

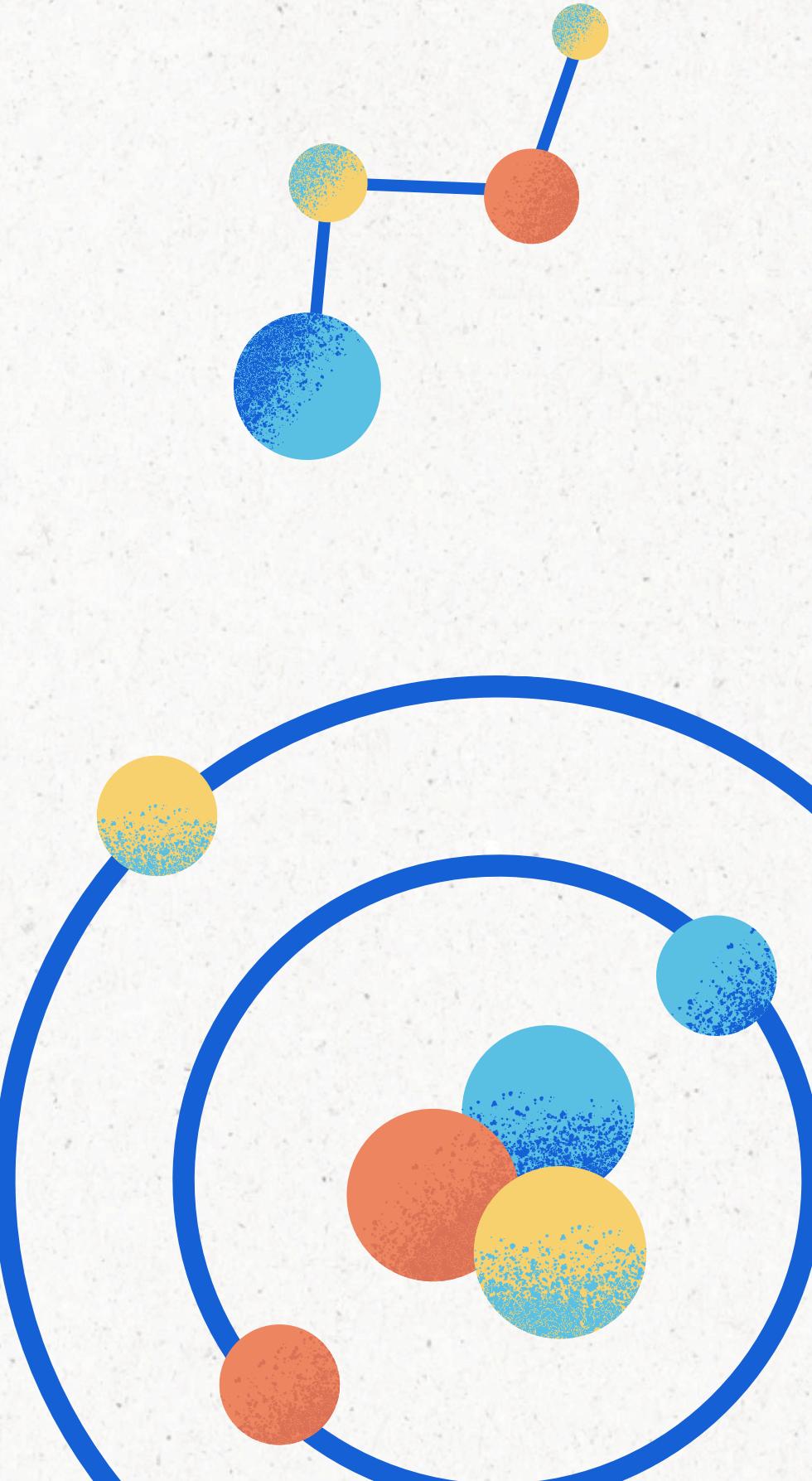
APPROACH

Models employed for Prediction Task

- K-Nearest Neighbors
- Decision tree
- Random Forest
- Gradient boosting
- Gaussian Naive Bayes
- Kernel SVM

