APPLIED MACHINE LEARNING
FINAL PROJECT

Prepared By:

Ahmed Yousry

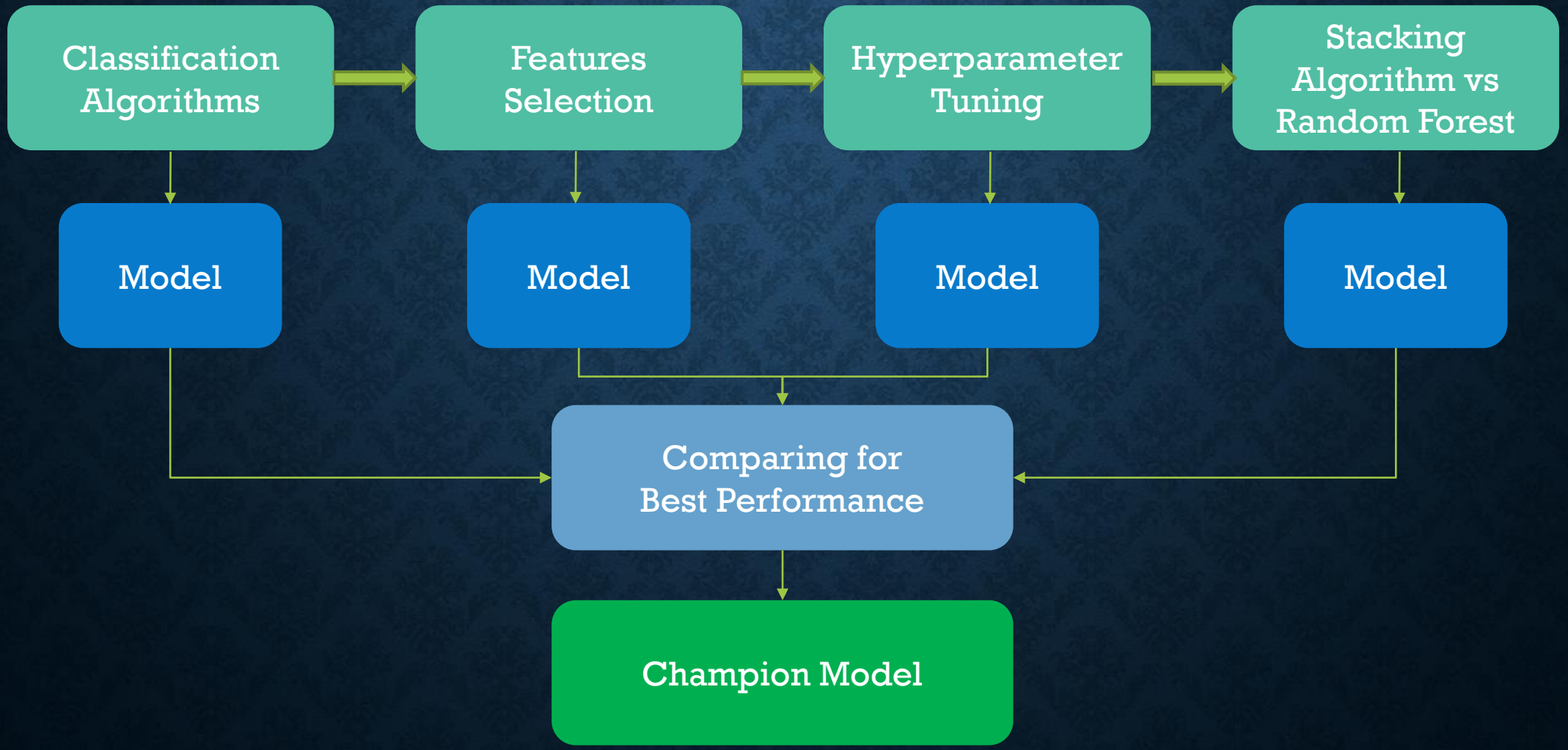
Bassel Emam

Mohammed El-Namoury

Supervised By:

