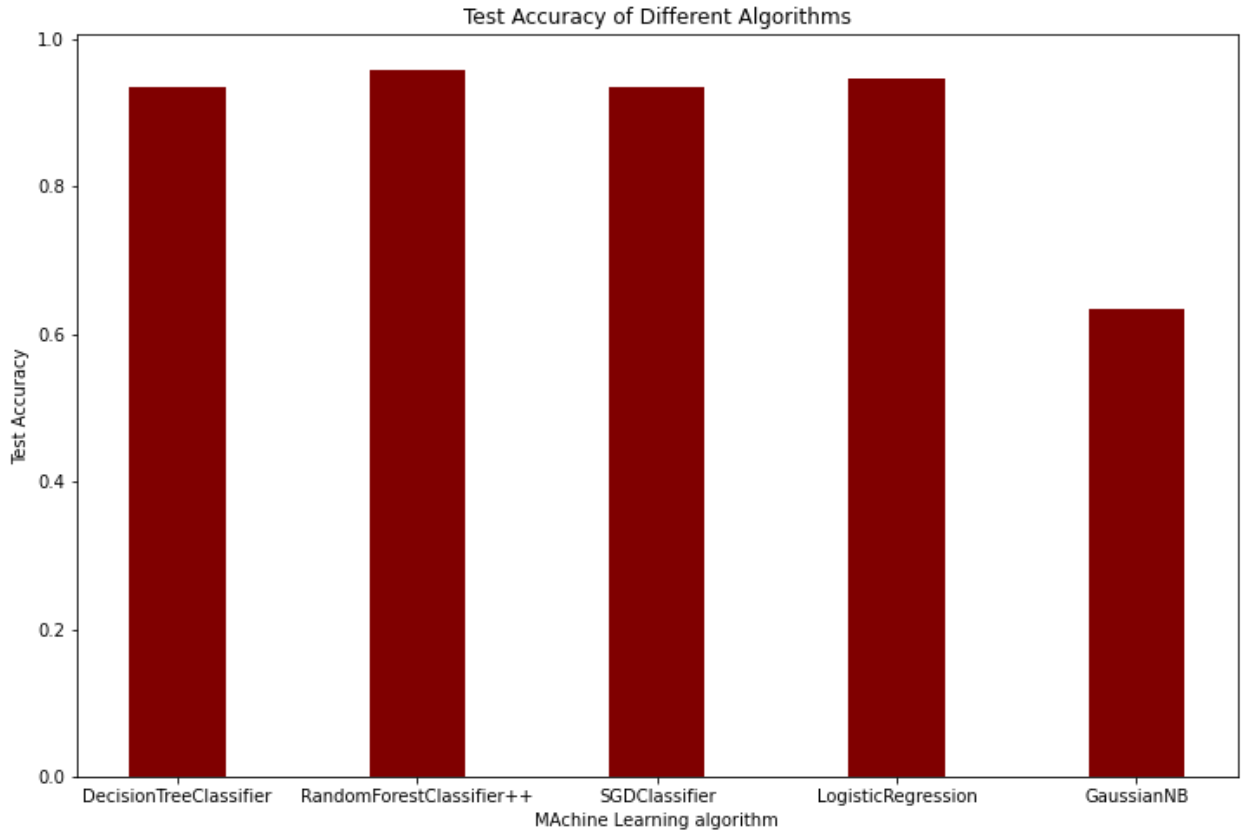
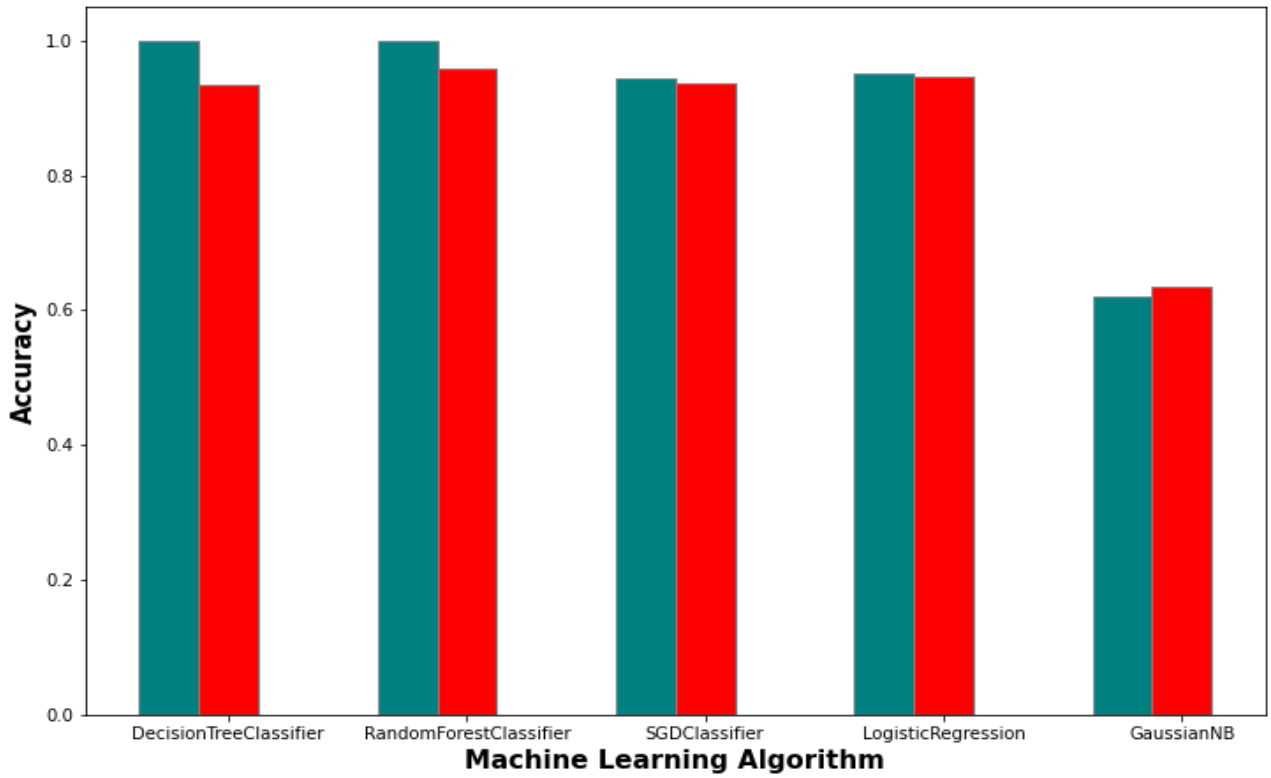
REPORT

