

# **PROJECT REPORT**

ON

**Call center voice clustering**

Submitted to the Savitribai Phule Pune University

in partial fulfilment of the requirements

for the award of the degree of

**BACHELOR OF ENGINEERING**

(COMPUTER ENGINEERING)

## **SUBMITTED BY**

Sanket sopan shirse : 71931103K

Shubham Ajit asabe : 71931072F

Rohan mandar salvi :71931098K

Vipul bhausahab ahire :71931071H

## **UNDER THE GUIDANCE OF**

prof. Manisha Bharati



**Department of Computer Engineering  
Indira College of Engineering and Management,**

**Parandwadi,Pune - 410506**

**Year: 2020 - 21 (Semester - II)**

**SAVITRIBAI PHULE PUNE UNIVERSITY**

**DEPARTMENT OF COMPUTER ENGINEERING**  
**INDIRA COLLEGE OF ENGINEERING AND MANAGEMENT**  
**PARANDWADI, PUNE - 410506**



**CERTIFICATE**

*This is to certify that the project work entitled*

**Forest Fire Prediction.**

Submitted by

Sanket sopan shirse

Exam No:71931103K

Shubham Ajit asabe

Exam No:71931072F

Rohan mandar salvi

Exam No:71931098K

Vipul bhausahab ahire

Exam No:71931071H

*is a bonafide work carried out under the supervision of Prof Deepali Dhadwad and it is submitted towards the partial fulfilment of the requirement of University of Pune, for the award of the degree of Bachelor of Engineering(Computer Engineering).*

prof .Manisha Bharati  
Project Guide

Dr .vinod kumar  
HOD Computer Dept

Dr. Sunil Ingole  
Principal(ICEM)

Place: Pune

Date: \_ \_ / \_ \_ / \_ \_ \_ \_

(Prof. -----)  
Internal Examiner

(-----)  
External Examiner

**SPONSERSHIP CERTIFICATE**

## Acknowledgment

We are presenting this project report on “Call center voice clustering ” as part of the curriculum of B.E. Computer Engineering with immense pleasure. We started working on it with a zeal and enthusiasm but then quickly realized that for satisfactorily completing the project not only require conviction and perseverance, but without due help and guidance such a task becomes futile. We wish to thank all the people who gave us an unending support right from the stage the idea was conceived. It gives us great pleasure, on the completion of this project, to acknowledge and appreciate all those who were there to help us.

We express our sincere and profound thanks to all our teachers **Dr. Vinod Kumar** (Head of Department), **Prof. Sumit Harale** (Project Coordinator) for their valuable Guidance. We would like to thank our college ICEM, Parandwadi, for the boost that it has provided. We wish to thank **Prof. Deepali Dhadwad**(Computer Dept.) for her/his student-like enthusiasm and her/his guidance from time to time. We heartily thank for all her/his help and valuable time. Her/his invaluable advice has helped us bring this work to completion. Besides, We take this opportunity to express our sincere gratitude to the Principal **Dr. Sunil Ingole**(ICEM, Pune) for providing a good environment and facilities to complete this project.

We would like to thank all our internal guides for providing the resources for project stage. We also acknowledge the research work done by all researchers in this field. And last but not least, all our friends, who have helped us directly or indirectly throughout the project.

Sanket sopan shirse  
Shubham Ajit Asabe  
Rohan mandar salvi  
Vipul bhausaheb ahire  
(B.E. Computer Engg.)

## Abstract

This work presents a study related to the methods of text mining applications, more specifically, data clustering, in call center's databases, whose texts are in the Portuguese language. The main objective is to identify new and useful knowledge, based on customers' claims. Through the information agreement, it will be possible to identify better ways to help the customer, increasing their satisfaction with company services as well as supplying the call center staff and other related areas with a set of procedures and information concerning the most common customer's questions.

**Keywords:** *Text Mining, Clustering, Call Center, Speech To Text.*

# Contents

Certificate	i
Sponsorship Certificate	iii
Acknowledgement	iv
Abstract	v
Contents	vi
List of Figures	x
List of Tables	x
Abbreviations	xi
<b>1 Introduction</b>	<b>1</b>
1.1 Overview . . . . .	1
1.2 Motivation . . . . .	2
1.3 Problem Definition and Objectives . . . . .	2
1.4 Project Scope & Limitations . . . . .	2
1.5 Methodologies of Problem Solving . . . . .	3
<b>2 Literature Review</b>	<b>4</b>
<b>3 Software Requirement Specifications</b>	<b>7</b>
3.1 Purpose . . . . .	7
3.2 Requirements Specification . . . . .	7
3.2.1 HARDWARE REQUIREMENTS . . . . .	7
3.3 SOFTWARE REQUIREMENTS . . . . .	8
3.3.1 FUNCTIONAL REQUIREMENTS . . . . .	8
3.3.2 NONFUNCTIONAL REQUIREMENTS . . . . .	8
3.3.3 PERFORMANCE REQUIREMENTS . . . . .	9
3.3.4 Feasibility Study . . . . .	10

3.3.5	Technical Feasibility . . . . .	10
3.3.6	Operational Feasibility . . . . .	10
3.3.7	Operational Feasibility . . . . .	11
3.3.8	Economic Feasibility . . . . .	12
3.4	Assumption . . . . .	13
3.5	Constraints . . . . .	13
3.6	Usability . . . . .	13
<b>4</b>	<b>System Design</b>	<b>15</b>
4.1	System Architecture . . . . .	15
4.2	Mathematical Model . . . . .	17
4.3	Data Flow Diagram . . . . .	18
4.4	UML Diagrams . . . . .	20
4.4.1	Use Case Diagram . . . . .	20
4.4.2	Class Diagram . . . . .	21
4.4.3	Activity Diagram . . . . .	22
4.4.4	Sequence Diagram . . . . .	23
4.4.5	State Chart . . . . .	25
4.4.6	Component Diagram . . . . .	26
4.5	Algorithmic Strategy . . . . .	27
4.6	Time Complexity of Proposed System . . . . .	27
<b>5</b>	<b>Project Plan</b>	<b>29</b>
5.1	Project Estimate . . . . .	29
5.1.1	Reconciled Estimates . . . . .	29
5.1.2	<b>Cost Estimate</b> . . . . .	29
5.1.3	<b>Time Estimates</b> . . . . .	29
5.1.4	<b>Man Month Utilization:</b> . . . . .	30
5.1.5	Project Resources . . . . .	30
5.2	Risk Management . . . . .	31
5.2.1	Risk Identification . . . . .	32



5.2.2	Risk Analysis . . . . .	34
5.3	Overview of Risk Mitigation ,Monitoring,Management . . . . .	35
5.4	Project Schedule . . . . .	36
5.4.1	Project Task Set . . . . .	36
5.5	Task network . . . . .	38
5.5.1	Time line Chart . . . . .	38
5.6	Team Organization . . . . .	40
5.6.1	Team Structure . . . . .	40
5.6.2	Management reporting and communication . . . . .	41
<b>6</b>	<b>Project Implementation</b>	<b>42</b>
6.1	Overview of Project Modules . . . . .	42
6.2	Tools and Technologies Used . . . . .	42
6.3	Algorithmic Details . . . . .	44
6.3.1	Algorithm 1: . . . . .	45
6.3.2	Algorithm 2: K-Means . . . . .	46
<b>7</b>	<b>Software Testing</b>	<b>47</b>
7.1	Types of Tasting . . . . .	47
7.2	Test cases and Test Results . . . . .	47
<b>8</b>	<b>Results</b>	<b>50</b>
<b>9</b>	<b>Conclusions</b>	<b>51</b>
9.1	Conclusions . . . . .	51
9.2	Future Work . . . . .	51
9.3	Applications . . . . .	52
	<b>Appendices</b>	<b>54</b>
<b>A</b>	<b>Appendix</b>	<b>54</b>
A.1	Appendix A . . . . .	54

<b>B</b>	<b>Appendix B</b>	<b>55</b>
B.1	I2CT Certificates . . . . .	55
B.2	I2CT Certificates . . . . .	55
<b>C</b>	<b>Appendix C</b>	<b>56</b>
C.1	Plagiarism Report . . . . .	56

## List of Figures

1	System Architecure . . . . .	16
2	Level 0 DFD . . . . .	18
3	Level 1 DFD . . . . .	19
4	Level 2 DFD . . . . .	20
5	Use case Diagram . . . . .	21
6	Class Diagram . . . . .	22
7	Activity Diagram . . . . .	23
8	Sequence Diagram . . . . .	24
9	Component Diagram . . . . .	26
10	Risk Table . . . . .	34
11	Risk Impact definitions . . . . .	35
12	Risk Impact definitions . . . . .	35
13	Task Network . . . . .	38
14	Timeline of Literature review . . . . .	39
15	Timeline of Project Stage 1 . . . . .	39

## List of Tables

1	Hardware Requirements . . . . .	7
2	Software Requirements . . . . .	8
3	Risk Probability definitions . . . . .	34

# Abbreviations

CNNs : Convolutional Neural Networks

CCTV : Circuit Television

YOLO : You Only Look Once

DSDM : Dynamic Systems Development Model

RAD : Rapid Application Development

VSS : Vector space similarity algorithm

SDLC : Systems Development Life Cycle

URL : Unified resource locator

LOC : Line of code

SRS : Software requirement specification

GUI : Graphical user interface

UML : Unified modelling language

STP : Software test plan

# **1 Introduction**

Many companies who give telephonic attendance services for its customers need solutions guided to the information technology that make possible the register of their customers contacts, the aid and support to the attendants. For such, new access technologies to the knowledge bases and flow of information management can increase the productivity of the attendants and, as consequence, improve the quality of the given services. Many times, the attendance can be made by the Internet, bringing some advantages such as: costs reduction with attendance, user satisfaction increasing and knowledge of the customers interests, gotten through the automatic storage of the typed information. Text mining tools directed to knowledge bases creation, support and operation, and for attendance centers are applicable for the Web cases email or telephone attendance. The technologies presented in these tools make possible to create knowledge bases for queries through the search of one or more words in documents, allow to combine resources of search in texts with resources of interaction with support (attendance it the Web: email, chats, virtual rooms, etc), and make possible the creation of knowledge bases for queries in natural language. Thus, the objective of this work is to show some methods of text mining applied in call center's databases, especially in customers' complaints of a electrical energy company, where the results serve to show that techniques are extremely useful during the company services evaluation processes, as also to be used in decision taking. For this, a text mining tool called Insight Discoverer Clusterer by Temis was used, that contains the documents grouping functionality or clustering, allowing the classification of the same ones. With this study we have created conditions that can to evidence problems that are occurring with frequency, as for example problems of invoicing, financial income, supply and attendance quality, etc, who are gotten by the analysis of the results.

