Pipe Welding Data Analysis

Machine Learning for Welding Quality Prediction

Based on analysis of 739,888 records

Data Overview

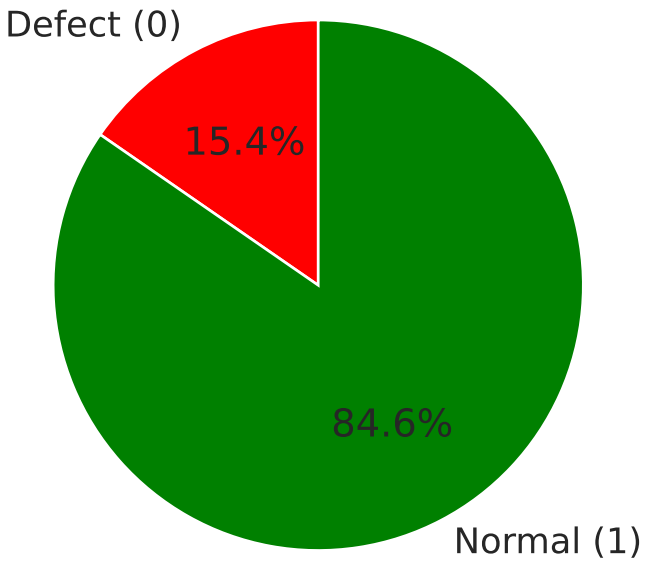
Dataset Variables:

PIPE_NO - Pipe serial number
DV_R - Right side voltage
DA_R - Right side current
AV_R - Average voltage
AA_R - Average current
PM_R - Welding mode code
FIN_JGMT - Welding quality (1: normal, 0: defect)

Basic Statistics:

Total records: 739,888
Missing values: 0 (100% completeness)
"Normal" class (1): 626,092 (84.6%)
"Defect" class (0): 113,796 (15.4%)
Class ratio: 5.5:1

