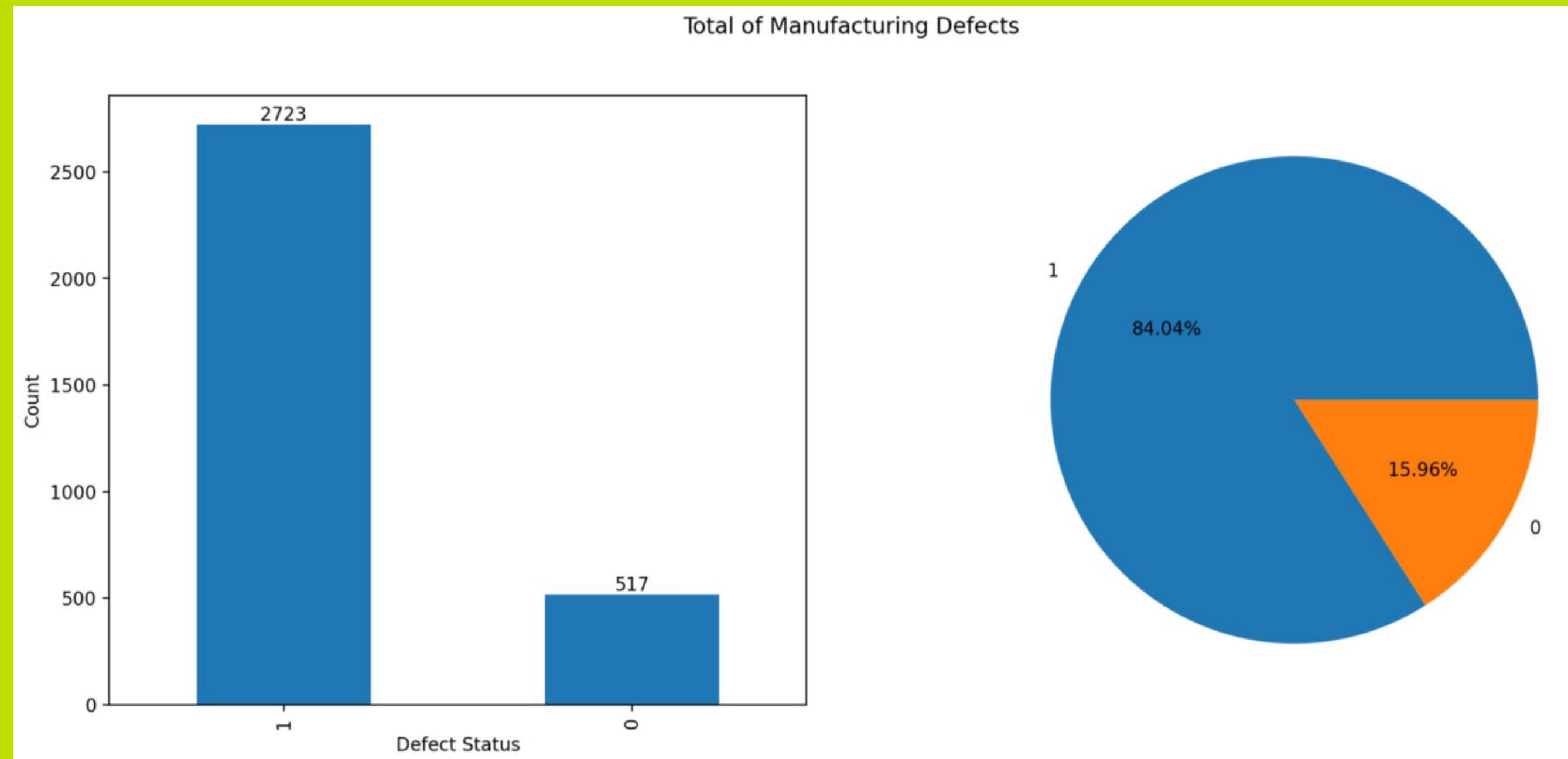
4

Model Deployment

