SITS/Computer Engineering/Projects/UG/2019-20/B13

A PROJECT REPORT ON

Text Segmentation With Feature Similarity For Exam Assessment Using Machine Learning

SUBMITTED TO THE SAVITRIBAI PHULE PUNE UNIVERSITY, PUNE IN THE PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE AWARD OF THE DEGREE OF

BACHELOR OF ENGINEERING (COMPUTER ENGINEERING)

SUBMITTED BY

Aditya Singh	Exam No.: B150574204
Pranav Kapse	Exam No.: B150574267
Rahul Nagpal	Exam No.: B150574307
Shivam Kumar	Exam No.: B150574335

Guide
Ms. P. V. Ambekar

Sinhgad Institutes

DEPARTMENT OF COMPUTER ENGINEERING
SINHGAD INSTITUTE OF TECHNOLOGY AND SCIENCE
PUNE, 411041

SAVITRIBAI PHULE PUNE UNIVERSITY
2019 - 2020

Sinhgad Institutes

C E R T I F I C A T E

This is to certify that *Mr. ADITYA SINGH Exam No B150574204*, *Mr. PRANAV KAPSE Exam No B150574267*, *Mr. RAHUL NAGPAL Exam No B150574307*, *Mr. SHIVAM KUMAR Exam No B150574334* have successfully completed the Project Stage-II entitled *Text Segmentation With Feature Similarity For Exam Assessment Using Machine Learning* under my supervision, in the fulfillment of Bachelor of Computer Engineering of Savitribai Phule Pune University.

Date :
Place :

Ms. P. V. Ambekar
Guide

Mrs. G. S. Navale
Head of Department

External Examiner
-
-

Dr. R. S. Prasad
Principal
Sinhgad Institute of Technology and Science, Pune 41

ACKNOWLEDGEMENT

We express our gratitude to my guide Ms. P. V. Ambekar for her competent guidance and timely inspiration. It is our good fortune to complete our Project under her able competent guidance. This valuable guidance, suggestions, helpful constructive criticism, keeps interest in the problem during the course of presenting this "Text segmentation with feature similarity for Exam Assessment using Machine Learning." project successfully.

We would like to thank our Project Coordinator Dr. Geeta S. Navale and all the Teaching, Non-Teaching staff of our department.

We are very much thankful to Dr. Geeta S. Navale, Head, Department of Computer Engineering and also Dr. R. S. Prasad, Principal, Prof. S. A. Kulkarni, Vice principal, Sinhgad Institute of Technology and Science, Narhe for their unflinching help, support and cooperation during this project work.

We would also like to thank the Sinhgad Technical Educational Society for providing access to the institutional facilities for our project work.

Aditya Singh (B150574204)
Pranav Kapse (B150574267)
Rahul Nagpal (B150574307)
Shivam Kumar (B150574335)

Date:
Place: Pune

ABSTRACT

The need of green computing in order to reduce the excess use of paper to assess the theoretical answer is a serious demand. We therefore intend to provide a solution by building a model which helps in evaluating the theoretical answers online to reduce the human efforts. The project involves the use of machine learning, NLP, keyword extraction and matching aggregation for checking the similarity between the user answer and the specimen answer. The user written answer is tokenized into bag of words and the meaning of words are extracted and matched with the specimen answer for semantic analysis. The machine learning algorithm analyzes the answer and gives the percentage of similarity between the two answers, with this system we can automatically evaluate the theoretical answers easily and efficiently, thus reducing the use of paper.

KEYWORDS:- Machine learning, NLP(Natural Language Processing), Keyword extraction.

Contents

Certificate **i**

List of Abbreviations **vii**

List of Figures **viii**

List of Tables **ix**

1 Introduction **1**
- **1.1 Relevance** 1
- **1.2 Motivation** 2
- **1.3 Problem Definition and Objectives** 2
- **1.4 Objectives** 2
- **1.5 Schedule of Project Work** 2
- **1.6 Budget of the Project** 3

2 Literature Survey **5**
- **2.1 Literature Survey** 5
- **2.2 Gap Statement** 7
- **2.3 Summary** 7

3 Software Requirements Specification **9**
- **3.1 Assumptions and Dependencies** 9
- **3.2 Functional Requirements** 10
  - **3.2.1 System Feature 1** 10
  - **3.2.2 System Feature 2** 10
  - **3.2.3 System Feature 3** 10

3.3 External Interface Requirements 10
3.3.1 User Interfaces 10
3.3.2 Hardware Interfaces 10
3.3.3 Software Interfaces 11
3.3.4 Communication Interfaces 11
3.4 Nonfunctional Requirements 11
3.4.1 Performance Requirements 11
3.4.2 Safety Requirements 11
3.4.3 Security Requirements 11
3.4.4 Software Quality Attributes 12
3.5 System Requirements 12
3.5.1 Database Requirements 12
3.5.2 Software Requirements (Platform Choice) 12
3.5.3 Hardware Requirements 12
3.6 Analysis Models: SDLC Model to be applied 13

4 System Design **15**
4.1 System Architecture 15
4.2 Data Flow Diagrams 16
4.3 UML Diagrams 17
4.3.1 Use Case Diagram 17
4.3.2 Activity Diagram 18
4.3.3 Sequence Diagram 19

5 PROJECT PLAN **20**
5.1 Project Estimate 20
5.1.1 Reconciled Estimates 20
5.1.2 Project Resources 21
5.2 Risk Management 21
5.2.1 Risk Identification 22
5.2.2 Risk Analysis 22
5.2.3 Overview of Risk Mitigation, Monitoring, Management 23
5.3 Project Schedule 25
5.3.1 Project Task Set 25

5.3.2 Timeline Chart 27
5.4 Team Organization 27
5.4.1 Team Structure 27
5.4.2 Management, Reporting and Communication . . . 28

6 PROJECT IMPLEMENTATION **30**
6.1 Overview of Project Modules 30
6.2 Tools and Technologies Used 31
6.3 Algorithm Details 31
6.3.1 Measuring Similarity 31
6.3.2 Vectorization 32
6.3.3 Distance Computation 35

7 SOFTWARE TESTING **36**
7.1 Type of Testing 36
7.2 Test cases and Test Results 38

8 RESULTS **41**
8.1 Outcomes 41
8.2 Screen Shots 41

9 CONCLUSIONS **47**
9.1 Conclusions 47
9.2 Future Work 47
9.3 Applications 48

REFERENCES **49**

Appendix A **50**

Appendix B **51**

Appendix C **54**

Appendix C **57**

List of Abbreviations

NLP	**Natural Language Processing**
KBQA	**Knowledge Based Question and Answer**
BOW	**Bag Of Words**

List of Figures

1.1 Activity Sheet 3

3.1 WaterFall Model 13

4.1 System Architecture 15
4.2 DFD 0.0 16
4.3 DFD 1.0 16
4.4 DFD 2.0 16
4.5 Use Case Diagram 17
4.6 Activity Diagram 18
4.7 Sequence Diagram 19

8.1 Putty Login 42
8.2 Code for Seperation of Stop Words and Keywords 42
8.3 Output of Stop Words and Keywords 43
8.4 Code for Frequency of each word 44
8.5 Code for Frequency of each word 45
8.6 Output of Frequency 45
8.7 Code for percentage Similarity 46
8.8 Output of Similarity 46

List of Tables

2.1 Literature Review 8

7.1 Test Case table 40

Chapter 1

Introduction

The aim of this project is to develop an online system for evaluating the theoretical answers. The set of question and answers are stored in the database with which the answer written by the candidate is matched based on its semantic analysis. The user answer is tokenized into keywords and their meaning are extracted which is further evaluated through the machine learning algorithm to check the similarity between the original answer and the candidate answer. The answer is assessed and the percentage similarity is given as an output.

1.1 Relevance

Machine Learning is the future of computer technology. The preliminaries suggest that the matching aggregation and NLP has a strong capability to implement the knowledge based question and answer the system will be able to solve some of the oldest pedagogical practices and also reduce the burden on the checker. This work will also be able to set some benchmark and can even be used as one to be compared with.

1.2 Motivation

Nowadays the usage of paper is increasing exponentially which in turns affects the environment.
To reduce the use of paper and to save the environment we are developing an online system which evaluates the theoretical answers online.
To reduce the human efforts and for faster generation of results Machine Learning Algorithm is used.

1.3 Problem Definition and Objectives

To develop an online examination system to assess theoretical answers by using NLP.

