CAR ACCIDENT SEVERITY PREDICTION

-GANAPATHI BALAJI

CONTENT

- Problem Definition
- Exploratory Data Analysis
- Feature Engineering
- Model Process
- Model Report
- Summary Conclusion

MODEL PROCESS

- Label Encoding for Categorical features
- Creating Training, Test set by balancing severity ratio
- Training by applying different Classification Techniques
- Model Evaluation and Hyper parameters tuning
- Model Score Report

MODEL SCORES REPORT

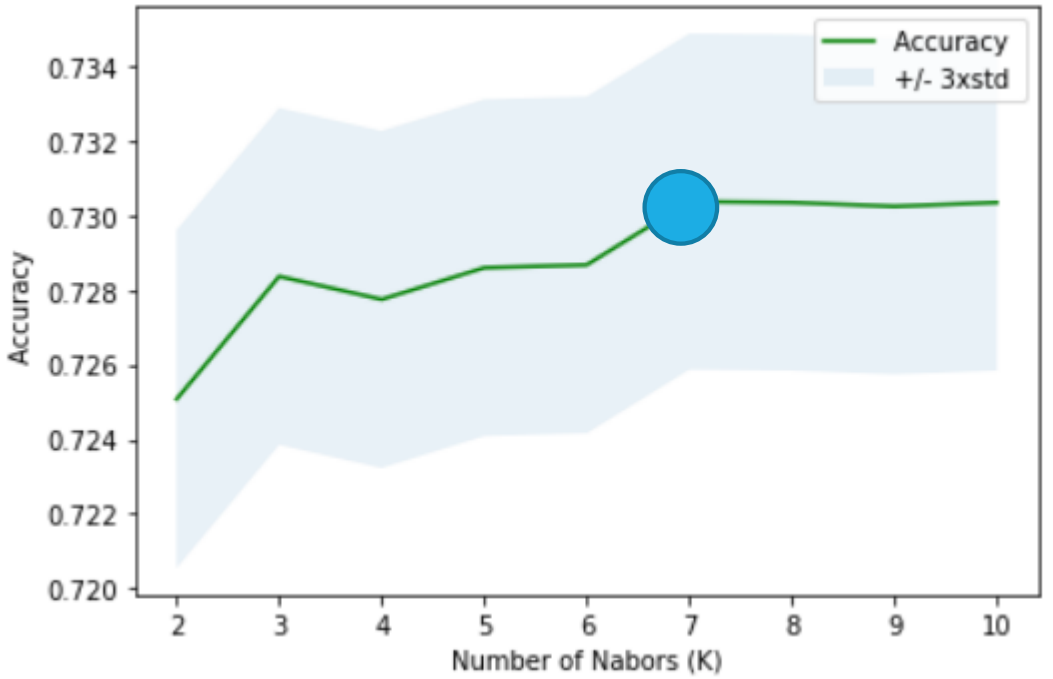
- I have used trained the data on 3 different algorithms
- I did Hyper parameter tuning by changing the weights and getting almost same model scores
- I was Training on SVM also it is taking lot of time so not putting that here

	Algorithm	F1-score	Jaccard	LogLoss
0	KNN	0.837984	0.730371	NA
1	Decision Tree	0.838124	0.730551	NA
2	LogisticRegression	0.837691	0.730037	0.556894

CLASSIFICATION EVALUATION

- At k value 7 is where we are getting better accuracy of 0.73

	precision	recall	f1-score	support
1	0.72	0.99	0.84	27297
2	0.90	0.11	0.20	11638
micro avg	0.73	0.73	0.73	38935
macro avg	0.81	0.55	0.52	38935
weighted avg	0.78	0.73	0.65	38935



DECISION TREE EVALUATION

- At k value 4 is where we are getting better F1-score of 0.73

Decision Trees is giving best results at Dept = 4 using entropy

[49]:
```
yhat=drugTree.predict(X_test)
yhat_prob=drugTree.predict_proba(X_test)

report=report.append({"Algorithm": "Decision Tree" , "Jaccard": jacc

print (classification_report(y_test, yhat))
```

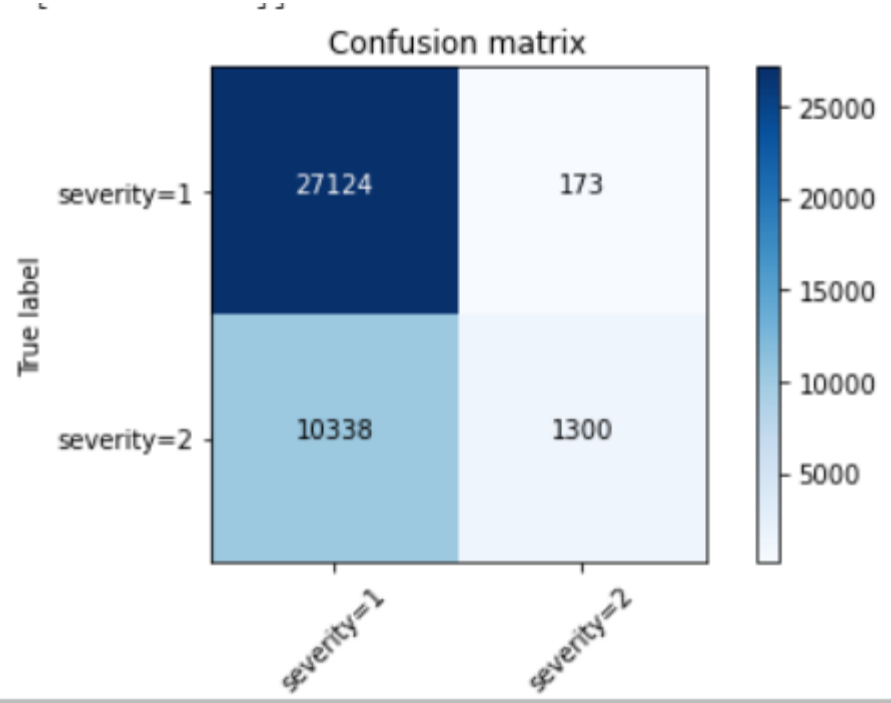
```
              precision    recall  f1-score   support

           1       0.72      0.99      0.84     27297
           2       0.90      0.11      0.20     11638

   micro avg       0.73      0.73      0.73     38935
   macro avg       0.81      0.55      0.52     38935
weighted avg       0.78      0.73      0.65     38935
```

DECISION TREE EVALUATION

- At k value 4 is where we are getting better F1-score of 0.84



Logistic Regression is performing best with regularization val = 0.03

[54]:
```
yhat=LR.predict(X_test)
yhat_prob=LR.predict_proba(X_test)

report=report.append({"Algorithm": "LogisticRegression" , "Jaccard": jaccar
print (classification_report(y_test, yhat))
```

	precision	recall	f1-score	support
1	0.72	0.99	0.84	27297
2	0.88	0.11	0.20	11638
micro avg	0.73	0.73	0.73	38935
macro avg	0.80	0.55	0.52	38935
weighted avg	0.77	0.73	0.65	38935

SUMMARY CONCLUSION

- Useful and informative models built to predict accident severity
- Value in guiding public traffic polices to focus on important factors to prevent accident injuries
- Accuracy of model has room for improvement, more insights could be gained
  - Collision type be further processed and used in model
  - Accident address be grouped based on injury occurrence ratio and used in model
  - Accident trend by dates

Thank You