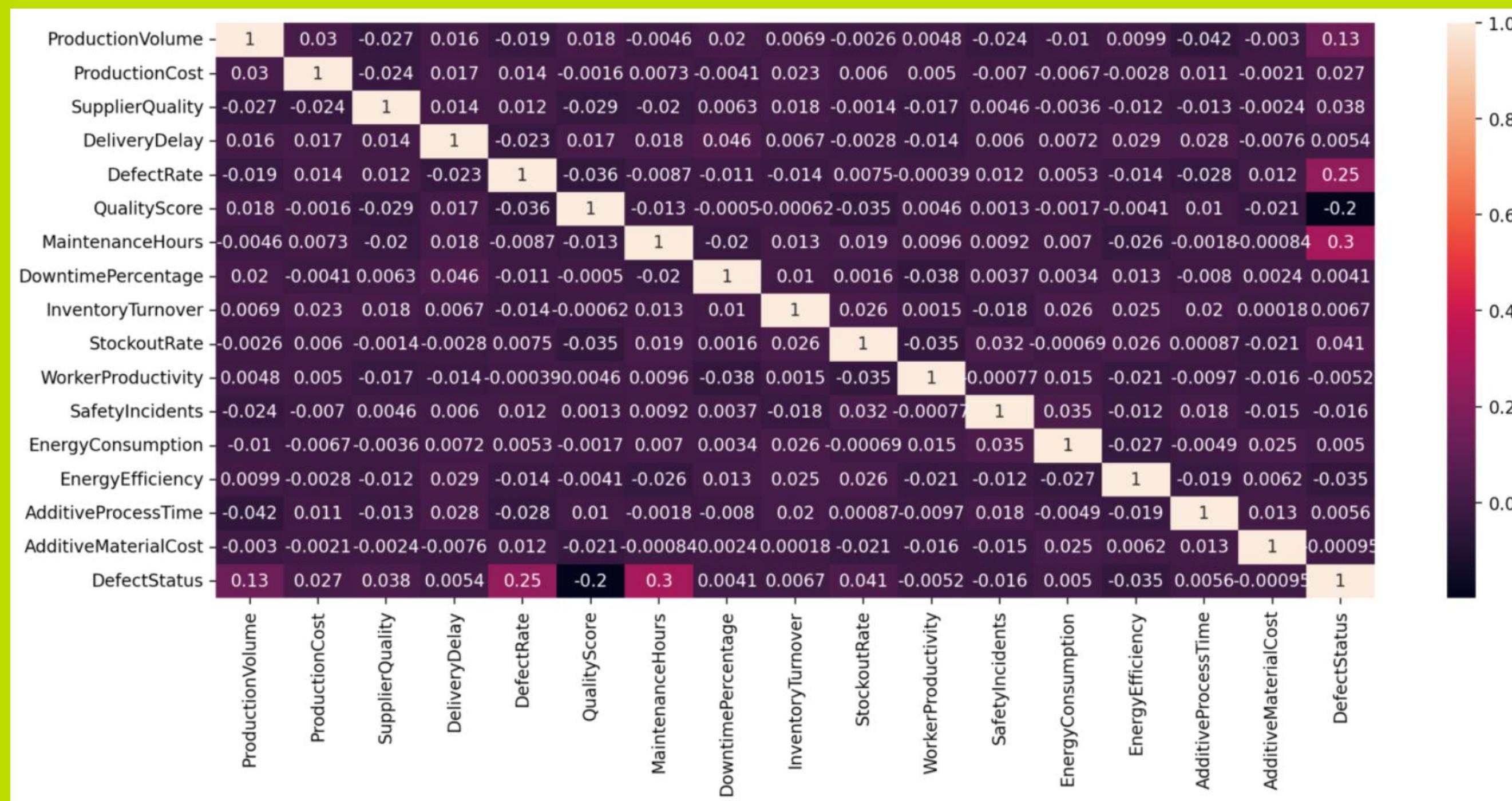
- Streamlit
- FastAPI