## **1.1 Overview**

The main objective is to identify new and useful knowledge, based on customers' claims. Through the information agreement, it will be possible to identify better ways to help the customer, increasing their satisfaction with company services as well as supplying

the call center staff and other related areas with a set of procedures and information concerning the most common customer's questions.

## **1.2 Motivation**

In today's world , many customer send request to call center there is a lot of waiting for thousands of people at a time ,to make this quick and efficient we are going to work on this project This was the problem statement given by TCS in their hackathone

## **1.3 Problem Definition and Objectives**

In today's global marketplace, many companies have turned to a call center model to assist, streamline and maximize customer service and sales needs at scale. With one eye focused on providing excellent support and the other on efficiency, an ideal call center needs to strike that perfect balance of care and resources. While call centers require an increase in overhead, they also sacrifice valuable "face to face" customer interaction. The current state of call center analytics provides employers the ability to improve service quality and doing so with the bottom line in mind.

## **1.4 Project Scope & Limitations**

Successful centers use advanced call center analytics software to monitor and review performance, not only from a customer lens, also from the employee's perspective, as well as a business-owner lens.

Each of these approaches offers its own advantages and together satisfies each angle. The key to choosing the correct analytics combination lies in understanding the approaches, and how they can be used to improve your call center. Here are the six most common approaches to analytics.

## **1.5 Methodologies of Problem Solving**

Our model was using the TF\*IDF algorithm is used for weighing a keyword in any content and assign the importance to that keyword based on the number of times it appears in the document. The value of K defines the Number of clusters that is the 5 categories we are choosing.

## 2 Literature Review

Interpreting speech signals by making a speech to text translation is an active research area especially in current machine learning/deep learning literature. The speech to text translation of call center recordings is an important and specialized application for speech to text translation. Detecting silence in audio recordings can be a pre-processing step in order to optimize processing speed by not-considering audio parts not having significant information. In this work, such a preprocessing framework for detecting silence parts in an audio signal is considered. It is shown that further statistical analysis on the silence distributions results in detecting interesting audio features which can help in finding audio recordings which do not have actual speech sound but a fax machine tone sequence. This foundation can be directly implemented in call center management software and makes it possible to discriminate between a normal conversation recording and a fax sound recording. [1].

Over the past decades, a tremendous amount of research has been done on the use of machine learning for speech processing applications, especially speech recognition. However, in the past few years, research has focused on utilizing deep learning for speech-related applications. This new area of machine learning has yielded far better results when compared to others in a variety of applications including speech, and thus became a very attractive area of research. This paper provides a thorough examination of the different studies that have been conducted since 2006, when deep learning first arose as a new area of machine learning, for speech applications. A thorough statistical analysis is provided in this review which was conducted by extracting specific information from 174 papers published between the years 2006 and 2018. The results provided in this paper shed light on the trends of research in this area as well as bring focus to new research topics. [2].

Speech processing is considered to be one of the most important application area of digital signal processing. Speech recognition and translation systems have consisted into two main systems, the first system represents an ASR system that contains two levels which are level one the feature extraction level As well as, level two the



classification technique level using Data Time Wrapping (DTW), Hidden Markov Model (HMM), and Dynamic Bayesian Network (DBN). The second system is the Machine Translation (MT) system that mainly can be achieved by using three approaches which are (A) the statistical-based approach, (B) rule -approach, and (C) hybrid-based approach. In this study, we made a comparative study between classification techniques from ASR point of view, as well as, the translation approaches from MT point of view. The recognition rate was used in the ASR level and the error rate was used to evaluate the accuracy of the translated sentences. Furthermore, we classified the sample text audio files into four categories which were news, conversational, scientific phrases, and control categories. [3].

Considerable attention has been paid to acquisition device recognition over the past decade in the forensic community, especially in digital image forensics. In contrast, acquisition device clustering from speech recordings is a new problem that aims to merge the recordings acquired by the same device into a single cluster without having prior information about the recordings and training classifiers in advance. In this paper, we propose a method for mobile phone clustering from speech recordings by using a new feature of deep representation and a spectral clustering algorithm. The new feature is learned by a deep auto-encoder network for representing the intrinsic trace left behind by each phone in the recordings, and spectral clustering is used to merge recordings acquired by the same phone into a single cluster. The impacts of the structures of the deep auto-encoder network on the performance of the new feature are discussed. Different features are compared with one another. The proposed method is compared with others and evaluated under special conditions. The results show that the proposed method is effective under these conditions and the new feature outperforms other features. [4]

This study presents the development of a voice activity detection (VAD) system tested on call center telephony data obtained from our local site. The concept of bag of audio words (BoAW) combined with a naïve Bayes classifier was applied to achieve the task. It was formulated as a binary classification problem with speech as the

positive class and silence/background noise as the negative class. All the processing was performed on the Mel-frequency cepstral coefficients (MFCCs) extracted from the audio recordings. The results which are presented as accuracy score and receiver operating characteristics (ROC) indicate an excellent performance of the developed model. The system is to be deployed within our call center to aid data analysis and improve overall efficiency of the center. [5]

## 3 Software Requirement Specifications

### 3.1 Purpose

The purpose of the design phase is to Provide service menus for locking in preferred customers, such as offering discount charge menus and added-value services, Analyze how both electric power income and related income change, and investigate the contents of the services, Measure whether customers offered services are satisfied with those services and reflect the results in the service menu planning.

### 3.2 Requirements Specification

Requirement Specification provides a high secure storage to the web server efficiently. Software requirements deal with software and hardware resources that need to be installed on a serve which provides optimal functioning for the application. These software and hardware requirements need to be installed before the packages are installed. These are the most common set of requirements defined by any operation system. These software and hardware requirements provide a compatible support to the operation system in developing an application.

#### 3.2.1 HARDWARE REQUIREMENTS

The hardware requirement specifies each interface of the software elements and the hardware elements of the system. These hardware requirements include configuration characteristics.

**Table 1:** Hardware Requirements

System	Mitigation
Hard Disk	100 GB
Monitor	15 VGA Color
Mouse	Logitech
RAM	1 GB

### 3.3 SOFTWARE REQUIREMENTS

The software requirements specify the use of all required software products like data management system. The required software product specifies the numbers and version. Each interface specifies the purpose of the interfacing software as related to this software product.

**Table 2:** Software Requirements

Operating system	Wondows XP/7/10
Development Framework	Flask Frameword
Programming Language	JAVA
Dataset	Mysql
IDE	Eclipse

#### 3.3.1 FUNCTIONAL REQUIREMENTS

The functional requirement refers to the system needs in an exceedingly computer code engineering method.

The key goal of determinant “functional requirements” in an exceedingly product style and implementation is to capture the desired behavior of a software package in terms of practicality and also the technology implementation of the business processes.

#### 3.3.2 NONFUNCTIONAL REQUIREMENTS

All the other requirements which do not form a part of the above specification are categorized as Non-Functional needs. A system perhaps needed to gift the user with a show of the quantity of records during info. If the quantity must be updated in real time, the system architects should make sure that the system is capable of change the displayed record count at intervals associate tolerably short interval of the quantity of records dynamic. Comfortable network information measure may additionally be a non-functional requirement of a system.

**The following are the features::**

1. Accessibility.
2. Availability.
3. Backup.
4. Certification.
5. Compliance
6. Configuration Management.
7. Documentation
8. Disaster Recovery
9. Efficiency(resource consumption for given load)
10. Interoperability

### **3.3.3 PERFORMANCE REQUIREMENTS**

Performance is measured in terms of the output provided by the application. Requirement specification plays an important part in the analysis of a system. Only when the requirement specifications are properly given, it is possible to design a system, which will fit into required environment. It rests largely with the users of the existing system to give the requirement specifications because they are the people who finally use the system. This is because the requirements have to be known during the initial stages so that the system can be designed according to those requirements. It is very difficult to change the system once it has been designed and on the other hand designing a system, which does not cater to the requirements of the user, is of no use.

The requirement specification for any system can be broadly stated as given below:

- The system should be able to interface with the existing system
- The system should be accurate
- The system should be better than the existing system

The existing system is completely dependent on the user to perform all the duties.

### **3.3.4 Feasibility Study**

Preliminary investigation examines project feasibility; the likelihood the system will be useful to the organization. The main objective of the feasibility study is to test the Technical, Operational and Economical feasibility for adding new modules and debugging old running system. All systems are feasible if they are given unlimited resources and infinite time. There are aspects in the feasibility study portion of the preliminary investigation.

- Technical Feasibility
- Operation Feasibility
- Economical Feasibility

### **3.3.5 Technical Feasibility**

The technical issue usually raised during the feasibility stage of the investigation includes the following:

- Does the necessary technology exist to do what is suggested?
- Do the proposed equipments have the technical capacity to hold the data required to use the new system?
- Will the proposed system provide adequate response to inquiries, regardless of the number or location of users?
- Can the system be upgraded if developed?

Are there technical guarantees of accuracy, reliability, ease of access and data security?

### **3.3.6 Operational Feasibility**

The technical issue usually raised during the feasibility stage of the investigation includes the following:

- Does the necessary technology exist to do what is suggested?
- Do the proposed equipments have the technical capacity to hold the data required to use the new system?
- Will the proposed system provide adequate response to inquiries, regardless of the

number or location of users?

- Can the system be upgraded if developed?

Are there technical guarantees of accuracy, reliability, ease of access and data security?

### **3.3.7 Operational Feasibility**

#### **User-friendly:**

Customer will use the forms for their various transactions i.e. for adding new routes, viewing the routes details. Also the Customer wants the reports to view the various transactions based on the constraints. These forms and reports are generated as user-friendly to the Client.

#### **Reliability**

The package will pick-up current transactions on line. Regarding the old transactions, User will enter them in to the system. .

#### **Security**

The web server and database server should be protected from hacking, virus etc.

#### **Portability**

The application will be developed using standard open source software (Except Oracle) like Java, tomcat web server, Internet Explorer Browser etc these software will work both on Windows and Linux o/s. Hence portability problems will not arise.

