**MAKERERE  UNIVERSITY**

**COLLEGE OF COMPUTING AND INFORMATION SCIENCES**

**(YEAR II) RECESS TERM**

**SOFTWARE REQUIREMENT SPECIFICATION DOCUMENT (SRS)**

**FOR PYTHON PROJECT**

**PROJECT MEMBERS**

|  |  |  |
| --- | --- | --- |
| **NAME** | **REGISTRATION NUMBER** | **STUDENT No** |
| Nagaba Angel | 17/U/726 | 217000189 |
| Okello Marvin Kevin Ochira | 17/U/9569/PS | 217017015 |
| Karungi Lydia | 17/U/4676/PS | 217002012 |
| Wepukhulu Bruno | 17/U/10891/PS | 217012574 |

**PROJECT COORDINATOR**: MR. KAMULEGEYA GRACE

## INTRODUCTION

**PURPOSE**

A system requirements specifications(SRS) document describes what the system will do and how it will be expected to do it . This document will include a use case diagram which shows how the system interacts with the outside world.

**Document conventions**

This document is edited using the following formats, font size 12, font style times new roman, line spacing 1.0. it was generated based on the guidelines provided by our recess coordinator Mr. Grace Kamulegeya.

**Intended audience and Reading suggestions**

The intended audience for this project includes, the data scientists(students working on the project), Road authorities and road users(drivers, riders, pedestrians, passengers).

**Project scope**

The purpose of this project is to aim at reducing the death rate caused by accidents. The project is based on the data that is collected by the police in the UK transport department every time an accident happens with the help of the report form this data can be found at [http://docs.adrn.ac.uk/888043/mrdoc/pdf/888043\_stats19-road-accident-injury-statistics-report-form.pdf](http://docs.adrn.ac.uk/888043/mrdoc/pdf/888043_stats19-road-accident-injury-statistics-report-form.pdf#inbox/_blank)

We shall analyse and visualise the data to come up with insights that will help us to recommend policies to help reduce the deaths caused by accidents. The insights will also help us to advise road users based on the occurrence causes of accidents.

**The users of the Data Science Pipeline include the following:**

* Data scientists

These are the developers of the system. They extract the relevant data from the data collected by the police and clean it, visualize it, model the data and interpret it.

* Data clerks in the Police.

These enter the collected data into the system. From they get the data from the Report forms in which information is recorded when an accident occurs.

* System Administrators

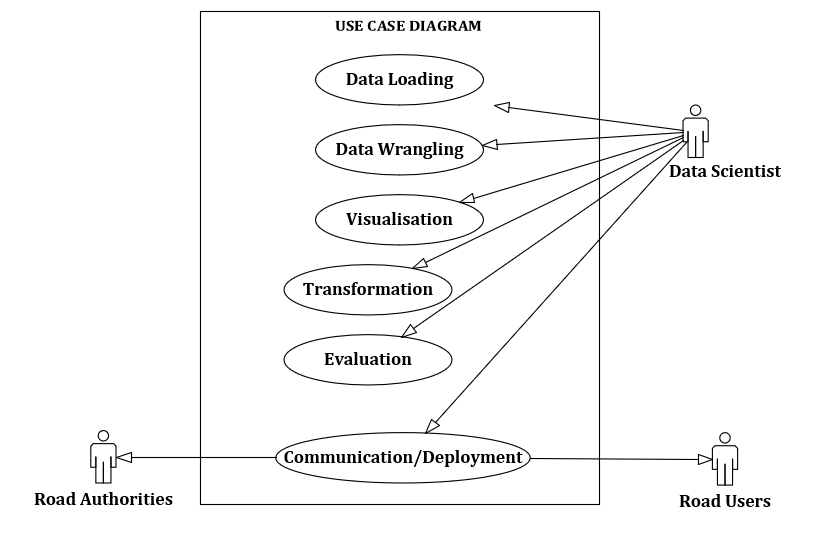
These have access rights to the system and carry out analysis of the data.

* System testers

These test the functionalities of the system, how correct the data output is and whether all needs of the users are fully satisfied.

