

Defect Analysis Dashboard

Dataset: Manufacturing Defect Data

Goal: Identify key drivers of defects

Model: Random Forest

0.27%

AVG DEFECT RATE

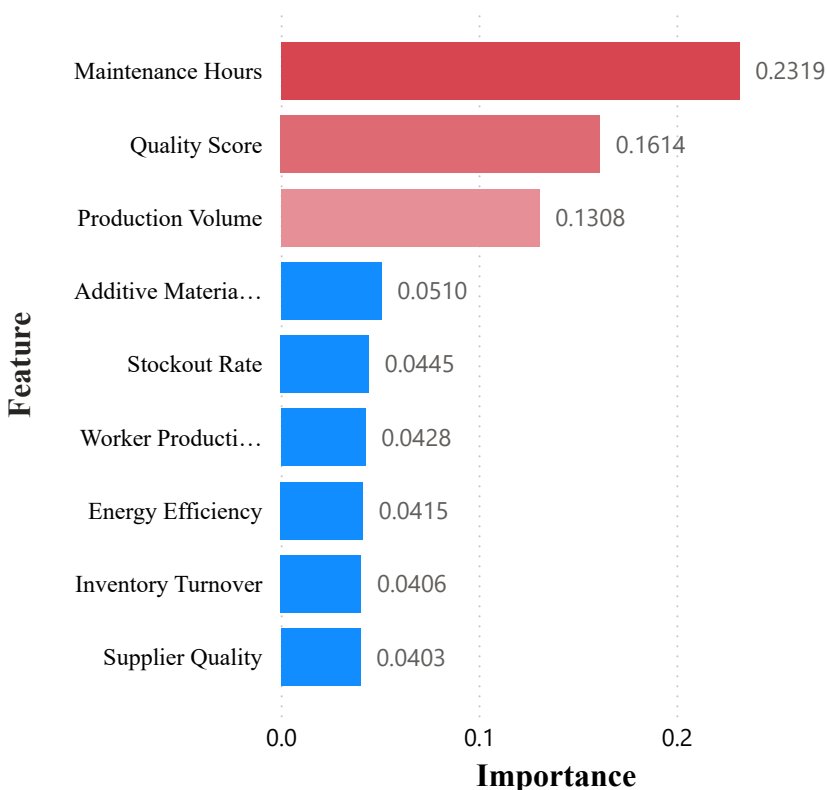
11.48

AVG MAINTENANCE HOURS

0.86

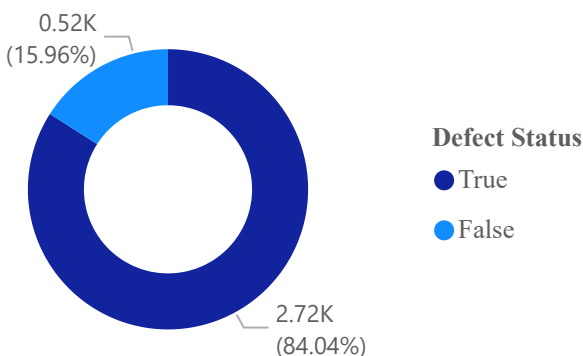
MODEL ACCURACY

Key Drivers in Manufacturing Defects



- Maintenance hours is a strong predictor of defect rate,
- Aim for product to achieve maximum quality with least amount of maintenance hours
- Higher volume may increase risk of defects. Aim to keep production volume moderate.

Proportion of Defects by Defect Status



Classification Report of Random Forest Model

Defect Status	Precision	F1-Score	Recall	Support
No Defect	0.52	0.56	0.59	143
Defect	0.93	0.92	0.91	829
Macro Avg	0.73	0.74	0.75	972
Weighted Avg	0.87	0.86	0.86	972

This dashboard uses a **Random Forest Classifier** trained with synthetic oversampling (**SMOTE**) to predict whether a manufactured product will be defective based on operational and material factors.

The model achieved an **accuracy** of **86%**, with particularly high performance in identifying defective items (**Recall = 92%**) while improving minority class recall from **28% to 59%** as it accounts for the class imbalance. This model is balanced and production-ready.