1.4 Objectives

- **To calculate the similarity percentage between user answer and specimen answer.**
- **To tokenize the data for analysis and matching.**
- **To identify the similarity between user and specimen answer according to their semantic properties.**

1.5 Schedule of Project Work

Major Tasks in the Project stages are:

- **Task 1: Requirement Gathering**
- **Task 2: Literature survey**
- **Task 3: Mathematical modeling**
- **Task 4: Feasibility testing**
- **Task 5: UML diagrams**
- **Task 6: Database design**
- **Task 7: GUI design**
- **Task 8: Functionality implemented**

- **Task 9: Testing**
- **Task 10: Reporting**

Sr.	List of Activities	W1	W2	W3	W4	W5	W6	W7	W8	W9	W10	W11	W12
1	Group Formation												
2	Domain Submission												
3	Title Submission												
4	Project Title Finalization												
5	Synopsis Submission												
6	Finalization of Problem Statement and Objectives												
7	Listing of Reference papers												
8	Literature Review												
9	Summary of Papers												
10	Proposal												
11	Report												
12	Final Submission of Report												

Figure 1.1: Activity Sheet

1.6 Budget of the Project

The budget of a project is calculation or estimation of all the efforts and costs required to implement the project. For this project, the budget has been calculated by using CoCoMo model. The basic CoCoMo model was used in Organic mode.

The basic CoCoMo equations are -

$$E = a_b(KLOC)^{b_b} \tag{1.1}$$

$$D = c_b(E)^{d_b} \tag{1.2}$$

$$SS = E/D \tag{1.3}$$

Estimated size of the project = 3 KLOC
So, using equations 1.1 & 1.2, we get
$E = 2.4(3)^{1.05} = 7.60PM$
$D = 2.5(7.60)^{0.38} = 5.40M$

$SS = 7.60/5.40 = 1.40P$

Here, E is Effort (measured in Person Months)

D is Deployment Time (measured in Months)

SS is Staff Size (units is Persons)

Hence, Total Effort required is 13 person months(approx.) yielding a Development Time of 5.40 months and a Staff Size of 2 persons.

As, the team size is 4 persons, the development time of 5.40 months can be speeded up and calculated as follows:

Persons	**D**
2	**1/5.40**
4	**1/x**

So, $x = 2 * 5.40/4 = 2.7$ **Hence, the project will require 3 month (approx.) to complete (theoretically).**

This concludes the first chapter of the report. All the initial details regarding the project were discussed. The idea and theme of the project is now clear and the further proceedings are scheduled successfully. The budgeting is done and the effort required to implement the project is known. In the next chapter, there will be a discussion about the literature survey and the related details of the project.

Chapter 2

Literature Survey

The first chapter tells about the description of this project. It gives an idea about how the project is distributed in parts and the techniques that will be used to implement the project. The problem statement gives a brief idea about the project and the objectives gives a step-wise execution process of the project. This chapter includes the related work studied in relation with this project. These papers are close to the objectives of the project and the observations of these research papers are analyzed in the project.

2.1 Literature Survey

Yunshi Lan et.al.[1] proposed a sequence matching-based solution to Knowledge Based Question Answers(KBQA). Yunshi also explored the use of a "matching-aggregation" framework to match candidate answers with questions. The method that proposed by Yunshi is able to attain state-of-art performance on datasets .Yunshi working on two datasets that are Web Questions and Simple Questions. This paper also overcomes limitation of existing neural network-based method for Knowledge Based Question Answers(KBQA).

Vitor Rolim et. al. [2] explained the topic-based model for combining textual data extracted from online discussion forums to other external source which helps to identify the strength and weakness of

student and help to create profile based on the similarity. Vitor proposed the adoption of natural language processing for extracting the data and given focus on keyword extraction.

Diellza et.al [3] proposed emerging technique Natural language processing in today's era and how it is useful in establishing machine which is capable of translating between linguistic pair. Diellza give two classifier 'Rule-based' or parts of Speech (POS) which helps in identifying feature of language from large text collection. In this paper Diellza also explained the graph based label Propagation for projecting POS across different languages i.e. for that also which do not have annotated data. Bensik explained that how raw text can be used to generate spellcheck dictionaries and Biemann proposed the Chinese-whispers algorithm to find rare used words.

Xue Hanet.al [4] described NLP based BPM method to automatically synchronize and transform different business process representations with less time and high efficiency. Xue tells that how NLP process works for semantic analysis. NLP is the ability of a computer program to understand human language as it spoken the semantic analysis of a natural language content starts by reading all of word in content to capture real meaning of any text it identifies text elements and assigns them to their logical and grammatical. In this paper Xue explained NLP pipeline method to increase performance of dependency parsing.

Oliver et.al.[5] explained how the duplication question on stack overflow benefit the software development community. Oliver Analyzed the duplicate question from two perspective, first we analyzed the experience of the user who post the duplicate question and second comparing the contents of duplicates to determine the degree of similarity. Oliver followed the data filtration, data extraction and tokenization of text approach for the identification of duplicate, which is very useful and useable in this project. Oliver also explained some future work like developing more precise technique for similarity and another technique for sentiment analysis to grapple and delivering quantative measure of

duplicates.
Josep Carmona et.al.[6] explained how NLP has prospective in increasing the benefits of BPM practices at different levels. In this paper Josep provides NLP techniques that facilitate the automation of particular tasks. Also this paper overcomes the previous limitation that provides open-source BPM datasets to both academia and industrial application.

2.2 Gap Statement

There is no such system available which evaluate the theoretical answer for online examination. Now a days we only have the online MCQ examination so similarly we are developing a system which can evaluate the theoretical answer which will reduce the effort.

2.3 Summary

This section gives the important findings obtained during reviews as seen in Section 2.1 and are depicted in Table 7.1

Table 2.1: Literature Review

Refrence Number	**Highlights**	**Observations**
1.	Sequence matching-based solution to Knowledge Based Question Answers(KBQA)	Complex questions are still challenge to KBQA and need strong logic and reasoning for the same.
2.	Using topic model-based approach to extract students weakness and strength based on Latent Dirichlet Allocation(LDA)	This paper uses classification algorithm to categorize the forum message as question and answer .
3.	The NLP resources are quite useful when it comes to building a machine capable of translating between linguistic pairs.	This paper achieved semantic role labeling,spatial expression,opinion summarization,topic linking and visualization plugins.
4.	Described NLP based BPM method to automatically synchronize and transform different business process representations with less time and high efficiency	NLP pipelines uses POS tagging for great performance effect on downstream tasks such as dependency parsing.
5.	Analyzing the Duplicate questions from different perspective and exploring technique to benefit developer community	Analyzed similarity between Text using NLP and TF-IDF
6.	NLP potential in raising the benefits of BPM practices at different levels	NLP techniques that facilitate the automation of particular tasks.

Chapter 3

Software Requirements Specification

The previous chapter focused on the work carried out by previous researchers, the highlights and observations. The current chapter gives a detailed explanation of the software requirement specifications.

3.1 Assumptions and Dependencies

Hardware Failure: Hardware failure is the norm rather than the exception. An cloud instance may consist of hundreds or thousands of server machines, each storing part of the file system's data. The fact that there are a huge number of components and that each component has a non-trivial probability of failure means that some component of cloud is always non-functional. Therefore, detection of faults and quick, automatic recovery from them is a core architectural goal of cloud.

Streaming Data Access: Applications that run on cloud need streaming access to their data sets. They are not general purpose applications that typically run on general purpose file systems. Cloud is designed more for batch processing rather than interactive use by users. The emphasis is on high throughput of data access rather than low latency of data access. POSIX imposes many hard requirements that are not needed for applications that are targeted for cloud. POSIX semantics

in a few key areas has been traded to increase data throughput rates.

Large Data Sets Applications that run on cloud have large data sets. A typical file in cloud is gigabytes to terabytes in size. Thus, cloud is tuned to support large files. It should provide high aggregate data bandwidth and scale to hundreds of nodes in a single cluster. It should support tens of millions of files in a single instance.

3.2 Functional Requirements

3.2.1 System Feature 1

Perimeter Security Guarding access to the cluster through network security, firewalls, and, ultimately authentication to confirm user identities.

3.2.2 System Feature 2

Data Security Protecting the data in the cluster from unauthorized visibility through masking and Encryption both at rest and in transit.

3.2.3 System Feature 3

Defining what authenticated users and applications can do with the data in the cluster through file system ACLs and fine-grained authorization.

3.3 External Interface Requirements

3.3.1 User Interfaces

No external user interface.

