Question Answer Generator Using NLP

SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS FORTHE AWARD OF DEGREE OF

BACHELOR OF TECHNOLOGYIN

INFORMATION TECHNOLOGY

Submitted By: Group 4

Project Team Members

Adarsh Kumar [1803814] Akhil Kathuria [1803816]
Ayush Kumar [1803821] Gautam Taneja [1803827]

Under the supervision of

Guide Name: Mrs. Reeta Bhardwaj

Designation: Associate Professor

DEPARTMENT OF INFORMATION TECHNOLOGY

DAV INSTITUTE OF ENGINEERING & TECHNOLOGY

Jalandhar – 144008

November 2021

ABSTRACT

Learning through the internet becomes popular that facilitates learners to learn anything, anytime, anywhere from the web resources. Assessment is most important in any learning system. An assessment system can find the self-learning gaps of learners and improve the progress of learning. The manual question generation takes much time and labour. Therefore, automatic question generation from learning resources is the primary task of an automated assessment system. This paper presents a survey of automatic question generation and assessment strategies from textual and pictorial learning resources. The purpose of this survey is to summarize the state-of-the-art techniques for generating questions and evaluating their answers automatically.

There are lots of examination portals that are deployed over several servers which are used to conduct online examination for various purposes among which some may include conducting a test for entrance examinations, or olympiads at national and international level and while some portals are designed to conduct a test for placement purposes. But what we have seen is that mostly all the portals are designed to conduct tests that contain multiple choice questions. Here our aim is not to work on the technology that is already existing, rather some technology that is veryrare. Here we talk of the descriptive online examination system. Multiple choice questions are easy to deal as they have a question, a few options and a field in the same question that stores the correct option in the database. While in the case of descriptive questions it is not so. It brings in or uses the concepts of Natural Language Processing or NLP to assign marks to answers. Answers are nothing but strings and the job of the model is to do some operations on the answer string such that it can assign the correct marks to answers written by the examinee. The data is basically collected from a descriptive online examination system. Further, it is analyzed and the designed model assigns accurate marks to the answers for the question.

An automatic system is proposed to generate different kinds of questions and answers from the input text. Question answer generation systems have been an interesting field of research for over decades. From generating questions for educational purposes to preparing answers to questions that could be asked in a legal proceeding, the purpose of question answer

ACKNOWLEDGEMENT

Achievement of this task would have been impossible without the help and constant guidance of certain people. We hereby take the opportunity of thanking a few people for their support, guidance and encouragement for the successful completion of our project.

I would like to express my special thanks of gratitude to our principal, Dr. Manoj Kumar who gave us the golden opportunity to do this wonderful project, which also helped us in doing a lot of research and we came to know about so many new things.

I would like to thank Dr. Dinesh Kumar, Hon'ble H.O.D.(IT), DAVIET. It's due to his efforts that we are able to experience a platform like DAVIET for our undergraduate degree.

I would like to express our sincere thanks to Mrs. Reeta Bhardwaj, who provided us with valuable guidance and support. She provided us with invaluable advice and helped us in difficult periods. Her motivation contributed tremendously to the successful completion of the project.

I would like to thank our group members Mr. Ayush Kumar, Mr. Gautam Taneja and Mr. Adarsh Kumar for making such a valuable effort and making good team spirit among all the members.

Each of us would also like to thank our respective parents who have been a constant motivation for us during the course of this project.

TABLE OF CONTENTS

List of Figures

List of Tables

[Chapter -1] Introduction

1.1 Project Overview

An examination is an effective method for testing one's understanding of any topic. Answering questions helps to improve the process of learning and is an integral part of academics. This project falls under internet-based industry automation. This project aims to build an automated examination system using machine learning, natural language processing (NLP), python environment, flask framework, and web technologies to provide an inexpensive alternative to the current examination system. Automatic question generation for textual inputs is valuable in academics where answering questions helps students to learn and improve their understanding of their field of study. Automatic question generation finds application in dialog systems or virtual assistants where asking questions is an important part of interactions between humans and machines. Apart from the education domain, Quiz Question Generation for the entertainment industry is also gaining importance. There are millions of people playing quiz games as a fun activity and at times serious money-making options. This requires the automated generation of millions of questions each day manually. The Automated Question Generation system aims at automating this process of question generation by providing an end-to-end system that will take the data which may be in the form of comprehension (sentence, chapters) as input and provide a question as output. Conducting examination and answer sheet evaluation are hectic testing tools for assessing academic achievement, integration of ideas, and ability to recall, but are expensive, resourceful, and time-consuming to generate questions and evaluate responses manually. Manual evaluating of answer sheets take up a significant amount of instructors' valuable time and hence is an expensive process. We implement a model to automatically generate questions with their respective answers and assess user responses.

