Our proposed approach will work with data collected from legacy CNC, in particular to bridge the connection to the edge network and allow for downtime prediction. To validate the effectiveness of the predictive approach we explore the dataset from [1] which is an open-source predictive maintenance dataset. The dataset consists of features that include:

* Air temperature [K]
* Process temperature [K]
* Rotational speed [rpm]
* Torque [Nm]
* Tool wear [min]
* Machine Failure

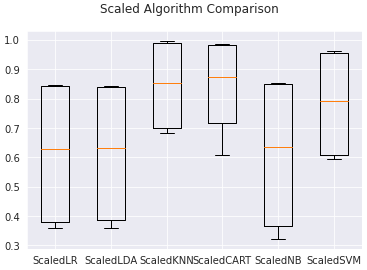
Our aim was to train a collection of machine learning models that can predict machine failure cases to help in preventive maintenance of the legacy machines. The machine failure is set to true when any of the following failures are encountered: Tool wear failure, heat dissipation failure, power failure, overstrain failure and random failure. Thus, the dataset is a strong characterization of the real-life scenarios that lead to CNC machine failures. The dataset comprises of 10,000 data points.

**Data Preprocessing:**

The data was initially inspected for missing values and the ‘Type’ feature associated with product ID were represented with letters L, M and H which denote the quality variants. In order to incorporate it to our machine learning models, we performed ordinal transformation on the data to further drive the machine learning models to seek patterns that might be a result of information embedded in the ordinality. The data label ‘Machine Failure’ was imbalanced as a result we performed synthetic minority oversampling technique to reach at a balanced training set.

**Machine Learning Models:**

A host of classifier-based models were experimented with linear machine learning (ML) models included Logistic Regression and Linear Discriminant Analysis. We also included nonlinear algorithms such as Decision Tree Classifiers, Support Vector Classifiers, Gaussian Naïve Bayes and K-Nearest Neighbors. Initially we performed analysis on this set of ML. The evaluation metric we considered was the f-beta score. The f-beta score takes into account both precession and recall with that places extra weight on the recall. The box plot of the ML models is given in Figure 1.



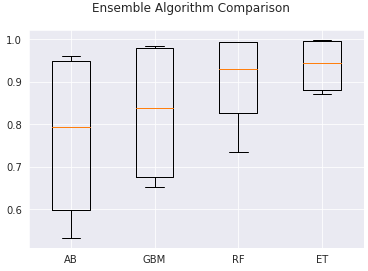
***Figure 1: Box plot of f-beta score for both linear and non-linear machine learning models***

We see that among both the linear and non-linear models KNN and CART (Decision Trees) perform the best in terms of f-beta score a summary of the f-beta scores is presented in Table I.

**Table 1:** Linear and Non-Linear ML prediction model generation

|  |  |  |
| --- | --- | --- |
| ML Model | Mean f-beta Score | Std dev f-beta Score |
| K-Nearest Neighbor | 0.845 | 0.14 |
| Decision Tree | 0.841 | 0.14 |
| Support Vector Machine | 0.783 | 0.17 |
| Linear Discriminant Analysis | 0.614 | 0.22 |
| Logistic Regression | 0.612 | 0.22 |
| Naïve Bayes | 0.607 | 0.24 |

As part of the experiment, we also applied ensemble modeling to improve our predictions. The ensembling approaches we used to predict machine failure were Ada-Boost classifier, Gradient Boost classifier, Random Forest classifier and Extra Tree classifier. The best performance was demonstrated by Random Forest and Extra Tree with Extra Tree classifier reaching a f-beta score of 0.937 and a standard deviation of 0.06. The comparison plot of the results is presented in the Figure 2.



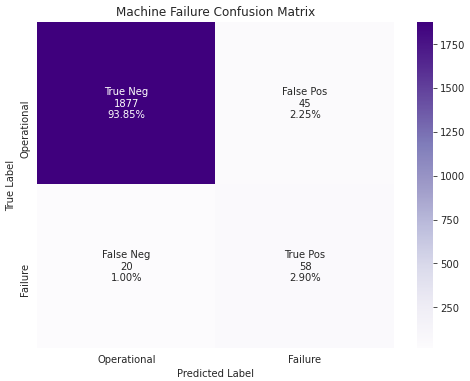
***Figure 2: Box plot of f-beta scores for ensemble learning models***

Table to further breaks down the results of all the models with mean f-beta score and corresponding standard deviations.

**Table 2:** Ensemble ML prediction mean f-beta scores and standard deviation

|  |  |  |
| --- | --- | --- |
| ML Model | Mean f-beta Score | Std dev f-beta Score |
| Extra Tree | 0.937 | 0.05 |
| Random Forest | 0.903 | 0.09 |
| Gradient Boosting | 0.828 | 0.15 |
| Ada-Boost | 0.772 | 0.18 |

Therefore, the overall best performance was that of Extra Tree ensembling model. In Figure 3 we present the confusion matrix for the machine failure prediction as part of the our frame work to allow prediction of down time.



***Figure 3: Confusion matrix for Extra Tree ensemble learning model on the machine failure prediction***