Road Accident-Severity Prediction

HASLA M

Sep 9, 2020

Outline

Introduction

- One of the important safety challenge in the modern world is to prevent or reduce road accidents.
- We need to analyse and find the main causes of these accidents.
- It will be helpful if we can predict possiblity of an accident, also the severity of the accidents.
- We can generate warnings or alerts to the public for taking particular precautions.
- This work will be helpful to the public of Seattle, especially, for the one who drives a vehicle.

Data aquisition and cleaning

- The dataset includes all types of collisions in Seattle from 2004-2020, by SDOT Traffic Management Division, Traffic Records Group, Seattle.
- It contain 40 attributes.
- The target attribute is SEVERITYCODE - code that corresponds to the severity of the collision.
- Many attribute contain missing values.
- Some of the attribute are irrelavent and are dropped.
- Attributes with object types are converted into int type by encoding each value to numerical codes.

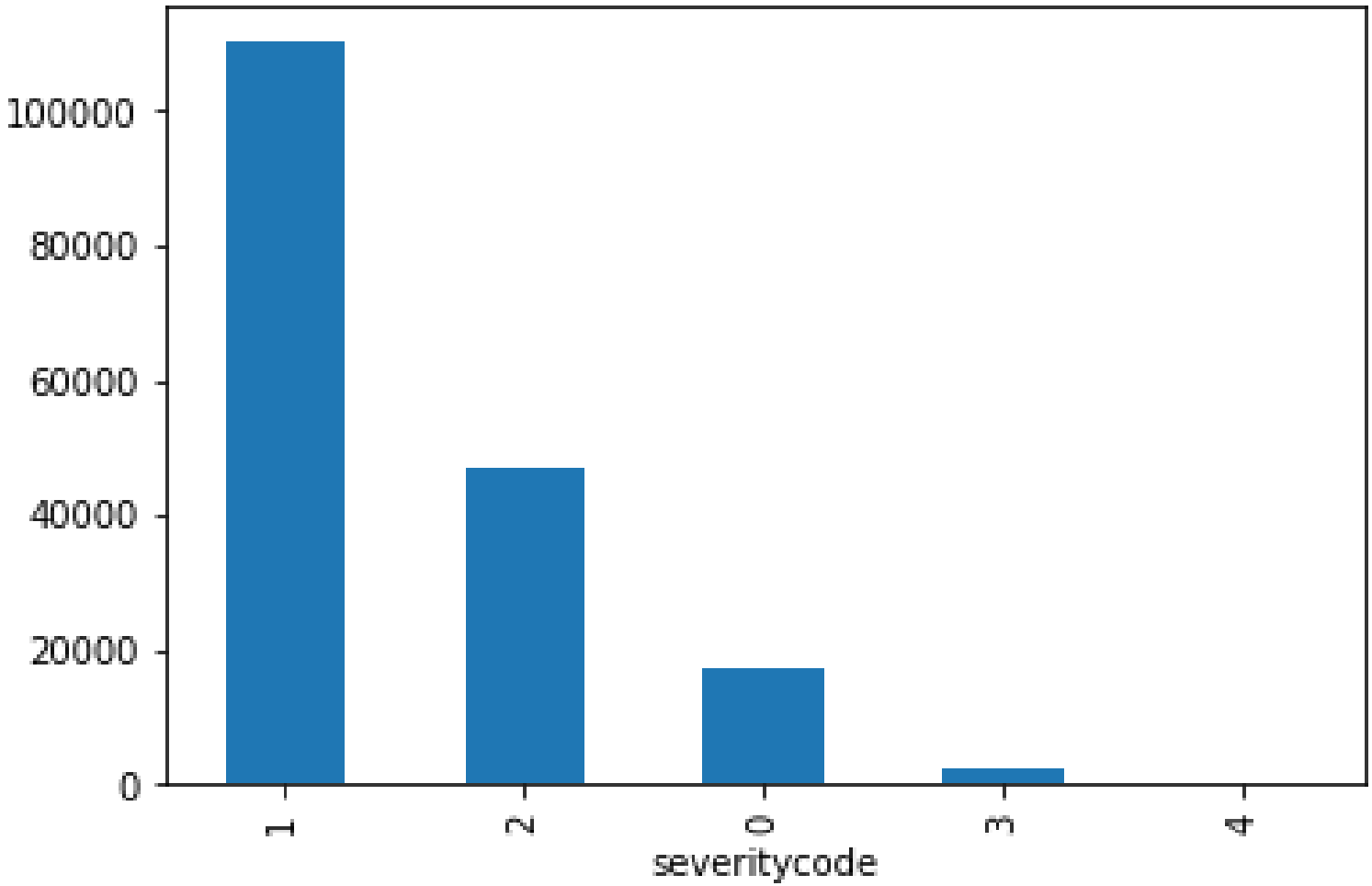
Correlation Analysis

	ADDRTYPE	SEVERITYCODE	COLLISIONTYPE	INJURIES	SERIOUSINJURIES	FATALITIES	JUNCTIONTYPE	SDOT_COLCODE	UNDERINFL
ADDRTYPE	1.000000	0.209163	0.231114	0.165353	0.034163	0.008504	0.295614	0.048061	0.077182