MODEL SELECTION TO PREDICT MANUFACTURING DEFECTS

Random Forest vs Logistic Regression Model

0.86

Random Forest Accuracy

0.74

Logistic Regression Accuracy

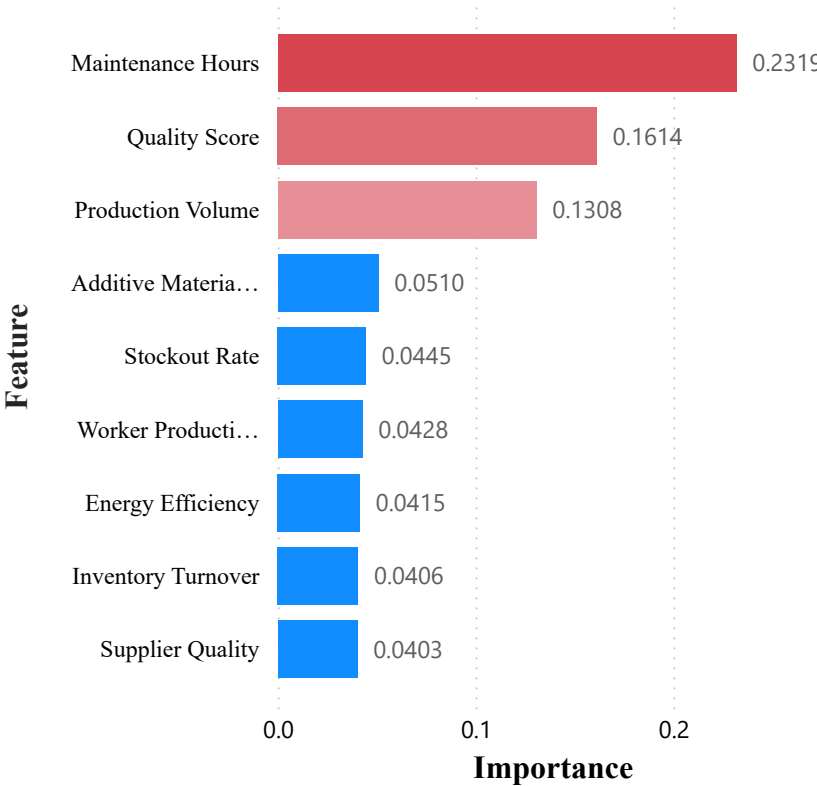
Classification Report of Random Forest Model

Defect Status	Precision	F1-Score	Recall	Support
No Defect	0.52	0.56	0.59	143
Defect	0.93	0.92	0.91	829
Macro Avg	0.73	0.74	0.75	972
Weighted Avg	0.87	0.86	0.86	972

Classification Report of Log Regression Model

Defect Status	Precision	F1-score	Recall	Support
Defect	0.93	0.83	0.75	806
Macro Avg	0.65	0.66	0.73	972
No Defect	0.37	0.49	0.71	166
Weighted Avg	0.83	0.77	0.74	972

Random Forest Importance by Feature

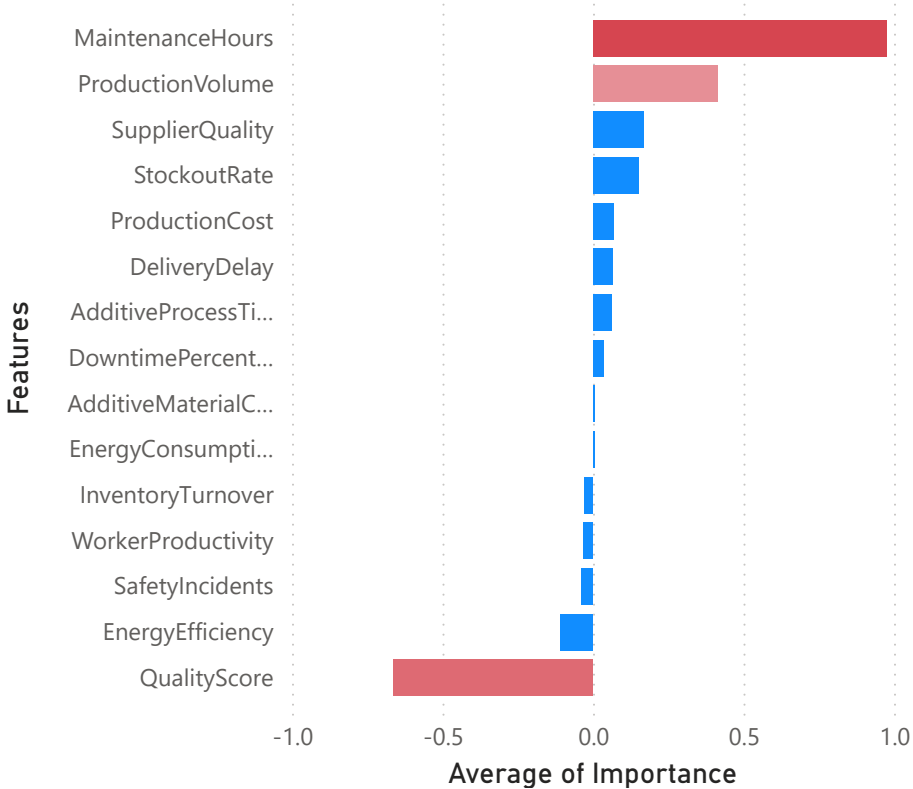


Random Forest a better predictive model because it has a higher accuracy score (86% > 74%) and precision score for both defects and non-defects than the logistic regression model.

Random Forest improved

- Precision for minority class by 15%
- Recall for defects by 16%
- F1 scores --> better balance and predictive power

Log Regression Importance by Feature



Logistic Regression model is more interpretable: higher quality score is linked to lower defect rates while higher maintenance hours is linked to higher defect rates.