University of Pisa
Msc in Artificial Intelligence and Data Engineering
Data Mining and Machine Learning

Application Data Mining Project

Leonardo Turchetti

Academic Year 2020/2021

Indice

1 Introduction

The introduction of the car ha enhanced the lives of many individuals and societies, but there is another side to the coin. In recent decades, for most of the world's population, the weight of road-traffic injury and deaths is become very substantially in terms of societal and economic costs.
Road traffic crashes causes the death of approximately 1.35 million of people around the world each year and are the leading cause of death for children and young adults. The WHO estimate that road traffic crashes cost most countries 3% of their Gross Domestic Product.

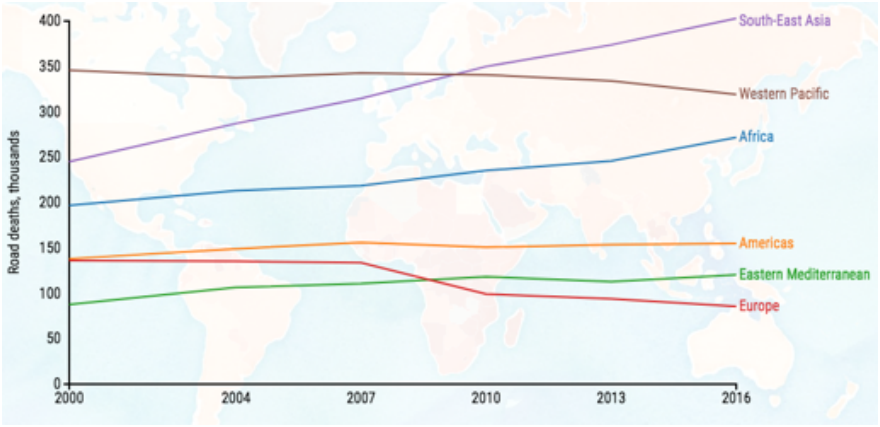

Figure 1: Tendency of the car accident people deaths

1.1 Importance of Forecasting

Traffic accident prediction is become an important and challenging problem in the domain of intelligent traffic safety management system. This effort will benefit the responsible authorities in black spot improvement and road safety to prevent or reduce any damages caused by accidents.
Accurate predictions can provide information for emergency responders to evaluate and estimate the potential impacts and implements efficient accidents management procedures.

2 Application

The main purpose of this application is to let an user know the severity of a car accident. It use a trained machine learning model and csv file containing a car accident for predicting the class of severity and it does not store any information about the user.

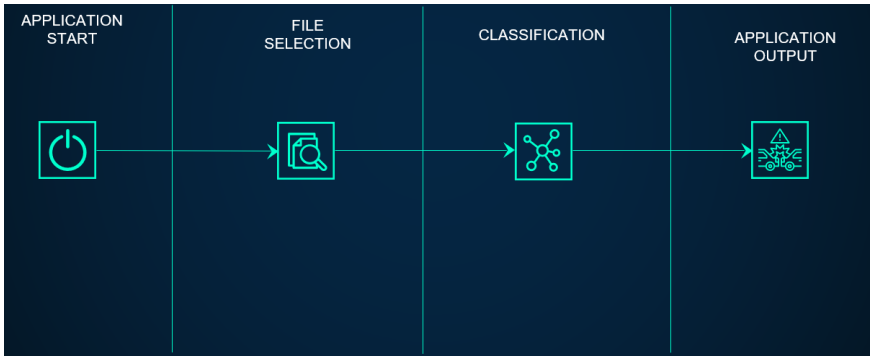

Figure 2: Application Flow

2.1 Use Case

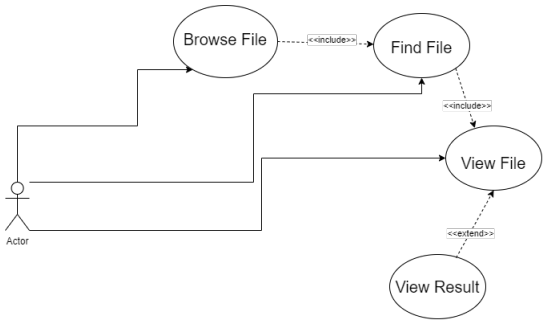

Figure 3: Use Case Diagram

2.2 Input

The application takes in input a csv file that is passed from the user using the prompt and this file containing an array of int values that represent the features of a car accident.
The features are divided in 3 different macrocategories:
1)Coordinates, that are information about the space (Latitude and Longitude) and the time (Minuts)
2)Weather, that are information about some weather paramaters like Temperature, Humidity, Pressure, Wind Speed, etc.
3)Boolean Parameters, that represent the presence or not presence of some elements near the accident area

$$X = [f0, .., f25]$$

where fi represent the ith feature of the car accident

```
(base) MBP-di-Valenina:Desktop valeninaberardi$ python Applicazione.py
/Users/valeninaberardi/opt/anaconda3/lib/python3.8/site-packages/sklearn/base.py:329: User
Warning: Trying to unpickle estimator DecisionTreeClassifier from version 0.22.2.post1 whe
n using version 0.23.2. This might lead to breaking code or invalid results. Use at your o
wn risk.
  warnings.warn(
/Users/valeninaberardi/opt/anaconda3/lib/python3.8/site-packages/sklearn/base.py:329: User
Warning: Trying to unpickle estimator RandomForestClassifier from version 0.22.2.post1 whe
n using version 0.23.2. This might lead to breaking code or invalid results. Use at your o
wn risk.
  warnings.warn(
Insert the file to analyze: /Users/valeninaberardi/Desktop/gravity4.csv
The result is: 4
```

Figure 4: Prompt vision of the application

2.3 Output

The application give in output on the prompt to the user a string format that represent the prediction of the severity class of the car accident that the user has passed to the application before. The severity of a car accident is expressed by a number and there are 4 different possible values.

2.4 Application Code

The following is how the application code was implemented

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt

#LOAD THE MODEL
filename = '/Users/valeninaberardi/Desktop/joblib_model.sav'
import joblib
model = joblib.load(filename)
#ASK FOR THE FILE TO ANALYZE
fileinput = input("Insert the file to analyze: ")
#READ THE FILE
rowToAnalyze = pd.read_csv(fileinput)
#PREDICT
prediction = model.predict([rowToAnalyze.iloc[0]])
#SHOW THE RESULT
print("The gravity of the car accident is: " + ...
    prediction[0].astype(str))
```

3 GitHub

https://github.com/Leonardo-Turchetti/DataMining