This youth can be utilized for overall development and prosperity of the nation through the ideological education system. The comprehensive concept of education aims at making individuals capable of becoming more responsible, productive and having critical understanding of social, economic and political aspects. It is through the education process that students learn to analyze their experiences, to critically examine the various aspects of life, lean to doubt, to question, to investigate and to think independently. The knowledge, skills and attitudes of the individuals should be modified in the due course of the education. And this modification should be assessed by using comprehensive means of evaluation and examination. The examination system of India is criticized from time to time due its defects. The University Education Commission (1948) was quite vocal regarding the defects of the examination system of our country. After this Secondary Education Commission (1952-53) and Kothari Commission highlighted the defects of examination system and

gave recommendations for the improvement of this system. But even after so many reforms, the prevalent examination system still has many defects as listed below:

1.2 Objectives

The main objective is that this tool will be a great help to the People using AI using NLP to manage the Test-taking task conveniently. The manual question generation takes much time and labour. Therefore, automatic question generation from learning resources is the primary objective of this automated assessment system.

The objectives of our project are as follows :

1. To develop both subjective and objective question answers.
2. To implement automatic computation of answers..
3. To implement automatic result evaluation.
4. To test and compare the proposed system with existing system.
5. To incorporate the transparency in test system.

1.3 Problem Formulation

AI-based examination systems add a step towards creating an effortless conducting of examinations. The working of the AI examination system ensures the workload management of professors and teachers of creating and marking the question paper manually. It also helps in better evaluation of students based on their knowledge.

It generates random questions for every student which makes the system even more precise and all the users will get different question sets so that the chances of cheating will be minimal. On another hand, it will help to save paper and transportation costs of papers to examination halls. It will help in resource management as this will cut corners on stationery products, it will also have greater efficiency with respect to time as it will produce instantaneous results.

A few problems that are faced in the current system are:

- Teachers have to set question papers manually.
- We are not able to generate different sets of question papers.
- Manual checking takes a lot of time and has chances of human error too.

- Student Record maintenance is manually done.
- Chances of cheating.
- Paper Wastage.

❖ Cheating

Cheating in examination at all levels of education in India is well known and from time to time the media presented such incidences. Cheating is done by students because examination demands memorization of large number of facts and figures and this demand put a lot of stress on the minds of the students. And thus, they use unfair means such as slips, mini photocopies of their help books, Bluetooth headphones concealed under their caps or turbans etc.

❖ Subjectivity

Subjectivity is another major defect of our education system. It occurs at three levels namely first at the level of paper setting by the examiner, second at the level of student who writes answer according to his/her own subjective nature and third at the level of evaluator who is evaluating the answer given by the student. Different persons have different views regarding the answer of the same question. The same answer evaluated by different persons have different marks.

❖ Theoretical in nature

The knowledge, skill and attitude of the students should be assessed to properly evaluate his/her performance in the course, but the present examination system only focus on the knowledge aspect of the students. The students are encouraged to read books and the examination are based on bookish knowledge. The theoretical nature of examination is responsible for the lack of skills even after the completion of degrees with good grades by different doctors, engineers, teachers and other professionals.

❖ Poor Content Coverage by Papers

The examination system is based upon the optional nature of question means a student has the option to answer one question from two or more questions and so on. This optional nature of paper setting is responsible for poor content coverage. The students deliberately ignore major portion of the content and prepared according to the option available.

1.4 Existing System

Offline System:

The offline examination system relies on manual work from printing to transporting the paper to the examination hall, then invigilation and the most tedious task of checking the answer sheets which is a huge mess for any examiner which sometimes leads to resource loss. Also, we hear news about paper leaks and answer sheets being lost in the transporting process.

