Data Science IBM

Leonardo Inza

Introduction/Business Problem

- The problem is that sometimes we are making a trip and in the middle of the road there was a collision and you have to wait two hours until all the accident area had taken care. So is there a way to predict this accident?
- The main idea is to predict an accident by some data that include weather, road condition, location, severity of the accident, type of coallision, etc.

Data

The Data atributes will be:

- SHAPE: x and y location

  - ADDRTYPE: type of address can be Alley, Block or Intersection

  - SEVERITYCODE: tells the severity of the collision can be: 3—fatality, 2b—serious injury, 2—injury, 1—prop damage, 0—unknown.

  - COLLISIONTYPE: type of collision

Data

- PERSONCOUNT: number of people that were involved in the accident.

- PEDCOUNT: number of pedestrians that were involved in the accident.

- PEDCYLCOUNT: number of bicycles that were involved in the accident.

- VEHCOUNT:number of vehicles that were involved in the accident.

- JUNCTIONTYPE: type of junction at the place of the accident.

- UNDERINFL: driver that was found involved was under the influence of drugs or alcohol?

Data

- WEATHER: type of wather during accident.

  - ROADCOND: condition of the road during accident.

  - LIGHTCOND: condition of the light during accident.

  - HITPARKEDCAR: in the accident a car that was parked was hit? (Y/N)

Data

- The idea of the data is to predict the severity of an accident, all variables except SEVERITYCODE are the independent variables and SEVERITYCODE y the dependent.
- First stage is the cleaning data, then design a machine learning model, train and validate it, and after that will display a graph to explain the conclusion.

Data Preparation

-   Delete columns that will not be use
-   Identify and delete rows nulls in X, COLLISIONTYPE, JUNCTIONTYPE, UNDERINFL, WEATHER, ROADCOND, LIGHTCOND
-   Convert String to categorical type
-   Split data to training and test
-   Scale data

Methodology

To make this proyect we are going to follow this steps:

- Data preparation: prepare data for our analysis
- Analysis: design deep learning model and train it until the accuracy is more than 70%
- Results: explain the results of our analysis
- Conclusion: brief conslusion of our work

Hypotesis: We are going to make a deep learning model to see if we can predict trafict accidents, the idea is to see if the model can be at least 70 percent accuracy with data that dosen´t know. I will split data in training and test datasets and test data is the information that is new for the deep learning model.

Methodology

- After data was ready to be used we make the ANN
    - Input layer was 14 nodes
    - Output layer was 1 node and predict values from 0 to 1
    - 1 hidden layer

- Then we evaluate the model with a conffusion matrix

Results

We made a plot of the confussion matrix to explain the results.

- In the plot we can see the percentage of all posible options in the confussion matrix
    - if we sum 65% with 10% we get 75% that the model predict one value and the reality was correct
    - So we can say that our model have an accuracy of it
    - If we sum 21% with 4% is 25% that the model predict one value and the reality was the other values.
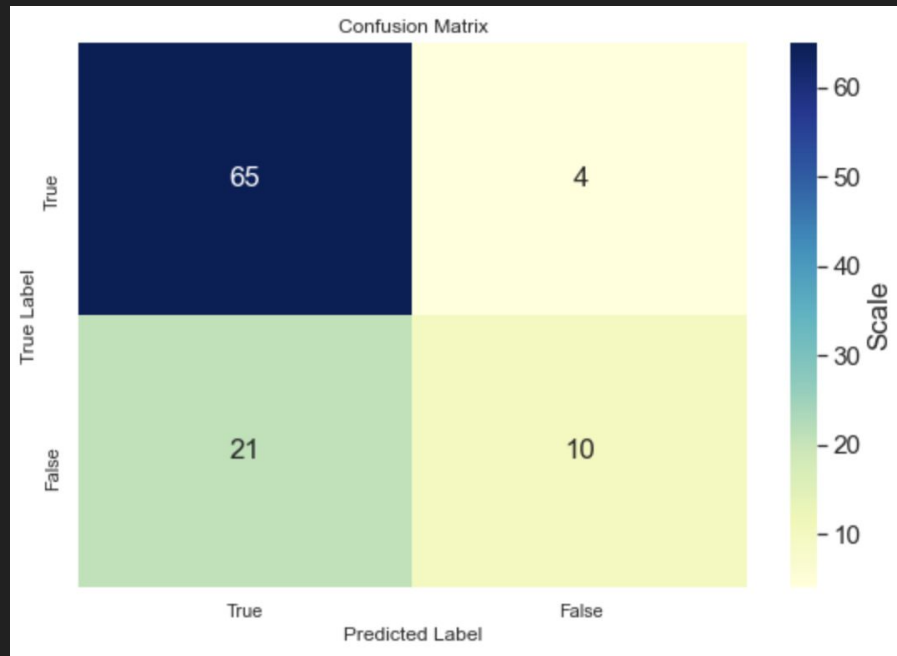
Conclusion

- In this proyecto we started with a data that wasn´t prepared for any machine learning model, so we had to prepare them, then we design a artificial neuran network to predict future accidents, and then we explain the results.

- Our objective was to achive a model that had 70% of accuracy and we get 75% so the results are greats, we end up deploying a model that can give us a trend of a posible accident, imagine of going on a trip and have a deep learning model that can help us avoid posible incidents, now we have it!