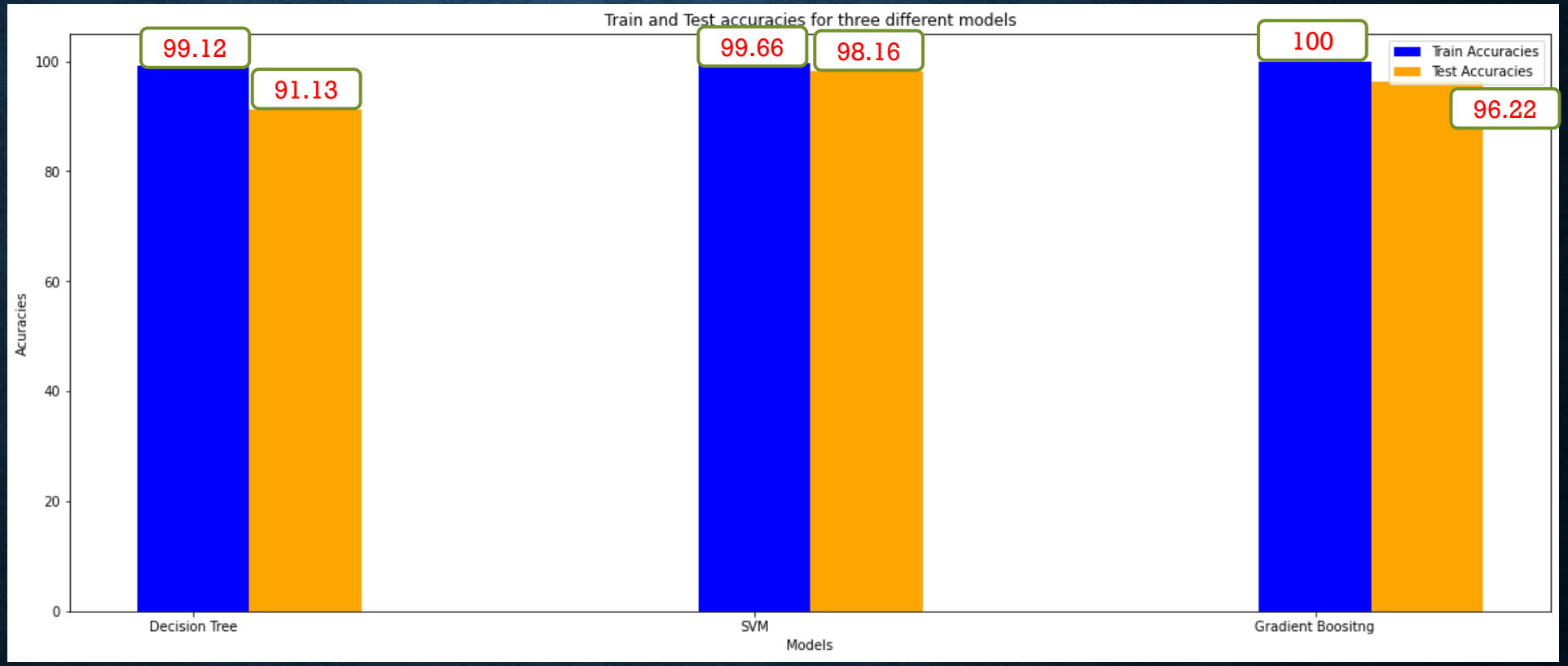
Prof. Murat Simsek

PROJECT ROADMAP

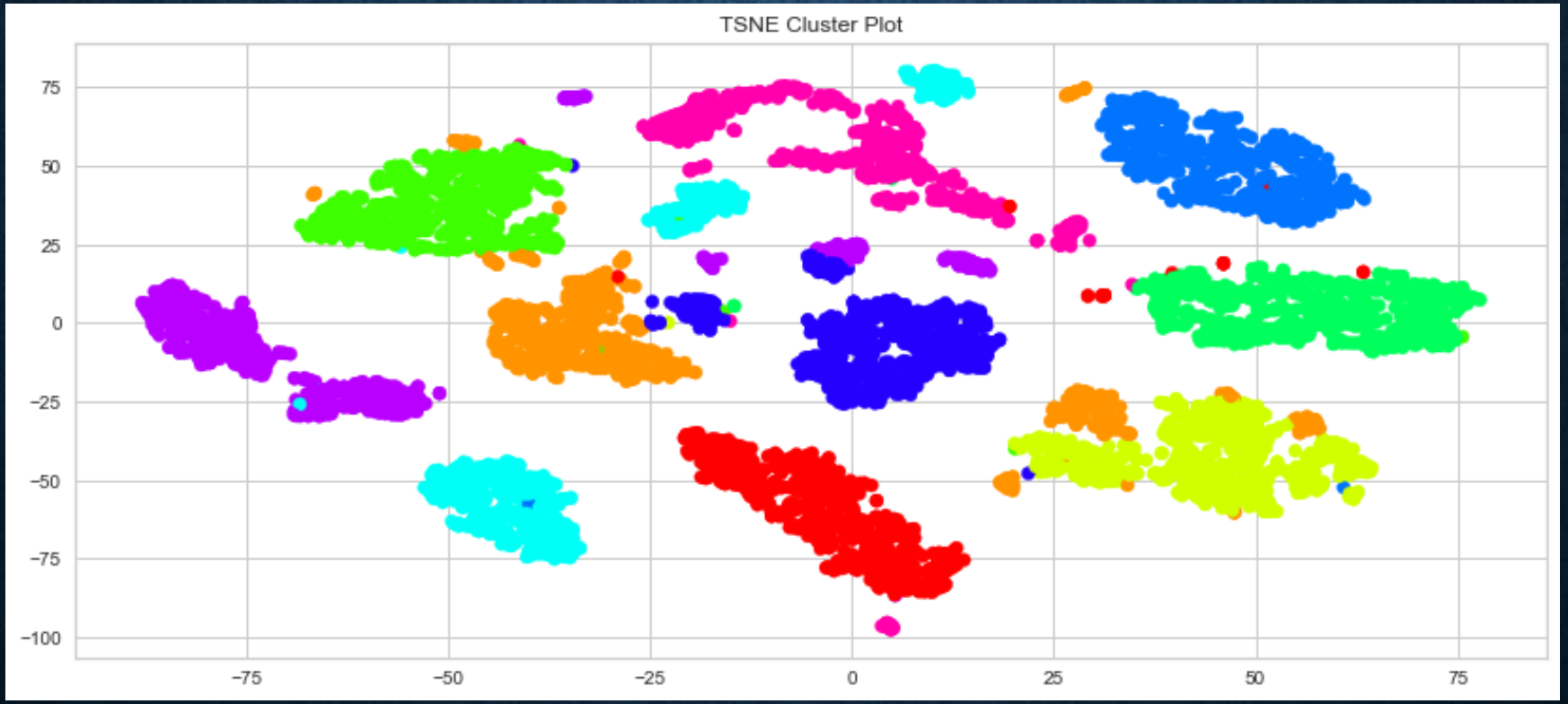


CLASSIFICATION ALGORITHMS BASE-LINE

The Base-Line accuracy is 98.16 %

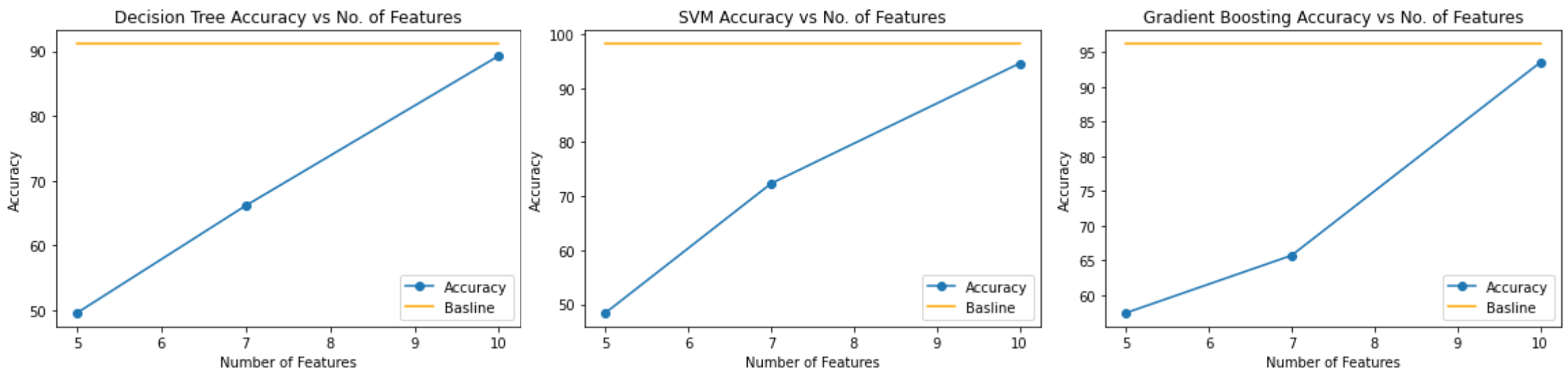


T-SNE 2D PLOTS BEFORE FEATURE SELECTION



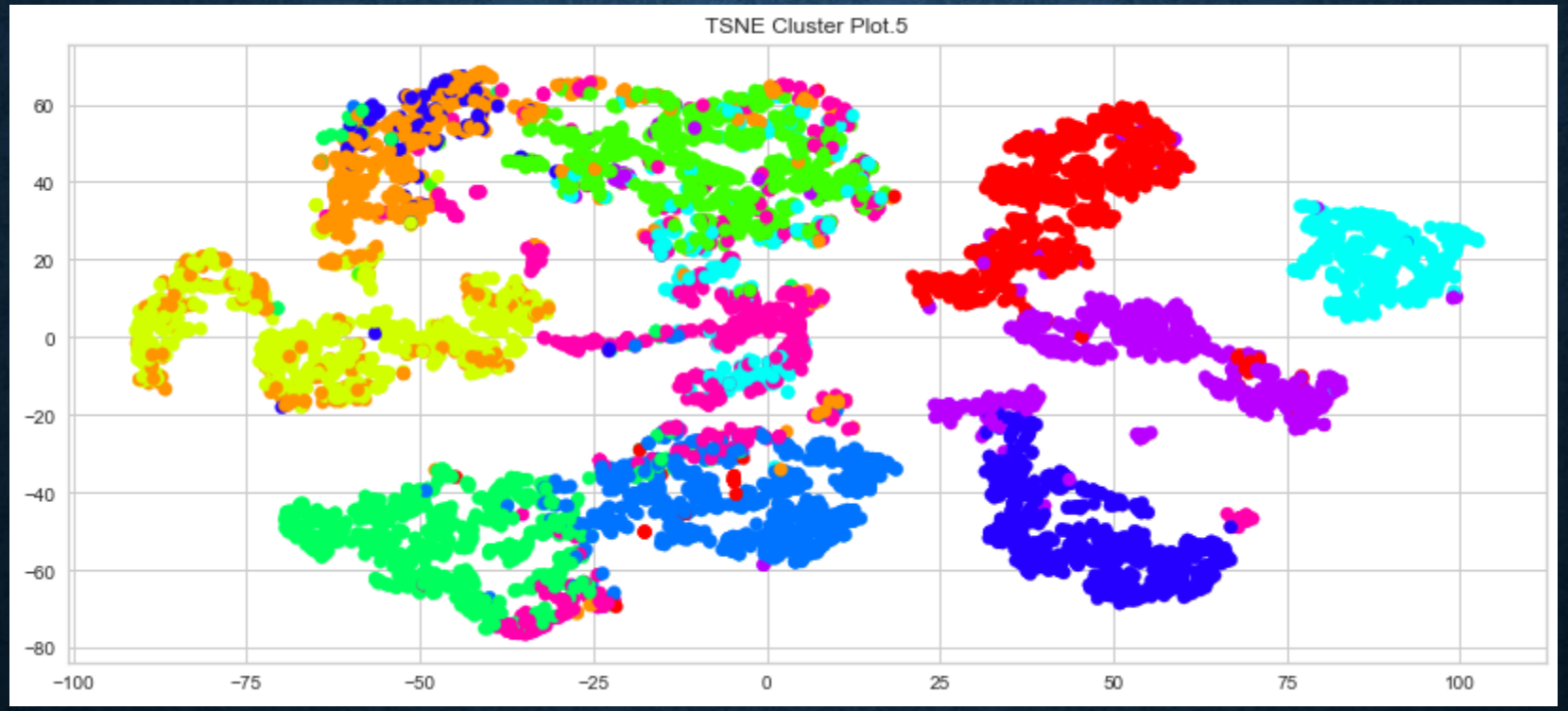
FEATURES SELECTION