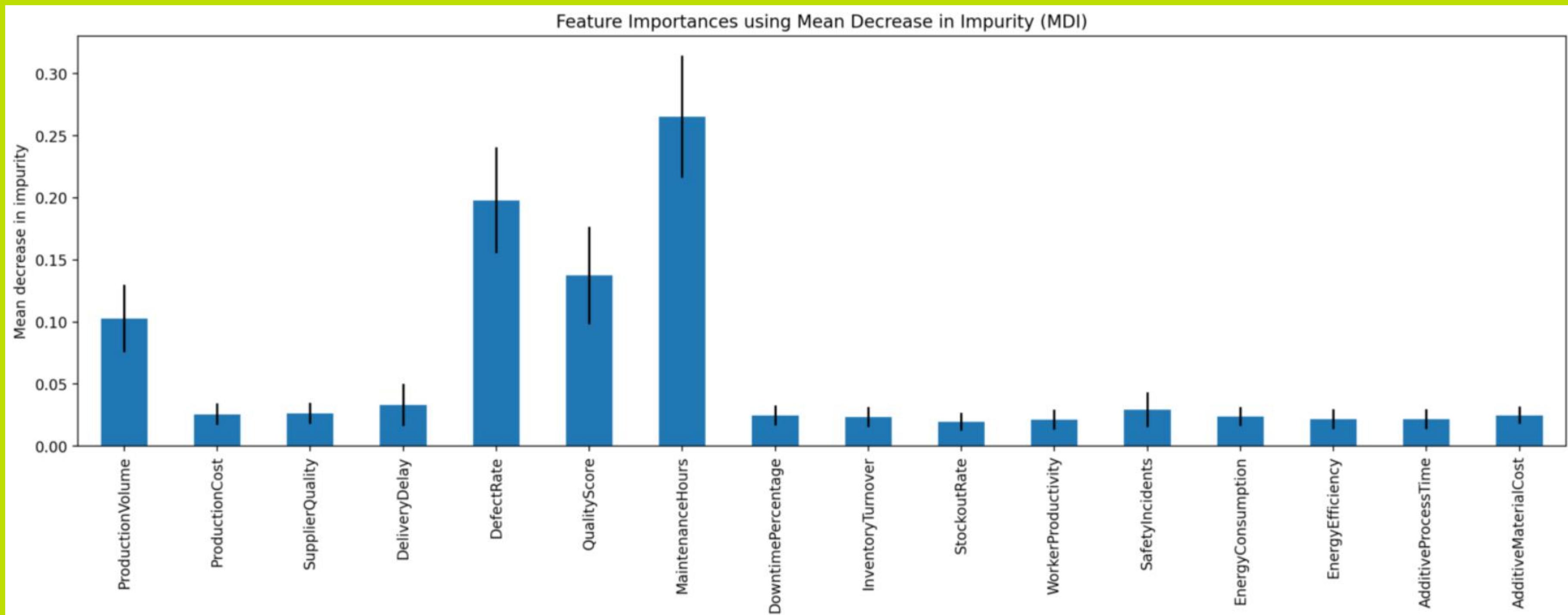
Dataset Summary