**Use Case Diagram**

Ause case diagram helps to describe the system’s behaviour and how a system interacts with outside world. The use case diagram consists of actors (who interacts with the system) and use cases(this shows what the actors do with the system). A use case basically helps in understanding of the system form the end user’s point of view.



**Source: Generated by the group using Microsoft Visio**

**Roles of the Actors in the Use case Diagram**

|  |  |  |  |
| --- | --- | --- | --- |
| Use Case | Actors | Purpose | Justification |
| Data Loading | Data Scientist | Extract data from the files into the development environment | This is done so a to have data in an environment where they can be processed |
| Data Wrangling | Data Scientist | To attain data that is free from missing values, duplication, irregularities and in their right formats | The data has to be cleaned so as to have accurate and consistent operation of algorithms |
| Visualisation | Data Scientist | To have an overview on the behaviour of data | This enables one to understand the behaviour of data. |
| Transformation | Data Scientist | To prepare data such that they can be fed to the machine learning algorithms to produce the best results | This is done so as to have more accurate learning for the machine learning algorithms. |
| Evaluation | Data Scientist | To know which machine learning algorithm performs best. | This is done to know which machine learning algorithm performs better. |
| Communication/Deployment | Data Scientist  Road Authorities  Road Users | To communicate to the users of the system and the roads of the outcome of the analysis | The communication of the results aids the actors in better decision making when using or planning for the roads. |

**Use Case Detailed Description**

|  |  |
| --- | --- |
| **Use case name:** | Data Loading |
| **Priority:** | High |
| **Actors:** | Data scientist |
| **Description:** | The data scientist loads the data from the csv files having all the data into the development environment using python libraries. |
| **Precondition:** | The data has to be in csv files, preloaded by data clerks from the stats19-road-accident-injury-statistics-report-form |
| **Conclusion:** | This stage concludes when the data from the csv files are loaded successfully into the development environment and are assigned to variables that hold the data ready to be manipulated. |
| **Assumptions:** | The raw data from the forms has already been entered into csv files by the data clerks |

|  |  |
| --- | --- |
| **Use case name:** | Data Wrangling |
| **Priority:** | High |
| **Actors:** | Data scientist |
| **Description:** | The csv files may have data with irregularities such as missing values, incorrect formats and others, so in this stage the data scientist makes sure that data is cleaned using data cleaning techniques, so as the data is ready for analysis |
| **Precondition:** | The data must be loaded in the development environment |
| **Conclusion:** | This stage is concluded when the data no longer has irregularities and duplications |

|  |  |
| --- | --- |
| **Use case name:** | Visualisation |
| **Priority:** | High |
| **Actors:** | Data Scientist |
| **Description:** | The data scientist at this stage manipulates the data to create visualisations in order to understand the behaviour of data |
| **Precondition:** | The data must be clean |
| **Conclusion:** | This stage is concluded when the data scientist comes up with visualisations of different dimensions of the data. |

|  |  |
| --- | --- |
| **Use case name:** | Transformation |
| **Priority:** | High |
| **Actors:** | Data Scientist |
| **Description:** | at this stage, the data is split into two categories where one will be used to train the machine learning algorithms and the other set to test the accuracy of the algorithm. |
| **Precondition:** | The data must be clean and split into training data and test data |
| **Conclusion:** | The stage is concluded when the machine learning algorithms are successfully trained and tested |

|  |  |
| --- | --- |
| **Use case name:** | Evaluation |
| **Priority:** | High |
| **Actors:** | Data Scientist |
| **Description:** | The data scientist compares the performances of the machine learning algorithms used, and determines which one performs best. |
| **Precondition:** | The machine learning algorithms must be trained and tested. |
| **Conclusion:** | This stage concludes with the determination of the best machine learning algorithm. |

|  |  |
| --- | --- |
| **Use case name:** | Communication/Deployment |
| **Priority:** | High |
| **Actors:** | Data Scientist, Road Authorities, Road Users |
| **Description:** | The Data Scientist displays or communicates the findings or the analysis of the dataset to the Road authorities and road users, who use this information to influence their decision making for the use of the roads to reduce the death rate caused by road accidents. The data scientist displays the visualisations he generated to aid the other actors in understanding the results. |
| **Conclusion:** | This stage is concluded with display of information through visualisations to the other actors who interface with the system. |