Predicting Traffic Accident Severity

1. Introduction

1.1. Background

Every year the lives of approximately 1.35 million people are cut short as a result of a road traffic crash. Between 20 and 50 million more people suffer non-fatal injuries, with many incurring a disability as a result of their injury.

Road traffic injuries cause considerable economic losses to individuals, their families, and to nations. These losses arise from the cost of treatment as well as lost productivity for those killed or disabled by their injuries, and for family members who need to take time off work or school to care for the injured. Road traffic crashes cost most countries 3% of their gross domestic product.

1.2. Who is at risk?

1.2.1. Socioeconomic status

More than 90% of road traffic deaths occur in low- and middle-income countries. Road traffic injury death rates are highest in the African region. Even within high-income countries, people from lower socioeconomic backgrounds are more likely to be involved in road traffic crashes.

1.2.2. Age

Road injuries are the leading cause of death for children and young adults aged 5-29 years.

1.2.3. Sex

From a young age, males are more likely to be involved in road traffic crashes than females. About three quarters (73%) of all road traffic deaths occur among young males under the age of 25 years who are almost 3 times as likely to be killed in a road traffic crash as young females.

1.3. Risk factors

1.3.1. The Safe System approach: accommodating human error

The Safe System approach to road safety aims to ensure a safe transport system for all road users. Such an approach considers people's vulnerability to serious injuries in road traffic crashes and recognizes that the system should be designed to be forgiving of human error. The cornerstones of this approach are safe roads and roadsides, safe speeds, safe vehicles, and safe road users, all of which must be addressed in order to eliminate fatal crashes and reduce serious injuries.

1.3.2. Speeding

- An increase in average speed is directly related both to the likelihood of a crash occurring and to the severity of the consequences of the crash. For example, every 1% increase in mean speed produces a 4% increase in the fatal crash risk and a 3% increase
- in the serious crash risk. The death risk for pedestrians hit by car fronts rises rapidly (4.5 times from 50 km/h to 65 km/h)..
- In car-to-car side impacts the fatality risk for car occupants is 85% at 65 km/h.

1.3.3. Driving under the influence of alcohol and other psychoactive substances

- Driving under the influence of alcohol and any psychoactive substance or drug increases the risk of a crash that results in death or serious injuries.
- In the case of drink-driving, the risk of a road traffic crash starts at low levels of blood alcohol concentration (BAC) and increases significantly when the driver's BAC is ≥ 0.04 g/dl.
- In the case of drug-driving, the risk of incurring a road traffic crash is increased to differing degrees depending on the psychoactive drug used. For example, the risk of a fatal crash occurring among those who have used amphetamines is about 5 times the risk of someone who hasn't.

1.3.4. Nonuse of motorcycle helmets, seatbelts, and child restraints

- Correct helmet use can lead to a 42% reduction in the risk of fatal injuries and a 69% reduction in the risk of head injuries.
- Wearing a seatbelt reduces the risk of death among drivers and front seat occupants by 45 - 50%, and the risk of death and serious injuries among rear seat occupants by 25%.
- The use of child restraints can lead to a 60% reduction in deaths.

1.3.5. Distracted driving

There are many types of distractions that can lead to impaired driving. The distraction caused by mobile phones is a growing concern for road safety.

- Drivers using mobile phones are approximately 4 times more likely to be involved in a crash than drivers not using a mobile phone. Using a phone while driving slows reaction times (notably braking reaction time, but also reaction to traffic signals), and makes it difficult to keep in the correct lane, and to keep the correct following distances.
- Hands-free phones are not much safer than hand-held phone sets, and texting considerably increases the risk of a crash.

1.3.6. Unsafe road infrastructure

The design of roads can have a considerable impact on their safety. Ideally, roads should be designed keeping in mind the safety of all road users. This would mean making sure that there are adequate facilities for pedestrians, cyclists, and motorcyclists. Measures such as footpaths, cycling lanes, safe crossing points, and other traffic calming measures can be critical to reducing the risk of injury among these road users.