	5 Features	7 Features	10 Features	Base-Line Accuracy
Decision Tree	49.61 %	66.17 %	89.30 %	**91.13 %**
SVM	48.41 %	72.34 %	94.56 %	**98.16%**
Gradient Boosting	57.42 %	65.71 %	93.53 %	**96.22 %**

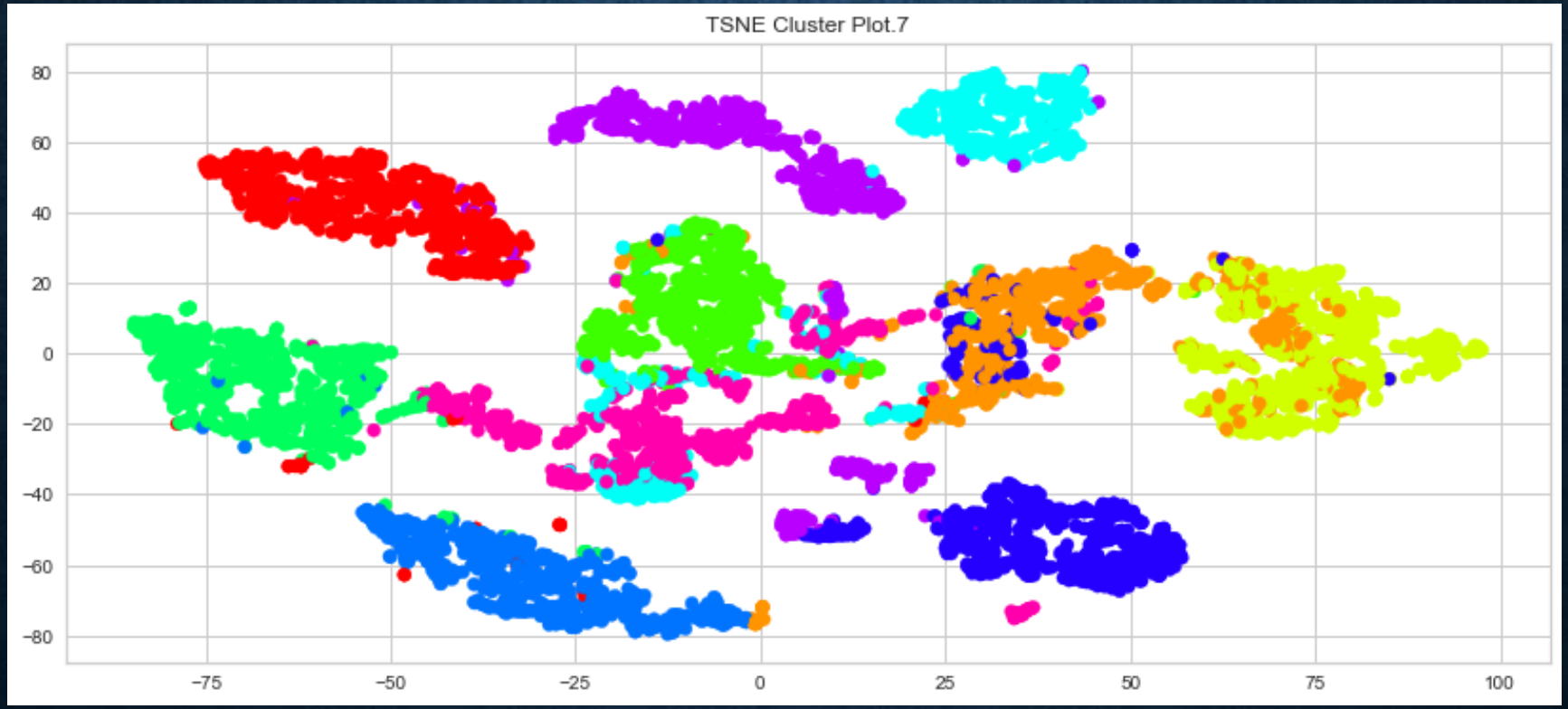


The accuracies of Feature Selection Phase are all below the base-line accuracy for each model. Therefore, no update for the base-line accuracy.