#### **Availability**

This software will be available always.

## **Maintainability**

This software will be available always.

The system uses the 2-tier architecture. The 1st tier is the GUI, which is said to be front-end and the 2nd tier is the database, which uses sqllite, which is the back-end. The front-end can be run on different systems (clients). The database will be running at the server. Users access these forms by using the user-ids and the passwords.

### **3.3.8 Economic Feasibility**

The computerized system takes care of the present existing system's data flow and procedures completely and should generate all the reports of the manual system besides a host of other management reports.

It should be built as a web based application with separate web server and database server. This is required as the activities are spread throughout the organization customer wants a centralized database. Further some of the linked transactions take place in different

The technical issue usually raised during the feasibility stage of the investigation includes the following:

- Does the necessary technology exist to do what is suggested?
- Do the proposed equipments have the technical capacity to hold the data required to use the new system?
- Will the proposed system provide adequate response to inquiries, regardless of the number or location of users?
- Can the system be upgraded if developed?

Are there technical guarantees of accuracy, reliability, ease of access and data security?

Two different data sources are used in social recommendation problem, which are 1) User-Item relation, & 2) User-user relation (social network). An example is shown



in figure ??(a). In social network graph there are trust scores between different users' while user-item relation is similar to that of query-URL relation matrix, binary relations connect users and items. Figure ??(b) shows single and consistent graph formed using above two different graphs.

Using this graph, diffusion for each user (Heat source) is started, and top N items are recommended to the user. In here, during diffusion there are two ways to diffuse heat from users to items, the first one is within user-item bipartite graph and the other is passing through the social network graph. The first route captures the intuition that similar users will see/view similar items & the second route reflects social interactions and influences between users. Hence, proposed DRec diffusion method fuses these two data sources together for social recommendation.

### 3.4 Assumption

The only assumption is query submitted by user must be present in the dataset, if it is not then no suggestions will be provided. The designed system works off-line based on predefined dataset.

### 3.5 Constraints

The dataset used is from 2006, it is difficult to compare the performance of the system with the current systems in market, hence the comparison is heuristic based.

### 3.6 Usability

Usability is a non-functional requirement of the system that specifies how easy the system is to use or how user-friendly the system is. It specifies how the system functionality is to be perceived by the user and how efficient it is in carrying out user's tasks. There are several factors that decide usability of the system such as ease of learning, task efficiency, understandability, subjective satisfaction, etc. The recommendation framework designed is just a preprocessing step, user will not be aware of it, user will just insert a query to the system, and system suggests recommendations,

user will not be able to see or judge what's going on in between, due to this reason factors such as ease of learning or understandability do not apply here. As it doesn't require user-item rating matrix and uses heat diffusion, end user experiences improved speed and better results.

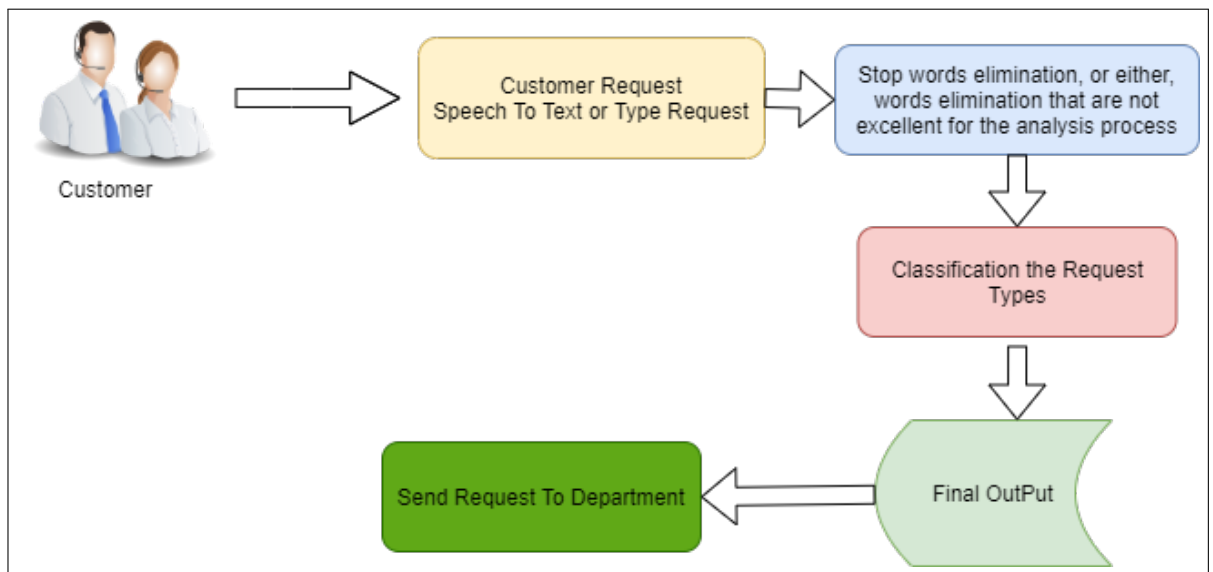
## 4 System Design

In this System Customer Request in Text Format To Service Call Center for classify the Request into different categories listed below:

1. New vehicle purchase enquiries ( i.e., Enquiry on latest or future or existing product features, price, availability, closest showroom to drop in for purchase or exchange, etc.)
2. Test drive requests ( i.e., Request for booking test drives, follow up Request with customers to schedule the same, confirmation that test drive has been done as per schedule or with delayed schedule, etc.)
3. Breakdown ( i.e., Customer Send Request TO contact center to report vehicle break down and providing his location details for repair or breakdown assistance, Road assistance mechanic reaching out to customer and reaching location with preliminary input on vehicle condition )
4. Feedback ( Feedback collected post sales /service on vehicle delivery on customer sales/ service experience)
5. Vehicle Quality ( Complaints of the vehicle parts not functioning properly, repetitive complaints, etc. - except breakdowns and failures.

### 4.1 System Architecture

The architecture of the proposed system is shown in figure 1. It has four modules explained as follows:



**Figure 1:** System Architecture

1. New vehicle purchase enquiries ( i.e., Enquiry on latest or future or existing product features, price, availability, closest showroom to drop in for purchase or exchange, etc.)
2. Test drive requests ( i.e., Request for booking test drives, follow up Request with customers to schedule the same, confirmation that test drive has been done as per schedule or with delayed schedule, etc.)
3. Breakdown ( i.e., Customer Send Request TO contact center to report vehicle break down and providing his location details for repair or breakdown assistance, Road assistance mechanic reaching out to customer and reaching location with preliminary input on vehicle condition )
4. Feedback ( Feedback collected post sales /service on vehicle delivery on customer sales/ service experience)
5. Vehicle Quality ( Complaints of the vehicle parts not functioning properly, repetitive complaints, etc. except breakdowns and failures

## 4.2 Mathematical Model

1. Identify the Users,

$$U = \{u1, u2, u3, \dots\}$$

Where 'U' is main set of Users like u1, u2, u3, ...

2. D be the set of Data.

$$D = \{D1, D2, D3, \dots\}$$

3. input video

$$Q = \{V1, V2, V3, \dots\}$$

Where 'V' is main set of Video v1, v2, v3, ...

4.  $SYS = \{DX, DF, AP, BG\}$

- DX = It Data Extractor which extract the data from the dataset.
- DF = find fire images using YOLOV3 in database.
- AP = Filter the results of YOLOV3 using Algorithm.
- BG = It generate the different frame

5. Identify the processes as P.

$$P = \{P1, P2, P3, P4\}$$

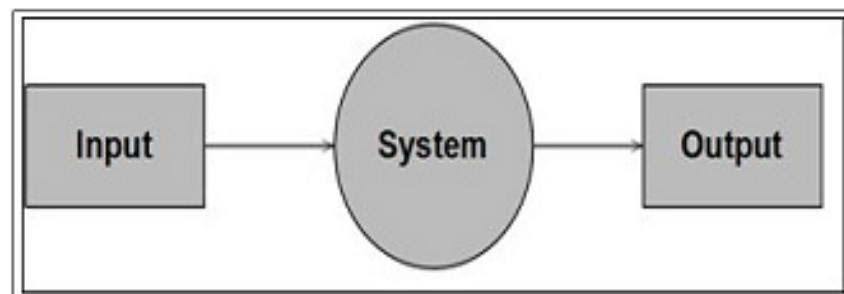
- $P1 = \{e1, e2\}$  where,  
 $\{e1 = i \mid i, \text{ database designing from the dataset}\}$   
 $\{e2 = j \mid j, \text{ show all clicks through data from the database}\}$
- $P2 = \{e1, e2, e3, e4\}$  where,  
 $\{e1 = i \mid i, \text{ Take the video from the user t}\}$   
 $\{e2 = i \mid i, \text{ Search query using volov3}\}$   
 $\{e3 = j \mid j, \text{ Filter the results of volov3 using Apriori}\}$   
 $\{e4 = j \mid j, \text{ Generate the frame}\}$   
 Graph  $G = \{E, V\}$  where,  
 \*  $V = \{v1, v2, v3, \dots\}$  be set of vertex  
 \*  $E = \{(v1, v2), (v2, v3)\}$  be set of edges

- $P3 = \{e1, e2\}$  where,  
 $\{ e1 = i \mid i, \text{Find out the fire and gun images in video } \}$

### 4.3 Data Flow Diagram

A data flow architecture represents graphical view of flow of data through an information system and modelling its process aspects. This are a preliminary step used to create an overview of the proposed system which can be elaborated later. Data flow architecture (Data Flow Diagram) can also be used for the visualization of data processing of system.

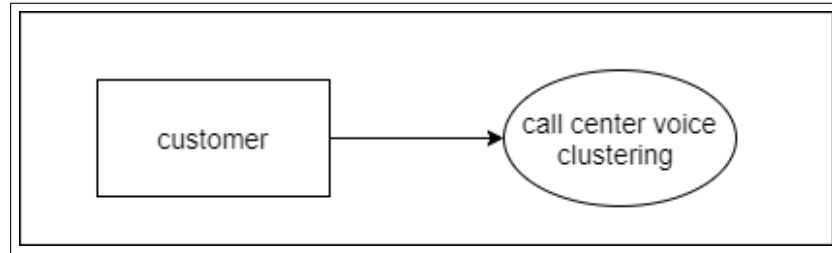
**DFD Level 0:** This is called fundamental level DFD for proposed system. It represents the entire system element as a single bubble with inputs and outputs. Input is the query submitted by user and output is nothing but the suggested queries. Figure 2 is the Level 0 DFD for the system.