3.3.2 Hardware Interfaces

There is no single hardware requirement set for installation.

3.3.3 Software Interfaces

Putty .

3.3.4 Communication Interfaces

No External Communication interfaces.

3.4 Nonfunctional Requirements

3.4.1 Performance Requirements

- **Ease of development**
- **Easy management at scale**
- **Advanced job management**
- **Multitenancy**

3.4.2 Safety Requirements

- **Data protection with snapshot and mirroring**
- **Automated self-healing**
- **Insight into software/hardware health and issues**

3.4.3 Security Requirements

- **Strong authentication and authorization**
- **Kerberos support**
- **Data confidentiality and integrity**

3.4.4 Software Quality Attributes

- **Portability**
- **Correctness**
- **Reliability**
- **Availability**

3.5 System Requirements

3.5.1 Database Requirements

- **MongoDB**

3.5.2 Software Requirements (Platform Choice)

- **Ubuntu (Stable Version)**
- **Pyhton 3**
- **JSON**
- **Cloud Server**
- **API**
- **Stream Processing Enviornment**

3.5.3 Hardware Requirements

- **2GB RAM**
- **50GB SDD**
- **1 CORE CPU**

3.6 Analysis Models: SDLC Model to be applied

Waterfall Model

The waterfall model Fig 3.1 is a breakdown of project activities into linear sequential phases, where each phase depends on the deliverables of the previous one and corresponds to a specialization of tasks. The approach is typical for certain areas of engineering design. The advantages of waterfall development are that it allows for departmentalization and control. A schedule can be set with deadlines for each stage of development and a product can proceed through the development process model phases one by one.

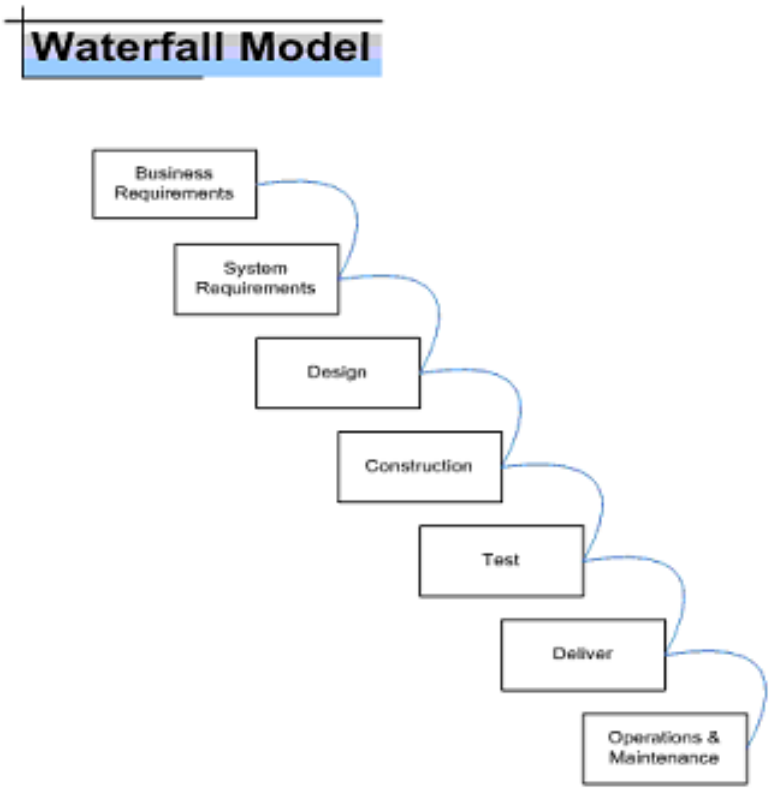


Figure 3.1: WaterFall Model

The sequential phases in Waterfall model are

- **Requirement Gathering and analysis : All possible requirements of the system to be developed are captured in this phase and documented in a requirement specification document.**
- **System Design : This system design helps in specifying hardware and system requirements and helps in defining the overall system architecture.**
- **Implementation : The inputs from the system design, the system is first developed in small programs called units, which are integrated in the next phase. Each unit is developed and tested for its functionality, which is referred to as Unit Testing.**
- **Integration and Testing : All the units developed in the implementation phase are integrated into a system after testing of each unit. Post integration the entire system is tested for any faults and failures.**
- **Deployment of system : Once the functional and non-functional testing is done; the product is deployed in the customer environment or released into the market.**
- **Maintenance : There are some issues which come up in the client environment. To fix those issues, patches are released. Also to enhance the product some better versions are released. Maintenance is done to deliver these changes in the customer environment.**

Chapter 4

System Design

This chapter contains the system design and diagrams which explain the flow of the project.

4.1 System Architecture

Figure 4.1: System Architecture

System Architecture used for showing the internal processing of the

module .It shohws how the machine is taking theh input and processing that input in a way and producing output as the text similarity.It is shown in Fig.4.1

4.2 Data Flow Diagrams

Figure 4.2: DFD 0.0

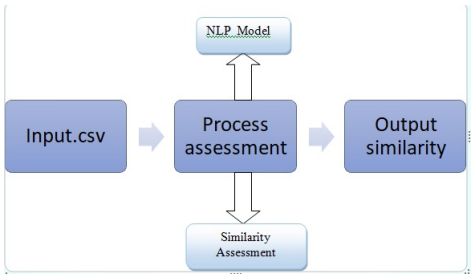


Figure 4.3: DFD 1.0

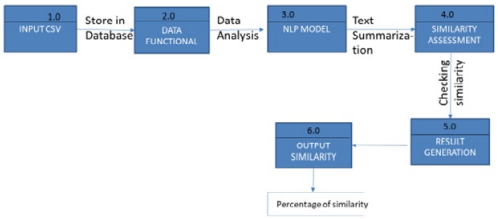


Figure 4.4: DFD 2.0

Data flow diagrams are used to graphically represent the flow of data in a business information system.The Fig.4.2 shows the Level1 of

Data Flow Diagram. DFD describes the processes that are involved in a system to transfer data from the input to the file storage and reports generation.The Fig.4.3 shows the Level 2 of Data Flow Diagram.The Fig.4.4 shows the overall structure of data flow diagram.It is simply merge of all data flow diagram levels.

4.3 UML Diagrams

4.3.1 Use Case Diagram

The use case diagram as shown in figure 4.3.1 summarizes the relationships between use cases and actors

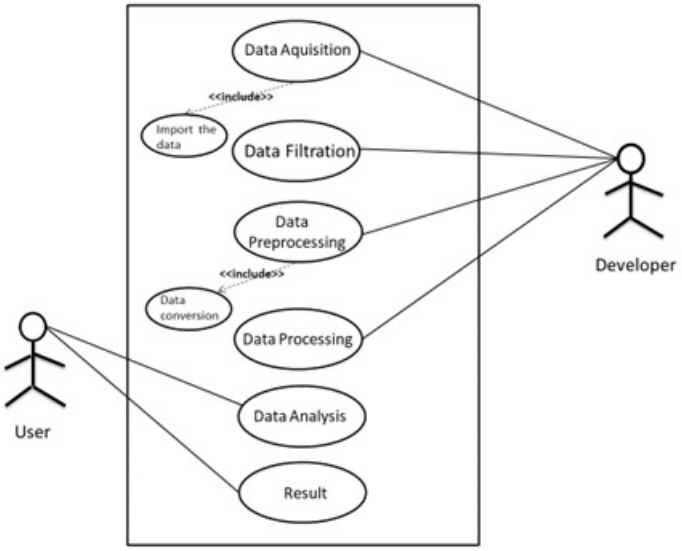


Figure 4.5: Use Case Diagram

A use case diagram at its simplest is a representation of a user's interaction with the system that shows the relationship between the user and the different use cases in which the user is involved. There are three main components in any Use case diagram Use case, Connectors and System boundary.

In above Fig.4.5 User connects with the Data Analysis and Result use cases and also developer connects with four use cases: Data Acquisi-

tion, Data Filtration, Data Preprocessing and Data Processing as shown in diagram. Data Acquisition has included relationship import the data i.e. to perform Data Acquisition, importing data should be compulsory. Similarly for Data Preprocessing has included relationship, Data conversion.

4.3.2 Activity Diagram

The diagram used to represent the flow from one operation to another operation of the system is called as activity diagram.