1.3.7. Unsafe vehicles

Safe vehicles play a critical role in averting crashes and reducing the likelihood of serious injury. There are several UN regulations on vehicle safety that, if applied to countries’ manufacturing and production standards, would potentially save many lives. These include requiring vehicle manufacturers to meet front and side impact regulations, to include electronic stability control (to prevent over-steering) and to ensure airbags and seatbelts are fitted in all vehicles. Without these basic standards the risk of traffic injuries – both to those in the vehicle and those out of it – is considerably increased.

1.3.8. Inadequate post-crash care

Delays in detecting and providing care for those involved in a road traffic crash increase the severity of injuries. Care of injuries after a crash has occurred is extremely time-sensitive: delays of minutes can make the difference between life and death. Improving post-crash care requires ensuring access to timely prehospital care and improving the quality of both prehospital and hospital care, such as through specialist training programs.

1.3.9. Inadequate law enforcement of traffic laws

If traffic laws on drink-driving, seatbelt wearing, speed limits, helmets, and child restraints are not enforced, they cannot bring about the expected reduction in road traffic fatalities and injuries related to specific behaviors. Thus, if traffic laws are not enforced or are perceived as not being enforced it is likely they will not be complied with and therefore will have very little chance of influencing behavior. Effective enforcement includes establishing, regularly updating, and enforcing laws at the national, municipal, and local levels that address the above-mentioned risk factors. It includes also the definition of appropriate penalties.

1.4. What can be done to address road traffic injuries?

Road traffic injuries can be prevented. Governments need to take action to address road safety in a holistic manner. This requires involvement from multiple sectors such as transport, police, health, education, and actions that address the safety of roads, vehicles, and road users. Effective interventions include designing safer infrastructure and incorporating road safety features into land-use and transport planning, improving the safety features of vehicles, improving post-crash care for victims of road crashes, setting and enforcing laws relating to key risks, and raising public awareness.

2. DATA

2.1. Data Source

The data can be found in the following Kaggle data.

2.2. Feature Selection

Five different data sets, consisting of all the recorded different accidents in France from 2005 to 2016 were collected from Kaggle. Mainly these data sets consist of different characteristics such as time, place, type of collision, weather, light conditions and types of intersection. The places dataset consists such as gradient, shape and category of the road, traffic rule and infrastructure and surface condition. Similarly, in user data set, users related data is found such as vehicle users, accident of users, reason of travelling, severity of accident, pedestrians, and use of safety equipment. In addition to that, the vehicle data sets possess type of vehicle involve in accident. These data sets possess irrelevant features which was removed or cleaned for the initial analysis and selected the relevant features which reduces from 54 to 28. From the characteristic's dataset following feature were selected such as lighting, localization, type of intersection, atmospheric conditions, type of collisions, department, time and the coordinates. In addition to these features, two new features were added, they are I) date to perform a seasonality analysis of the accident severity, and II) weekend indicating if the accident occurred during the weekend or not. Similarly, from places dataset, the following features were selected such as road category, traffic regime, number of traffic lanes, road profile, road shape, surface condition, school nearby and infrastructure. In users data set new features were added such as total number of people involved in the accident as number of users, in accident whether pedestrians were involved or not; if yes 1 otherwise 0, and the age of the users (between 17-31 year old). Similarly, severity of accidents was identified as nature of accident, if light or unscathed then it is 0 otherwise hospitalized, heavy wounded or death is 1.

2.3. Data Cleaning

Data cleaning is the process of removing or adding data which is relevant for the analysis. In this project, following steps were used to clean the data.

I) The NaN, 0 or outlier, which is related with longitude, latitude, and road number were dropped.

II) Replaced the missing values such as atmospheric conditions, type of collision, road category and the surface conditions by other values.

