Road Traffic Conditions – Predictive Analysis

1. **Introduction:**

I have to calculate the probability of Fatal Auto accidents. There are multiple dependent and independent factors to consider. Among them are Road Conditions, Visibility, Time of the Day, Day of the Week, Number of Vehicles Involved, the Road Type, Speed Limit (in Kilometers), the Weather and Road Number (Identification).

I will create the best possible model to perform Predictive Analysis on the test case.

1. **Data Description:**

I will use the data of the Example Dataset of the week 1 to make this project. It consists of 37 columns (including weather and road conditions, speed...) and around 190,000 rows of different accidents.

1. **Methodology:**

My approach was to code in Python and running the object using Jupyter Notebook on an IBM Watson platform. First of all, I have removed the irrelevant columns taking into account the correlation between variables. Afterwards, I have removed or filled the NaN cells and I have realized that there were ten columns that required data cleansing from a Python object into an Integer.

These columns were ADDRTYPE, COLLISIONTYPE, WEATHER, ROADCOND, UNDERINFL, LIGHTCOND, INATTENTIONIND, PEDROWNOTGRNT, SPEEDING and HITPARKEDCAR. In addition, I assigned arbitrary numeric values to each of these four columns.

To calculate the accuracy, I calculated the F1 Score and Jaccard Similarity.

1. **Results:**

I applied logistic regression model, support vector machines (SVM), decision tree and knn models to the dataset, using f1-score and jaccard score as the tuning and evaluation metric. The results all had the same problems. Most of the real values of injuries (1) are predicted as 0. These results were not acceptable.

|  |  |  |
| --- | --- | --- |
|  | F1 Score | Jaccard Score |
| Logistic Regression | 0.68 | 0.70 |
| Support Vector Machine | 0.69 | 0.71 |
| KNN | 0.7 | 0.69 |
| Decision tree | 0.66 | 0.71 |

1. **Conclusions:**

In this study, I analyzed the relationship between different factors and the cars accidents´ gravity. I identified Collision address type (ADDRTYPE), Collision type (COLLISIONTYPE), The total number of people involved in the collision (PERSONCOUNT), the total number of people involved in the collision (PEDCOUNT), The number of bicycles involved in the collision (PEDCYLCOUNT) and whether or not the pedestrian right of way was not granted (PEDROWNOTGRNT) among the most important features that affect a car accidents gravity. I built classification models to predict whether a car accident is grave or not severe. These models can be very useful in helping drivers to be carefull in dangerous places and conditions and reduce the number of fatal accidents.