**Figure 2:** Level 0 DFD

**DFD Level 1:** This is called as Advanced level DFD for proposed system. It represents systems entire process activities and inputs, outputs in DFD Level 0 will remain same for DFD level 1. The DRec algorithm will accept user query, perform all required operations and returns final suggestions to the user. It consists of user validation, Data extractor, Graph generator, Heat diffusion model and last is the heat value calculation by the random jump. The user is validated, then the user enters a

query, using the given query data is extracted from the database, and using a graph generator Query-URL bipartite graph is generated, after applying heat diffusion on it, heat values are calculated by random jump, and those queries having the highest value of heat are suggested to the user. Figure 3 is the Level 1 DFD of the system.



**Figure 3:** Level 1 DFD

**DFD Level 2:** This is called as Advanced level DFD for proposed system. It represents systems entire process activities and inputs, outputs in DFD Level 0 will remain same for DFD level 2. The DRec algorithm will accept user query, perform all required operations and returns final suggestions to the user. It consists of user validation, Data extractor, Graph generator, Heat diffusion model and last is the heat value calculation by the random jump. The user is validated, then the user enters a query, using the given query data is extracted from the database, and using a graph generator Query-URL bipartite graph is generated, after applying heat diffusion on it, heat values are calculated by random jump, and those queries having the highest value of heat are suggested to the user. Figure 4 is the Level 2 DFD of the system.

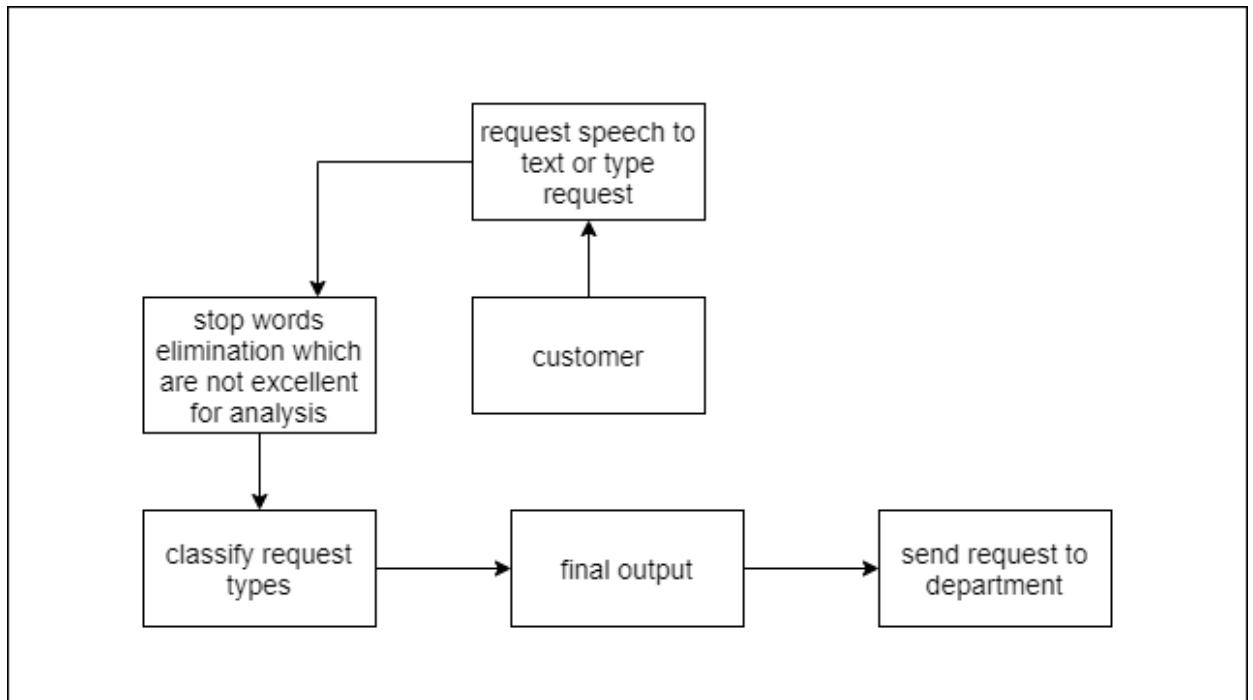


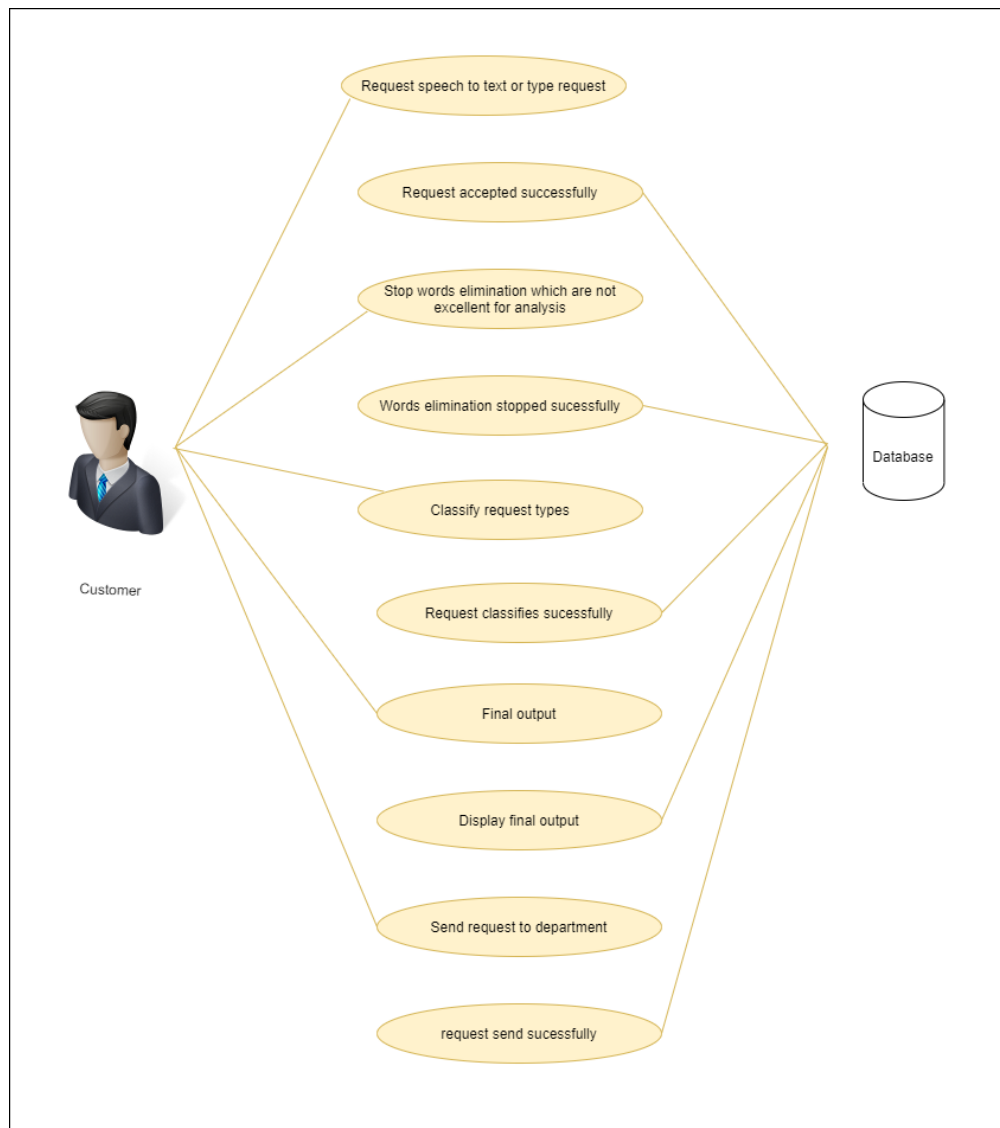
Figure 4: Level 2 DFD

## 4.4 UML Diagrams

### 4.4.1 Use Case Diagram

Use case diagram is a list of steps, typically defining interactions between a role (known in UML as an actor) and a system, to achieve a goal. In systems engineering, use cases are used at a higher level than within software engineering, often representing missions or stakeholder goals. The detailed requirements may then be captured in SysML or as contractual statements. A use case diagram represents all different types of users and all the various ways in which they can interact with the system. Thus, it helps providing a high level view of the system. It is a graphical representation of what the system must accomplish. Use case diagrams help in capturing user requirements, validating design, and generating test cases. Use case diagram for current system is shown in figure 5.

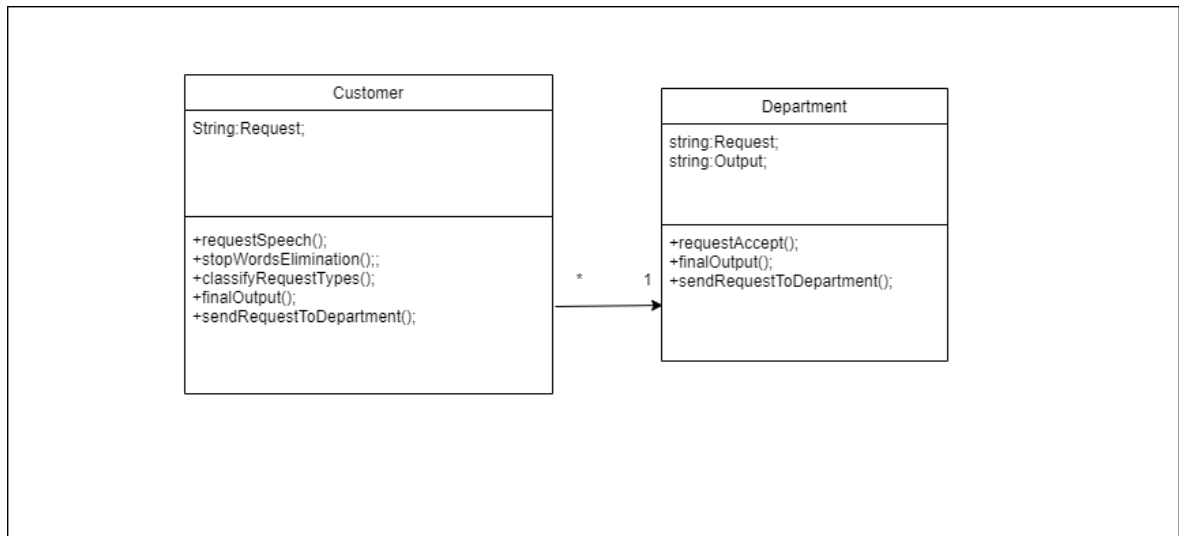




**Figure 5:** Use case Diagram

#### 4.4.2 Class Diagram

Class diagram describes the structure of a system by showing the system's classes, their attributes, and the relationships among the classes. Proposed system contains seven different types of classes and each posses their own attributes and methods. Main Classes of the proposed system are RandomJump, DBconnection, DFS, Bipartitegraph and HDModel each have different functionalities. Class diagram for proposed system is in figure 6.