III) The school values were replaced with value of 1 for 100 and departmental were divided by 10.

IV) Define date feature with the datetime type.

3. Data Analysis

The data visualization was primarily used to define the target variables. From the graph, it is confirmed that the data set is balanced with more cases of lower severity. Then, seasonality plot was used to define the global trend of the accidents yearly, monthly, and day of the week.

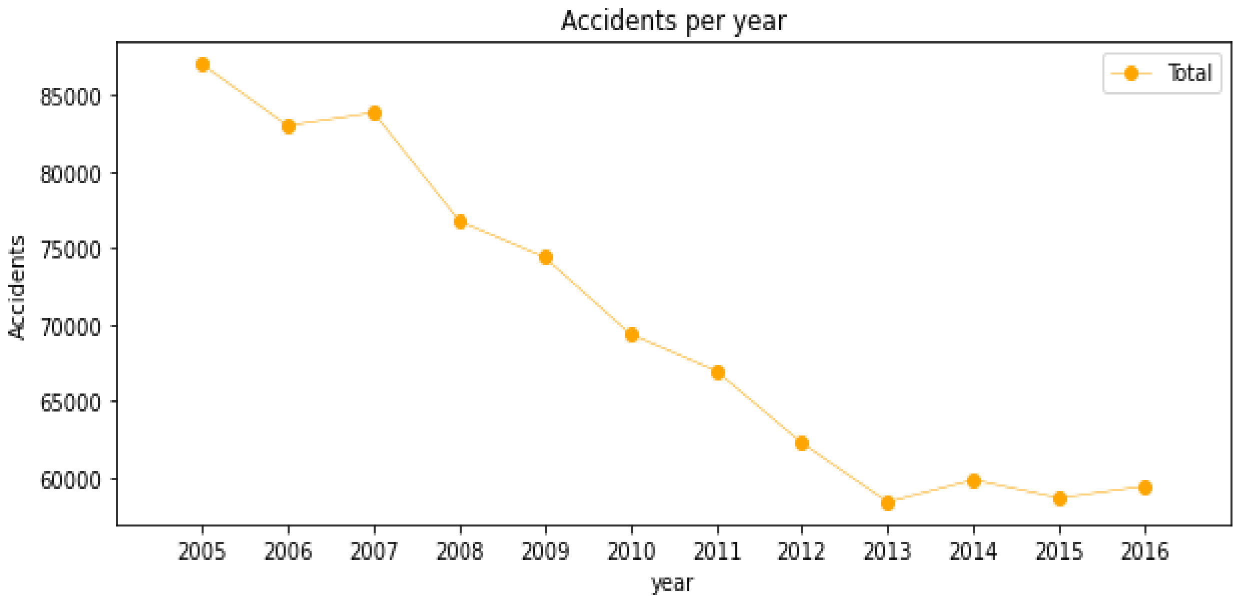

Figure 1: Total number of accidents per year

From the figure 1. We can conclude that the traffic accident decreased by over the years from 2005 to 2016. From the year 2013 -2016, the accidents seem to be a normal.

