Madan Mohan Malaviya University of Technology

**Deoria Road, Singhariya, Gorakhpur (U.P.) India**

Established by U.P. Act no. 22 of 2013 of U.P. Government (Formerly Madan Mohan Malaviya Engineering College



# COMPUTER SCIENCE AND ENGINEERING DEPARTMENT

**Mid-Semester Project Report (BCS-40)**

**on**

**“Real Time Railways Twitter Complaint System”**

**Submitted by:**

### Gaurav Pandey (2018021052)

**Kumari Vani (2019022010)**

**Meera Gupta (2019022011)**

***Under the guidance of***

**Dr. P. K. SINGH**

**Professor, CSED MMMUT, Gorakhpur**

# TABLE OF CONTENT

1. [Introduction 3-5](#_bookmark0)
   1. [Background 03](#_bookmark1)
   2. [Purpose 04](#_bookmark2)
   3. [Scope of the Project 04](#_bookmark3)
   4. [Significance 04](#_TOC_250006)
2. Requirements Analysis and Feasibility Study 05-08
   1. Requirement Analysis… 05
      1. [Information Gathering 05](#_TOC_250005)
      2. Functional Requirement… 05
      3. [Non-Functional Requirements… 06](#_TOC_250004)
         1. [Essential Requirements… 06](#_TOC_250003)
         2. Optimal Requirements 06
   2. Software Requirement Specification 06
      1. Software Requirements… 06
      2. Hardware Requirements… 07
   3. [Feasibility Study 07](#_TOC_250002)
      1. [Operational Feasibility 07](#_TOC_250001)
      2. [Technical Feasibility 07](#_TOC_250000)
      3. [Economic Feasibility 0](#_bookmark4)8
      4. [Behavioural Feasibility 0](#_bookmark5)8
3. System Analysis and Design……………………………….10-13
   1. System Analysis…………………………………………10
      1. Existing System Description………………………….10
      2. Proposed System……………………………………...10
   2. System Design…………………………………………...11
      1. Data Flow Diagram…………………………………....11

## INTRODUCTION

The project reduces the work complexity of scanning through thousands of useless data to find particular information but from here we can directly find the relevant tweets that needs attention. To solve the problem, we will be using a Machine Learning (Naive Bayes) model.

In machine learning, Naïve Bayes classifiers are a family of simple "probabilistic classifiers" based on applying Bayes' theorem with strong (naïve) independence assumptions between the features. They are among the simplest Bayesian network models. Naïve Bayes has been studied extensively since the 1960s. It was introduced (though not under that name) into the text retrieval community in the early 1960s, and remains a popular (baseline) method for text categorization, the problem of judging documents as belonging to one category or the other (such as spam or legitimate, sports or politics, etc.) with word frequencies as the features.

With appropriate pre-processing, it is competitive in this domain with more advanced methods including support vector machines. It also finds application in automatic medical diagnosis.

Also, the machine learning model will be flexible and can be applied on various other local authorities like Nagar Nigam, State electricity board and various other authorities. To do that we just have to filter the tweets on the basis of that particular authority. It will reduce the operational time.

### BACKGROUND:

The Indian Railways currently receives close to 5,000 Tweets a day. The Tweets are monitored 24×7 and responded to by their level of seriousness or urgency. By this logic a complaint of sexual harassment or threat to personal safety feature on top as do medical emergencies and dealt with very quickly. Second come complaints against the Railway management, uncooperative personnel, meals, hygiene or matters that need immediate attention within trains or in stations. Last come complaints and requests that are not of immediate importance and can be solved later.

### PURPOSE:

* + - We try to implement and improve the current working model to a better machine learning model in the future development and also try to work with new cutting edge technology stack.
    - We have tried to achieve increased fault-tolerance. We also want to increase the accuracy and reliability of the current Railway Twitter Complaint Registration System.
    - By the use of machine learning, we have tried to increase the operational efficiency as the tweets that were not attended due to lack of manual labour, will now be addressed as our model will not skip any tweet, and will classify the tweet into feedback and emergency, which will lead to better decision making by the Indian Railway.
    - We have also increased Data Security as manual labour is not involved so no unwanted user can read all such information.
    - We tried to automate all this process so that the human resource of the government can be saved and same resource can be utilized in a better way on something else.
    - This software package can be readily used by non-programming personal avoiding human handled chance of error.

### SCOPE OF THE PROJECT:

This project has a large scope as it has the following features which help in making it easy to use, understand and modify it:

* + - Automation of Feedback and complaint addressing. No Need to do Paper Work.
    - To save the environment by using paper free work.
    - To increase the accuracy and efficiency of the complaint addressing procedure. Management of passenger data.
    - This software package can be readily used by non-programming personal avoiding human handled chance of error.