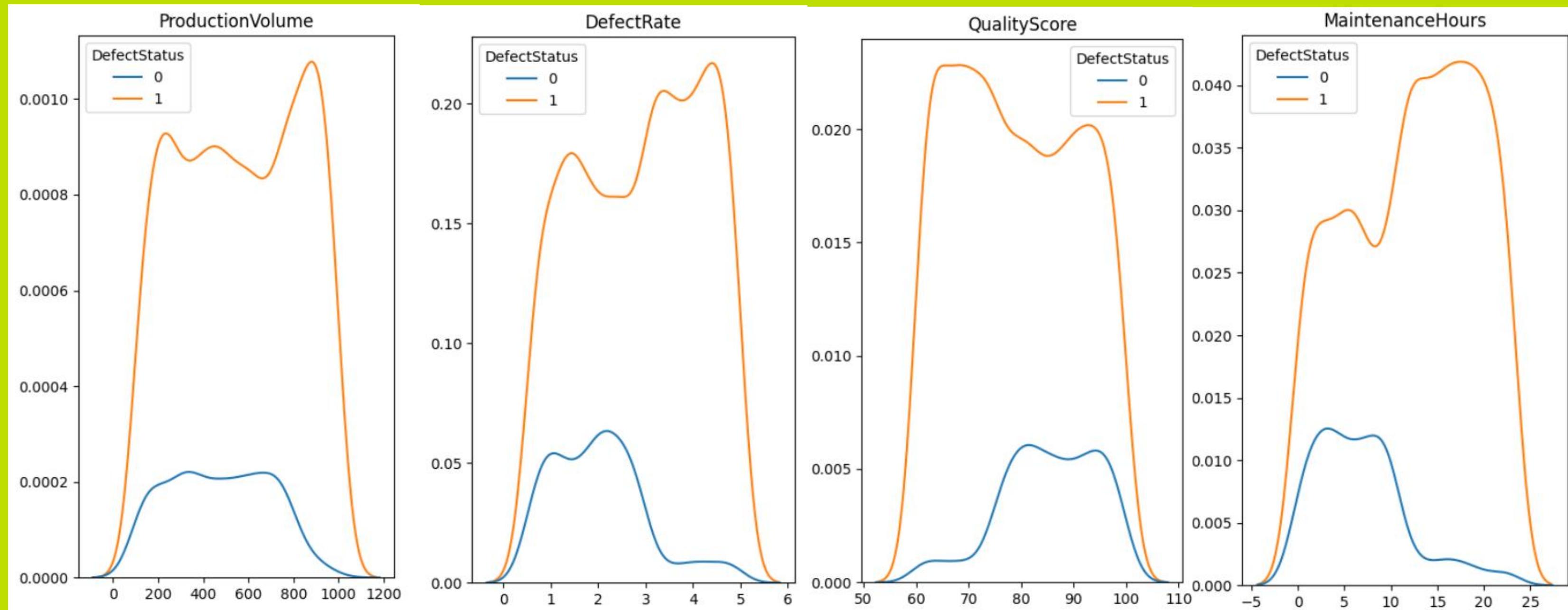
Correlation Matrix