BENEFITS

ZERO TECHNICAL ISSUES

With offline examination, there will never be any concern of internet or electricity.

LOWER CHANCE OF CYBER FRAUD

When schools conduct the exams at physical centers, invigilators play a vital role in regulating activities and ensuring no fraud or cheating takes place.

DISADVANTAGES

WASTES TIME AND RESOURCES

You need appropriate and distinctive resources, exam centers, furniture and other accessories to conduct offline examination.

IT IS LESS COMFORTABLE FOR STUDENTS

Take a census and more students will agree on the online mode of examinations, considering its accessibility and effectiveness.

Online System:

The online examination system does not feature descriptive questions. They are good at evaluating the answers but they have little to no scope for the descriptive ones and the analysis is not well implemented to get meaningful results. Even the most famous of them just have a simple system of storing the correct options in the database and just matching the correct option with it to calculate the result.

BENEFITS

- EASILY MANAGEABLE

  Online examinations are comparatively manageable than offline exams as a dedicated team handles the entire process, even providing a secure connection and webcam. These facilities ensure that teachers have to be less concerned about fraudulent practices during an online examination.

- ENVIRONMENT FRIENDLY

Online exams are paperless. Thus, it saves paper and, consecutively, the environment. You do not have to worry about printing and distributing papers, saving time for teachers as well.

- COST-EFFECTIVE

For both students and school, online exams are quite affordable as they do not have to travel to a location to give exams. The school also saves money on printing, distribution, and other costs with online examination.

- THEY ARE SAFE

Due to the ongoing pandemic, offline exams are better for students, parents, and teachers alike. From the comfort of their home, they can take exams and graduate to a higher grade without worrying about contagion.

DISADVANTAGES

- THE QUESTIONS NEED TO BE TRICKY

  A student can always take help from the internet or a book to fill their answer sheets while giving exams at home. Thus, examiners must be tricky with their questions. Having multiple sets of question papers that you can rotate within a group of students is one way to go.

- THEY DO NOT GO WELL WITH DESCRIPTIVE QUESTIONS

  Online exams can only incorporate MCQs. If you want to hold a subjective exam, it's wise to choose offline exams.

1.5 Proposed System

In the proposed model we are taking the online examination system to a new level by enabling the examinee to write descriptive answers which will get evaluated on their own i.e. automating the entire offline examination system with the efficiency of computing having no human error involved, this can be done using NLP or Natural Language Processing. The evaluated answers will be stored in the database and they can be viewed anytime and a particular student profile will be maintained for better evaluation of the student.

This will be a huge boost to the online examination system as this will allow it to overcome its biggest con and it will also help the online examination system to stretch its paw even in the half-yearly or annual examination conducted by schools or colleges for evaluating the profile of the student. This will have instant benefits like the system will relieve the burden of the teachers and professors of checking copies and in return, they can be more productive with their time in teaching

things, this will also eliminate biasing in answer script checking and will have leased space for any human error as copies would not be scanned and the entire marks will be allotted according to the way answers are written by the examinee while he was on and there will be little to no space for acquisition, it will help in resource management as this will cut corners on stationery products, it will also have greater efficiency with respect to time as it will produce instantaneous results and will be more secure and reliable.

1.5 Features of the Project

1. Secure login portal: The portal will have two logins, one for the admin and one for the students. Admin's login portal will have authority to set the test and will also have the option to download the result of all the student after they have attempted the test. Student's login portal has the option of attempting test, after the teacher have already set the test.

2. Generating questions and answers: Random questions and answers will be generated automatically for every user using AI. NLP is used for generating question with material (Text) provided by the teacher and it will generate answer as well from the same material with the corresponding questions. It uses various processes like tokenization, morphology, part of speech tagging, distributional similarity, chunking, name entity recognition.

3. Evaluation: Results of student performance will be available just after the test submission.
    - User Interface design is very simple and easy to understand. Only Authenticated user are able to access website. Ajax is used to provide real-time experience to user.
    - Less number of mouse clicks and keystrokes are required to accomplish this task.
4. Maintaining record: When all the users have submitted their test, a CSV file containing records of users will be generated.