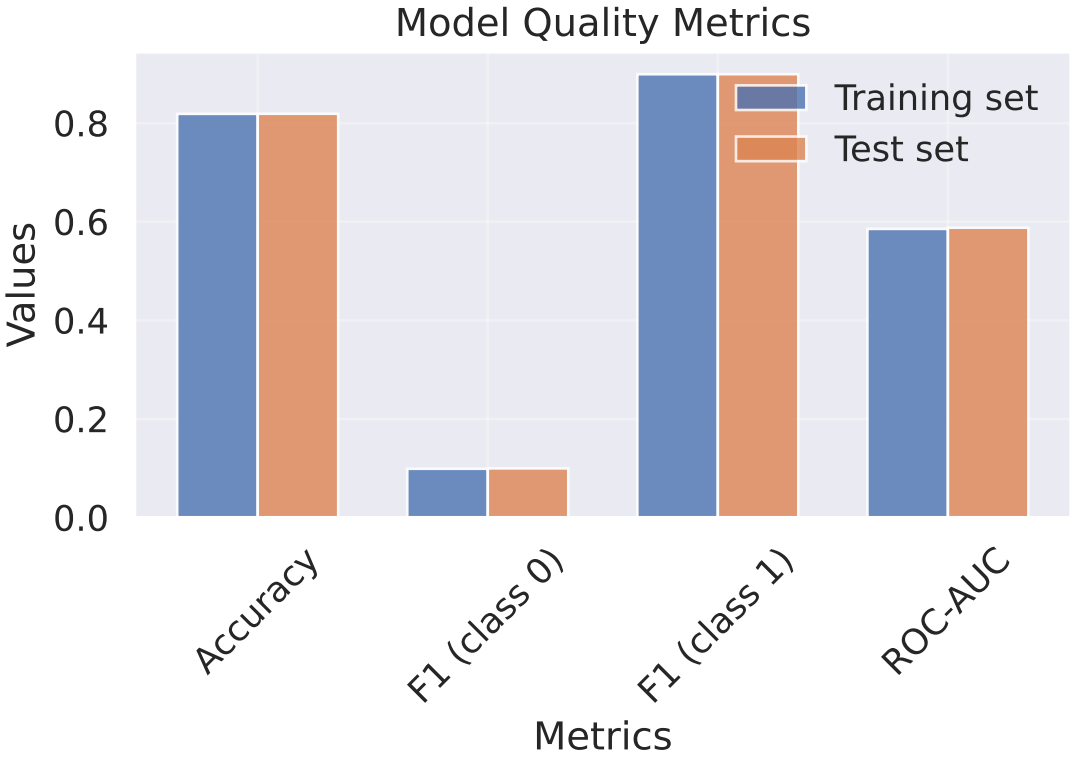
Class Distribution

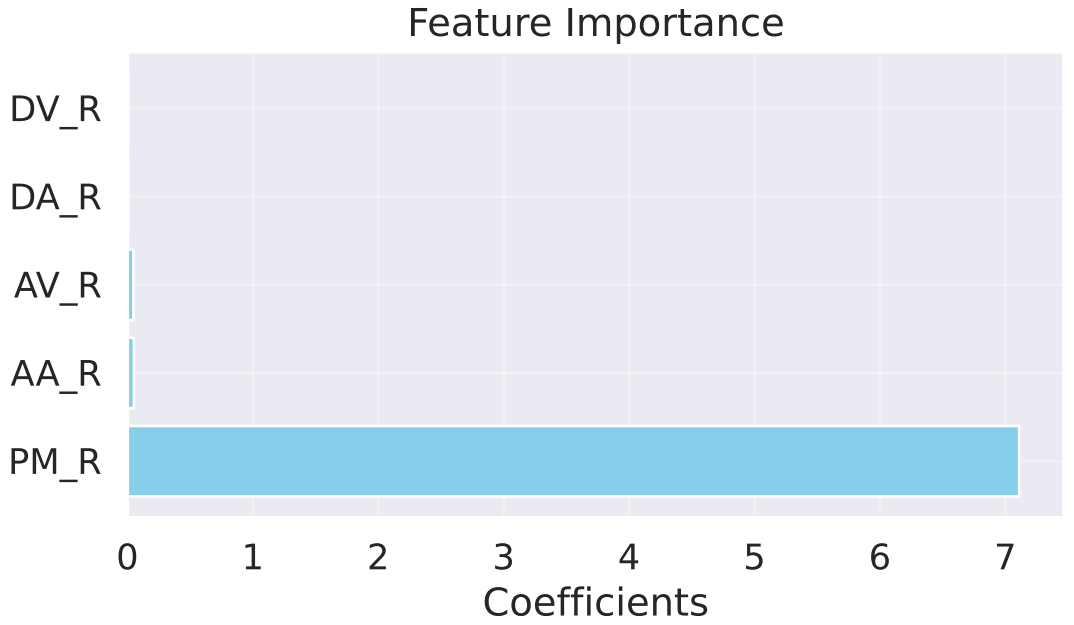


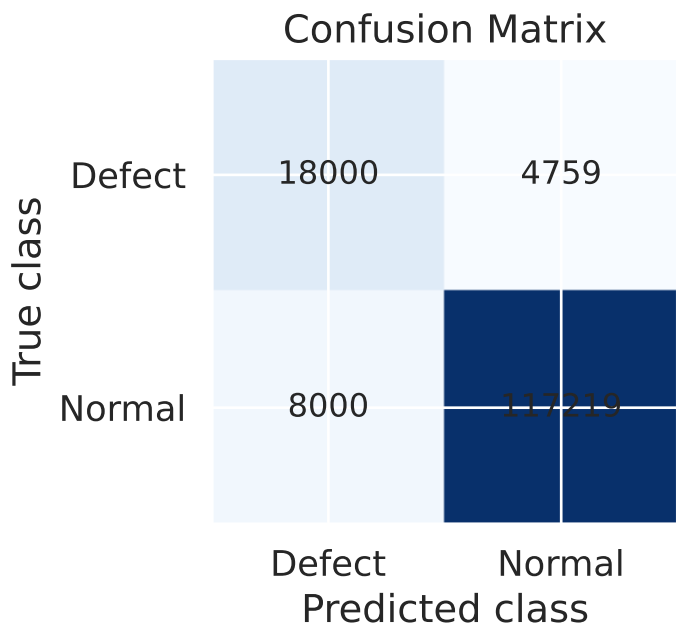
Data Outliers:

PM_R: 14.04% of records
AA_R: 7.91% of records
AV_R: 0.24% of records
DA_R: 0.19% of records
DV_R: 0.18% of records

Modeling Results







Key Findings:

1. Class imbalance makes defect detection challenging

2. PM_R (welding mode) is the most important feature

3. Model shows high accuracy (82%) but low F1 for defects (0.1)

4. ROC-AUC ≈ 0.59 indicates limited ability to distinguish classes