SEVERITYCODE	0.209163	1.000000	0.460696	0.700391	0.280070	0.168462	0.176663	0.311601	0.518682
COLLISIONTYPE	0.231114	0.460696	1.000000	0.233704	0.101766	0.042020	0.204633	0.378235	0.403133
INJURIES	0.165353	0.700391	0.233704	1.000000	0.279368	0.067180	0.119387	0.138529	0.167160
SERIOUSINJURIES	0.034163	0.280070	0.101766	0.279368	1.000000	0.173007	0.008052	0.086668	0.053116
FATALITIES	0.008504	0.168462	0.042020	0.067180	0.173007	1.000000	-0.002087	0.045834	0.030762
JUNCTIONTYPE	0.295614	0.176663	0.204633	0.119387	0.008052	-0.002087	1.000000	0.153773	0.115806
SDOT_COLCODE	0.048061	0.311601	0.378235	0.138529	0.086668	0.045834	0.153773	1.000000	0.260193
UNDERINFL	0.077182	0.518682	0.403133	0.167160	0.053116	0.030762	0.115806	0.260193	1.000000
WEATHER	-0.065416	0.112097	0.039913	-0.036641	-0.009453	-0.005765	-0.059323	-0.023557	0.273222
ROADCOND	-0.025085	0.250832	0.170047	0.018521	0.001953	-0.004540	-0.011316	0.072950	0.460852
LIGHTCOND	-0.010892	0.222471	0.141959	0.023051	0.016723	0.007310	-0.017136	0.102368	0.439445
SEGLANEKEY	0.038785	0.097485	0.167047	0.059399	0.031577	0.005112	0.016764	0.202097	0.016853
CROSSWALKKEY	0.168820	0.167778	0.235657	0.100689	0.055903	0.031851	0.040911	0.187265	0.031643