**Figure 6:** Class Diagram

#### 4.4.3 Activity Diagram

An activity is particular operation of the system. An activity diagram is intended to represent stepwise work-flow of activities or actions that can take place in the system. It shows overall flow of control and models computational and organizational processes. Activity diagrams are used to model dynamic aspects of the system. Activity diagram for the system is shown in figure 7.

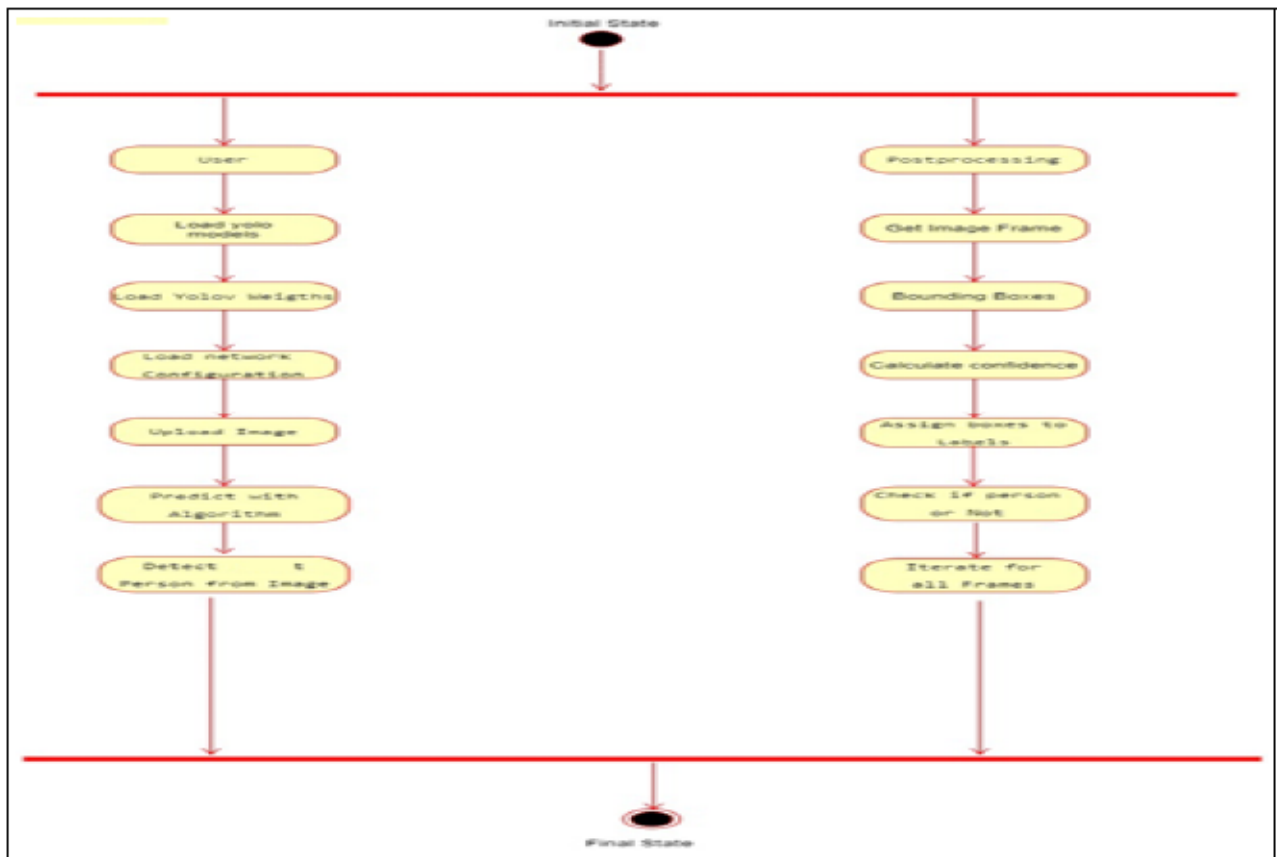
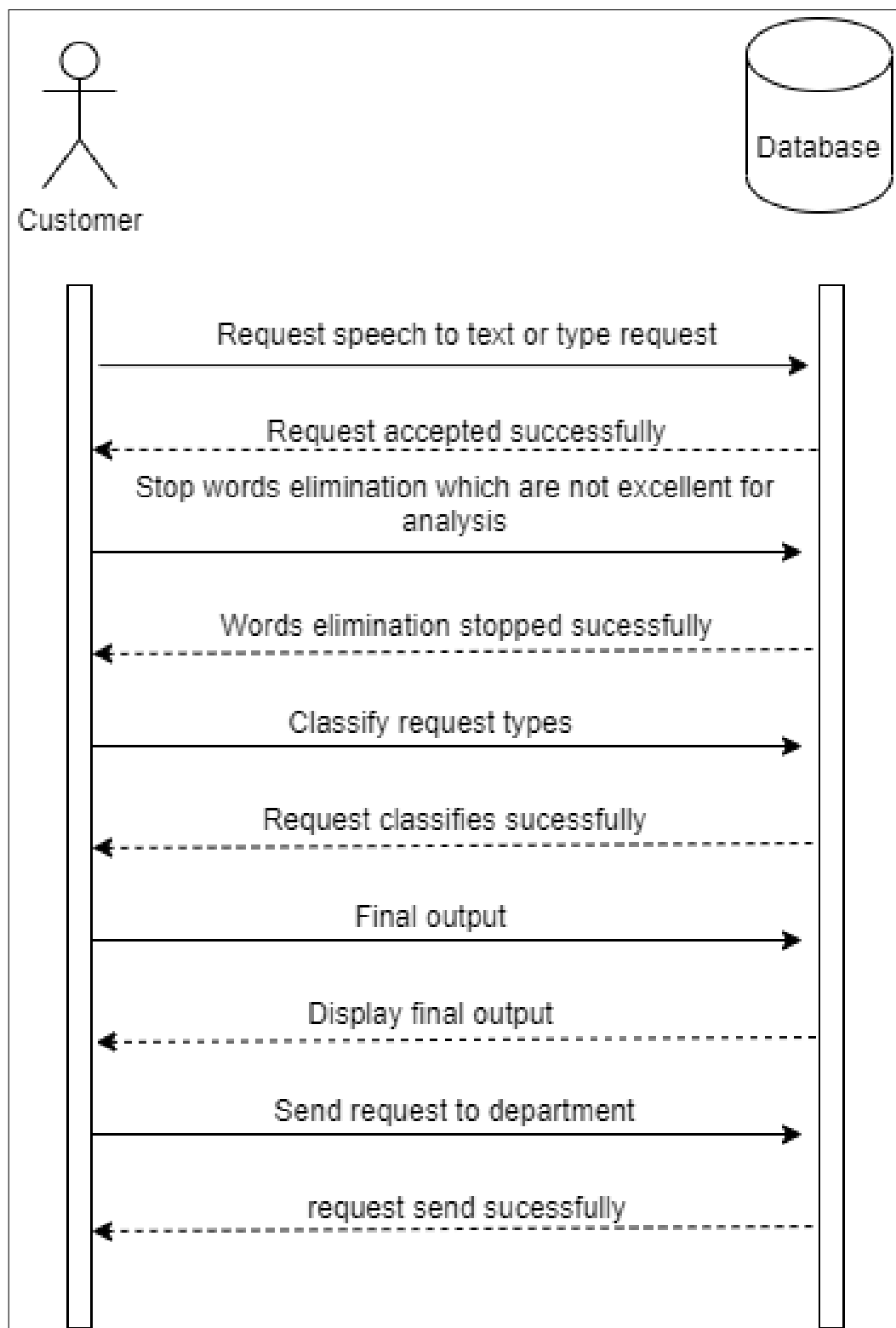


Figure 7: Activity Diagram

#### 4.4.4 Sequence Diagram

Sequence diagram shows how objects communicate with each other in terms of a sequence of messages. It also indicates the lifespans of objects relative to those messages. There are mainly three different objects User, Recommender system and Database. User enters query, recommender system extracts suggestions from the database using diffusion model and provide results to the user. Figure 8 is sequence diagram for the proposed system.