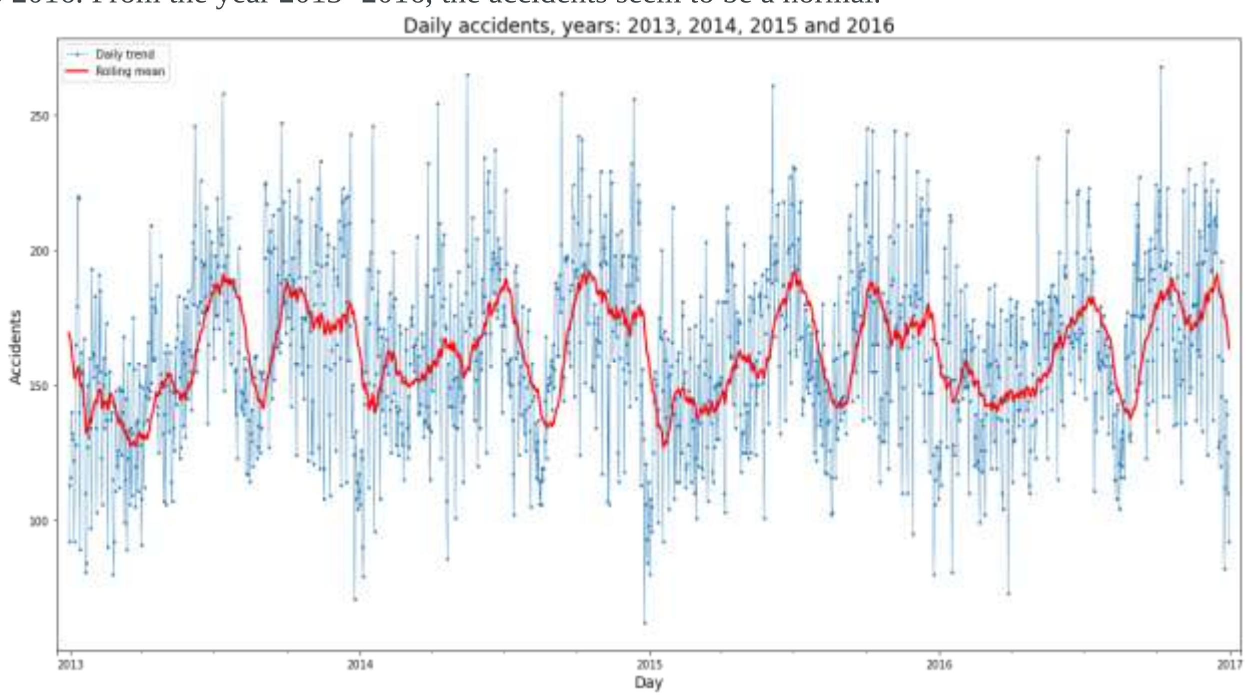

Figure 2: Total number of accidents per year with mean of 30 days

.