1. Natural Language Understanding

Option 2. Binary Classifier Using the status variable build a binary classifier to predict the probability whether a project will be “closed” or “canceled/distressed”

- Algorithms used by me with final Test accuracy :
  1. DecisionTreeClassifier - 0.934032569693624
  2. RandomForestClassifier - 0.9585978470880486
  3. SGDClassifier – 0.9359646701628485
  4. LogisticRegression - 0.9472812586254485
  5. GaussianNB - 0.6340049682583494

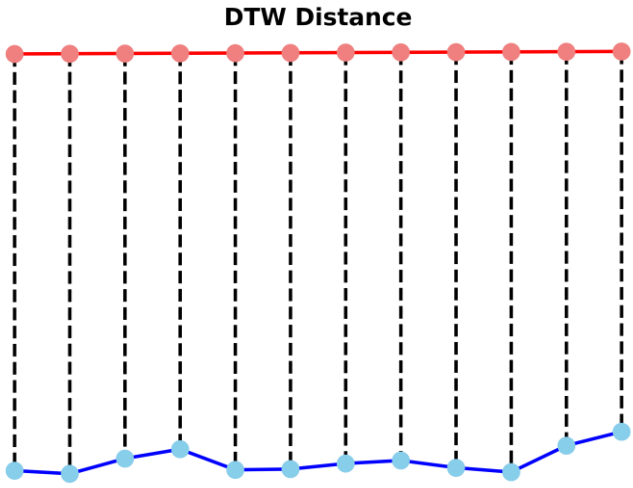


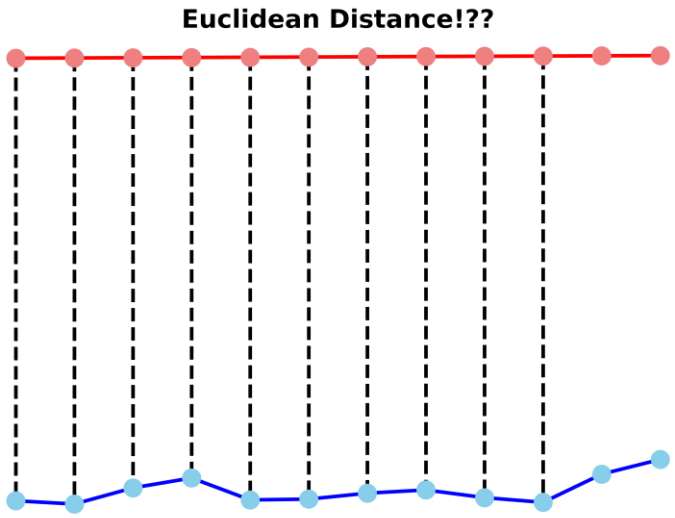


2. Time-Series Analysis

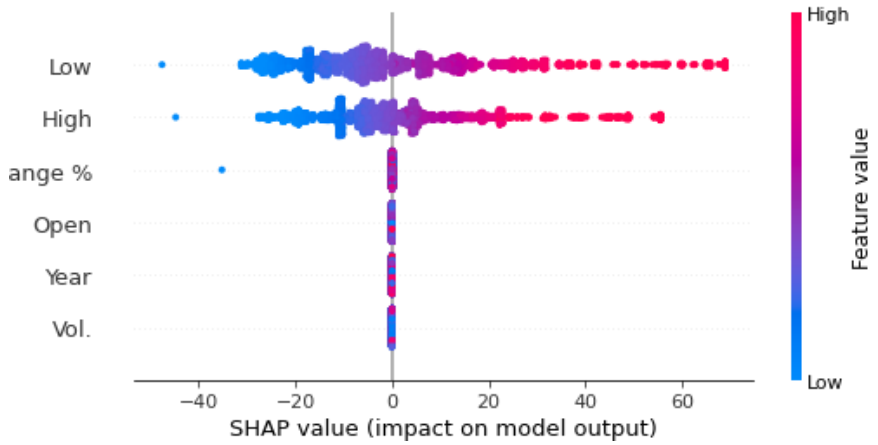
Option 1. Feature Importance. Dynamic Time Warping and/or XGBoost/Shapley Value hybrid model approach to quantify which factors influence the target positively or negative

1. **Dynamic Time Wraping:**





2. **Feature importance with XGBoost/ Shapley Value hybrid model approach to quantify which factors influence the target positively or negative**



Conclusion:

With more time and resouces I would have use hyper parameter tuning for binary classifiers and would have improved the accuracy of algorithms more.