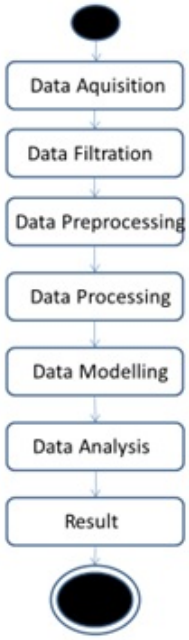


Figure 4.6: Activity Diagram

Acitivity diagram is a basically a flowchart to represent the flow from one activity to another activity. The activity can be described as an operation of the system as shown in Fig.4.6.

4.3.3 Sequence Diagram

The order in which the interactions between object takes place is depicted by sequence diagram.

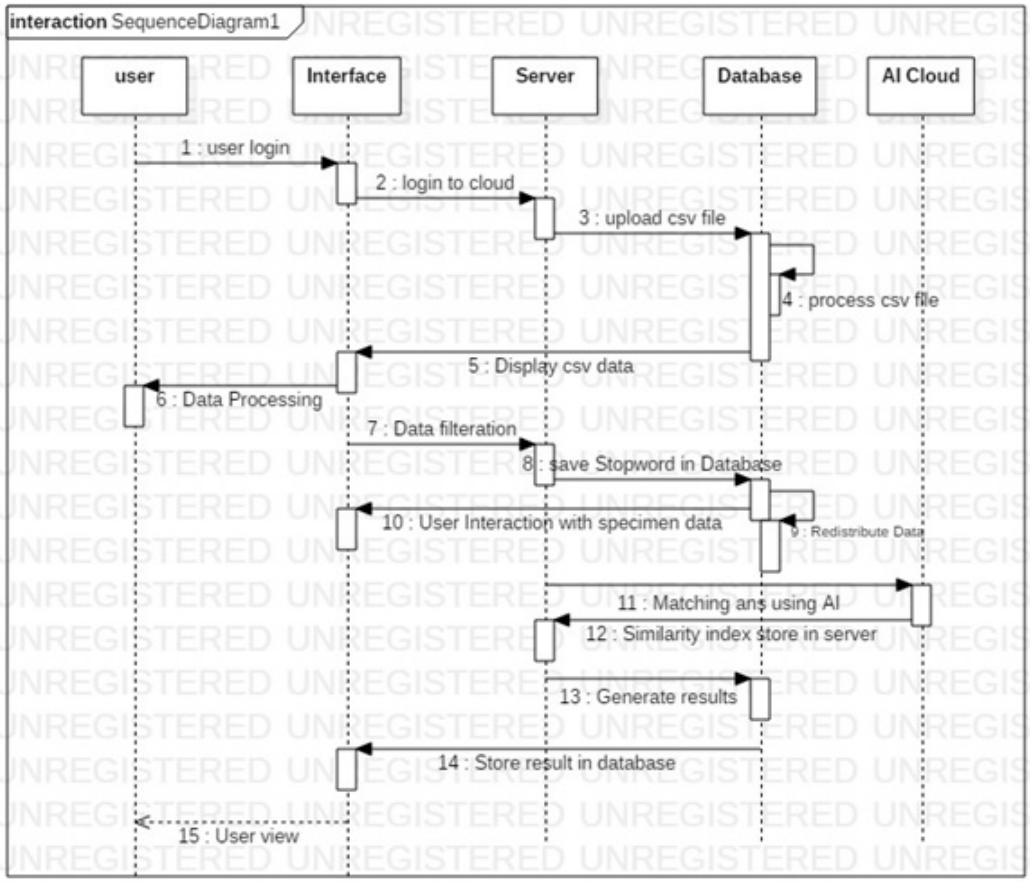


Figure 4.7: Sequence Diagram

Sequence diagram is used to show the all the process processing by the machine sequentially i.e, in a particular order Fig 4.7 shows which process is processed by whom and who is resppnsible for input and output .

Chapter 5

PROJECT PLAN

5.1 Project Estimate

Project Estimation is a critical part of project planning, involving a quantitative estimate of project costs,resources or duration. Estimating project costs are crucial to the success of overall project. Budget management and estimation are two major challenges of the project. A project estimation typically includes a breakdown of the tasks, resources, billing rates, and schedule for a project.

5.1.1 Reconciled Estimates

Reconciled estimation is an estimation type that compares the actual cost against the estimated cost.Reconciled estimates ensure that the differences between the two costs are appropriate.A reconciliation is usually organized by Work Breakdown Structure (WBS) and covers all aspects of project documentation. The process typically focuses on specific changes in scope, basis of estimate, and schedule and risks and involves clearly stating the differences between the two estimates and the rationale for those differences.

5.1.2 Project Resources

1. **RSS Feed : The news data source from where the raw data will be collected.**
2. **Hardware : Any Processor which is greater than 1.7 GHz And have greater than 2 GB Ram Hard Disk Must Be Greater than 100GB**
3. **Tools : Pyhton, Anaconda, MongoDB, Putty, FeedParser, Pymongo, Pycharm**

5.2 Risk Management

- **Easy data ingestion: Copying data to and from the Cloud cluster is as simple as copying data to a standard file system using Direct Access NFS. Applications can therefore ingest data into the Cloud cluster in real time without any staging areas or separate clusters just to ingest data.**
- **Existing applications work: Due to the Cloud platform's POSIX compliance, any non- Java application works directly on SERVER without undergoing code changes. Existing toolsets, custom utilities and applications are good to go on day one.**
- **Multi-tenancy: Support multiple user groups, any and all enterprise data sets, and multiple applications in the same cluster. Data modellers, developers and analysts can all work in unison on the same cluster without stepping on each other's toes.**
- **Business continuity: SERVER provides integrated high availability (HA), data protection, and disaster recovery (DR) capabilities to protect against both hardware failure as well as site-wide failure.**
- **High scalability: Scalability is key to bringing all data together on one platform so the analytics are much more nuanced and accurate. SERVER is the only platform that scales all the way to a trillion files without compromising performance.**

- **High performance: The Distribution for Cloud was designed for high performance, with respect to both high throughput and low latency. In addition, a fraction of servers are required for running Cloud distributions, leading to architectural simplicity and lower capital and operational expenses.**

5.2.1 Risk Identification

- **Security Information and Event Management (SIEM): Analyze and correlate large amounts of real-time data from network and security devices to manage external and internal security threats, improve incident response time and compliance reporting.**
- **Application Log Monitoring: Improve analysis of application log data to better manage system resource utilization, security issues, and diagnose and preempt production application problems.**
- **Network Intrusion Detection: Monitor and analyze network traffic to detect, identify, and report on suspicious activity or intruders.**
- **Fraud Detection: Use pattern/anomaly recognition on larger volumes and variety of data to detect and prevent fraudulent activities by external or internal parties.**
- **Risk Modelling: Improve risk assessment and associated scoring by building sophisticated machine learning models on Cloud that can take into account hundreds or even thousands of indicators.**

5.2.2 Risk Analysis

Risk analysis is the process of assessing the likelihood of an adverse event occurring within the corporate, government, or environmental sector. Risk analysis is the study of the underlying uncertainty of a given course of action and refers to the uncertainty of fore caste cash flow streams, the variance of portfolio or stock returns, the probability of a project's success or failure, and possible future economic states. Risk analysts often work in tandem with forecasting professionals to minimize future negative unforeseen effects.

5.2.3 Overview of Risk Mitigation, Monitoring, Management

Risk id	**1**
Risk Description	**Analyze and correlate large amounts of real-time data from network and security devices to manage external and internal security threats, improve incident response time and compliance reporting.**
Category	**Security Information and Event Management.**
Source	**Cloud Cluster**
Probability	**Low**
Impact	**Low**
Response	**Mitigate**
Strategy	**High Performance**
Risk Status	**Rarely Occurred**

Risk id	**2**
Risk Description	**Improve analysis of application log data to better manage system resource utilization, security issues, and diagnose and preempt production application problems.**
Category	**Application Log Monitoring**
Source	**Main Node**
Probability	**High**
Impact	**High**
Response	**High**
Strategy	**Multi Tenancy**
Risk Status	**Occurred**

Risk id	**3**
Risk Description	**Monitor and analyze network traffic to detect, identify, and report on suspicious activity or intruders.**
Category	**Network Intrusion Detection**
Source	**Name Node**
Probability	**Low**
Impact	**Low**
Response	**Mitigate**
Strategy	**High Scalability**
Risk Status	**Identified**

Risk id	**4**
Risk Description	**Use pattern/anomaly recognition on larger volumes and variety of data to detect and prevent fraudulent activities by external or internal parties.**
Category	**Fraud Detection**
Source	**Cloud Server**
Probability	**Low**
Impact	**Low**
Response	**Mitigate**
Strategy	**Easy Data Ingestion**
Risk Status	**Identified**

Risk id	**5**
Risk Description	**Improve risk assessment and associated scoring by building sophisticated machine learning models on Cloud that can take into account hundreds or even thousands of indicators.**
Category	**Risk Modeling**
Source	**Cloud Cluster**
Probability	**High**
Impact	**High**
Response	**Continuous**
Strategy	**Existing Application Work.**
Risk Status	**Occurred**

5.3 Project Schedule

The project schedule is the tool that communicates what work needs to be performed, which resources of the organization will perform the work and the time frames in which that work needs to be performed. The project schedule should reflect all of the work associated with delivering the project on time.