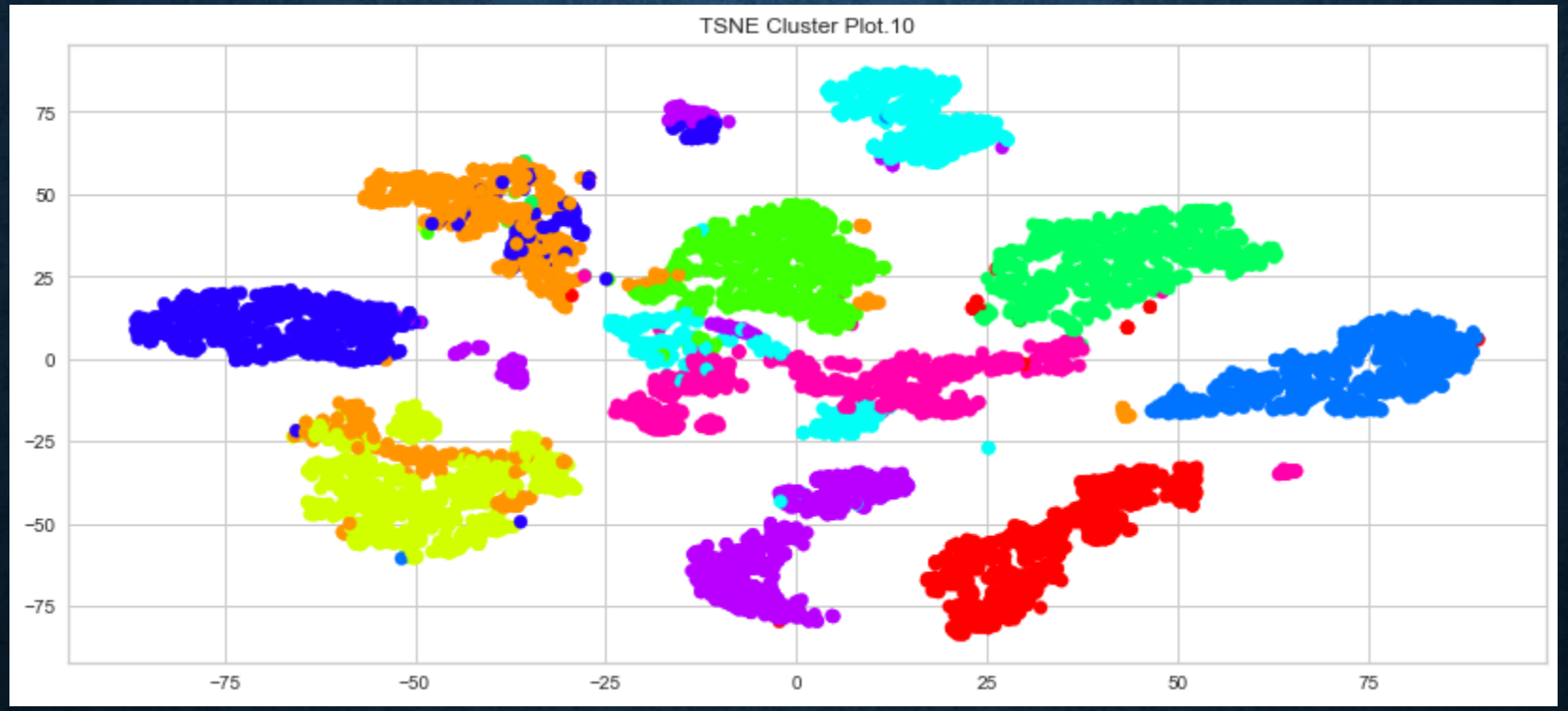
T-SNE 2D PLOTS WITH 5 FEATURES



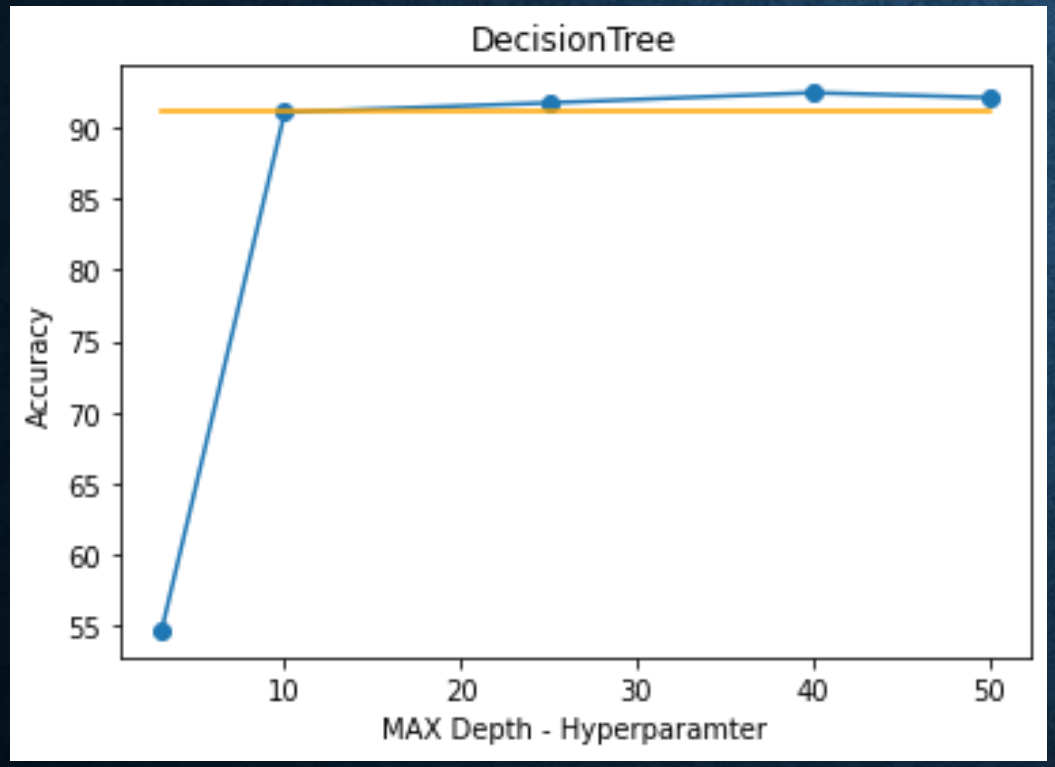
T-SNE 2D PLOTS WITH 7 FEATURES

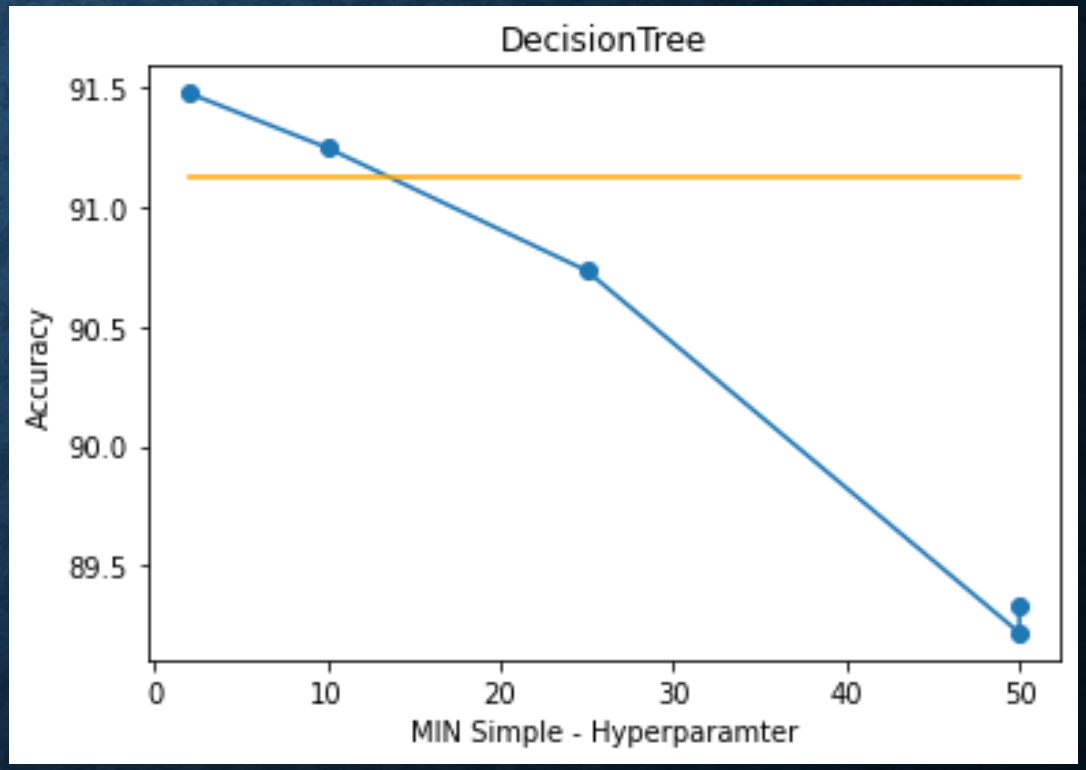


T-SNE 2D PLOTS WITH 10 FEATURES



