**Car Accident Severity Prediction**

**Xue Wenchang**

**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. According to preliminary estimates from National Highway Traffic Safety Administration (NHTSA), 36,120 people died in motor vehicle crashes in [2019](https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812946), down 1.2 percent from 36,560 in 2018. Among all countries, the USA has the largest economic burden of road injuries of $487 billion, followed by China ($364 billion) and India ($101 billion); according to a research journal published by THE LANCET.

There is a lack of awareness amongst travelers regarding the risks they might be facing while taking certain routes, crossing certain areas, driving at a specific speed, driving on a specific road, and being inattentive while driving, etc. High-accident-prone areas are seldom inspected with regards to road maintenance, and deployment of additional emergency services personnel, causing additional damage caused by road accidents.

Reducing traffic accidents is an important public safety challenge around the world. With the help of modern machine learning technology, people may be able to predict the severity given some basic information related to the accident so that police can better handle the accident when they arrive at the scene.

This project aims to predict whether an accident that happens under a specific set of circumstances will be an accident limited to property damage or if it will include some form of physical injury to the driver and/or the passengers. The severity of an accident is the label used to train the model, and the proposed model can be used to predict the severity given some basic information related to the accident so that police can better handle the accident when they arrive at the scene.

**Data**

Data used in this project is provided by Coursera. It can be downloaded from <https://s3.us.cloud-object-storage.appdomain.cloud/cf-courses-data/CognitiveClass/DP0701EN/version-2/Data-Collisions.csv>

It contains 194,673 cases of accidents with related information. It will be used to build the machine learning model.

**Methodology**

Tensorflow will be used to build a neural network in order to predict the severity of the accident based on some given attributes of the accidents. The reason to use the neural network to tackle the problem is that it can automatically find out the relations between the input and output with high reliability if the model is correctly constructed and properly trained and tuned.

Only keep the data whose "STATUS" is "Matched" since "Unmatched" data are unreliable

The following attributes are kept and others are dropped to simplify the problem.

The attributes are 'SEVERITYCODE', 'ADDRTYPE', 'INATTENTIONIND', 'UNDERINFL', 'SDOT\_COLCODE', 'SPEEDING'. The place and condition of the accident is very related to the severity. Also, drivers’ mental and physical conditions play a vital part in an accident.

Empty cells are filled to process the data.

The Sample Data is divided into 3 parts:

74% for training

20% for test

16%for validation

The setting of the neural network is shown below

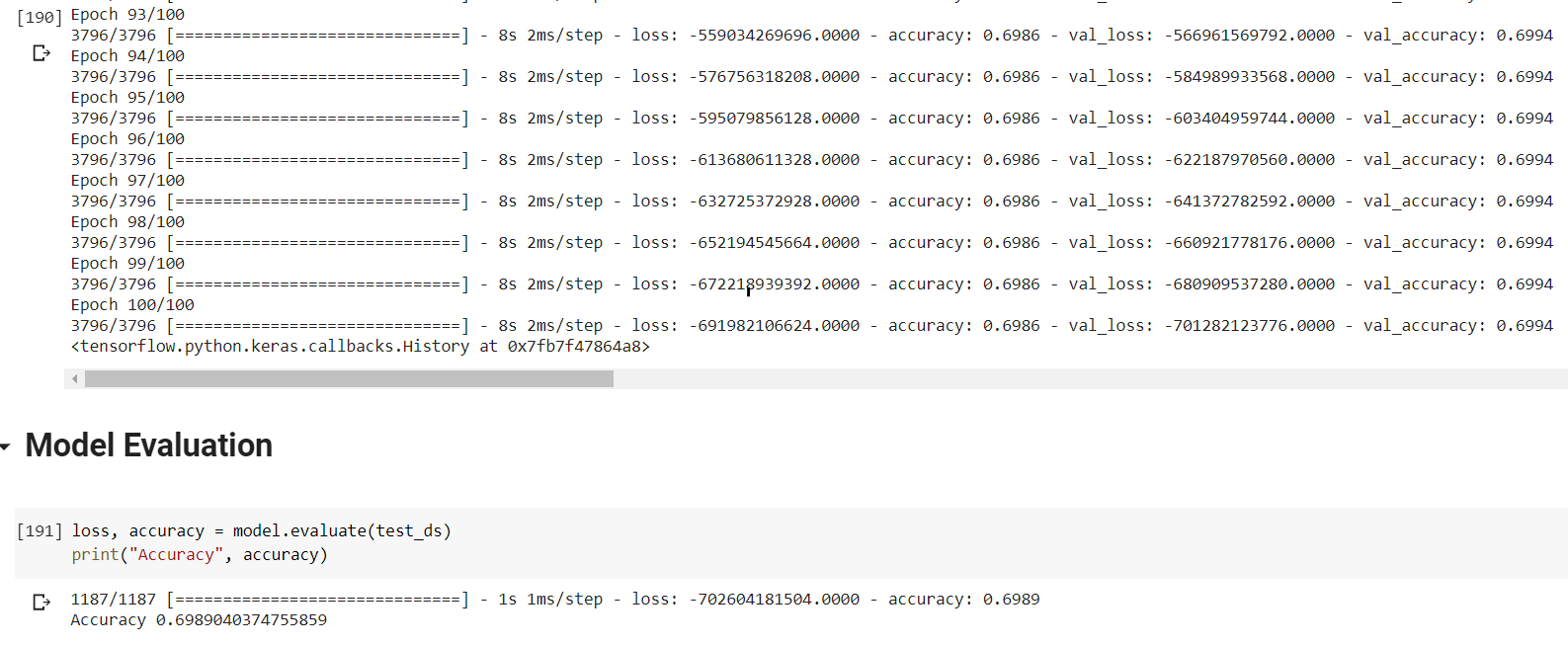
Feature layer

Dense layer (128)

Dense layer (128)

Dropout layer (.1)

Dense layer (1)

**Result**

The accuracy of the model prediction is 69.89%, which is close to the distribution of the data. The training process shows the model does not converge, which indicates it does not work.

**Discussion**

The model fails to predict the severity of the accident correctly based on given features as the accuracy is very close to the distribution of the data. The reason for failure may be the incorrect choice of features. Or the severity of the accidents has no relation with the attributes in the data. In other words, the severity cannot be predicted based on the given data.

**Conclusion**

A neural network is constructed with Tensorflow to predict the severity of the accidents using the given data. However, the training is not successful as the model does not converge. Further study needs to be done to find out the best approach to tackle the problem.