5.3.1 Project Task Set

Project development is the end-to-end process of conceptualizing and delivering a project given a set of resources and constraints. This typically involves following stages.

- **Task 1: Requirement Gathering**
- **Task 2: Literature Survey**
- **Task 3: Mathematical Modeling**
- **Task 4: Feasibility testing**
- **Task 5: UML Diagrams**

- **Task 6: Database Design**
- **Task 7: GUI Design**
- **Task 8: Functionality Implemented**
- **Task 9: Testing**
- **Task 10: Reporting**

5.3.2 Timeline Chart

Month	**Plan**
June	Domain Selection
July	Title Finalization
August	Technology Selection

Month	**Plan**
September	Literature Survey
October	Presentation,UML
November	Report Writing
December	Paper Writing
January	Module1(Data Acqusition)
Febuary	Module2(Data Filtering)
March	Module3(Data Analysis)
April	Module4(Prediction)
May	Module5(Data Storage)
June	Final Project Execution

5.4 Team Organization

Proper project team organization is one of the key constraints to project success. If the project has no productive and well-organized team, there's an increased probability that this project will be failed at the very beginning because initially the team is unable to do the project in the right manner. Without right organization of teamwork, people who form the team will fail with performing a number of specific roles and carrying out a variety of group/individual responsibilities. Hence, when you plan for a new project, first you must take care of the best project team organization through team building activities.

5.4.1 Team Structure

Team structure refers to the composition of an individual team or of a multi-team system. Team structure is an integral part of the teamwork

process.The different roles played by the team members contributes in the requirement gathering, development, testing and deployment of project.

1. **Team Leader:**

   A team leader is someone who provides direction, instructions and guidance to a group of individuals, who can also be known as a team, for the purpose of achieving a certain goal. An effective leader will know her team members' strengths, weaknesses and motivations.

2. **Team Member:**

   A successful project management team takes preparation and planning. Project team members are persons who work on one or multiple stages of a project. Team member roles will vary depending on the individual project. Team members have a variety of roles to contribute toward achieving the project's aims and goals.

3. **Analyzer:**

   Collect the necessary information required to start projects .Develop project strategy plans based on logical framework approaches.Create and manage documentation and reports for projects.Verify data and information and analyze it to suit the direction of a project.

4. **Test Manager:**

   The role of the software test manager is to lead the testing team. Test Manager plays a central role in the Team.The Test Manager takes full responsibility for the project's success. The role involves quality test advocacy, resource planning management, and resolution of issues that impede the testing effort.

5.4.2 Management, Reporting and Communication

- **Management:**

  Teamwork in the workplace is an important factor for project success. As a result, developing an effective project team is one of the

primary responsibilities of a project manager.Teamwork is important because it creates human synergy. It amplifies the results of each member of your team such that the overall result is greater than the individual contributions made by each member.

- **Reporting:**

  A project management report is a summary overview of the current status of the project. It is a formal record of the state of a project at a given time. ... It is provided to all project stakeholders to help keep them up to date on the progress of the project and any pressing challenges the project may be facing.

- **Communication:**

  Successful project management communication is about being there for everyone, being in touch with the real challenges of the project, understanding the real issues within the team who must deliver the project as well as understanding the issues of the sponsors who the team delivers the project for.

Chapter 6

PROJECT IMPLEMENTATION

6.1 Overview of Project Modules

This Chapter focuses on all modules that are included in this project.There are total 5 modules as follows:

1. **Data Acqusition : This Module import the Data from the csv file which will be the specimen answer.This data will be use for matching the user answer .In this after importing the data, it is processed for further filteration process.**

2. **Data Filtering : This Module helps in filtering the data i.e saving Stopwords in Database and then Redistribution of Data takes place for matching the data with specimen data.**

3. **Data Analysis :This modules helps in matching the data of user with specimen answer.Semantic analysis is used in this module for the comparison purpose.In this the data received after removing stopwords is matched with specimen answer.**

4. **Prediction :In this module data is predicted on the basis of algorithms and analysis techniques.User data and specimen data are compared in this module. After comparing this module predicts required output.**

5. **Datact Storage.A : This Module stores the required information regarding of project aspects.This module helps in retrieval of result information, adding new data to the storage.In this project MongoDb database is used for the predefined data storage.**

6.2 Tools and Technologies Used

1. **Pyhton**
2. **Anaconda**
3. **MongoDB**
4. **Putty**
5. **FeedParser**
6. **Pymongo**
7. **Pycharm**

6.3 Algorithm Details

6.3.1 Measuring Similarity

Measure of similarity can be qualitative and/or quantitative. In qualitative, the assessment is done against subjective criteria such as theme, sentiment, overall meaning, etc. In the quantitative, numerical parameters such as length of the document, number of keywords, common words, etc. are compared. The process is carried out in two steps, as mentioned below:

- **Vectorization: Transform the documents into a vector of numbers. Following are some of the popular numbers (measures): TF (Term Frequency), IDF (Inverse Document Frequency) and TF*IDF.**

- **Distance Computation: Compute the cosine similarity between the document vector. As we know, the cosine (dot product) of the same vectors is 1, dissimilar/perpendicular ones are 0, so the dot product of two vector-documents is some value between 0 and 1, which is the measure of similarity amongst them.**

6.3.2 Vectorization

Characterize each text as a vector. Each text has some common and some uncommon words compared to each other. To account for all possibilities, a word set is formed which consists of words from both the documents. There are various methods by which words can be vectorised, meaning, converted to vectors (array of numbers). A few of the prominent ones are explained below.

Frequency Count Method:

A simplest way to create the vectors is to count number of times each word from the common word set, occurs in individual document.FreqDist counts the number of occurrence of a word in the given text. So, in the above code snippet text1_count_dict has word-count pairs of all the words from the common word_set, along with their individual counts. Following table shows few words with their frequencies:

```
from nltk.probability import FreqDist

word_set = set(text1).union(set(text2))

freqd_text1 = FreqDist(text1)
text1_count_dict = dict.fromkeys(word_set,0)
for word in text1:
    text1_count_dict[word] = freqd_text1[word]

freqd_text2 = FreqDist(text2)
text2_count_dict = dict.fromkeys(word_set,0)
for word in text2:
    text2_count_dict[word] = freqd_text2[word]
```

FreqDist counts the number of occurrence of a word in the given text. So, in the above code snippet text1_count_dict has word-

count pairs of all the words from the common word_set, along with their individual counts. Following table shows few words with their frequencies:

	westbound	whether	windows	workers	worse	would	years
text1	1		1			1	
text2	1	1		1	1	1	1

These vectors, in a crude way, represent their respective texts and similarity can be assessed amongst them. This is the 'Containment Ratio' method mentioned above. TF-IDF is much better measure to represent a document.