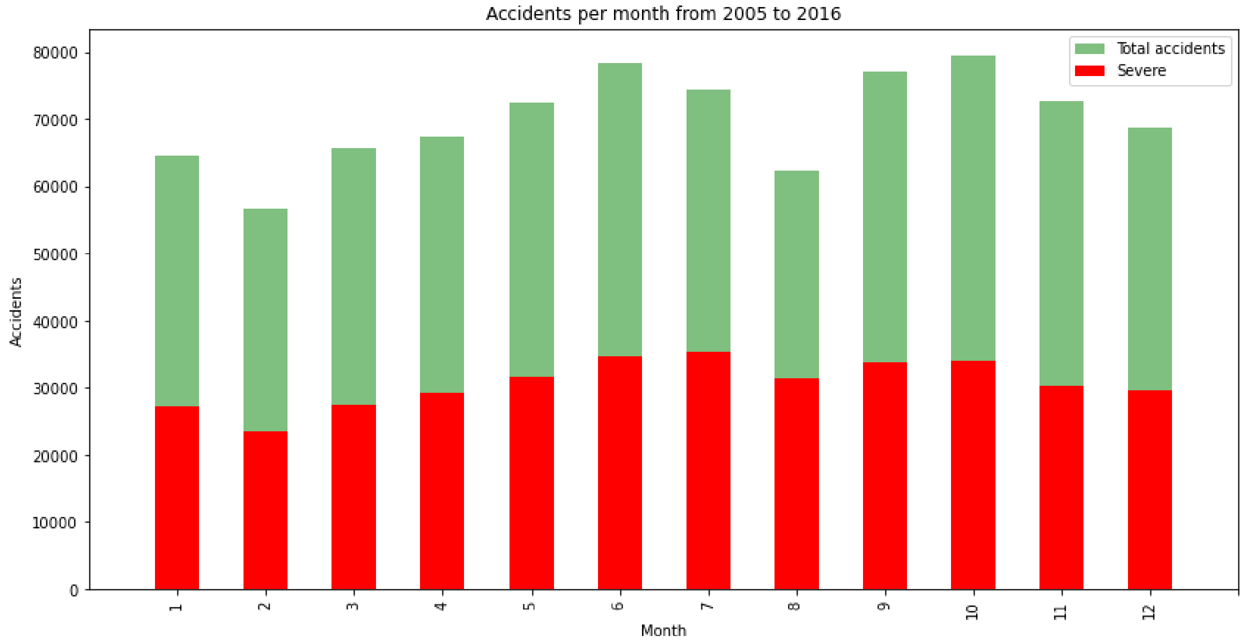

Figure 3: Number of accidents per month from 2005 to 2016

The traffic accident was decreased in trend over the years, however, the monthly from March and in September (Figure 2 and 3)

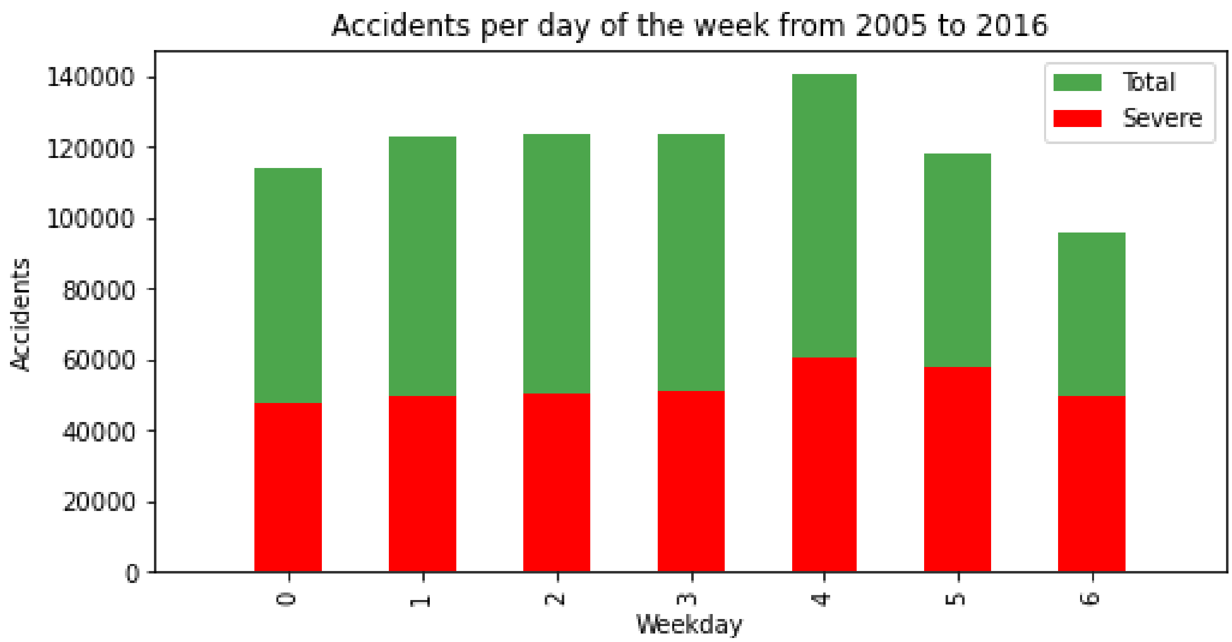

Figure 4: Number of accidents per day of the week from 2005 to 2016

In the figure 4, we can see the significant increase of accident on Friday and less accident recorded on Sunday.