Model performance evaluation metrics

- Matthew's Correlation Coefficient (MCC)
- Weighted F1 score
- Confusion Matrix



RESULTS



Model Performance

Sl. No	Model	MCC Train	MCC Test	Weighted F1 Train	Weighted F1 Test
1	K-Nearest Neighbors(KNN)	0.925	0.907	0.950	0.940
2	Decision Tree	0.971	0.745	0.980	0.850
3	Random Forest	1.000	0.753	1.000	0.850
4	Gradient Boosting	0.965	0.889	0.980	0.930
5	Gaussian Naive Bayes	0.504	0.503	0.740	0.750
6	Kernel SVM	0.945	0.917	0.970	0.950

Note : MCC(Matthew's Correlation Coefficient), higher the values higher is the classification performance

- The KNN and Kernel SVM models are performing better as MCC of train and test data are higher.
- Over-fitting is evident in Random forest, Decision Tree and Gradient Boosting as MCC of test data is lower compared to train data.
- The weighted F1 scores showed similar trends as of Matthew's correlation coefficient.

Model Performance after Hyperparameter Tuning

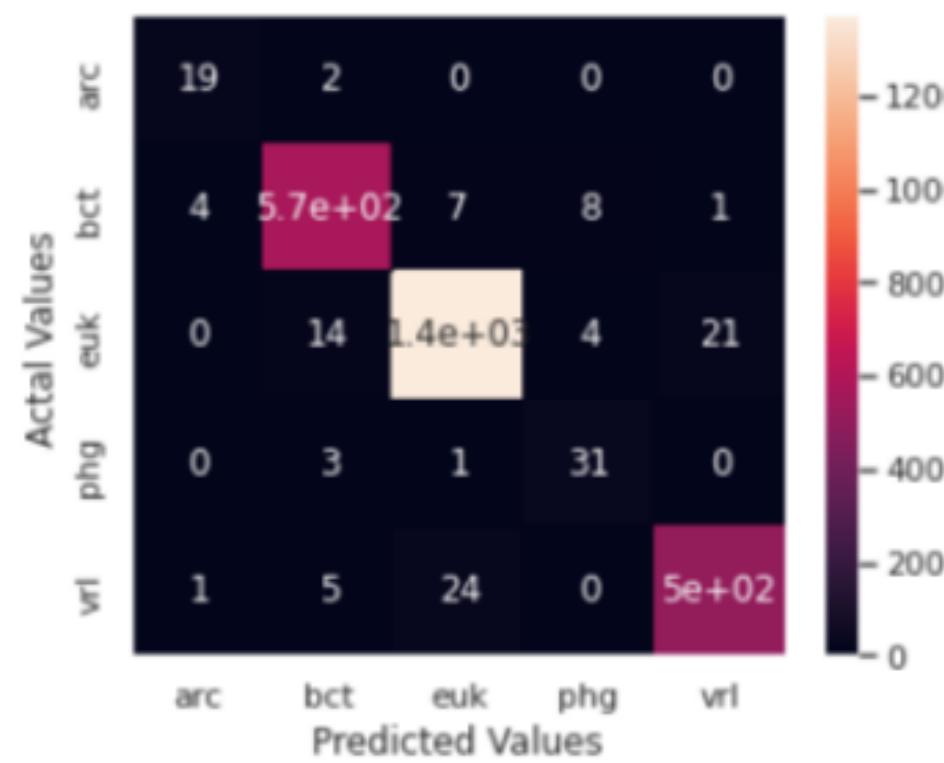
Sl. No	Model	MCC Train	MCC Test	Weighted F1 Train	Weighted F1 Test
1	K-Nearest Neighbors(KNN)	0.970	0.939	0.980	0.960
2	Decision Tree	0.970	0.753	0.980	0.850
3	Random Forest	0.974	0.753	0.980	0.850
4	Gradient Boosting	1.000	0.907	1.000	0.940
5	Gaussian Naive Bayes	0.504	0.503	0.740	0.750
6	Kernel SVM	0.971	0.932	0.980	0.960

Note : MCC(Matthew's Correlation Coefficient), higher the values higher is the classification performance