TF-IDF Method:

TF is document specific. It is a way to score the importance of words (or "terms") in a document based on how frequently they appear. If a word appears frequently in a document, it's important, it gets a high score. Although it is easy to compute, it is ambiguous ('green' the colour and 'green' the person's name is not differentiated).

```
# TF calculations
freqd_text1 = FreqDist(text1)
text1_length = len(text1)
text1_tf_dict = dict.fromkeys(word_set,0)
for word in text1:
    text1_tf_dict[word] = freqd_text1[word]/text1_length

freqd_text2 = FreqDist(text2)
text2_length = len(text2)
text2_tf_dict = dict.fromkeys(word_set,0)
for word in text2:
    text2_tf_dict[word] = freqd_text2[word]/text2_length
```

IDF is for the whole collection. It is a way to score how many times a word occurs across multiple documents. If a word appears in many documents, it's not a unique identifier, thus gets a lower score.

```
# IDF calculations
text12_idf_dict = dict.fromkeys(word_set,0)
text12_length = 2 # 2 documents
for word in text12_idf_dict.keys():
    if word in text1:
        text12_idf_dict[word] += 1
    if word in text2:
        text12_idf_dict[word] += 1

import math
for word, val in text12_idf_dict.items():
    text12_idf_dict[word] = 1 + math.log(text12_length/(float(val)))
```

```
# TF-IDF Calculations
text1_tfidf_dict = dict.fromkeys(word_set,0)
for word in text1:
    text1_tfidf_dict[word] = (text1_tf_dict[word])*(text12_idf_dict[word])

text2_tfidf_dict = dict.fromkeys(word_set,0)
for word in text2:
    text2_tfidf_dict[word] = (text2_tf_dict[word])*(text12_idf_dict[word])
```

TFIDF of a word = (TF of the word) * (IDF of the word)

Word Embedding Method:

```
from gensim.models.doc2vec import TaggedDocument

taggeddocs = []
doc1 = TaggedDocument(words=text1, tags=[u'NEWS_1'])
taggeddocs.append(doc1)
doc2 = TaggedDocument(words=text2, tags=[u'NEWS_2'])
taggeddocs.append(doc2)

# build the model
model = gensim.models.Doc2Vec(taggeddocs, dm=0, alpha=0.025, size=20,
min_alpha=0.025, min_count=0)

# training
for epoch in range(80):
    if epoch % 20 == 0:
        print('Now training epoch %s' % epoch)
    model.train(taggeddocs)
    model.alpha -= 0.002  # decrease the learning rate
    model.min_alpha = model.alpha  # fix the learning rate, no decay
```

Once the words in the text are vectorised, the similarity score between them is nothing but the 'distance' between them.

6.3.3 Distance Computation

Following are the steps to compute the similarity of two texts using TF-IDF Method. It is computed using the dot product of given vectors v1 and v2.

```
# Compute Cosine distance
v1 = list(text1_tfidf_dict.values())
v2 = list(text2_tfidf_dict.values())
similarity = 1 - nltk.cluster.cosine_distance(v1,v2)
print("Similarity Index: {:4.2f} %".format(similarity*100))
```

For the given two news items the similarity score came to about 72.62 %.

```
similarity = model.n_similarity(text1,text2)
print("Similarity Index: {:4.2f} %".format(similarity*100))
```

In case of Word Embedding method, the Doc2Vec model itself can compute similarity of given texts. For the given two news items the similarity score came to about 79.06 %.

Chapter 7

SOFTWARE TESTING

7.1 Type of Testing

- **Black Box Testing :**

  In black box testing, the internal working of the system is tested as opposed to its internal structure. In this case, the tester is required to know the working of the software only, not how the working is carried out. The main intention of this testing is to check whether the system provides the expected output with the given input or not.

- **White Box Testing :**

  White box testing takes into account the internal mechanism of the system. It is used to verify whether the source code provides the expected results or not. Two methods of this testing are Unit Testing and Integration Level Testing.

- **Unit Testing :**

  Unit testing is carried out by the tester to check source code, different program modules to determine if they are properly written for the dissertation. This approach is used to find out the bug in the modules individually. All the different modules that form the system are checked individually rather than integrating them.

- **Integration Testing :**

Integration testing is the phase in which all the different components or modules in the software are combined and they are tested. This test is actually done to check whether the integration of the different modules is done properly or not. It occurs after the unit testing phase. Also it takes the input as the modules that are tested individually in the Unit Testing and integrate them and applies the methods that are defined for the testing. This provides the output whether the system is ready or not.

7.2 Test cases and Test Results

Test case ID	Test case summary	Prerequisites	Procedure	Expected Results	Actual Results	Status
1	Data Acquisition	To Connect the Web Link and Acquire data	API Class Object	Object Created	Object Created	Pass
			API Function Call	Function Calling	Function Called	Pass
			API Response	Response Received	Response received from the object	Pass
			Data feed collectionl	All data is collected in runtime object	Data collected	Pass
2	System	Check that the systems performance does not degrade	Performance	System performance should not degrade	System performance does not degrade	Pass

3	Analysis	Check that all the functions are working properly according to plan	Parameters	The output should be according to only selected parameters	The output is according to selected parameters	Pass
			Invalid Selection	If invalid parameter that are not available an error should come	Display of error message	Pass
			Erroneous data	Discard Erroneous data	Erroneous data Discarded	Fail
			Historical data	Collect all the Historical Data	Data collected	Pass
4	Database	Proper database queries are fired on the database	Retrieval of result information	Should provide proper authentication	Authentication is provided	Pass
			Adding new data	New Data persisted in mongo DB.	Data is uploaded in the expected time range	Pass

5	Application	Check that the application works properly with all system configuration and response time must not degrade	Generation of Analysis	Result generation should be done within API response time.	The output is in time limit.	Pass
			Adding new data	New Dara persisted in mongo DB.	Data is uploaded in the expected time range	Pass
			Values displayed on screen	The values displayed after analysis should be correct.	The values are correctly displayed	Pass

Table 7.1: Test Case table

Chapter 8

RESULTS

8.1 Outcomes

8.2 Screen Shots

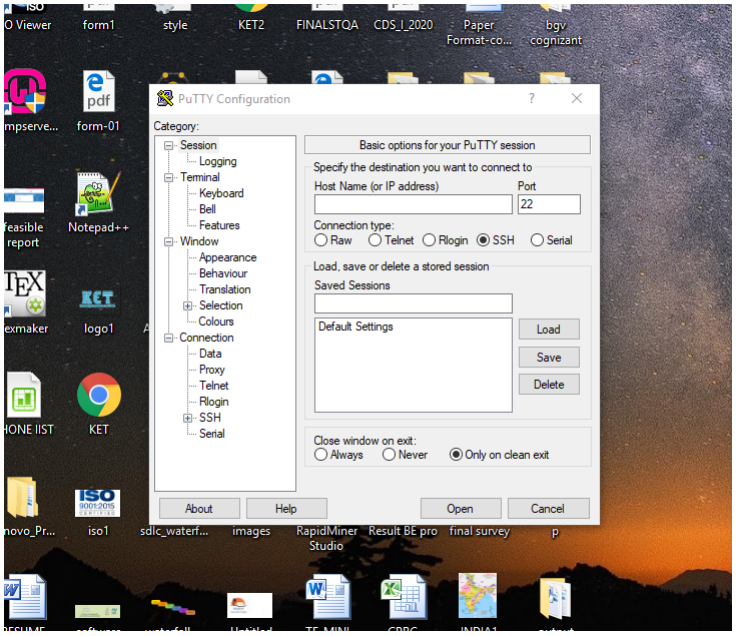
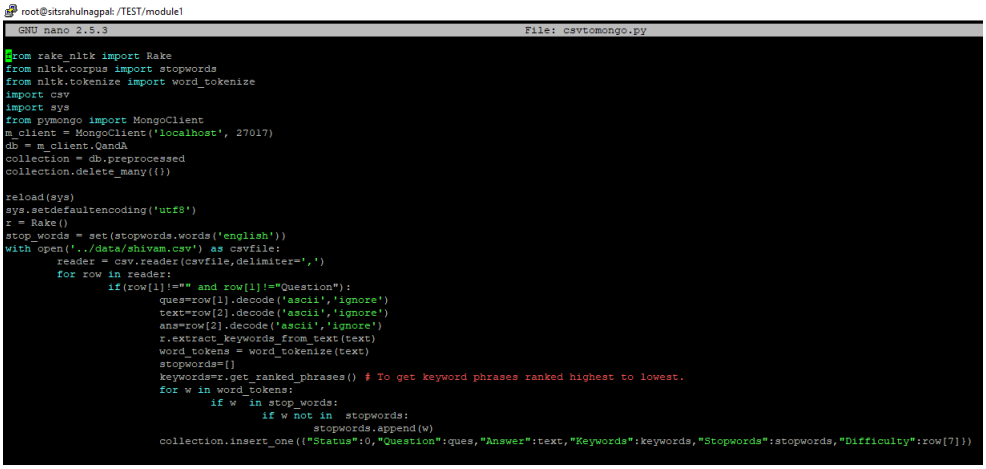


