Car Accident Severity

Introduction

Accidents in traffic lead to associated fatalities and economic losses every year worldwide and thus is an area of primary concern to society from loss prevention point of view. Modeling accident severity prediction and improving the model are critical to the effective performance of road traffic systems for improved safety.

Business Understanding

In an effort to reduce the frequency of car collisions in a community, an algorithm must be developed to predict the severity of an accident given the current weather, road and visibility conditions. When conditions are bad, this model will alert drivers to remind them to be more careful.
We can see that anyone who drives regularly could make use of the product. Transport departments will be benefitted as the losses would be cut down considerably.

Data understanding

The data was collected by the Seattle Police Department and Accident Traffic Records Department from 2004 to present.
The data consists of 37 independent variables and 194,673 rows. The dependent variable, “SEVERITYCODE”, contains numbers that correspond to different levels of severity caused by an accident from 0 to 4.
Severity codes are as follows:
0: Little to no Probability (Clear Conditions)
1: Very Low Probability — Chance or Property Damage
2: Low Probability — Chance of Injury
3: Mild Probability — Chance of Serious Injury
4: High Probability — Chance of Fatality
Furthermore, because of the existence of null values in some records, the data needs to be preprocessed before any further processing.

Data Preparation

Consider only the variables that would be more useful than the rest.
Following is the list of variables I have considered.

```
: features=['SEVERITYCODE','COLLISIONTYPE','PERSONCOUNT','VEHCOUNT','WEATHER','ROADCOND', 'LIGHTCOND','SPEEDING']
```

Check for the presence of Null values and either remove or replace them.

```
SEVERITYCODE
False    13846
Name: SEVERITYCODE, dtype: int64

COLLISIONTYPE
False    13709
True       137
Name: COLLISIONTYPE, dtype: int64

PERSONCOUNT
False    13846
Name: PERSONCOUNT, dtype: int64

VEHCOUNT
False    13846
Name: VEHCOUNT, dtype: int64

WEATHER
False    13703
True       143
Name: WEATHER, dtype: int64

ROADCOND
False    13704
True       142
Name: ROADCOND, dtype: int64

LIGHTCOND
False    13701
True       145
Name: LIGHTCOND, dtype: int64

SPEEDING
True     13086
False      760
Name: SPEEDING, dtype: int64
```

Apply the function value_counts() to get the frequency of the labels

```
df['WEATHER'].value_counts()
```

```
Clear                       8222
Raining                     2436
Overcast                    2078
Unknown                      779
Other                         76
Snowing                       53
Fog/Smog/Smoke                45
Sleet/Hail/Freezing Rain       8
Blowing Sand/Dirt              5
Severe Crosswind               1
Name: WEATHER, dtype: int64
```

This will help in deciding what values could be eliminated.

Modeling

K Nearest Neighbour (KNN) has been employed in my submission file.

After importing necessary packages and splitting preprocessed data into test and train sets, for each machine learning model, I have built and evaluated the model and shown the results as follow:

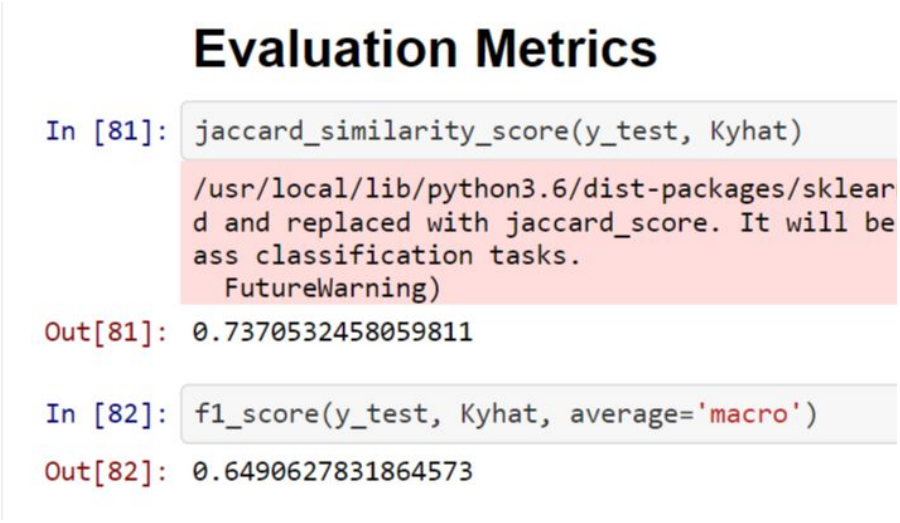
K Nearest Neighbor

```
In [79]: from sklearn.neighbors import KNeighborsClassifier
         k=25

In [80]: neigh= KNeighborsClassifier(n_neighbors=k).fit(X_train, y_train)
         neigh
         Kyhat= neigh.predict(X_test)
         Kyhat[0:5]

Out[80]: array([1, 1, 2, 1, 1])
```

Evaluation



Evaluation Metrics

```
In [81]: jaccard_similarity_score(y_test, Kyhat)

         /usr/local/lib/python3.6/dist-packages/sklear
         d and replaced with jaccard_score. It will be
         ass classification tasks.
           FutureWarning)

Out[81]: 0.7370532458059811

In [82]: f1_score(y_test, Kyhat, average='macro')

Out[82]: 0.6490627831864573
```

Results

Based on the above Evaluation metrics, KNN model is the pretty good at predicting the car accident severity.

Conclusion

Based on the dataset provided for this capstone from weather, road, light conditions, etc pointing to certain classes, we can conclude that particular conditions have a discernible impact on whether or not travel could result in any kind of injury or damage.