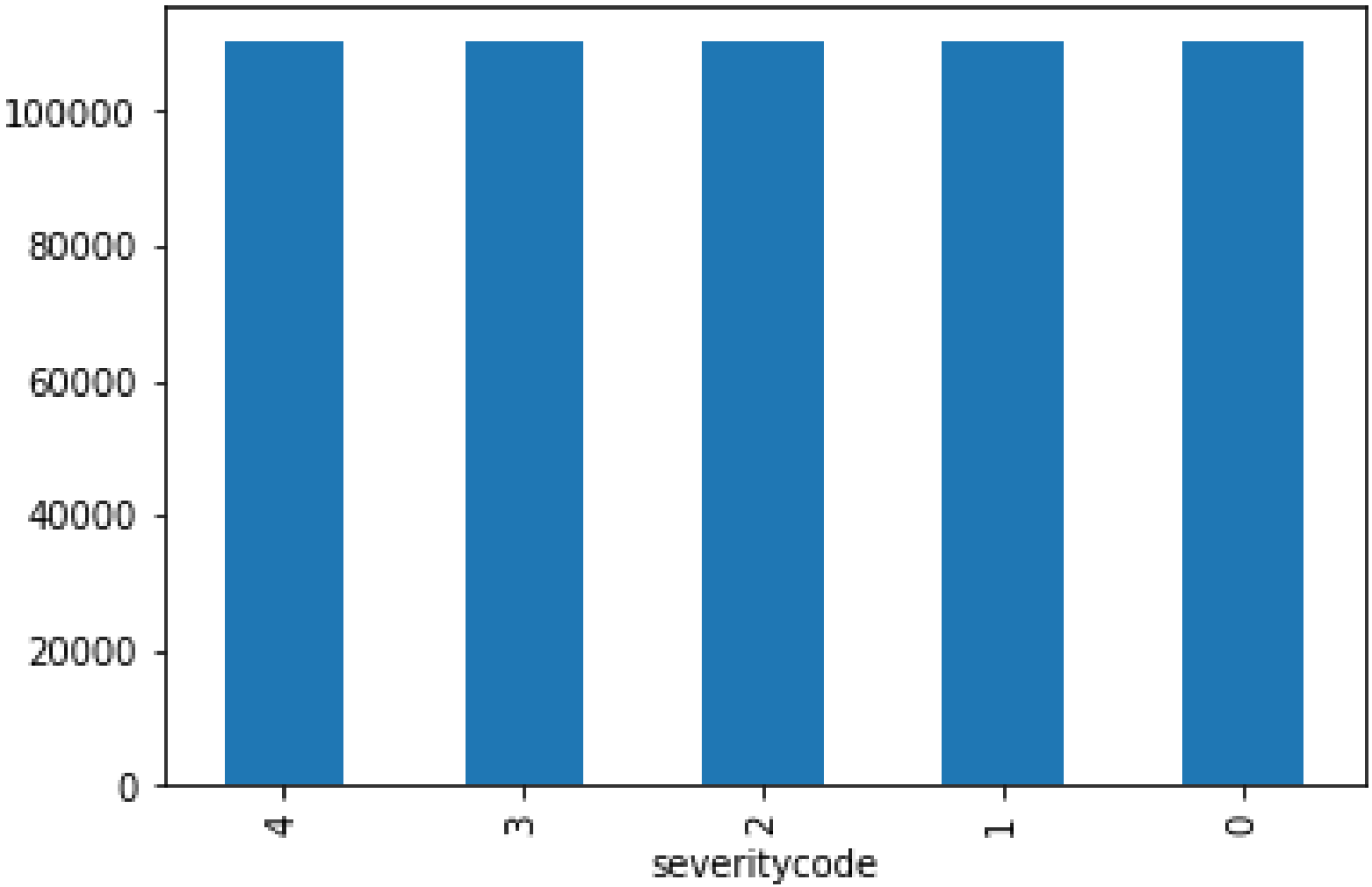
Model Building

- Train/test split done in 80:20.
- The dataset it highly imbalanced.
- The training set is balanced using SMOTE
- Models we built - KNN, SVM, Logistic regression

Train set before upsampling



Train set after upsampling



Evaluation

- Model have to be evaluated to measure and compare their performance.
- The evaluation metrics used are F1-score, Precision and Recall
- We have also used confusion matrix for each model