[Chapter -2] Requirement Analysis

2.1 Feasibility Study

The feasibility study includes consideration of all the possible ways to provide a solution to the given problem. The proposed solution should satisfy all the user requirements and should be flexible enough so that future changes can be easily done based on the future upcoming requirements. A feasibility study aims to objectively and rationally uncover the strengths and weaknesses of an existing business or proposed venture, opportunities and threats present in the natural environment, the resources required to carry through, and ultimately the prospects for success. In its simplest terms, the two criteria to judge feasibility are cost required and value to be attained.

I. Economic Feasibility:

Yes, this project is Economical Feasible.

This is a very important aspect to be considered while developing a project. We decided on the technology based on the minimum possible cost factor.

- AI hardware and software cost has to be borne by our team.
- Overall, we have estimated that the benefits the organization is going to receive from the proposed system will surely overcome the initial costs and the later on running cost for the system.

Hardware requirements are not that high but the effort required to make the software application adds significantly to the cost of developing this application. In Economic Feasibility study cost and benefit of the project is analysed. Means under this feasibility study a detail analysis is carried out what will be cost of the project for development which includes all required cost for final development like hardware and software resource required, design and development cost and operational cost and so on. After that it is analysed whether project will be beneficial in terms of finance for organization or not.

II. Technical Feasibility:

Yes, this project is Technical Feasible. In our opinion the team required to make this software application should be highly skilled and should have the knowledge of technologies mentioned

This included the study of function, performance, and constraints that may affect the ability to achieve an acceptable system. For this feasibility study. We studied complete functionality to be

provided in the system, as described in the System Requirement Specification (SRS). and checked if everything was possible using a different type of frontend and backend platforms.

III. Operational Feasibility:

Yes, this project is Operational Feasible.

No doubt the proposed system is fully GUI based that is very user friendly and all inputs to be taken all self-explanatory even to a layman. In Operational Feasibility degree of providing service to requirements is analysed along with how much easy product will be to operate and maintenance after deployment. Along with these other operational scopes are determining usability of product, determining suggested solution by software development team is acceptable or not etc.

2.2 Software Required Specification Document

1. Introduction

A software requirements specification (SRS) is a description of a software system to be developed. It lays out functional and non-functional requirements and may include a set of use cases that describe user interactions that the software must provide.

1.1 Purpose

The purpose of our project is to generate the question-answers automatically from a given text and the main focus for the project is on the teachers and students. It provides teacher with a platform that can help them to generate questions for a test and automatically evaluate the result of the student.

1.2 Scope

- The best thing about this software is that of generating Objective Type and Subjective Type questions in an easy way.

1.3 Definitions, Acronyms, and Abbreviations.

Question Answer Generator:
Question Answer generator software is provided for both objective type and subjective type questions.

Abbreviations of these terms used in SRS

MCQ	Multiple Choice Question
QB	Question Bank
RS	Random Selection
SQL	Structured Query Language
APPS	Application

Table No. 1 : Abbreviations of these terms used in SRS

2. The Overall Description

The main factor that can affect the system is the length of the data. If the user did not provide enough text to generate relevant number of questions.

2.1 Product Functions

- Dataset is required for generating question and answers using NLP (Natural Language Processing).
- Different login portals available for teachers and students.
- Objective as well as Subjective questions will be generated and the result will be evaluated using NLP.
- Result is evaluated and a file containing result of all the students is sent to teacher's portal.

2.2 User Characteristics

There are two types of users that are going to interact with system, i.e. students and teachers. User interface of the system is easily understandable by every age group. The end users might use the user interface model to get their desired outputs and the backend engineers (developers) that control, fix bugs and change the code or technique of the software program as per the requirement.

2.3 General Constraints

- Teacher will have complete authority over all test assessment.
- Teacher and student portals for a better interface.
- Passwords to enhance security.

2.4 Assumptions and Dependencies

- User should be aware about the operational functioning of the system.
- The Source data provided by the user for the generation of the questions-answer must be in meaningful and usable form.
- User should have enough storage to store the results of students.