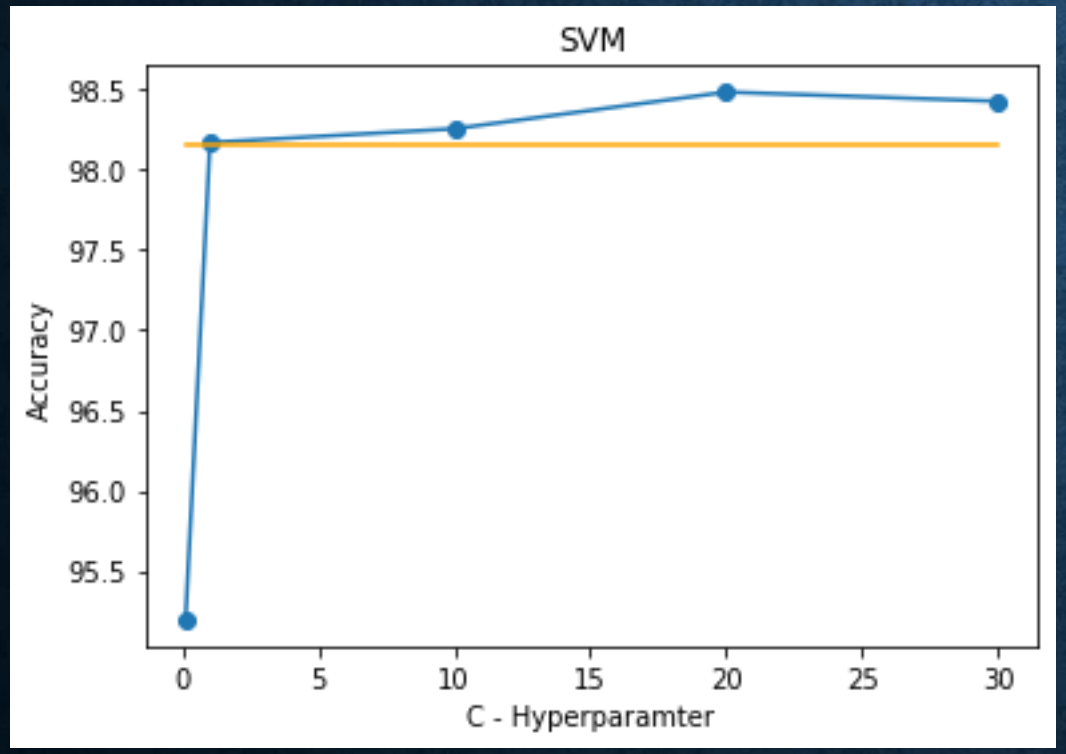
HYPERPARAMETER TUNING DECISION TREE

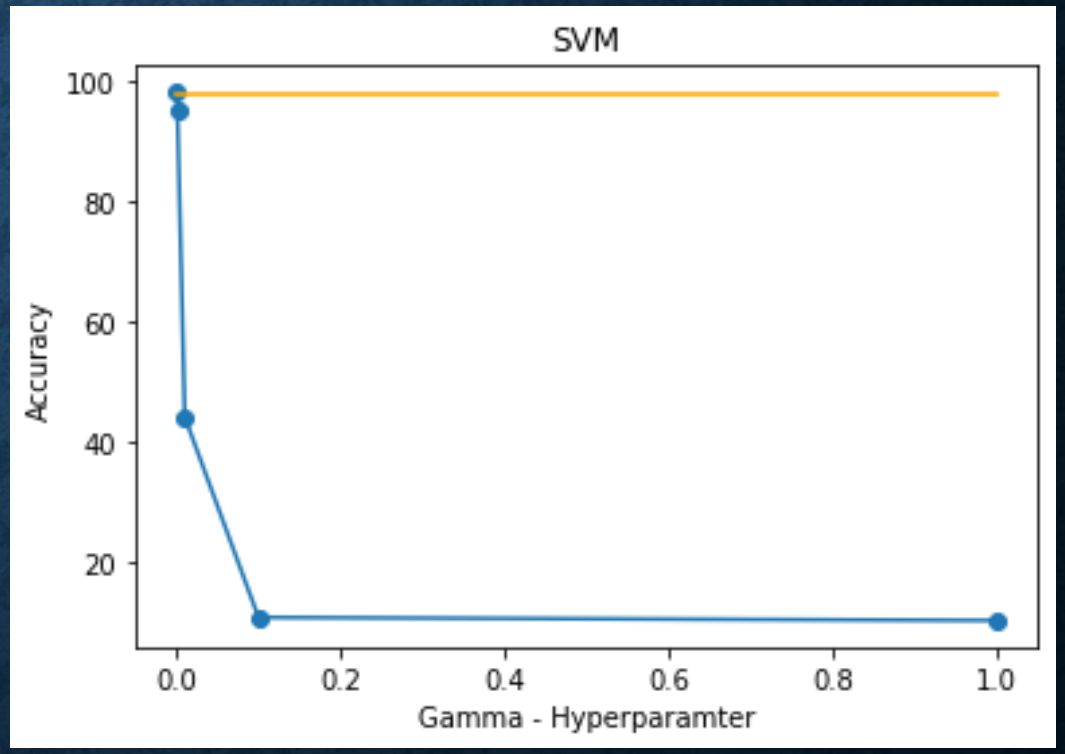




Decision Tree Test Accuracy using Max- Depth Hyperparameter tuning is 91.87 %

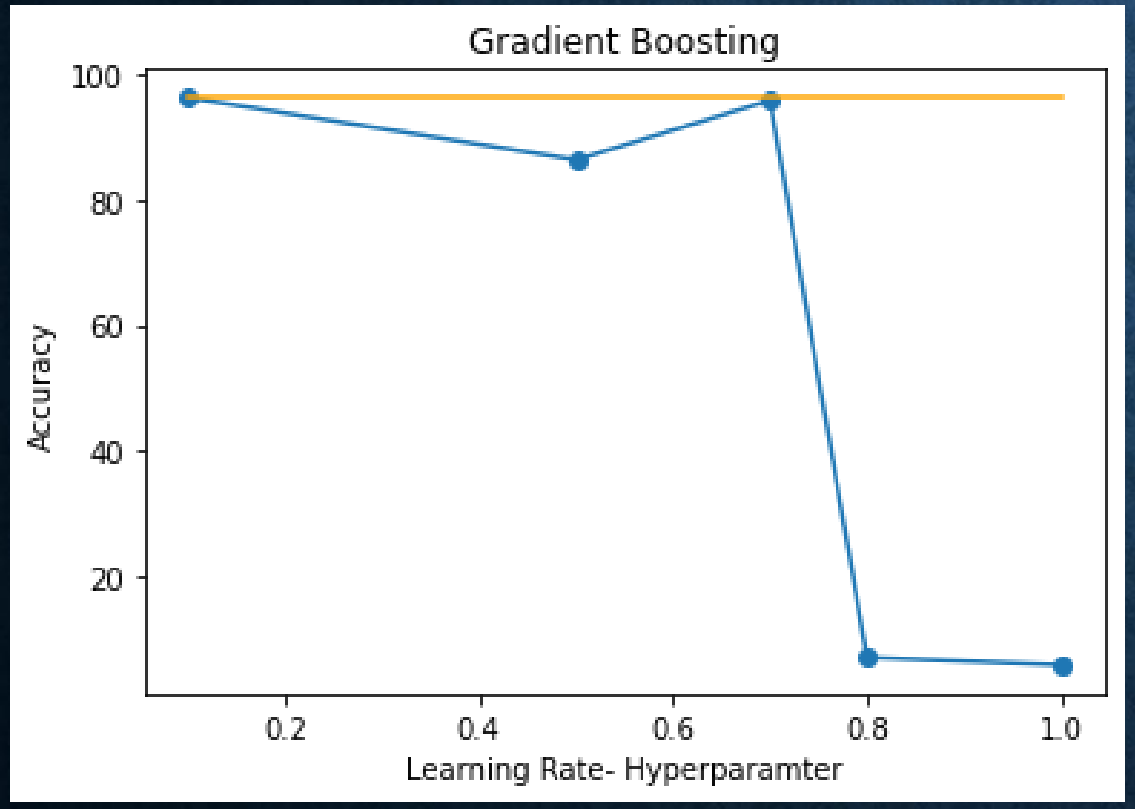
HYPERPARAMETER TUNING