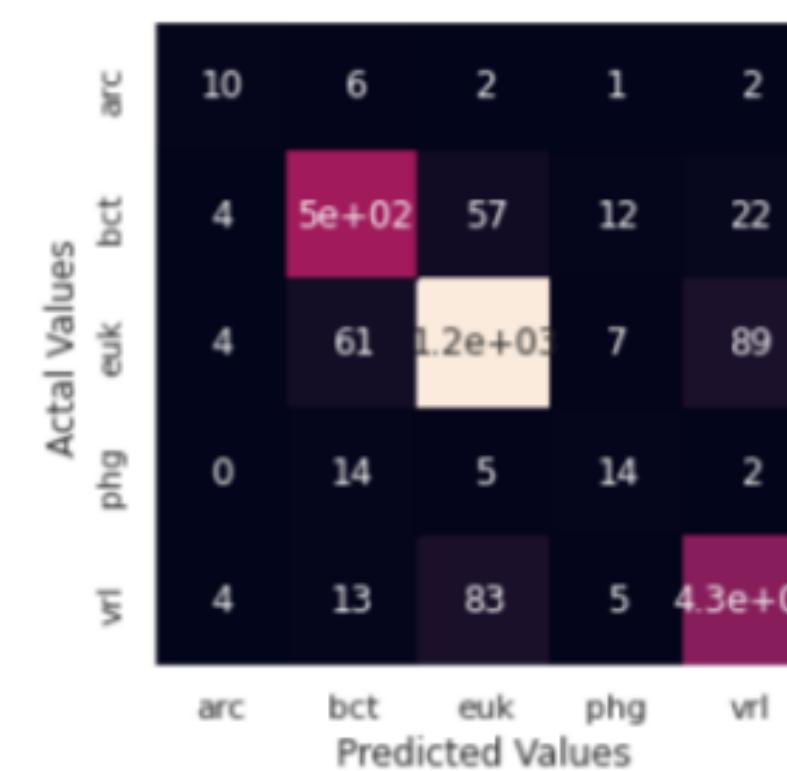
- Based on Test score KNN and Kernel SVM models were performing the classification task better than other models