Figure 8.1: Putty Login

Fig 8.1 shows Putty Configuration. It is a free and open-source terminal emulator, serial console and network file transfer application.It can also connect to a serial port.



```
root@sitsrahulnagpal: /TEST/module1
 GNU nano 2.5.3                                File: csvtomongo.py

from rake_nltk import Rake
from nltk.corpus import stopwords
from nltk.tokenize import word_tokenize
import csv
import sys
from pymongo import MongoClient
m_client = MongoClient('localhost', 27017)
db = m_client.QandA
collection = db.preprocessed
collection.delete_many({})

reload(sys)
sys.setdefaultencoding('utf8')
r = Rake()
stop_words = set(stopwords.words('english'))
with open('../data/shivam.csv') as csvfile:
        reader = csv.reader(csvfile,delimiter=',')
        for row in reader:
                if(row[1]!="" and row[1]!="Question"):
                        ques=row[1].decode('ascii','ignore')
                        text=row[2].decode('ascii','ignore')
                        ans=row[2].decode('ascii','ignore')
                        r.extract_keywords_from_text(text)
                        word_tokens = word_tokenize(text)
                        stopwords=[]
                        keywords=r.get_ranked_phrases() # To get keyword phrases ranked highest to lowest.
                        for w in word_tokens:
                                if w  in stop_words:
                                        if w not in  stopwords:
                                                stopwords.append(w)
                        collection.insert_one({"Status":0,"Question":ques,"Answer":text,"Keywords":keywords,"Stopwords":stopwords,"Difficulty":row[7]})
```

Figure 8.2: Code for Seperation of Stop Words and Keywords

Fig 8.2 shows the coding related part for seperation of stop words and keywords.

Figure 8.3: Output of Stop Words and Keywords

Fig 8.3 Shows the expected output of stop words and keywords which are seperated as shown.

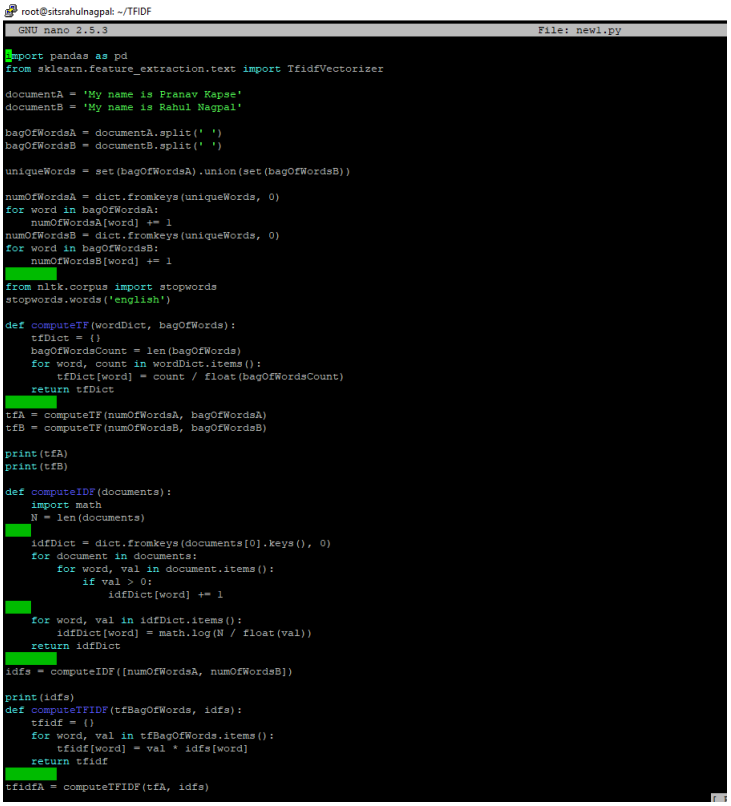


Figure 8.4: Code for Frequency of each word

Fig 8.4 Shows the coding related part for calculating the frequency of each word.

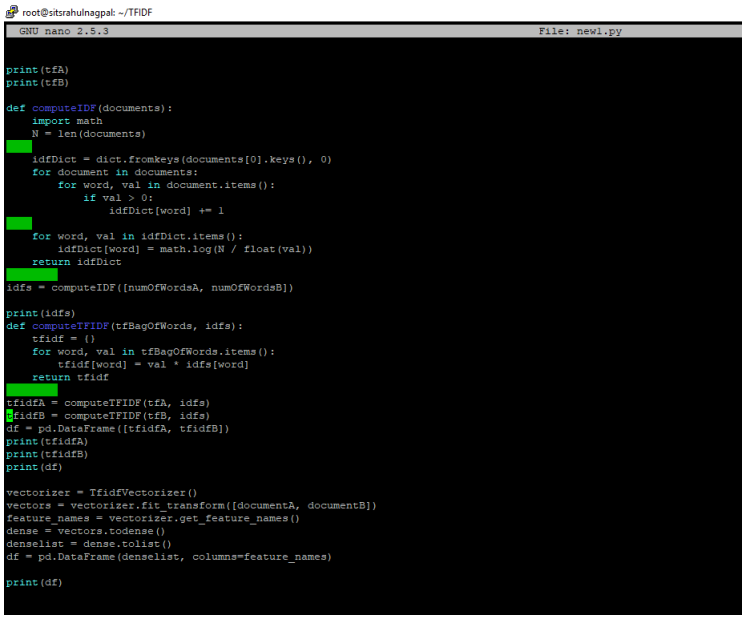


Figure 8.5: Code for Frequency of each word

Fig 8.5 Shows the coding related part for calculating the frequency of each word.

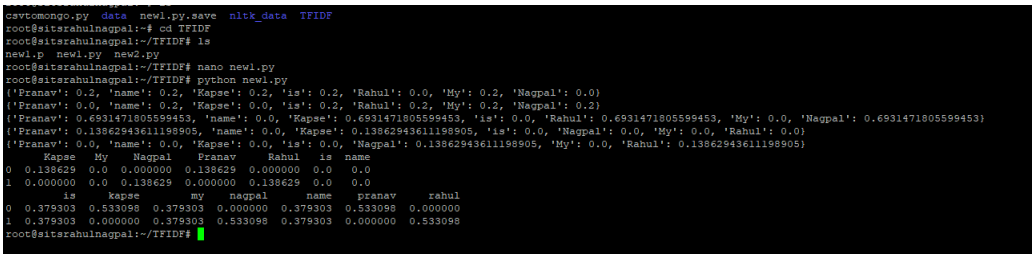


Figure 8.6: Output of Frequency

Fig 8.6 Shows the expected output of frequency part.It shows each word's frequency as shown.

```
root@sitsrahulnagpal: ~/TFIDF
  GNU nano 2.5.3                                                File: new2.py

import nltk, string
from sklearn.feature_extraction.text import TfidfVectorizer

#nltk.download('punkt') # if necessary...


stemmer = nltk.stem.porter.PorterStemmer()
remove_punctuation_map = dict((ord(char), None) for char in string.punctuation)

def stem_tokens(tokens):
    return [stemmer.stem(item) for item in tokens]

'''remove punctuation, lowercase, stem'''
def normalize(text):
    return stem_tokens(nltk.word_tokenize(text.lower().translate(remove_punctuation_map)))

vectorizer = TfidfVectorizer(tokenizer=normalize, stop_words='english')

def cosine_sim(text1, text2):
    tfidf = vectorizer.fit_transform([text1, text2])
    return ((tfidf * tfidf.T).A)[0,1]

ans="BE student should attend regular college"
ans1="regular college should be attended by rahul"
#ans3="vaseline is a trademark of unilever"

ans2="everyone from BE should enjoy attending regular college"

print str(cosine_sim(ans,ans1)*100)+" % Similar"
print str(cosine_sim(ans,ans2)*100)+" % Similar"
```

Figure 8.7: Code for percentage Similarity

Fig 8.7 Shows the coding related part for calculating and comparing user and specimen answers.

```
1  0.000000  0.0  0.138629  0.000000  0.138629  0.0   0.0
         is     kapse        my    nagpal      name    pranav
0  0.379303  0.533098  0.379303  0.000000  0.379303  0.533098  0
1  0.379303  0.000000  0.379303  0.533098  0.379303  0.000000  0
root@sitsrahulnagpal:~/TFIDF# nano new2.py
root@sitsrahulnagpal:~/TFIDF# python new2.py
60.29748160380572 % Similar
51.01490193104813 % Similar
root@sitsrahulnagpal:~/TFIDF#
```

Figure 8.8: Output of Similarity

Fig 8.8 Shows the expected output of similarity.It shows the similarity between user and specimen answers in terms of percentage.