SVM

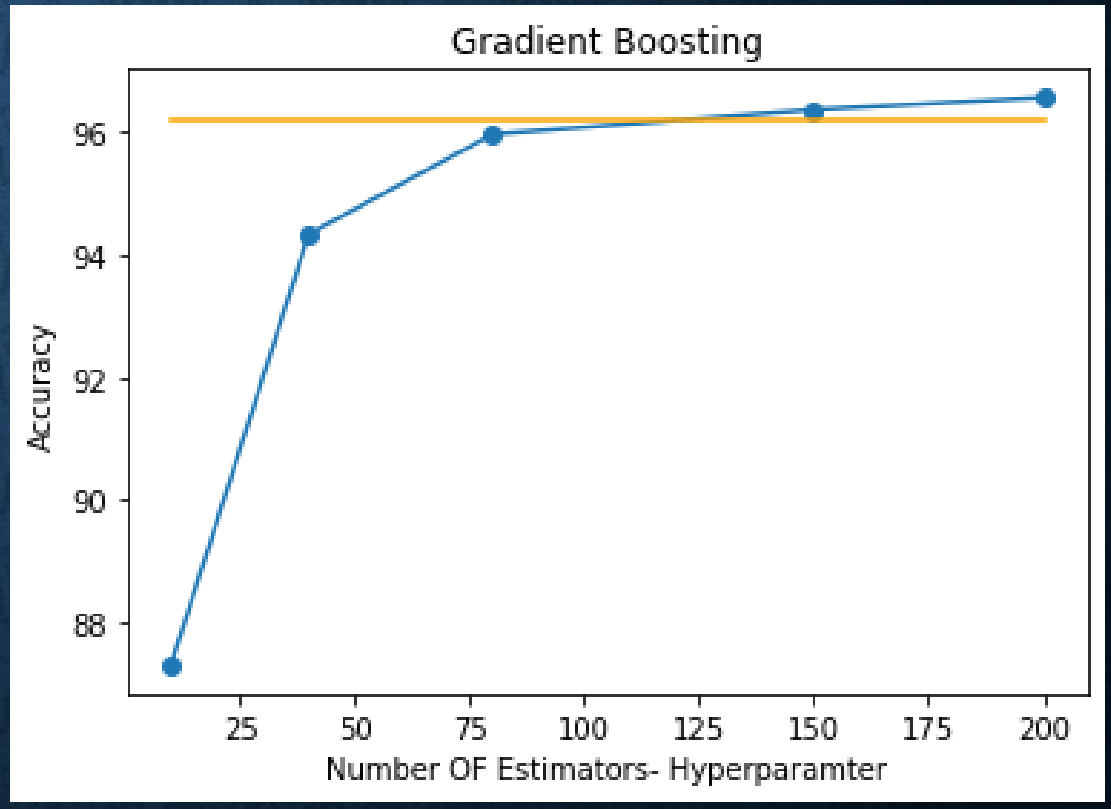




SVM Test Accuracy using C Hyperparameter = 20 is 98.39 %

HYPERPARAMETER TUNING
GRADIENT BOOSTING





Decision Tree Test Accuracy using Number of Estimators Hyperparameter = 200 is 96.56 %