Model wise Confusion Matrix

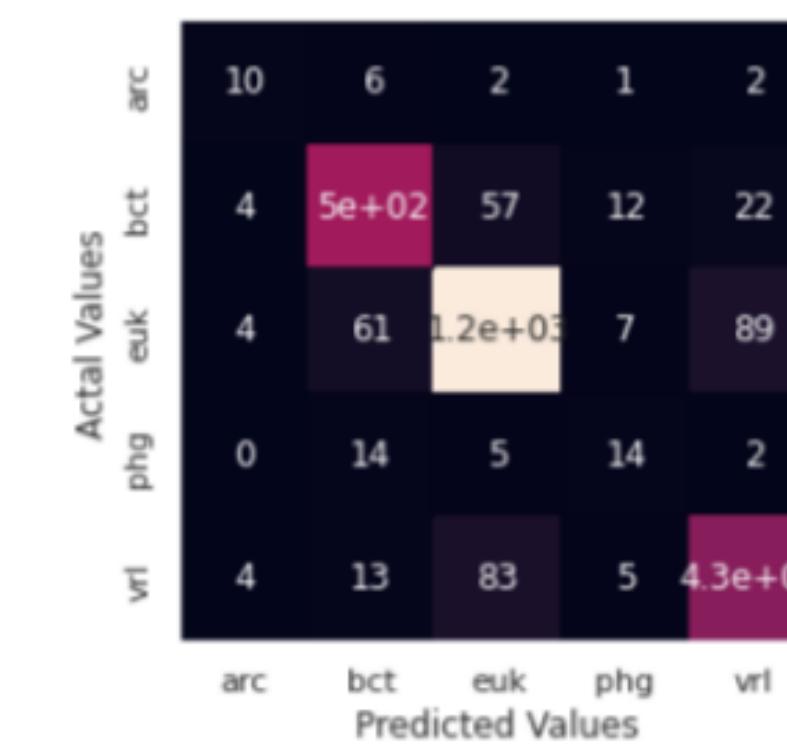
KNN



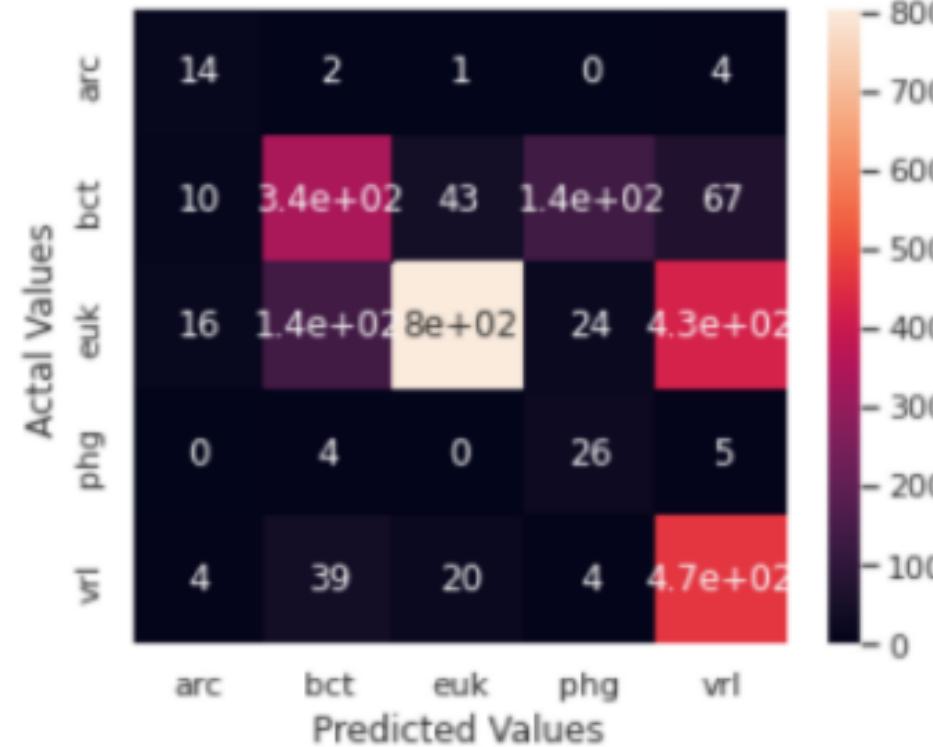
Decision Tree



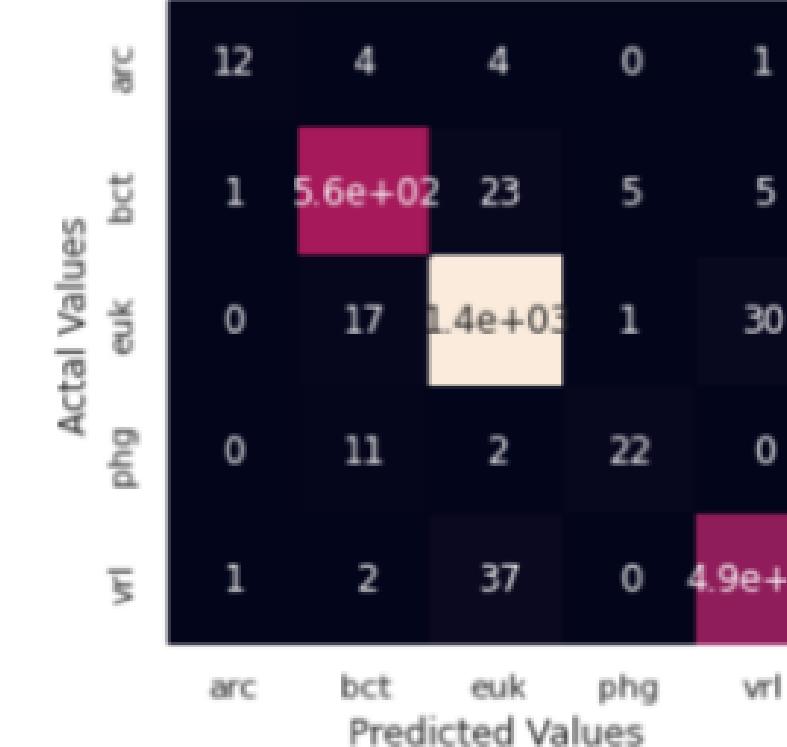
Random Forest



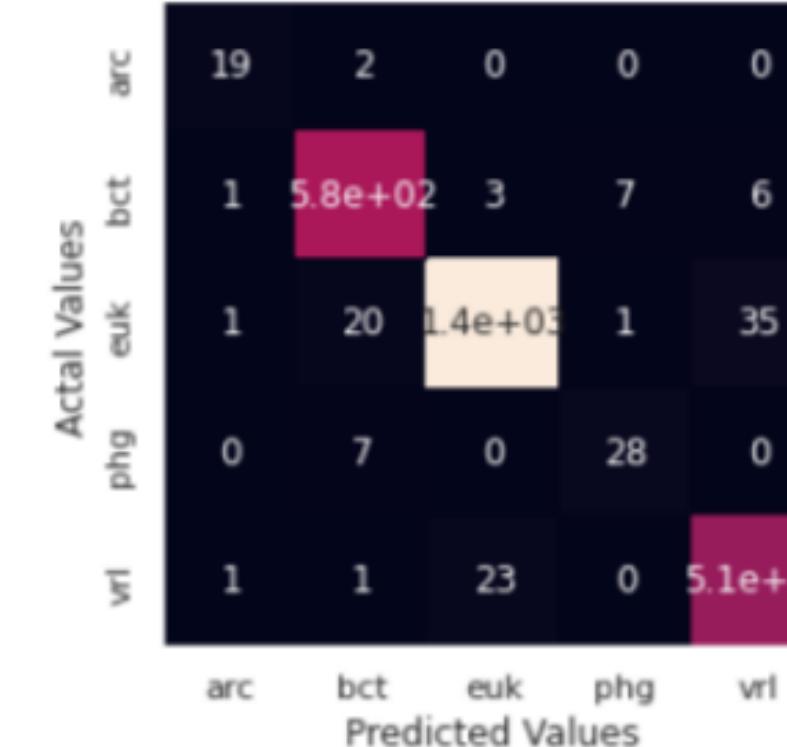
Naive Bayes



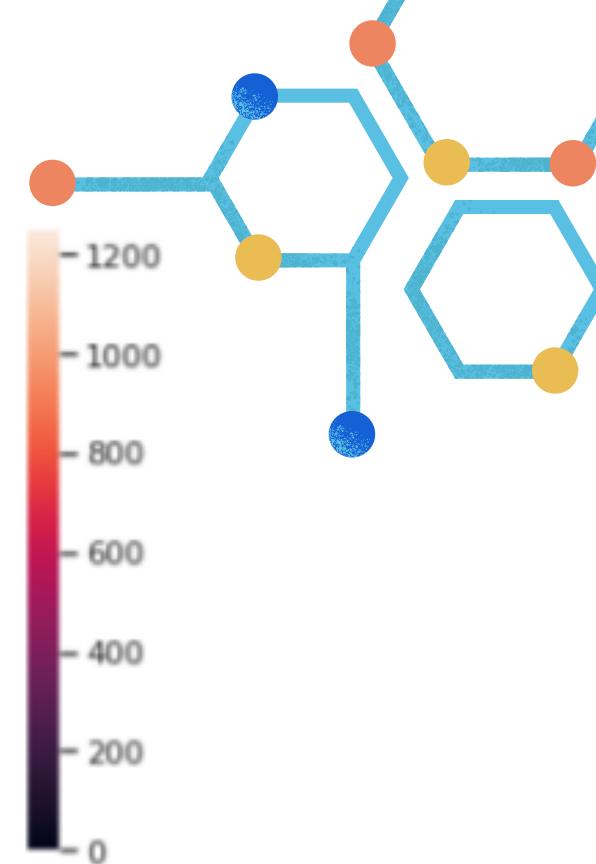
Gradient Boosting



Kernel SVM



Note: Diagonal Elements-Show the number of samples correctly classified for each class. Off Diagonal Elements-Represent misclassifications



- KNN and Kernel SVM represent better predictions compared to the other models.