### SIGNIFICANCE:

* + - Simplified Management of complaint addressing.
    - Job Posting.
    - Feedback and Emergency Results.

## REQUIREMENT ANALYSIS AND FEASIBILITY STUDY

### REQUIREMENT GATHERING:

In this phase, we try to find out what are the requirements that need to be fulfilled by the proposed system, what all aspects we need to cover so as to make this system worth of usage.

### INFORMATION GATHERING:

To collect all the information related to the project we have to specify that what is required and what are the specifications of the project. What we want to done with our project means what are the functionalities we want to provide. Information gathering is done in order to understand the problem statement and collect the useful information. Now for our project, we collect all the necessary information that is tweets. We also collect basic information about our user, PNR number of the train, tweet id.

### FUNCTIONAL REQUIREMENTS:

User: The user should have a twitter account. If the user does not have a twitter account he/she can sign up by filling some particular details like name, mail-id, etc. The user can tweet with @RailMinIndia if he/she have can complain or feedback regarding Indian Railway.

Dataset: A data set (or dataset, although this spelling is not present in many contemporary dictionaries) is a collection of data. Most commonly a data set corresponds to the contents of a single database table, or a single statistical data matrix, where every column of the table represents a particular variable, and each row corresponds to a given member of the data set in question. The dataset lists values for each of the variables, such as height and weight of an object, for each member of the data set. Each value is known as a datum. The data set may comprise data for one or more members, corresponding to the number of rows.

App module: This module contains admin module, login module, reply module. Admin module- The person who reply for the tweets.

Login module- The admin can only login into the app with particular login id and password. Reply Module- To reply the tweets in real time.

Web module: This module contains admin module, login module, reply module. Admin module- The person who reply for the tweets.

Login module- The admin can only login into the app with particular login id and password. Reply Module- To reply the tweets in real time.

### NON-FUNCTIONAL REQUIREMENTS:

### Essential Requirements:

* + - * + The system should have 97 percentage of reliability.
        + The system should have a backup schedule every day.
        + The data in the system should provide security using encryption method.
        + The system should update the scheme availabilities real time.

### Optional Requirements:

* + - * + The system should be able to process each request within 2 seconds.
        + The system can retrieve the data since created.

### SOFTWARE REQUIREMENT ANALYSIS:

* + 1. **Software Requirement:**
* MySQL Essentials (Server)
* Jupyter Notebook
* Apache Spark
* Visual Studio Code for Web Development
* Android Studio
* Microsoft PowerPoint 2010/2013/2018
* Microsoft Word 2010/2013/2018
* Windows 10 or Linux operating System

### Hardware Requirement:

* 2.7 GHz x86 Processor
* 4 GB of System Memory ( RAM )
* 512 GB of Hard drive Space
* Keyboard / Mouse for Data Input
* Monitor to Display Output

### FEASIBILITY STUDY:

Here, we have studied all the feasibility aspects of the project under consideration to check out if the project is feasible with the decided requirements and availability of information, technologies, and budget.

### Operational feasibility:

Operational feasibility refers to the measure of solving problems with the help of a new proposed system. It helps in taking advantage of the opportunities and fulfils the requirements as identified during the development of the project. It takes care that the management and the users support the project. The proposed Real-time Railways Twitter Complaint System is capable identifying the emergency tweets more quickly and is more flexible and efficient than the existing systems. The machine learning and big data has made the system more congenial to end user.

### Technical Feasibility:

The system as mentioned is developed using the latest technologies which include Kafka, Spark, Zookeeper, Xampp, Cloud, Flutter, Big data. It uses a Machine Learning model

trained with the dataset extracted from various sources. Furthermore, the Naive Bayes Algorithm is used for the classification of data into emergency and feedback.

### Economic Feasibility:

Cost-benefit analysis is very important in deciding whether the project is economically feasible or not. It is alone sufficient to save time and money. Through cost-benefit analysis, it is quite evident that the benefits of this system outweigh costs and thus the project is economically feasible.

### Behavioral Feasibility:

Behavioral feasibility determines how much effort will go into educating, selling and training the user staff on a candidate system. This project is evaluated to be behaviorally feasible as it is very user-friendly and hardly needs any extra efforts to educate user for its facility and functioning. This is a one-time setup project and once initially installed, requires no intervention of the user.

**3. SYSTEM ANALYSIS AND DESIGN**

**3.1 SYSTEM ANALYSIS:**

Systems analysis is a problem solving technique that decomposes a system into its component pieces for the purpose of the studying how well those component parts work and interact to accomplish their purpose". According to the Merriam-Webster dictionary, systems analysis is "the process of studying a procedure or business in order to identify its goals and purposes and create systems and procedures that will achieve them in an efficient way". Analysis and synthesis, as scientific methods, always go hand in hand; they complement one another. Every synthesis is built upon the results of a preceding analysis, and every analysis requires a subsequent synthesis in order to verify and correct its results.