STACKING VS RANDOM FOREST

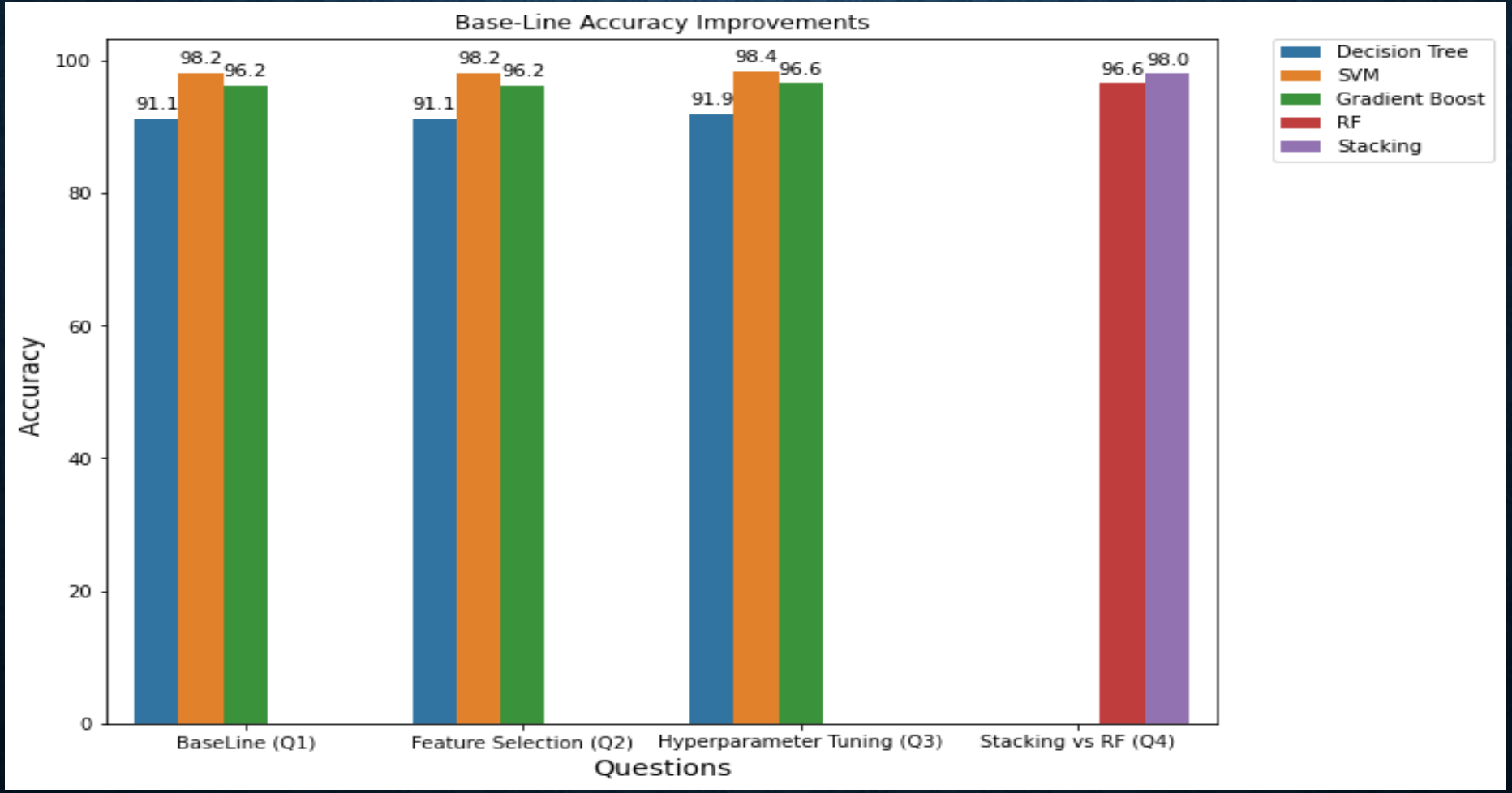
- ***Stacking***

  A combination of the three models (Decision Tree, SVM, and Gradient Boosting) has accuracy of **97.96 %**