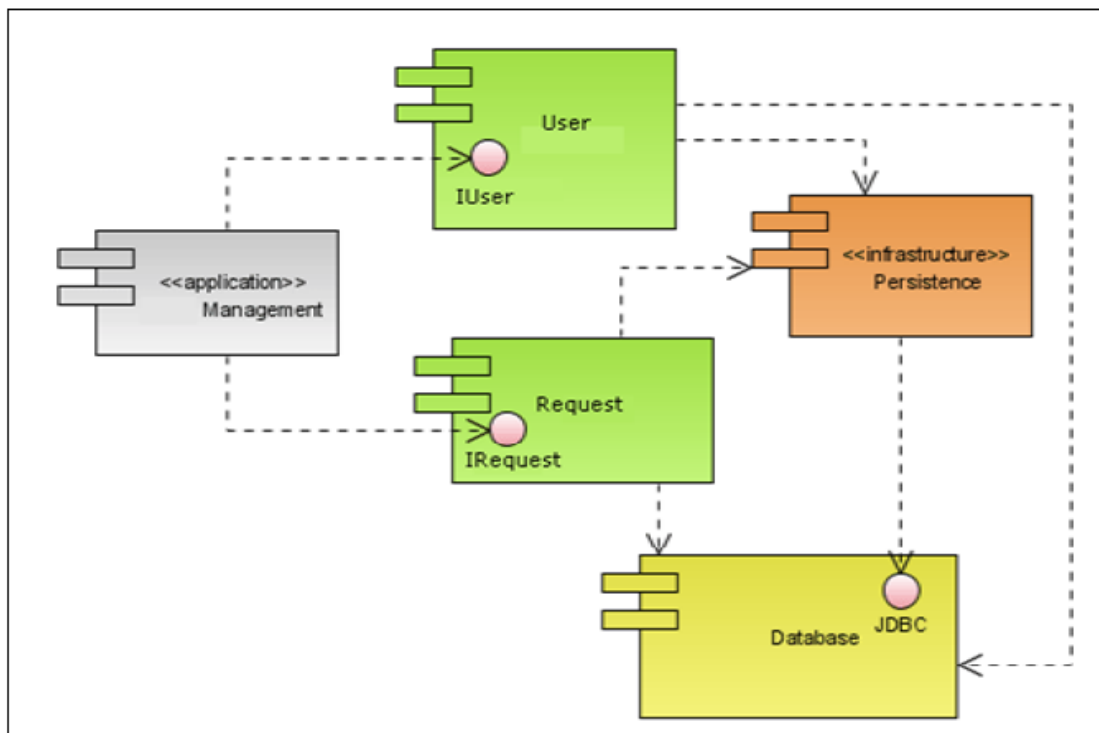
**Figure 8:** Sequence Diagram

#### **4.4.5 State Chart**

UML state chart describes the states and state transitions of the system. There are many different states through which system transits. First of all user enters a query, system extracts a query-URL bipartite graph from the database, Heat diffusion model is applied over the graph, heat values are calculated after doing random jump and the suggestions to the given query, having top heat values are supplied to user. Figure ?? shows the respective diagram.

#### 4.4.6 Component Diagram

Component diagram is different than other UML diagrams. Instead of depicting functionality of the system, a component diagram describes how a system is composed by combining different components together. A component diagram describes structural relationship between different components of the system, what are the required interfaces, etc. Components may include executable files, library files, database tables, etc. Component diagram for the system is shown in figure 9.



**Figure 9:** Component Diagram

## 4.5 Algorithmic Strategy

In our project we use TF\*IDF algorithm is an information retrieval technique that weighs a term's frequency (TF) and its inverse document frequency (IDF). Each word or term that occurs in the text has its respective TF and IDF score.

The product of the TF and IDF scores of a term is called the TF\*IDF weight of that term. Put simply, the higher the TF\*IDF score (weight), the rarer the term is in a given document and vice versa.

The TF\*IDF algorithm is used to weigh a keyword in any content and assign importance to that keyword based on the number of times it appears in the document. More importantly, it checks how relevant the keyword is throughout the web, which is referred to as corpus.

Gather words. Write your content. Run a TF\*IDF report for your words and get their weights. The higher the numerical weight value, the rarer the term. The smaller the weight, the more common the term. Compare all the terms with high TF\*IDF weights with respect to their search volumes on the web. Select those with higher search volumes and lower competition. Work smart.

A good rule of thumb is, the more your content "makes sense" to the user, the more weight it is assigned by the search engine. With words having a high TF\*IDF weight in your content, your content will always be among the top search results, so you can:

stop worrying about using the stop-words.  
successfully hunt words with higher search volumes and lower competition  
be sure to have words that make your content unique and relevant to the user, etc.

## 4.6 Time Complexity of Proposed System

The computation of  $e^{\alpha R}$  is very time consuming, so discrete approximation of  $e^{\alpha R}$  is used (Equation 3), where P is a positive integer. Two techniques are introduced to reduce the computational complexity (1) since  $f(0)$  is a vector,  $(I + \frac{\alpha}{p}R)^p f(0)$  is

calculated iteratively by applying the operator  $(I + \frac{\alpha}{p}R)$  to  $f(0)$ ; (2) for matrix  $R$ , a data structure is employed which only stores the information of non-zero entries, since it is a very sparse matrix. Thus, supposing a graph is connected by  $M$  edges (relationships between nodes), the complexity of executing the heat diffusion process is  $O(PM)$ , which represents the number of iterations  $P$  multiplied by the number of edges  $M$  in a graph. In most cases,  $P = 10$  is enough for approximating the heat diffusion equation. The complexity  $O(PM)$  shows that the heat diffusion algorithm enjoys very good performance in scalability since it is linear with respect to the number of edges in the graph.



## 5 Project Plan

### 5.1 Project Estimate

#### 5.1.1 Reconciled Estimates

#### 5.1.2 Cost Estimate

Cost of project

$$C = N * C_p$$

$$C = 3 * 8000$$

$$C = 24,000$$

The Cost of the project is approximately up to 24000/-

#### 5.1.3 Time Estimates

##### Line of Code (LoC):

Estimating LOC for this project is difficult at estimation stages this project is of innovative type project. Average estimation of this project is 10000 to 12000 line of code.

##### LOC based Estimation:

##### Efforts in Person in months

$$E = 3.2 * (KLOC)^{1.05}$$

$$E = 3.2 * 9.0^{1.05} \text{ to } 11.0 * 4.2^{1.05}$$

Function	Estimated KLOC
GUI design	1.1-1.3
Logical code	1.5-2.0
Location Based code	1.1-1.3
Directory matching code	1.0-1.3
Business logic	2.2-2.5
Testing	1.1-1.2
Re-correct Code	1.0-1.2
Total	9.0-10.11

#### 5.1.4 Man Month Utilization:

Estimation of the man month is divide into following sub activities:

1-Technical training of the team member: This will take nearly 1 months. This will include Advance java, mysql, serialization etc.

2-Research: Being an innovative project research for the project is an important part currently it seems to have 1 to 1.5 months

#### 5.1.5 Project Resources

- **Hardware Resources Required:**

1. Processor: Intel i3
2. Hard Disk: Minimum 100GB
3. RAM: 4GB

- **Software Resources Required:**

1. Platform: Windows7 and above.
2. Backend: Mysql 5.5.0
3. Front End: JAVA.

## 5.2 Risk Management

- In appropriate dataset -To overcome this risk we are trying to use well organized and complete dataset.
- Security- To overcome and improving security we use multilevel security like access permissions of user.

When solving problems we have to decide the difficulty level of our problem. There are three types of classes provided for that. These are as follows:

- P class
- NP-Hard Class
- NP Complete Class

### **P Class:-**

In computational complexity theory, P, also known as PTIME or  $DTIME(nO(1))$ , is a fundamental complexity class. It contains all decision problems that can be solved by a deterministic Turing machine using a polynomial amount of computation time, or polynomial time. P class problems are deterministic problems i.e. P class problems can be solve by deterministic Turing Machine.

### **Np Class:-**

In computational complexity theory Equivalently, the formal definition of NP is the set of decision problems solvable in polynomial time by a theoretical non-deterministic Turing machine. This second definition is the basis for the abbreviation NP, which stands for "nondeterministic, polynomial time." However, the verifier-based definition tends to be more intuitive and practical in common applications compared to the formal machine definition. The two definitions are equivalent because the algorithm

for the machine definition consists of two phases, the first of which consists of a guess about the solution, which is generated in a non-deterministic way, while the second phase consists of a deterministic algorithm that verifies or rejects the guess as a valid solution to the problem.

### 3. NP Hard Class:-

- NP-hardness stands for Non-deterministic polynomial-time hard.
- Informally, "at least as hard as the hardest problems in NP" are called as NP hard class problem.

#### 5.2.1 Risk Identification

For risks identification, review of scope document, requirements specifications and schedule is done.

1. Have top software and customer managers formally committed to support the project?

Ans: All the required software's are freely available and hence development will be possible.

2. Are end-users enthusiastically committed to the project and the system/product to be built?

ans: The end user will be developers itself.

3. Are requirements fully understood by the software engineering team and its customers?

Ans: Yes. All the requirements are fully understood by our team.

4. Have customers been involved fully in the definition of requirements?

Ans: This is academic level project. So that whatever requirement be specify it should be by our team members and our guide.

5. Do end-users have realistic expectations?

Ans: Yes.

6. Does the software engineering team have the right mix of skills?

Ans: Yes, we have.

7. Are project requirements stable?

Ans: All the basic requirements for this project are stable, from though some being variable but can be fulfilled.

8. Is the number of people on the project team adequate to do the job?

Ans: Yes.

9. Do all customer/user constituencies agree on the importance of the project and on the requirements for the system/product to be built?

Ans: Yes.

### 5.2.2 Risk Analysis

The risks for the Project can be analyzed within the constraints of time and quality.

ID	Risk Description	Probability	Impact		
			Schedule	Quality	Overall
1	Deadline Risk	medium	Low	High	High
2	Technical skill Risk	medium	Low	High	High
3	Hardware Failure Risk	high	high	medium	high
4	Accuracy Risk	medium	medium	low	high

**Figure 10:** Risk Table

Probability	Value	Description
High	Probability of occurrence is	> 75%
Medium	Probability of occurrence is	26 – 75%
Low	Probability of occurrence is	< 25%

**Table 3:** Risk Probability definitions

### 5.3 Overview of Risk Mitigation ,Monitoring,Management

Impact	Value	Description
Very high	> 10%	Schedule impact or Unacceptable quality
High	5 – 10%	Schedule impact or Some parts of the project have low quality
Medium	< 5%	Schedule impact or Barely noticeable degradation in quality Low Impact on schedule or Quality can be incorporated

Figure 11: Risk Impact definitions

Risk ID	1
Risk Description	Development Deadline Risk
Category	Development Environment.
Source	Software requirement Specification document.
Probability	Low
Impact	High
Response	Mitigate
Strategy	Team Work distribution and Task plan.
Risk Status	Occurred

Figure 12: Risk Impact definitions

Risk ID	2
Risk Description	Technical skill risk
Category	Requirements
Source	Software Design Specification documentation review.
Probability	Low
Impact	High
Response	Mitigate
Strategy	Self study and Internet will be best source for technology knowledge
Risk Status	Identified

Risk ID	3
Risk Description	Hardware Failure
Category	Requirements
Source	Software Design Specification documentation review.
Probability	Low
Impact	High
Response	Mitigate
Strategy	Check whether Hardware is accurately fit or not according to knowledge
Risk Status	Identified

## 5.4 Project Schedule

### 5.4.1 Project Task Set

Major Tasks in the Project stages are:

