**SCHOOL OF COMPUTING AND INFORMATION TECHNOLOGY**

**A Concept Paper for Nairobi traffic**

**School of Computing and Information Technology**

**Supervisor**

**MR. Grace Kamulegeya**

**Department of networks**

**School of Computing and Information Technology, Makerere University**

**GROUP MEMBERS.**

|  |  |
| --- | --- |
|  | **NAMES** |
| **1** | **LWANGA DARLINE** |
| **2** | **GYAGENDA ARAFAT** |
| **3** | **KISA JEREMIAH** |
| **4** | **KYADONDO SOLOMON** |

1. **BACKGROUND**

Nairobi is one of the most heavily congested cities in Africa. Each day thousands of Kenyans make the trip into Nairobi from towns such as Kisii, Keroka, and beyond for work, business, or to visit friends and family

Nairobi traffic is the data-set of tickets purchased from Mobiticket for the 14 routes from “up country” into Nairobi between 17 October 2017 and 20 April 2018. This dataset includes the variables: ride\_id, seat\_number, payment\_method, payment\_receipt, travel\_date, travel\_time, travel\_from, travel\_to, car\_type, max\_capacity.

The dataset was got from Uber and Mobiticket. The dataset is structured and from a known system

**2. Description of the Problem**

We would like to build a model that predicts the number of seats that Mobiticket can expect to sell for each ride, i.e. for a specific route on a specific date and time. There are 14 routes in this dataset. All of the routes end in Nairobi and originate in towns to the North-West of Nairobi towards Lake Victoria.

**3. Problem Statement**

The aim of the project is to create a predictive model using traffic data provided from Uber Movement and historic bus ticket sales data from Mobiticket to predict the number of tickets that will be sold for buses into Nairobi from cities in "up country" Kenya

1. **Objectives**

* The main objective is to find out how traffic patterns influence people’s decisions to come into the city by bus and which bus to take.
* Visualize data inform of graphical charts, figures, and bars.
* correlations in the data features.

**5. Methodology**

The data used to train the model will be historic hourly traffic patterns in Nairobi and historic ticket purchasing data for 14 bus routes into Nairobi from October 2017 to April 2018, and includes the place or origin, the scheduled time of departure, the channel used for the purchase, the type of vehicle, the capacity of the vehicle, and the assigned seat number

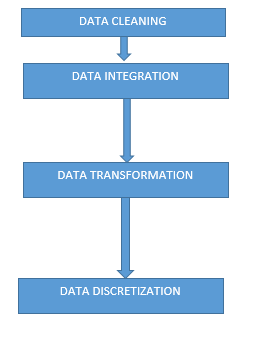
**6. Pre-processing and Data Preparation**

Data pre-processing is an important step in the data mining process. The phase “garbage in, garbage out” is particularly applicable to data mining and machine learning projects. Data pre-processing may affect the way in which outcomes of the final data processing can be interpreted. Data pre-processing includes cleaning, instance selection, normalization, transformation, feature extraction and selection.

**7. Data visualization and analysis**

Data visualization refers to the techniques used to communicate data or information by encoding it as visual objects ( points, lines or bars )contained in graphics. The goal is to communicate information clearly and efficiently to users.

A diagram showing data pipeline architecture



1. **References**

https://zindi.africa/competitions/traffic-jam-predicting-peoples-movement-into-nairobi/data