Predicting traffic incidents using historical data

Predicting severity of traffic incidents is useful for

- The General Public
  - We can alert drivers to be more alert in hazardous conditions
- First responders
  - We can provide insight on areas to be present that have a higher likelihood of severe accident

Data Acquisition and Cleaning

- The data we used is provided from coursera and has all collisions in Seattle as provided by SPD and recorded by Traffic Records from 2004 to present.
- Excess data from severity code was reduced to create a more accurate model
- Cleaned data removed excess data and converted written types to numerical data

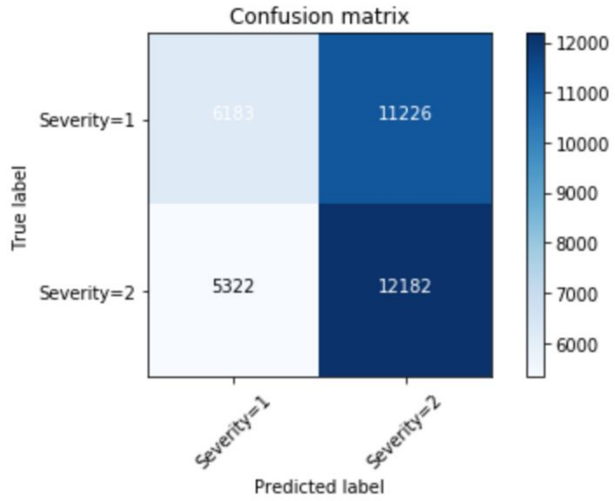
Using Logistic Regression to create model

```
n [21]: yhat_prob = LR.predict_proba(X_test)
        yhat_prob

Out[21]: array([[0.57133499, 0.42866501],
               [0.47089523, 0.52910477],
               [0.67367858, 0.32632142],
               ...,
               [0.47096955, 0.52903045],
               [0.47089523, 0.52910477],
               [0.47096955, 0.52903045]])
```

Confusion Matrix and f1-score

```
Confusion matrix, without normalization
[[ 6183 11226]
 [ 5322 12182]]
```



In [27]:
```python
print (classification_report(y_test, yhat))
```

```
              precision    recall  f1-score   support

           1       0.54      0.36      0.43     17409
           2       0.52      0.70      0.60     17504

   micro avg       0.53      0.53      0.53     34913
   macro avg       0.53      0.53      0.51     34913
weighted avg       0.53      0.53      0.51     34913
```

In [28]:
```python
from sklearn.metrics import log_loss
log_loss(y_test, yhat_prob)
```

Out[28]: 0.6849475402881953

Conclusion

- Due to lack of severity options we were unable to create an effective predictive model