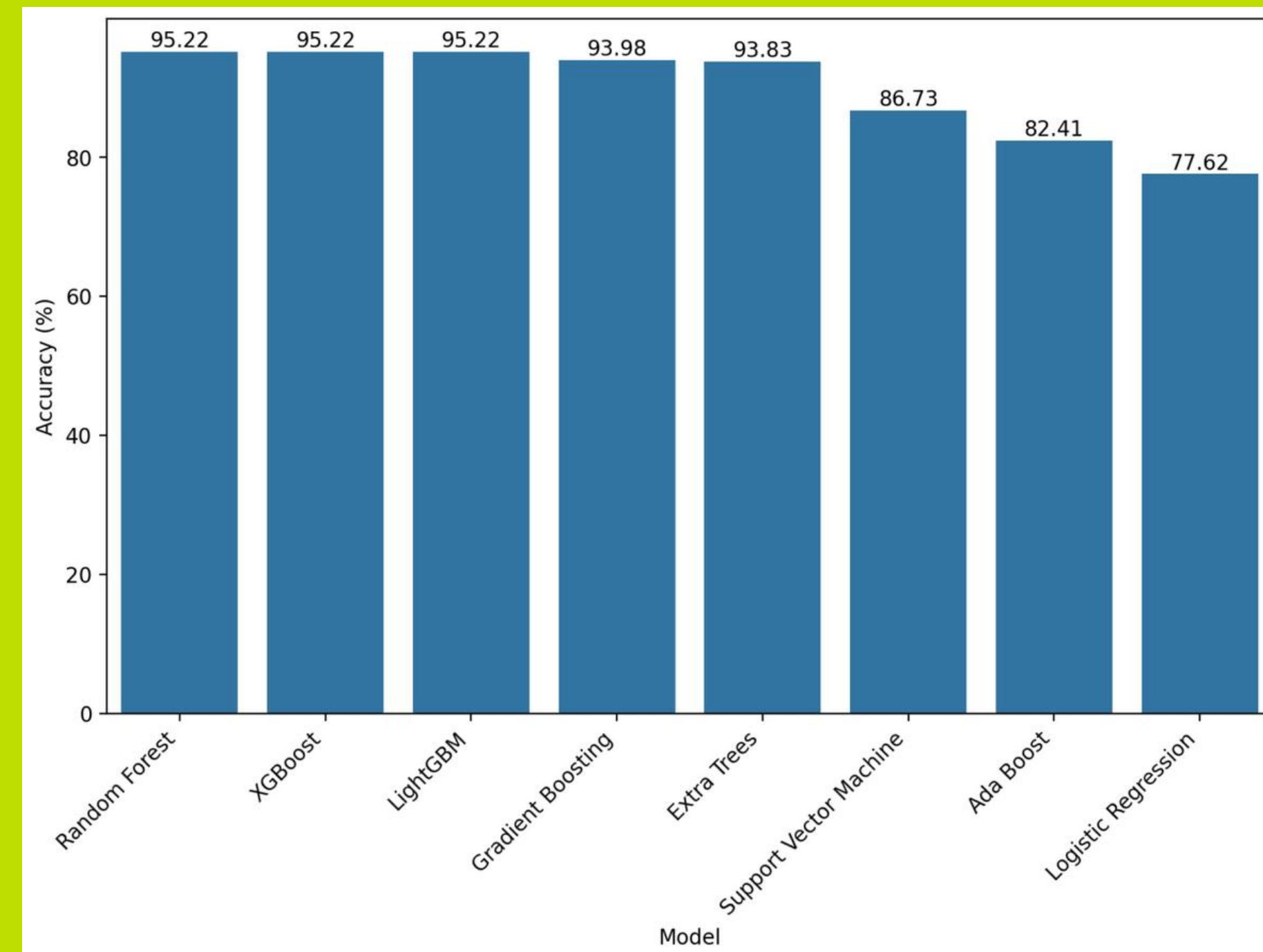
Feature Importances (Random Forest)