- Task 1.1: Checking Feasibility of product
- Task 1.2: Scope of Product

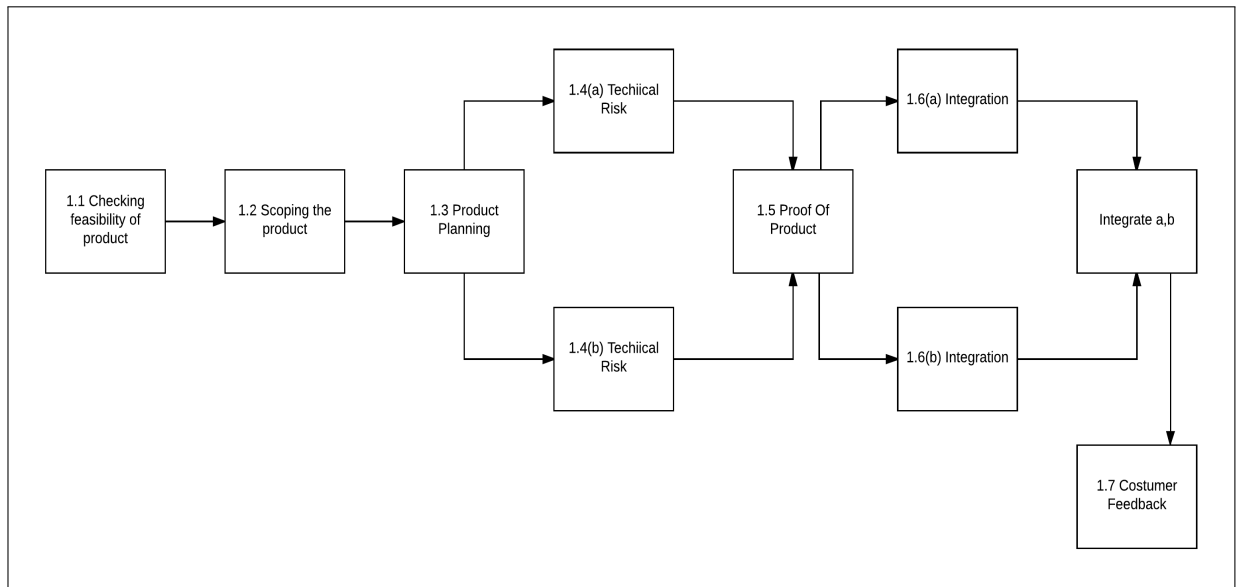


Risk ID	4
Risk Description	Weather API Risk
Category	Environmental Condition through OpenWeatherMap API.
Source	Software requirement Specification document.
Probability	Low
Impact	High
Response	Mitigate
Strategy	Capture Current as well as Future weather data.
Risk Status	Identified

- Task 1.3: Product Planing
- Task 1.4: Technical Risk
- Task 1.5: Proof of product
- Task 1.6: Implementation
- Task 1.7: Costumer Feedback

## 5.5 Task network

Project tasks and their dependencies are noted in this diagrammatic form.



**Figure 13:** Task Network

### 5.5.1 Time line Chart

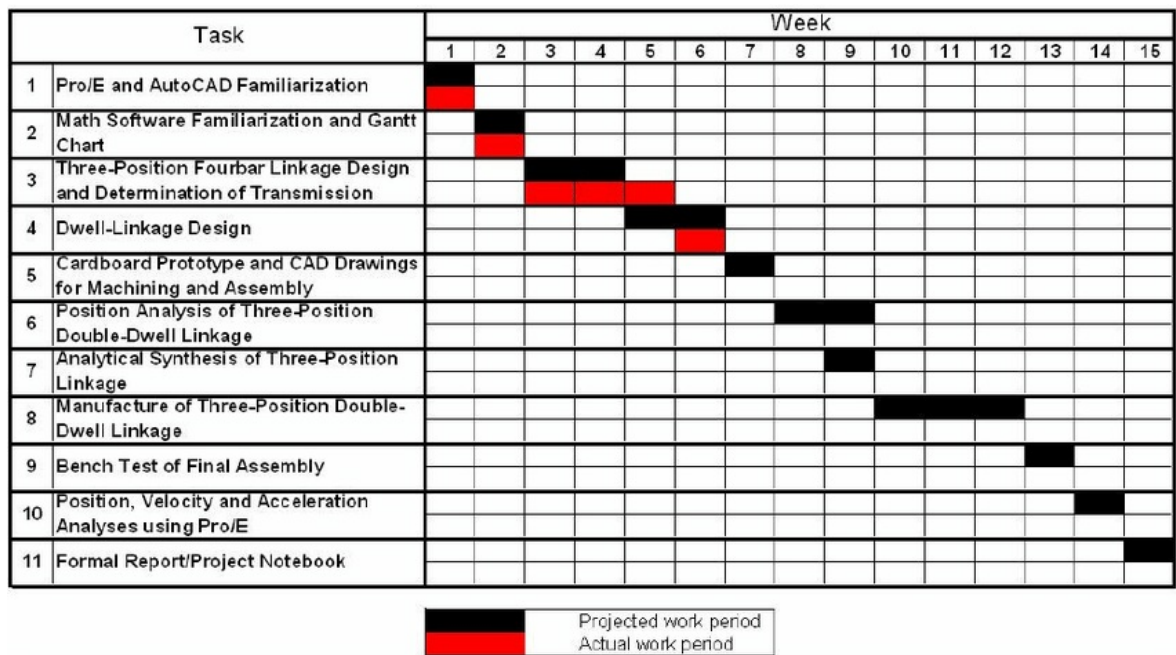


Figure 14: Timeline of Literature review

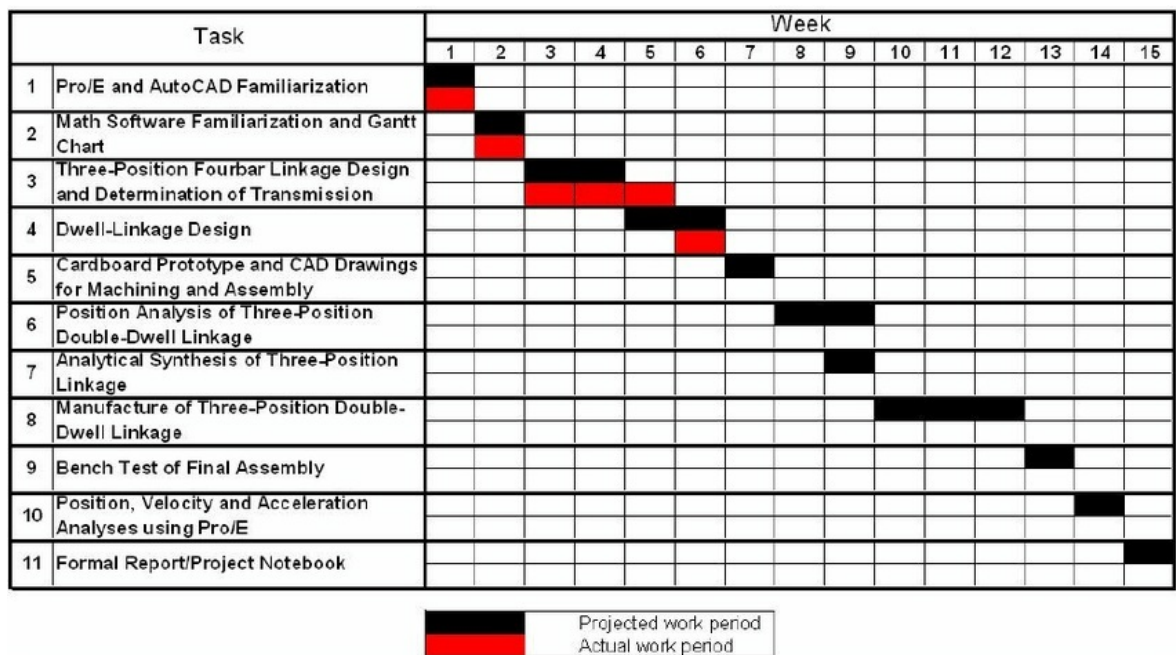


Figure 15: Timeline of Project Stage 1

## **5.6 Team Organization**

### **5.6.1 Team Structure**

The team structure for the project is identified. Roles are defined. Our team have three members. We select this topic after discussing with each other. All the members performing all the task whatever tasks are assign to the members. Team Member Details:-

Sanket sopan shirse Shubham Ajit asabe Rohan mandar salvi Vipul bhausahab ahire

### 5.6.2 Management reporting and communication

SR.NO	Reporting Date	Project Activity
1	22 June 2020	Decide project group member
2	29 June 2020	Submitted 3 Project Topic with IEEE Paper
3	13 Jul 2020	Discuss 5 point analysis of selected IEEE Paper
4	20 Jul 2020	3 Topics are presented and 1 topic selected
5	27 Jul 2020	Created and Submitted synopsis of a selected project
6	3 Aug 2020	Literature Survey and info gathering of a selected project
7	10 Aug 2020	30 percent project completion and presentation
8	31 Aug 2020	Draw UML diagram of a project
9	31 Aug 2020	50 percent project completion and presentation
10	7 Feb 2020	100 percent project completion and presentation
11	3 March 2021	Show the paper published
12	23 March 2021	Show the final report
13	6 April 2021	Show the final PPT
14	9 May 2021	Term 1st Project overview

## 6 Project Implementation

### 6.1 Overview of Project Modules

#### **Users:-**

First user open browser and enter url

- Registration.
- Login.
- Request speech to text or request.
- Stop word elimination which are not excellent for analysis
- Classify Request type.
- find out solution
- Send request to department.

#### **Postprocessing:-**

means server system it include,

- User request accept
- Word elimination stopping.
- display final output.
- Request send successfully .

### 6.2 Tools and Technologies Used

- **JAVA:**

Java is one of the most popular and widely used programming language and platform. A platform is an environment that helps to develop and run programs written in any programming language.

Java is fast, reliable and secure. From desktop to web applications, scientific

supercomputers to gaming consoles, cell phones to the Internet, Java is used in every nook and corner.

Java is a programming language and computing platform first released by Sun Microsystems in 1995. There are lots of applications and websites that will not work unless you have Java installed, and more are created every day. Java is fast, secure, and reliable. From laptops to datacenters, game consoles to scientific supercomputers, cell phones to the Internet, Java is everywhere!