**3.1.1 Existing System Description:**

The Indian Railways currently receives close to 5,000 Tweets a day. The Tweets are monitored 24×7 and responded to by their level of seriousness or urgency. By this logic a complaint of sexual harassment or threat to personal safety feature on top as do medical emergencies and dealt with very quickly. Second come complaints against the Railway management, uncooperative personnel, meals, hygiene or matters that need immediate attention within trains or in stations. Last come complaints and requests that are not of immediate importance and can be solved later. Whenever a passenger (Twitter user) tweets with the required tags, the application at Indian Railways receives a request. This request is automated by twitter and sent to Indian Railways. Once the application (A software) receives such request, it generates a ‘Ticket’ in the support team. There is a whole big team sitting in front of their Computers to help you resolve your complaints. So this Ticket gets assigned to one of the support team members (Let us call him ‘A’) and it pops up on his computer. As soon as the person ‘A’ receives the ticket, he has to act on it within a particular amount of time.

**3.1.2 Proposed System:**

The project reduces the work complexity of scanning through thousands of useless data to find particular information but from here we can directly find the relevant tweets that needs attention. To solve the problem, we will be using a Machine Learning (Naive Bayes) model. In machine learning, Naïve Bayes classifiers are a family of simple "probabilistic classifiers" based on applying Bayes' theorem with strong (naïve) independence assumptions between the features. The proposed system will divide the tweets in two categories that is emergency and feedback. The manual work done by the employees will be reduced The tweets will be replied in the real time.

This system will improve the –

Accuracy and reliability

Reduced operational time

Fault-tolerance

**3.2. SYSTEM DESIGN:**

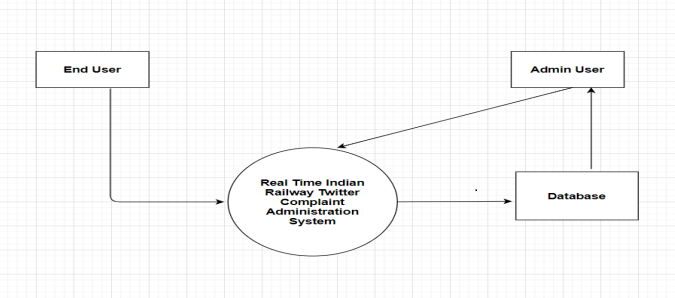
Systems design is the process of defining the architecture, modules, interfaces, and data for a system to satisfy specified requirements. Systems design could be seen as the application of systems theory to product development. System Analysis is the process that decomposes a system into its component pieces for the purpose of defining how well those components interact to accomplish the set requirements. The purpose of the System Design process is to provide sufficient detailed data and information about the system and its system elements to enable the implementation consistent with architectural entities as defined in models and views of the system architecture.

**3.2.1 Data Flow Diagram:**

A data flow diagram (DFD) maps out the flow of information for any process or system. It uses defined symbols like rectangles, circles and arrows, plus short text labels, to show data inputs, outputs, storage points and the routes between each destination.

