Data source:
Pleschberger, Martin, Zernig, Anja, & Kaestner, Andre. (2020). Equipment Sensor Data from Semiconductor Frontend Production (Version 1) [Data set]. Zenodo. http://doi.org/10.5281/zenodo.4282611

EDA:
Equipment 1 has 24 sensors (ranging from sensor 1 to sensor 24)
Equipment 2 has 31 sensors (ranging from sensor 25 to sensor 56)

Response has lot, wafer as features and response and class as targets.

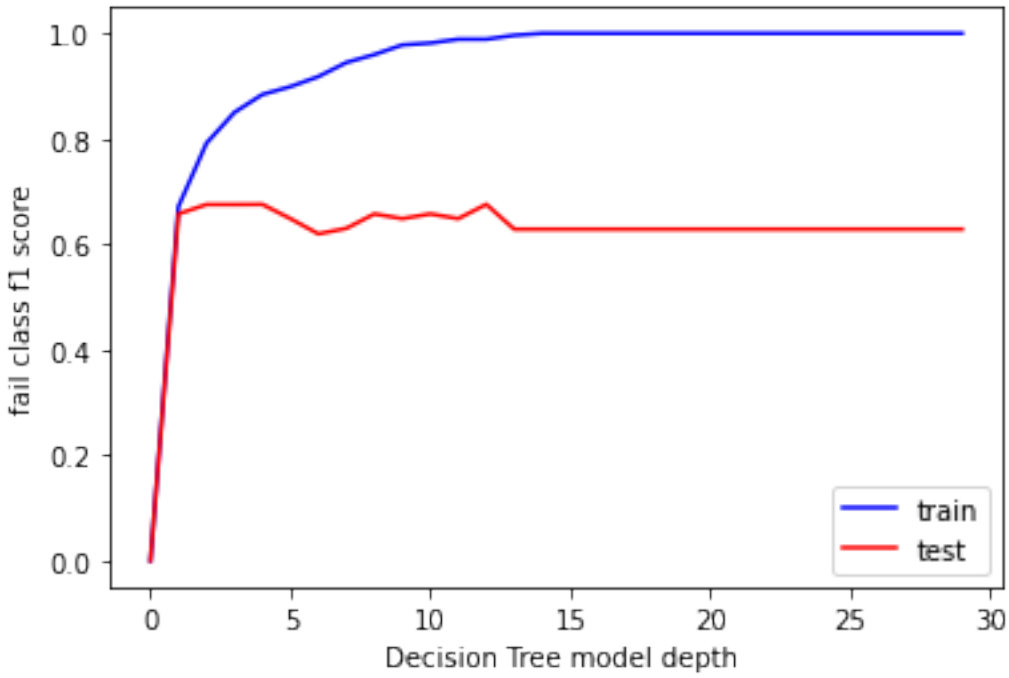
Response value > 0.75 equates to a Class of 1
Response value < 0.75 equates to a Class of 0

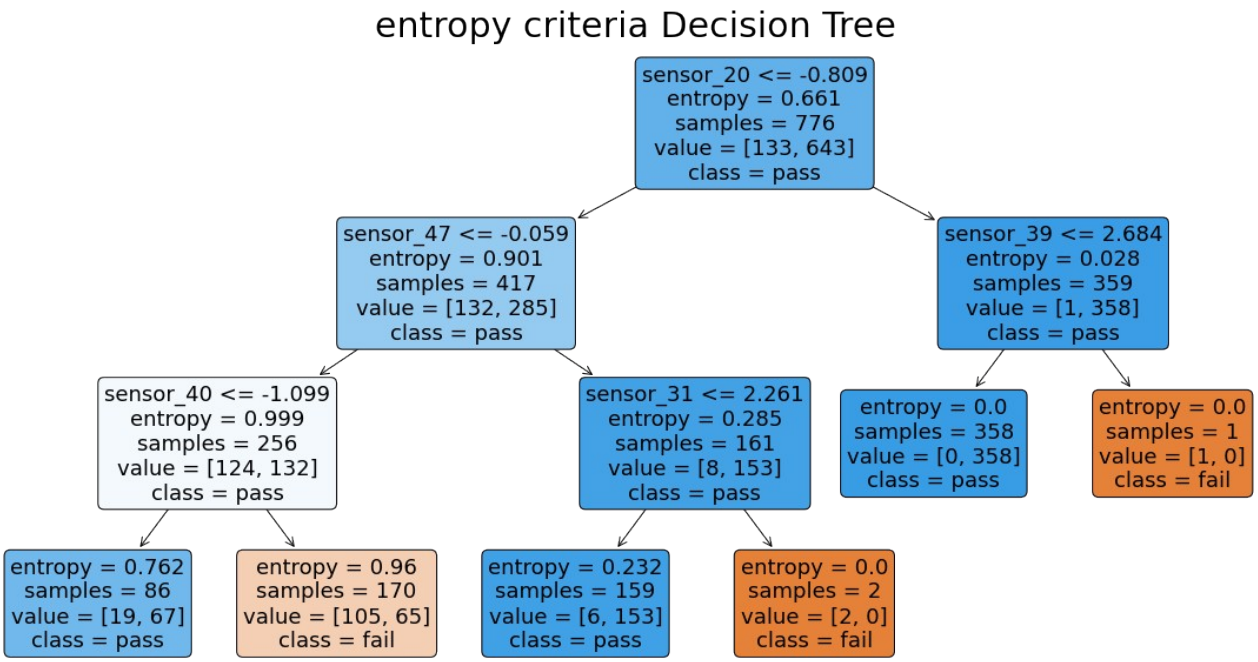
After merging data that is grouped by lot and wafer, 348 observations that were missing Equipment 1 sensor data were dropped.