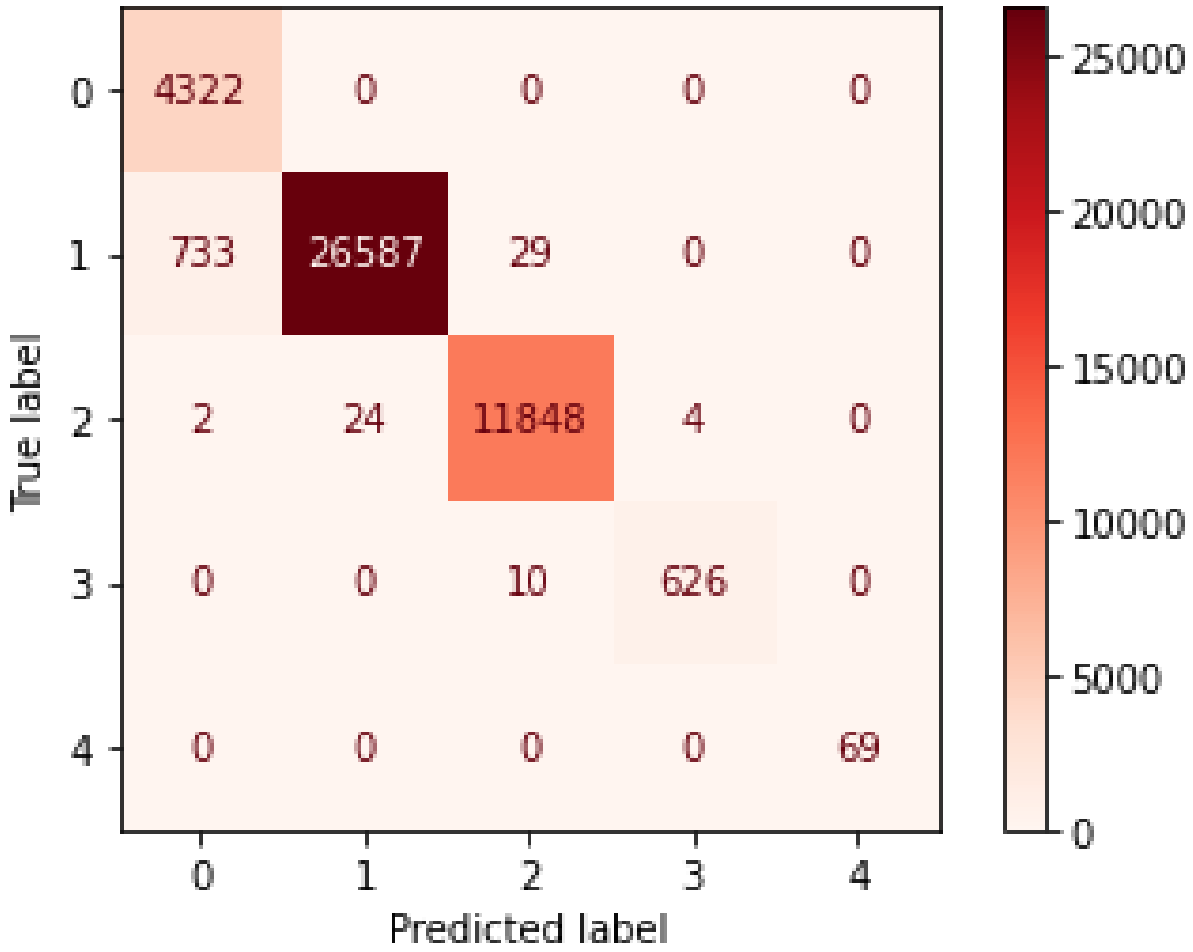
Result

	Algorithm	F1-score	Precision	Recall
0	KNN	0.9824	0.9843	0.9819
1	SVM	0.9829	0.9848	0.9824
2	LogisticRegression	0.9837	0.9856	0.9832

Confusion matrix - KNN

```
plot_confusion_matrix(neigh, X_test, y_test,cmap='Reds')
```