**Level 0 DFD:**

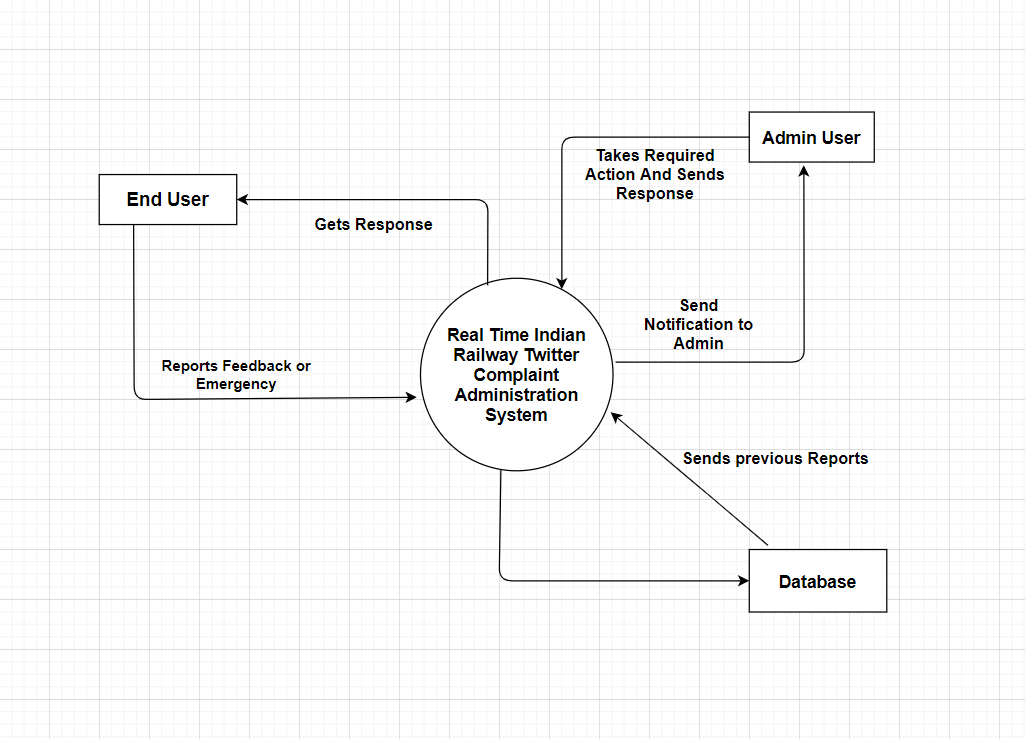
Its shows the abstract view of the system. Context Diagrams and DFD Layers and Levels. A context diagram is a top level (also known as "Level 0") data flow diagram. It only contains one process node ("Process 0") that generalizes the function of the entire system in relationship to external entities.



**0-Level DFD**

**Level 1 DFD:**

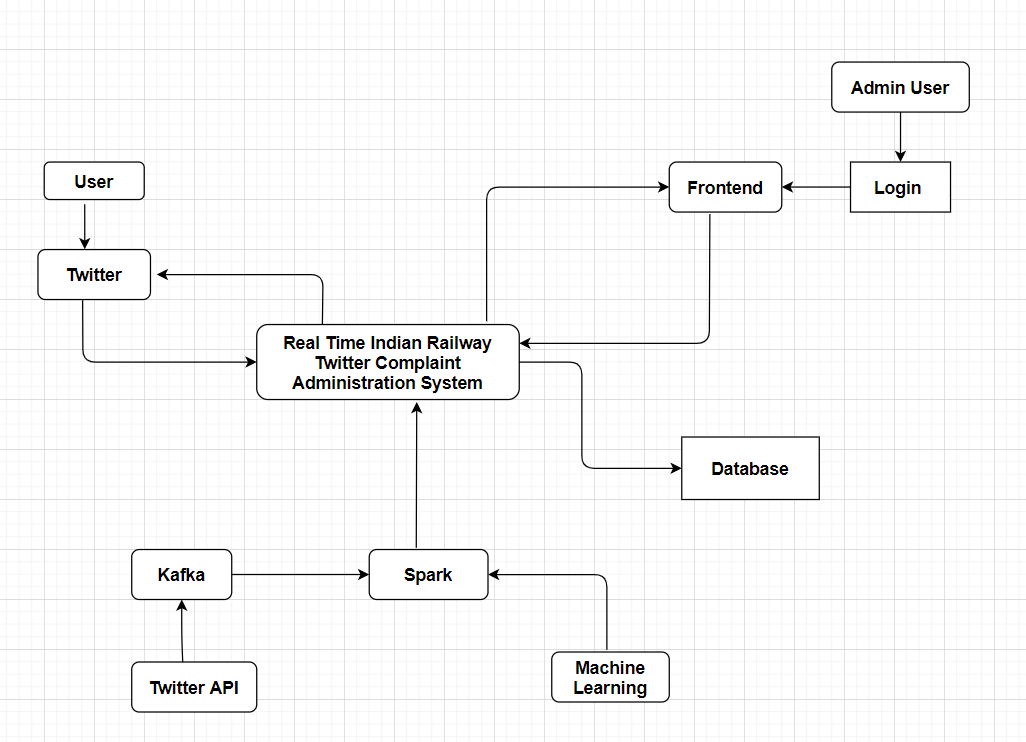
The next stage is to create the Level 1 Data Flow Diagram. This highlights the main functions carried out by the system. As a rule, we try to describe the system using between two and seven functions - two being a simple system and seven being a complicated system. This enables us to keep the model manageable on screen or paper. It explains the system in detailed view. When drawing Context Level DFD’s, we must first identify the process, all the external entities and all the data flows. We must also state any assumptions we make about the system. It is advised that we draw the process in the middle of the page. We then draw our external entities in the corners and finally connect our entities to our process with the data flows.



**1-Level DFD**

**Level 2 DFD:**

2-level DFD goes one step deeper into parts of 1-level DFD. It can be used to plan or record the specific/necessary detail about the system’s functioning.



**2-Level DFD**