Java is a general-purpose, concurrent, object-oriented, class-based, and the runtime environment(JRE) which consists of JVM which is the cornerstone of the Java platform. This blog on What is Java will clear all your doubts about why to learn java, features and how it works.

- **XAMPP:**

XAMPP stands for Cross-Platform (X), Apache (A), MySQL (M), PHP (P) and Perl (P). It is a simple, lightweight Apache distribution that makes it extremely easy for developers to create a local web server for testing purposes. Everything you need to set up a web server – server application (Apache), database (MySQL), and scripting language (PHP) – is included in a simple extractable file. XAMPP is also cross-platform, which means it works equally well on Linux, Mac and Windows. Since most actual web server deployments use the same components as XAMPP, it makes transitioning from a local test server to a live server is extremely easy as well. Web development using XAMPP is especially beginner friendly.

- **JDK:**

The Java Development Kit (JDK) is an implementation of either one of the Java Platform, Standard Edition, Java Platform, Enterprise Edition, or Java Platform, Micro Edition platforms released by Oracle Corporation in the form of a binary product aimed at Java developers on Solaris, Linux, macOS or Windows. The JDK includes a private JVM and a few other resources to finish the development

of a Java Application. Since the introduction of the Java platform, it has been by far the most widely used Software Development Kit (SDK).[citation needed] On 17 November 2006, Sun announced that they would release it under the GNU General Public License (GPL), thus making it free software. This happened in large part on 8 May 2007, when Sun contributed the source code to the OpenJDK.

- **Apache:**

Apache is the actual web server application that processes and delivers web content to a computer. Apache is the most popular web server online, powering nearly 54 percent of all websites.

- **MySQL:**

Every web application, howsoever simple or complicated, requires a database for storing collected data. MySQL, which is open source, is the world's most popular database management system. It powers everything from hobbyist websites to professional platforms like WordPress. You can learn how to master PHP with this free MySQL database for beginners course.

## 6.3 Algorithmic Details

In our project we use TF\*IDF algorithm is an information retrieval technique that weighs a term's frequency (TF) and its inverse document frequency (IDF). Each word or term that occurs in the text has its respective TF and IDF score.

The product of the TF and IDF scores of a term is called the TF\*IDF weight of that term. Put simply, the higher the TF\*IDF score (weight), the rarer the term is in a given document and vice versa.

The TF\*IDF algorithm is used to weigh a keyword in any content and assign importance to that keyword based on the number of times it appears in the document. More importantly, it checks how relevant the keyword is throughout the web, which is referred



to as corpus.

Gather words. Write your content. Run a TF\*IDF report for your words and get their weights. The higher the numerical weight value, the rarer the term. The smaller the weight, the more common the term. Compare all the terms with high TF\*IDF weights with respect to their search volumes on the web. Select those with higher search volumes and lower competition. Work smart.

A good rule of thumb is, the more your content “makes sense” to the user, the more weight it is assigned by the search engine. With words having a high TF\*IDF weight in your content, your content will always be among the top search results, so you can:

stop worrying about using the stop-words.  
successfully hunt words with higher search volumes and lower competition  
be sure to have words that make your content unique and relevant to the user, etc.

### **6.3.1 Algorithm 1:**

TF\*IDF is an information retrieval technique that weighs a term’s frequency (TF) and its inverse document frequency (IDF). Each word or term that occurs in the text has its respective TF and IDF score. TF\*IDF is used by search engines to better understand the content that is undervalued.

Step 1 :Tokenize the sentences.

Step 2 :Create the Frequency matrix of the words in each sentence.

Step 3: Calculate TermFrequency and generate a matrix.

Step 4: Creating a table for documents per words. Step 5 : Calculate IDF and generate a matrix.

Step 6: Calculate TF-IDF and generate a matrix.

Step 7: Score the sentences.

Step 8: Find the threshold.

### 6.3.2 Algorithm 2: K-Means

K-means is a centroid-based algorithm, or a distance-based algorithm, where we calculate the distances to assign a point to a cluster. In K-Means, each cluster is associated with a centroid.

**Input:**  $k$  (the number of clusters),  
 $D$  (a set of lift ratios)

**Output:** a set of  $k$  clusters

**Method:**

Arbitrarily choose  $k$  objects from  $D$  as the initial cluster centers;

**Repeat:**

1. (re)assign each object to the cluster to which the object is the most similar, based on the mean value of the objects in the cluster;
2. Update the cluster means, i.e., calculate the mean value of the objects for each cluster

**Until** no change;

## 7 Software Testing

### 7.1 Types of Tasting

Testing is an important part of software development life cycle. It is performed to ensure quality of the developed system. Testing includes a set of investigative activities that can be planned in advance and conducted systematically, to assure the stakeholder that system fulfils all the requirements gathered during requirement gathering phase. Software testing is one of the key elements in software projects that is often referred to as verification and validation. Verification refers to the set of activities that ensure that software correctly implements specified functionality. Validation refers to a set of activities built around traceability matrix which ensure that the functionality implemented by the system is traceable to customer requirements.

The software test plan (STP) is designed to test each module to measure its performance, to uncover bugs in the system, to set aright any flaws in logic that may be present, and to check logical flow from one module to another within system.

### 7.2 Test cases and Test Results

A strategy outlines what to plan, and how to plan it. A successful strategy is your guide through change, and provides a firm foundation for ongoing improvement. Unlike a plan, which is obsolete from the point of creation, a strategy reflects the values of an organization - and remains current and useful. When an organization tests its products or its tools, it tries to compare them against its expectations and values. By its nature, testing introduces change as problems are identified and resolved. A test strategy is necessary to allow these two impulses to work together. Furthermore, testing can never be said to be 'complete', and a core skill in testing is the justified management of conflicting demands; without a strategy, these judgements will be inconsistent to the point of failure.

Software development is a creative process. A test strategy is a vital enabler to this process keeping focus on core values and consistent decision-making to help achieve

desired goals with best use of resource. A good strategy stands as a clear counter to reactive, counter-productive test approaches.

Test case	Login Screen- Sign up
Objective	Click on sign up button then check all required/ mandatory fields with leaving all fields blank
Expected Result	All required/ mandatory fields should display with symbol “*”. Instruction line “* field(s) are mandatory” should be displayed
Test case	Create a Password >>Text Box Confirm Password >>Text Box
Objective	Check the validation message for Password and Confirm Password field
Expected Result	Correct validation message should be displayed accordingly or “Password and confirm password should be same” in place of “Password mismatch”.

Test Case ID	Test Case	Test Case I/P	Actual Result	Expected Result	Test case criteria(P/F)
001	Enter The Wrong username or password click on submit button	Username or password	Error comes	Error Should come	P
002	Enter the correct username and password click on submit button	Username and password	Accept	Accept	P

Test Case ID	Test Case	Test Case I/P	Actual Result	Expected Result	Test case criteria(P/F)
001	Enter the number in username, middle name, last name field	Number	Error Comes	Error Should Comes	P
001	Enter the character in username, middle name, last name field	Character	Accept	Accept	p
002	Enter the invalid email id format in email id field	<u>Kkgmail.com</u>	Error comes	Error Should Comes	P
002	Enter the valid email id format in email id field	kk@gmail.com	Accept	Accept	P
003	Enter the invalid digit no in phone no field	99999	Error comes	Error Should Comes	P
003	Enter the 10 digit no in phone no field	9999999999	Accept	Accept	P

## 8 Results

## 9 Conclusions

### 9.1 Conclusions

The performance of the tool and the quality of the obtained results was improved using a specific cartridge to Portuguese language. This facility contains particular properties that are very helpful. In accordance with the obtained results, it must be stressed that the sample of data, as well as the activities carried through the pre-processing step and clustering process are of fundamental importance to the knowledge extraction. In our experiments the traditional categorization process built during the company practice was confirmed, and could be validated. Although promising results have been achieved in this work, there are some issues that can be further investigated by the call center experts. But we can conclude that now is possible to implement an automatic classification system to online monitor the service quality. Finally it can be concluded that the obtained text mining results can be used in the power electric industry to: Understand the needs of customers while accessing preferred customers;

- Provide service menus for locking in preferred customers, such as offering discount charge menus and added-value services;
- Analyze how both electric power income and related income change, and investigate the contents of the services;
- Measure whether customers offered services are satisfied with those services and reflect the results in the service menu planning.

### 9.2 Future Work

The feature work also added customers' bills invoicing, probably one of the problems that more receive calls in all the existing topics in a company from the power electric industry.

## **9.3 Applications**

Successful centers use advanced call center analytics software to monitor and review performance, not only from a customer lens, also from the employee's perspective, as well as a business-owner lens.

Each of these approaches offers its own advantages and together satisfies each angle. The key to choosing the correct analytics combination lies in understanding the approaches, and how they can be used to improve your call center. Here are the six most common approaches to analytics:

### **1. Call Center Speech Analytics**

Speech Analytics is a fairly new and relatively rare field, but one that many early adopters are finding significant success with. Using a team of analysts to monitor calls in real time, a company can unearth inefficiencies in their current model, and make process improvements, such as moving to a call script, or developing systems for call center agents to utilize in order to achieve the desired call outcome.

### **2. Call Center Text Analytics**

The last several years have seen an explosion in the social media universe, and most forward thinking companies have developed a brand presence online. This paradigm shift has rendered text analytics ever more important, as we are no longer communicating with our customers through written documents, but also through email, secure messaging, Facebook, Twitter, and other text-centered media. Text analytics can review and monitor not only the messages sent to customers, but also the message they are sending to the company. This is vital in seeing any potential issues through the customer lens.

### **3. Predictive Analytics**

The modern predictive analysis engine is an invaluable tool in the call center environment. Using in-depth review of past performance in areas as diverse as call volume, service level, handle time, and customer satisfaction, predictive analysis



makes it possible to apply past solutions to upcoming problems. How many agents will we need staffed on Christmas Day? How will your new product rollout affect call volume on weekends? What will this change to your fee structure do to your customer satisfaction score? By analyzing past results, companies can plan and strategize for the future.

# Appendices

## A Appendix

### A.1 Appendix A

fire.pdf.pdf fire.pdf

## **B Appendix B**

### **B.1 I2CT Certificates**

### **B.2 I2CT Certificates**

## **C Appendix C**

### **C.1 Plagiarism Report**