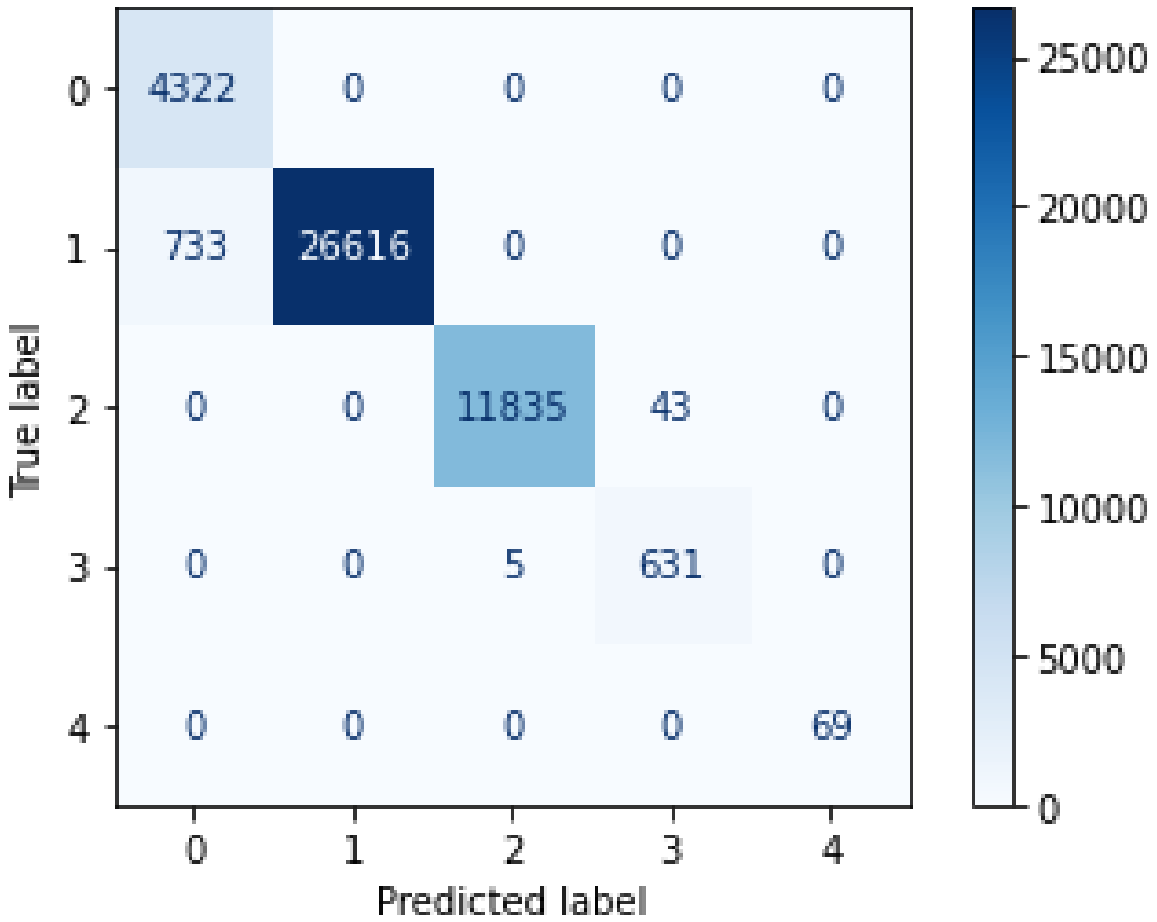
```
<sklearn.metrics._plot.confusion_matrix.ConfusionMatrixDis
```



Confusion matrix - SVM

```
plot_confusion_matrix(clf, X_test, y_test,cmap='Blues')
```

