Problem Statement 3

Take Dataset from Kaggle

https://www.kaggle.com/datasets/clmentbisaillon/fake-and-real-news-dataset

Apply SVM, KNN, Naive Bayes Classifier to detect fake or real news.
Show performance analysis of SVM, Naive Bayes and KNN Classifier based on precision and recall. Write your conclusion based on it.

handle at least one drawback of given classifier

Solution by Team Infinite

Importing Libraries:

Various Libraries are imported for Data Preprocessing, Text preprocessing , data visualization and model building

Dataset load

Both the dataset of real and fake news are loaded and later merged for further easy use

Data preprocessing

Dropping unwanted

Columns of dates are dropped as they do not contribute much in further prediction purpose

Removal of Duplicate

Finding the duplicate rows from the merged dataset and then removal of such rows

Text Preprocessing

Lowering the text

Text from the dataset is lower to lowercase so that uniformity of data is maintained

Tokenozation

Splitting the data in smaller units

Stopwords removal

Removal of words such is,the etc which are less contributing to model building

Merging Columns

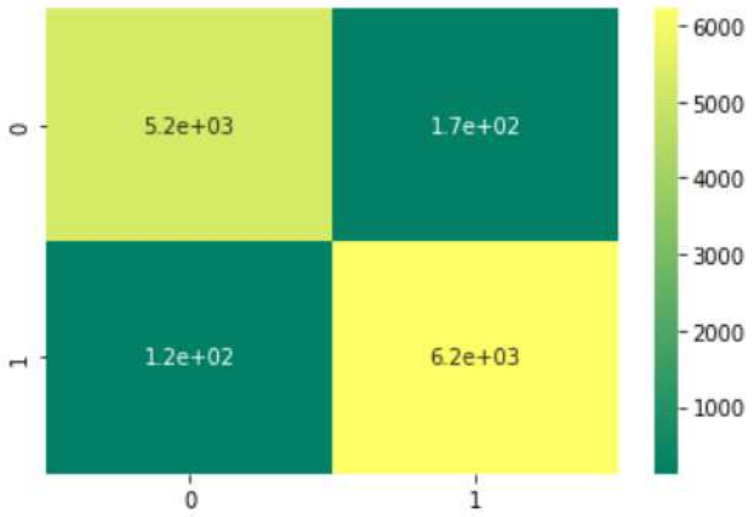
Later these colums are merged

Model building using ML models

Data is splitted into training and testing for model building

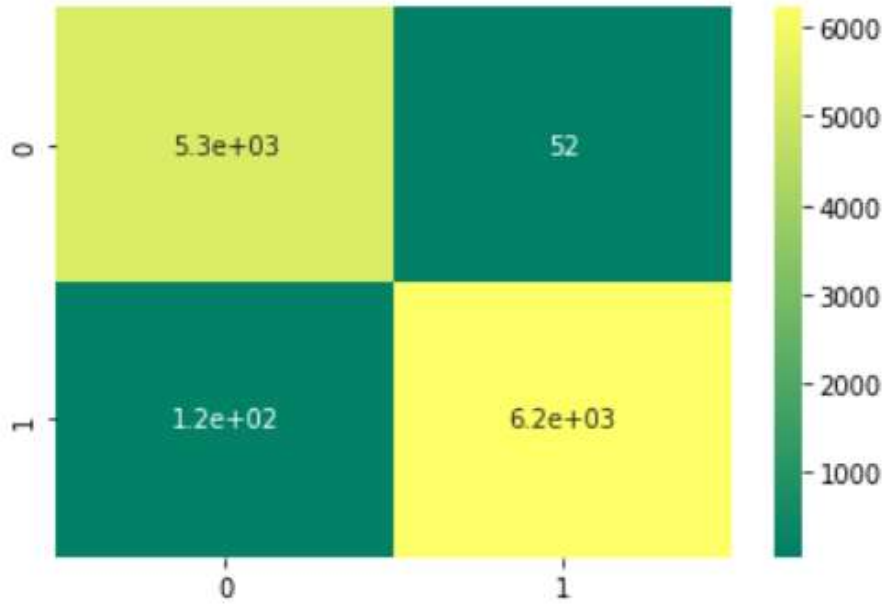
SVM model

	precision	recall	f1-score	support
Fake	0.98	0.97	0.97	5392
Real	0.97	0.98	0.98	6340
accuracy			0.97	11732
macro avg	0.97	0.97	0.97	11732
weighted avg	0.97	0.97	0.97	11732



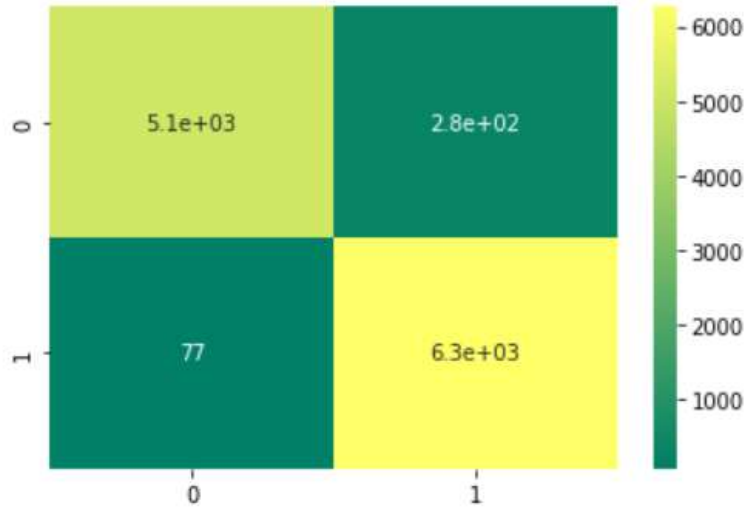
KNN

	precision	recall	f1-score	support
Fake	0.98	0.99	0.98	5392
Real	0.99	0.98	0.99	6340
accuracy			0.99	11732
macro avg	0.99	0.99	0.99	11732
weighted avg	0.99	0.99	0.99	11732



NaiveBayes

	precision	recall	f1-score	support
Fake	0.99	0.95	0.97	5392
Real	0.96	0.99	0.97	6340
accuracy			0.97	11732
macro avg	0.97	0.97	0.97	11732
weighted avg	0.97	0.97	0.97	11732



Comparison between models

	Train score	Test score
SupportVectorMachine	0.973076	0.974599
RandomizedSearch_KNeighborsClassifier	0.989990	0.985510
votingclassifier_NB	0.971103	0.969826

Conclusion

After doing the comparison we can conclude that KNN has given better results as compared to SVM and NaiveBayes with **0.98 training score and 0.98 test Score.**