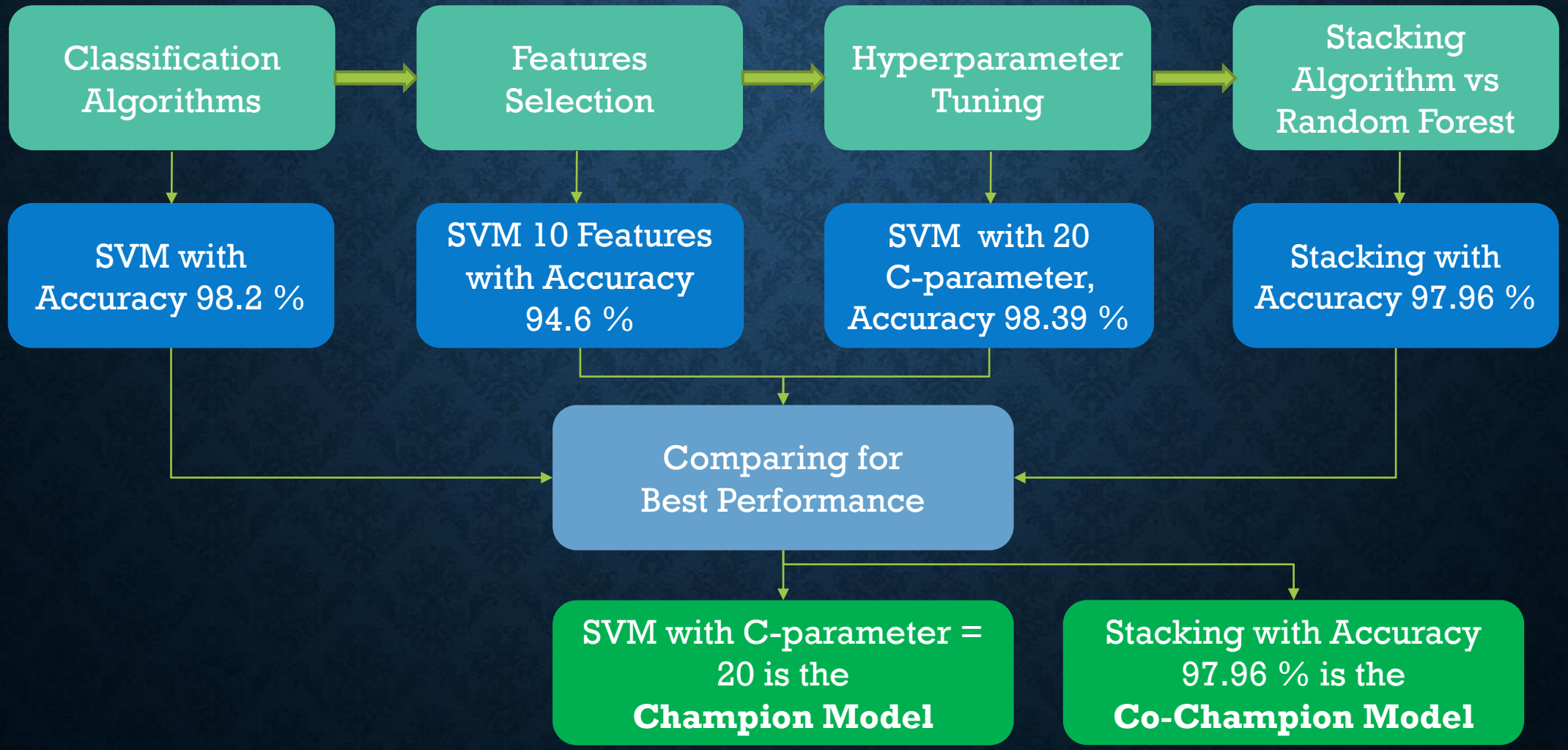
- ***Random Forest***

  A bagging algorithm of 100 decision tree has accuracy of **96.56 %**

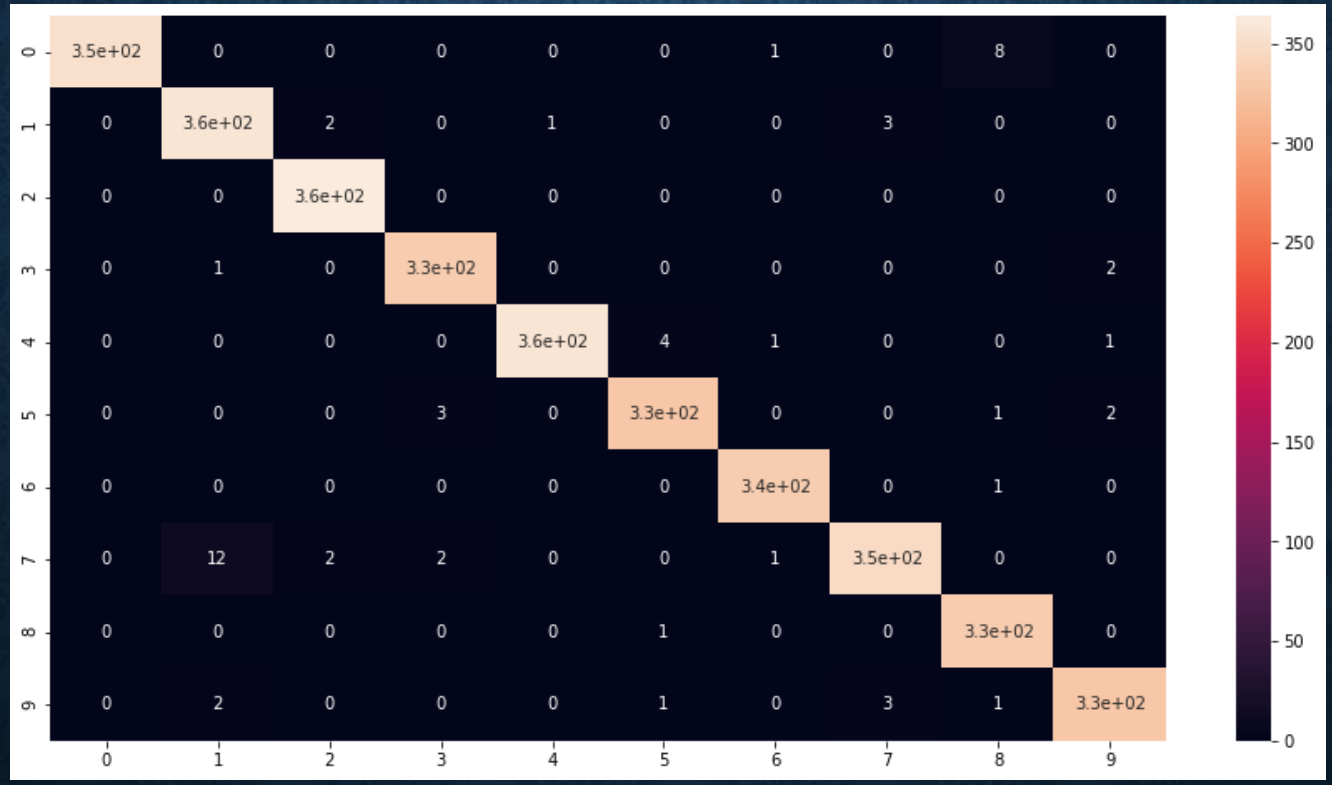
ACCURACY IMPROVEMENTS OVER DIFFERENT MODELS



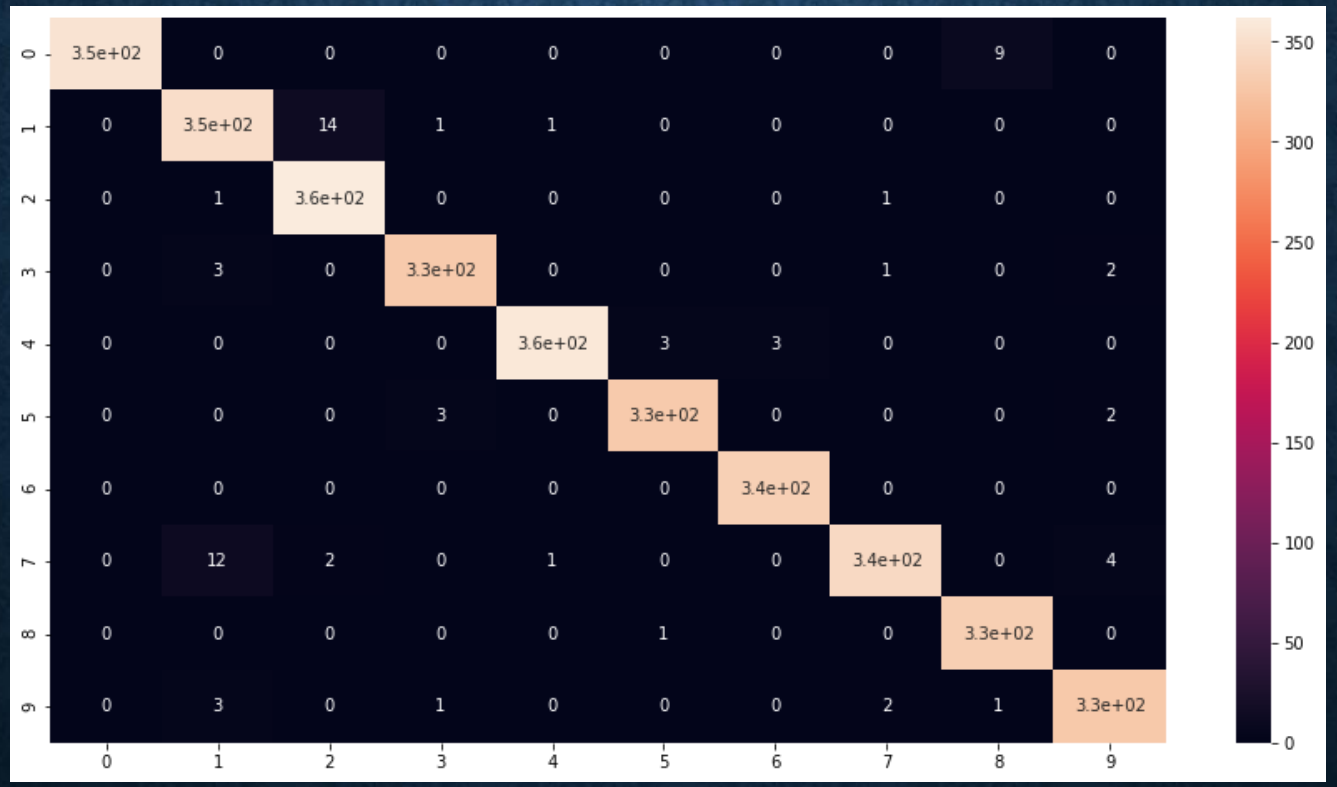
PROJECT ROADMAP



CONFUSION MATRIX CHAMPION MODEL (SVM)



CONFUSION MATRIX
CO-CHAMPION MODEL (STACKING)



CONCLUSION

- Question 2

  When we tried feature selection approach here the results came lower than using the full dataset and as shown in the graph as number of features increase the accuracy of the model increase as number of features increases it is a positive relationship

- Question3

  When we tried to different parameters in tunning the model the accuracy increased and svm came as the champion model here

- Question4

  The graph shows that the svm made it to the finals as the best model and using ensemble methods like stacking and random forest produced good results but svm still the winner and stacking came second and as shown in the graph question number two there wasn't any improvement so we sticked to the baseline models results

THANK YOU