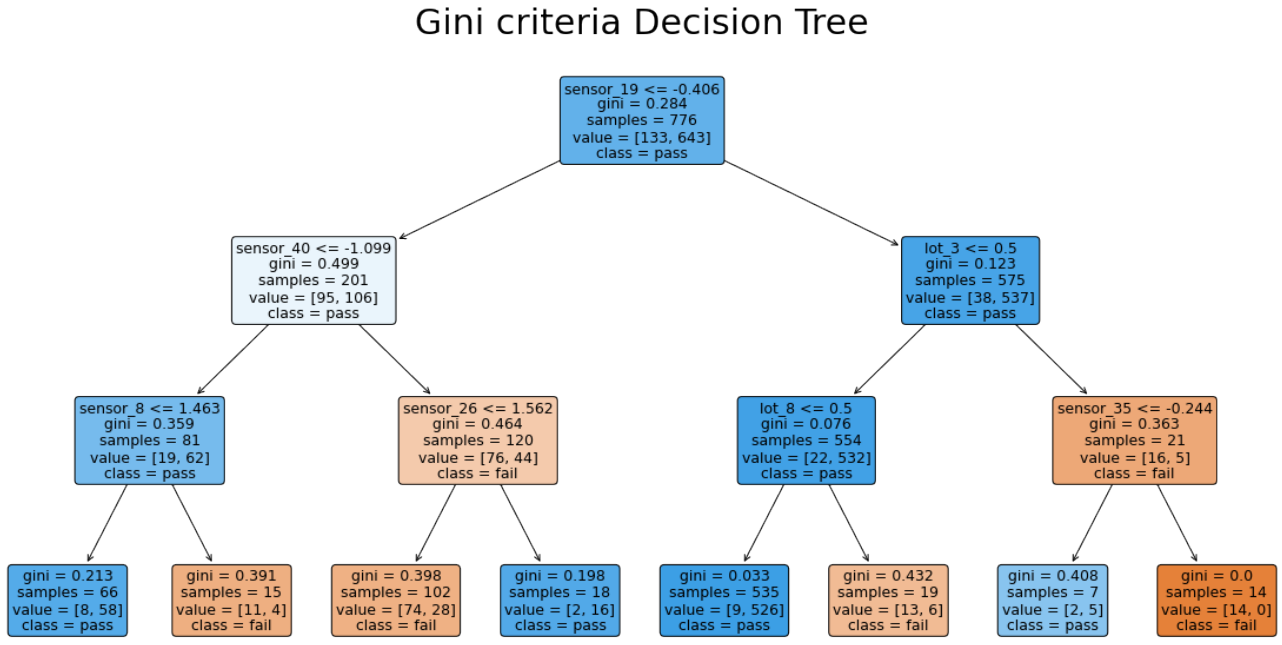
Lot and Wafer categorical features were encoded to be nominal data.

Models:
Logistic regression classifier, decision tree classifier and the random forest ensemble methods were evaluated.

Hyper parameters tuned: tree max_depth and decision criterion for Decision tree model and max_iter for Logistic regression classifier







Metrics table:

Dataset	Model	Hyper parameters	F1 Score
Train	Logistic Reg. Classifier	(max_iter:550)	0.70
Test	Logistic Reg. Classifier	(max_iter:550)	0.77
Train	Decision Tree Classifier	(max_depth: 3, Gini)	0.77
Test	Decision Tree Classifier	(max_depth: 3, Gini)	0.77
Train	Decision Tree Classifier	(max_depth: 3, Entropy	0.68
Test	Decision Tree Classifier	(max_depth: 3, Entropy)	0.72