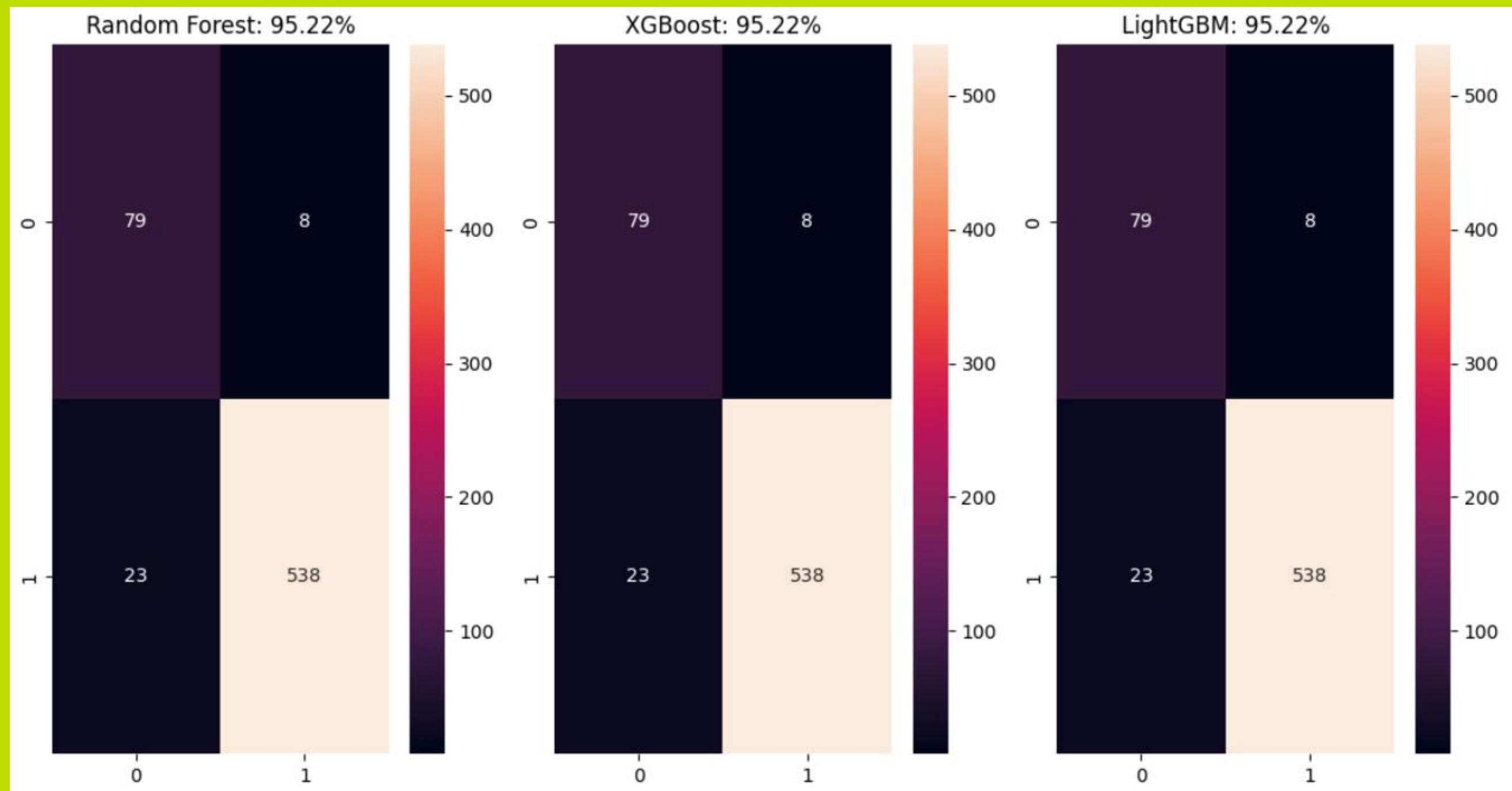
Exploratory Data Analysis



Model Accuracy Comparison



Confusion Matrices



Hyperparameter Tuning

```
param = {
    'n_estimators': [100, 200, 300],
    'max_depth': [10, 20, 30],
    'min_samples_split': [2, 5, 10]
}
rfc_grid = GridSearchCV(RandomForestClassifier(),
                        param,
                        refit = True,
                        cv = 5,
                        verbose=3)

rfc_grid.fit(x_train, y_train)
```

```
rfc_grid.best_params_
```

```
{'max_depth': 30, 'min_samples_split': 2, 'n_estimators': 100}
```

```
rfc_grid.best_score_
```

```
0.9570487087110864
```

Project Timeline



Week 1 & 2

- Project Planning & Setup (2 Weeks)



Week 12

- Documentation & Deployment (1 Week)



Week 13 - 14

- Evaluation & Iteration (2 Weeks)



Week 15

- Project Closure & Handover (1 Week)



Week 6 - 9

- Modeling & Analysis (4 Weeks)



Week 10 - 11

- Insights & Reporting (2 Weeks)

Project Budget

Item	Initial Investment
System Deployment (including personnel, data acquisition, infrastructure, software, testing & maintenance)	RM 50,000 - RM 200,000

Conclusion



Defect's Factors

- Number of units produced per day.
- Quality ratings of suppliers.
- Defects per thousand units produced.
- Hours spent on maintenance per week.

Defect's Prediction

- Random Forest model proved the most useful & accurate.
- The percentage of defects make the company choose the best decision.



**Thank You For
Your Time**