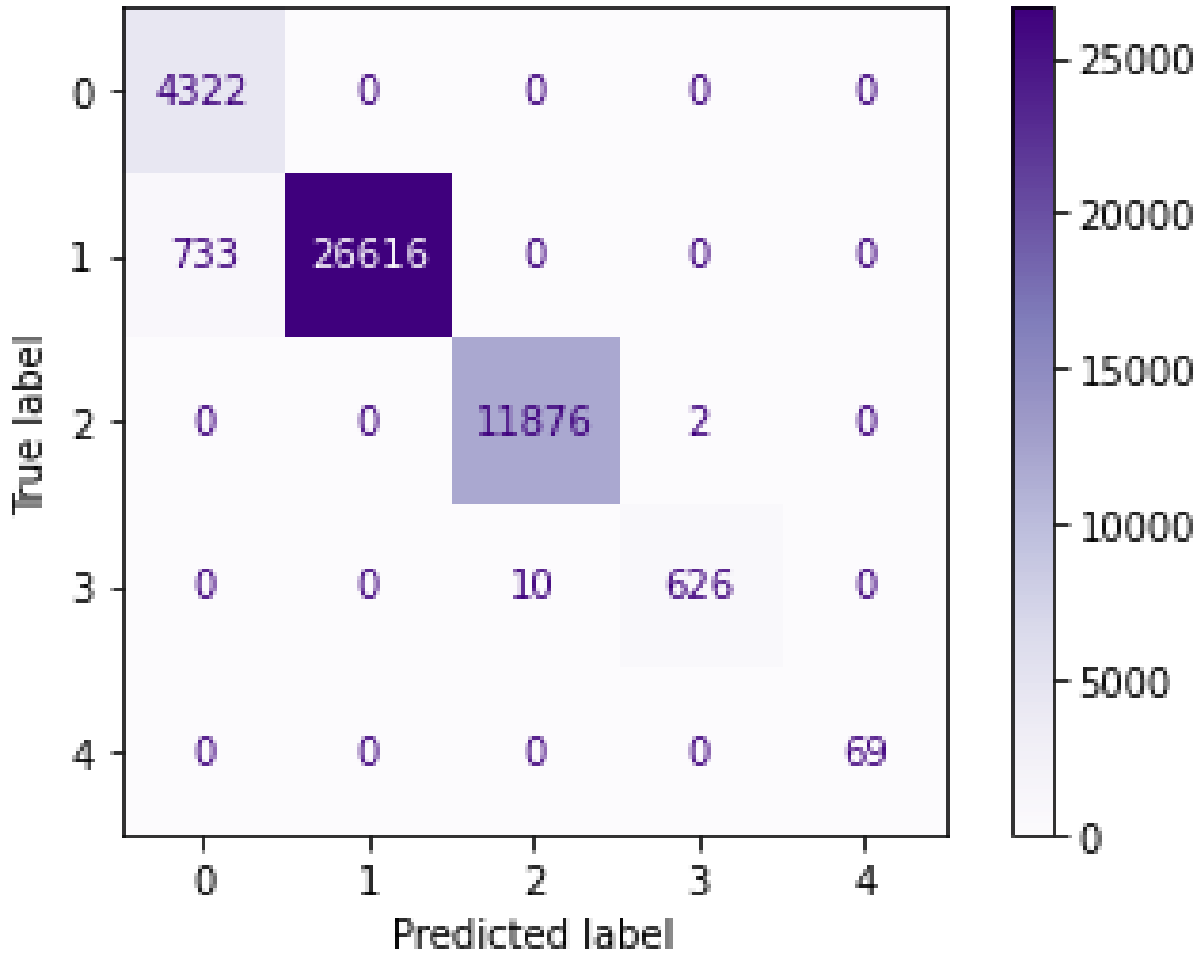
```
<sklearn.metrics._plot.confusion_matrix.ConfusionMatrixD
```



Confusion matrix - Logistic Regression

```
plot_confusion_matrix(LR, X_test, y_test,cmap='Purples')
```

```
<sklearn.metrics._plot.confusion_matrix.ConfusionMatrixDis
```



Conclusion

- Built model for road accident severity prediction using collision dataset provided by SDOT, Seattle.
- We used 11 best attributes to train and test the models.
- SVM model gives less misclassification as compared to other two models.
- It can be infered from the analysis that the consumption of alcohol or drugs, the location, the weather etc, have a high impact on accidents occurred.
- The model can be used for predicting the possibility of accident, and alerting the public, which will help them in reducing the severity.