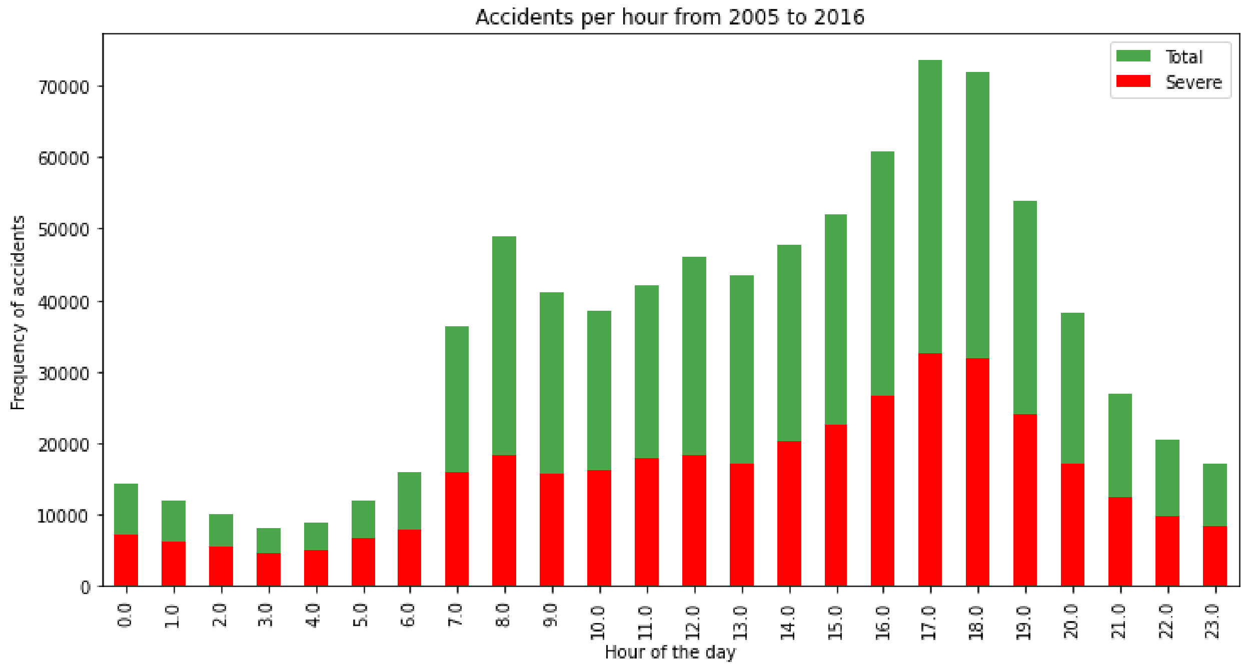

Figure 5: Number of accident per day with time

In figure 5, we can clearly observe that the traffic accident mainly occurs at 8 am and 5 pm, which is mainly the office time, which might result in the traffic congestion. The higher proportion of the accident was observed from noon-morning which is almost 51%, while from 7am to 8pm is 42%. There is clear observation of accident regarding the time of the season, but no significant correlation was observed with all features while analyzing the Pearson correlation coefficient.

4. Predictive modelling

The four different modelling approaches were used.

1. Decision tree (Random Forest)
2. Logistic regression
3. K- nearest neighbor
4. Supervised vector machine (SVM)

These algorithms are supervised learning predicting approach with certain accuracy and computational time. These two properties have been compared in order to determine the best suited algorithm fort his specific problem. Primarily, the data were split into training and test data sets by 80 and20 to create the validation set for the development of the models. The data was standardized with mean value 0 and unit variance. With the train and validation sets the best hyperparameters were selected and using the test set the accuracy and computational time for the development of the models were calculated. After evaluating the parameters for each algorithm these were the models.