Chapter 9

CONCLUSIONS

9.1 Conclusions

The report gives an over-view of the actual project. There were various aspects of the projects which were discussed briefly in the past chapters. The first chapter gave the introduction to the topic on and the gist of the project. The literature survey conducted on Six papers in support of the topic discussed in the first chapter is described. The outline the SRS document of the project is also presented which adds more clarity to the project scope and the technical details. Finally, the diagramatical explanation of the things which could be collected up to this point were also illustrated.

9.2 Future Work

With the help of this system we can check theoretical papers with faster speed and also we will be able to generate accurate result. And this leads in contribution to Green Computing.We are just developing this system for theoretical question and answers, so in future we will researching for Numerical questions and various types of different questions.This will help in different educational fields.

9.3 Applications

1. **Online portal Exam**
2. **Business Intelligence (BI) (Survey)**
3. **Marketing Strategy**
4. **Input for Survey Reports**

REFERENCES

[1] **Yunshi Lan, Shuohang Wang, and Jing Jiang 2019"Knowledge Base Question Answering With a Matching-Aggregation Model and Question-Specific Contextual Relations".**

[2] **Vitor Rolim, Maverick Ferreira, Anderson Pinheiro Cavalcanti 2019" Identifying students' weaknesses and strengths based on online discussion using topic modeling".**

[3] **Xue Han, Yabin Dang, Lijun Mei, Yanfei Wang, Shaochun Li, Xin Zhou 2019"A Novel Part of Speech Tagging Framework for NLP based Business Process Management".**

[4] **Diellza Nagavci Mati ,Jaumin Ajdari ,Bujar Raufi ,Mentor Hamiti ,Besnik Selimi 2019"A Systematic Mapping Study of Language Features Identification from Large Text Collection".**

[5] **Durham Abric, Oliver E. Clark, Matthew Caminiti, Keheliya Gallaba, and Shane McIntosh 2019"Can Duplicate Questions on Stack Overflow Benefit the Software Development Community?".**

[6] **Han van der Aa,Henrik Leopold,Jan Mendling,Josep Carmona 2018"Challenges and Opportunities of Applying Natural Language Processing in Business Process Management".**

Appendix A

Problem statement feasibility assessment using, satisfiability analysis and NP Hard,NP-Complete or P type using modern algebra and relevant mathematical models.

This project is in the domain of Machine Learning. The problems in the Machine Learning domain depend on various factors such as data size, hyperparameters, etc. The hyperparameters are more deciding than the data size in consideration. So, given proper tuning of the hyperparameters, the Deep Learning problems are approachable in polynomial time and hence are known to be NP problems. The majority of the time is spent in preprocessing of the dataset, followed by training of the models. The training time can be reduced by using GPU-enabled ML libraries.

Hence, it can be concluded that the project model falls in P type.

Appendix B : Publication

SINHGAD INSTITUTE OF TECHNOLOGY AND SCIENCE, PUNE

Sinhgad Institutes

Presents

International Conference on Future Intelligence in Science & Technology

31st May 2020

Certificate

This is to certify that Prof./Dr/Mr./Ms. Aditya Singh

of Savitribai Phule Pune University has participated / presented a paper entitled

Survey paper on text segmentation with feature similarity for exam assessment using Machine Learning. at

"International Conference on Future Intelligence in Science & Technology " held on 31st May 2020 at Sinhgad Institute of Technology and Science, Pune

Prof. S.A. Kulkurni
Organizing Secretory, SITS FIST 2020

Dr. R. S. Prasad
General Chair, SITS FIST 2020

Yashika Publications

Scopus

SINHGAD INSTITUTE OF TECHNOLOGY AND SCIENCE, PUNE

Sinhgad Institutes

Presents

International Conference on Future Intelligence in Science & Technology

31st May 2020

Certificate

This is to certify that Prof./Dr/Mr./Ms. Pranav Kapse

of Savitribai Phule Pune University *has participated / presented a paper entitled*

Survey paper on text segmentation with feature similarity for exam assessment using Machine Learning *at*

"International Conference on Future Intelligence in Science & Technology " held on 31st May 2020 at Sinhgad Institute of Technology and Science, Pune

Prof. S.A. Kulkurni
Organizing Secretory, SITS FIST 2020

Dr. R. S. Prasad
General Chair, SITS FIST 2020

Yashika Publications Scopus

SINHGAD INSTITUTE OF TECHNOLOGY AND SCIENCE, PUNE

Sinhgad Institutes

Presents

International Conference on Future Intelligence in Science & Technology

31st May 2020

Certificate

This is to certify that Prof./Dr/Mr./Ms. Rahul Nagpal

of Savitribai Phule Pune University *has participated / presented a paper entitled*

Survey paper on text segmentation with feature similarity for exam assessment using Machine Learning *at*

"International Conference on Future Intelligence in Science & Technology " held on 31st May 2020 at Sinhgad Institute of Technology and Science, Pune

Prof. S.A. Kulkurni
Organizing Secretory, SITS FIST 2020

Dr. R. S. Prasad
General Chair, SITS FIST 2020

Yashika Publications Scopus

SINHGAD INSTITUTE OF TECHNOLOGY AND SCIENCE, PUNE

Sinhgad Institutes

Presents

International Conference on Future Intelligence in Science & Technology

31st May 2020

Certificate

This is to certify that Prof./Dr./Mr./Ms. Shivam Pandey

of Savitribai Phule Pune University *has participated / presented a paper entitled*

Survey paper on text segmentation with feature similarity for exam assessment using Machine Learning *at*

"International Conference on Future Intelligence in Science & Technology " held on 31st May 2020 at Sinhgad Institute of Technology and Science, Pune

Prof. S.A. Kulkurni
Organizing Secretory, SITS FIST 2020

Dr. R. S. Prasad
General Chair, SITS FIST 2020

Yashika Publications

Scopus

Appendix C : Certificate

Project Competition Certificates

MIT ADT UNIVERSITY PERSONA FEST 20

PRISM OF EXPRESSION

The Biggest Techno - Cultural Extravaganza

MIT-ADT UNIVERSITY PUNE, INDIA

We, the Trustee and members of the Managing Committee of
the MIT Art, Design & Technology University, Pune have great pleasure in presenting this

Certificate of Participation

Mr./Ms. Pranav Kapse

of SITS NARHE has participated in the Innovative Ideas event

at "Persona Fest 2020" organized by

MIT Art, Design & Technology University, Loni Kalbhor, Pune, India,

held from 12th -15th, February 2020.

Prof. Dr. Kishore Ravande
Chairperson - Persona Fest 2020
MIT-ADT University, Pune, India

Prof. Dr. Mangesh T. Karad
Executive President
MIT-ADT University, Pune, India

Prof. Dr. Vishawanath D. Karad
Founder & President
MIT-ADT University, Pune, India

MIT ADT UNIVERSITY PERSONA FEST 20

PRISM OF EXPRESSION

The Biggest Techno - Cultural Extravaganza

MIT-ADT UNIVERSITY PUNE, INDIA

We, the Trustee and members of the Managing Committee of
the MIT Art, Design & Technology University, Pune have great pleasure in presenting this

Certificate of Participation

Mr./Ms. Rahul Nagpal

of SITS NARHE has participated in the Innovative Ideas event

at "Persona Fest 2020" organized by

MIT Art, Design & Technology University, Loni Kalbhor, Pune, India,

held from 12th -15th, February 2020.

Prof. Dr. Kishore Ravande
Chairperson - Persona Fest 2020
MIT-ADT University, Pune, India

Prof. Dr. Mangesh T. Karad
Executive President
MIT-ADT University, Pune, India

Prof. Dr. Vishawanath D. Karad
Founder & President
MIT-ADT University, Pune, India

MIT ADT UNIVERSITY
PERSONA FEST 20

PRISM OF
EXPRESSION
The Biggest Techno - Cultural Extravaganza

MIT-ADT UNIVERSITY
PUNE, INDIA

We, the Trustee and members of the Managing Committee of
the MIT Art, Design & Technology University, Pune have great pleasure in presenting this

Certificate of Participation

Mr./Ms. Shivam Kumar

of SITS NARHE has participated in the Innovative Ideas event

at "Persona Fest 2020" organized by

MIT Art, Design & Technology University, Loni Kalbhor, Pune, India,

held from 12th -15th, February 2020.

Prof. Dr. Kishore Ravande
Chairperson - Persona Fest 2020
MIT-ADT University, Pune, India

Prof. Dr. Mangesh T. Karad
Executive President
MIT-ADT University, Pune, India

Prof. Dr. Vishawanath D. Karad
Founder & President
MIT-ADT University, Pune, India

Appendix D : Plagiarism