1. Random Forest:

Decision tress: 10

Maximum depth of features: 12

2. **Logistic Regression**

C= 0.001

3. **K_Nearest Neighbour**

K=15

4. **Supervised Vector Machine**

Training size=75,000 samples

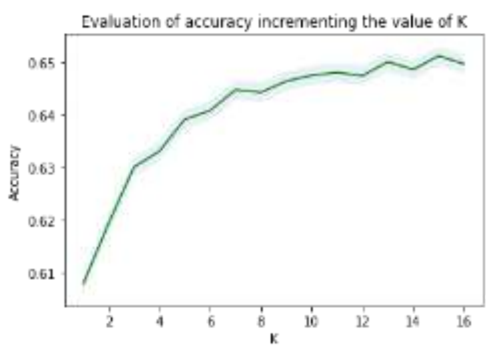

Figure 6 : KNN model

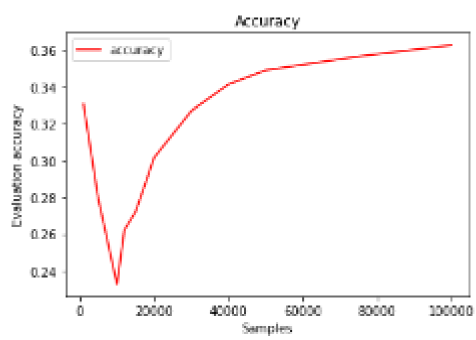

Figure 7: SVM model

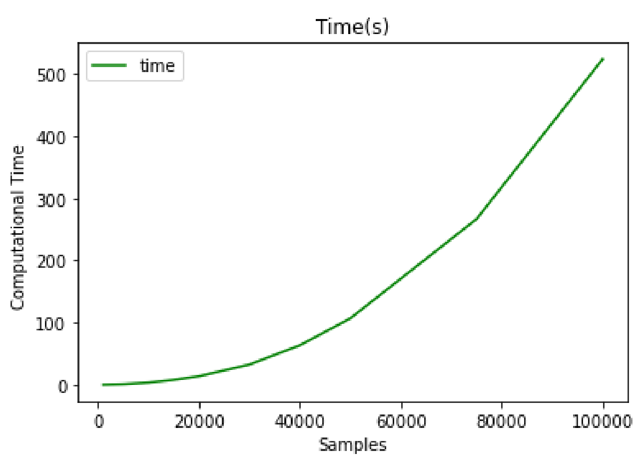

Figure 8: SVM model with computational time increases with samples size

The accuracy was found to be increased with size of training set (figure 6&7) while figure 8 shows the increase of computational time with sample size.

5. **Result**

Algorithm	Jaccard	F1-score	Precision	Recall	Time(S)
Random Forest	0.46987	0.63934	0.734314	0.566119	4.736231
Logistic Regression	0.40058	0.57202	0.665856	0.50137	5.833542
KNN	0.37563	0.54612	0.665315	0.463148	181.549
SVM	0.35691	0.5260	0.691429	0.424539	295.6989

From the above table, we can see that model Random Forest shows better performance in compared to the other models. With reference to the time, recall and precision, random forest considers as best model. However, logistic regression shows comparable performance.

6. **Conclusion**

From this study, I found that department, day and time, road condition features were main features for different severity of accident. I compared the four different models with severity of accidents and features. By identifying the features that favor the most the gravity of an accident, these could be tackled by improving road conditions or increasing the awareness of the population.

7. **References:**

1. World Health Organization (WHO). Global Status Report on Road Safety 2018. December 2018. [cited 2019 April 8]. Available from URL: https://www.who.int/violence_injury_prevention/road_safety_status/2018/en/external icon
2. Centers for Disease Control and Prevention (CDC), National Center for Injury Prevention and Control (NCIPC). Web-based Injury Statistics Query and Reporting System (WISQARS). [cited 2019 November 4]. Available from URL: http://www.cdc.gov/injury/wisqars
3. https://www.cdc.gov/injury/features/global-road-safety/index.html#:~:text=Each%20year%2C%201.35%20million%20people,on%20roadways%20around%20the%20world.&text=Every%20day%2C%20almost%203%2C700%20people,pedestrians%2C%20motorcyclists%2C%20and%20cyclists.