3. Specific Requirements

The basic specific requirements are:

- The project will work only when Python and SQLite is installed on your PC.
- Some libraries use some older version of Python.
- PC with good GPU must be used for faster execution of code.

3.1 External Interfaces

3.1.1 Interfaces

There are basically three modules involved in this project and they are:

1. User login
    - Teacher Login:-

Teacher can generate the test by providing some data. Test will be available to every student and after the student completes the test, teacher can either see the results on the website or can download the CSV file.

- Student Login:-

Each student will receive subjective or objective test with random questions and after they submit the test, result is evaluated and sent to teacher's portal.

2. Online generated test questions
   - Objective type questions:-

   Five objective questions will be generated.

   - Subjective type question:-

   Two objective questions will be generated.
3. Result declaration
   - CSV file format:-

   Teacher can download the result of all the students.

   - Online Portal:-

   Teacher can view the results online.

3.2 Functional Requirement

The system should specify the following requirement of administrator that will be required as.

1. This project requires dataset for generating questions and answers.
2. Based on this dataset the software will generate questions using NLP (Using Python and its libraries).
3. Its User Interface design presents a seamless blend of visual design, interaction design, and information architecture.
4. Automatic evaluation of result also reduces effort of teacher to manually evaluate them. Non Functional requirements

1. User Satisfaction: - The system is such that it stands up to the user expectations.
2. Response Time: - The response of all the operation is good. This has been made possible by careful programming.
3. Safety and Robustness: - The system is able to avoid or tackle disastrous action. In other words, it should be foul proof. The system safeguards against undesired events, without human intervention.
4. User friendliness: - The system is easy to learn and understand. A native user can also use the system effectively, without any difficulties.

3.4 Performance Requirements

This subsection specifies both the static and the dynamic numerical requirements placedon the software or on human interaction with the software.

3.3.1 Static Requirements

This software can get any length of the data from the user but there is a limit to the minimum data given to the software. There is some threshold data length in which relevant number of questions can be generated.

3.3.2 Dynamic Requirements

The more the data the more time will be taken by the software to generate questions. The software should process 50% of the text in first five seconds.

3.5 Logical Database Requirements

- The database can only be accessed by the admin.
- External user does not interact to the database.
- Invalid users are not allowed to enter.
- Incorrect data does not enter in database.

.

3.6 Design Constraints

3.5.1 Standards Compliance

Specify the requirements derived from existing standards or regulations. They might include:

- Does not accept incorrect data from user.
- After the test ends, result will be available to both student and teacher.
- Result can be downloaded in a CSV file.

3.6.1 Software Constraints

The necessary softwares required are:

- Python & its libraries
- SQLite
- NLP(its libraries)

3.6.2 Hardware Constraints

The minimum hardware requirements are as follows:

1. Processor: Intel Quad-core 1.7 GHZ Processor or above.
2. HD: Minimum 10 GB of HD.
3. RAM: Minimum 8 GB of RAM.

[Chapter -3] System Design

3.1 Product perspective:

In the fast-paced society, people prefer to use practical tools to improve work efficiency. As for online exam, convenient operation methods can reduce the workload of teachers and also make it easier for students to study. The convenience of online exams is that teachers can use Question Answer Generation System which is a powerful online exam system to upload test questions in batches and quickly create test papers. It only takes four steps, and the teacher can successfully create a test paper in a few minutes.

3.2 Product functions:-

- Dataset is required for generating question and answers using NLP (Natural Language Processing).
- Different login portals available for teachers and students.
- Objective as well as Subjective questions will be generated and the result will be evaluated using NLP.
- Result is evaluated and a file containing result of all the students is sent to teacher's portal.

3.3 User characteristics:-

There are two types of users that are going to interact with system, i.e. students and teachers. User interface of the system is easily understandable by every age group. The end users might use the user interface model to get their desired outputs and the backend engineers (developers) that control, fix bugs and change the code or technique of the software program as per the requirement.

3.4 Constraints:-

Few constraints to run the project:-

- The project should have all the libraries installed that are required for Natural Language Processing.
- The project needs to install third-party products.
- The source data provided should be efficient to generate required number of questions.

3.5 Flow chart/DFDS

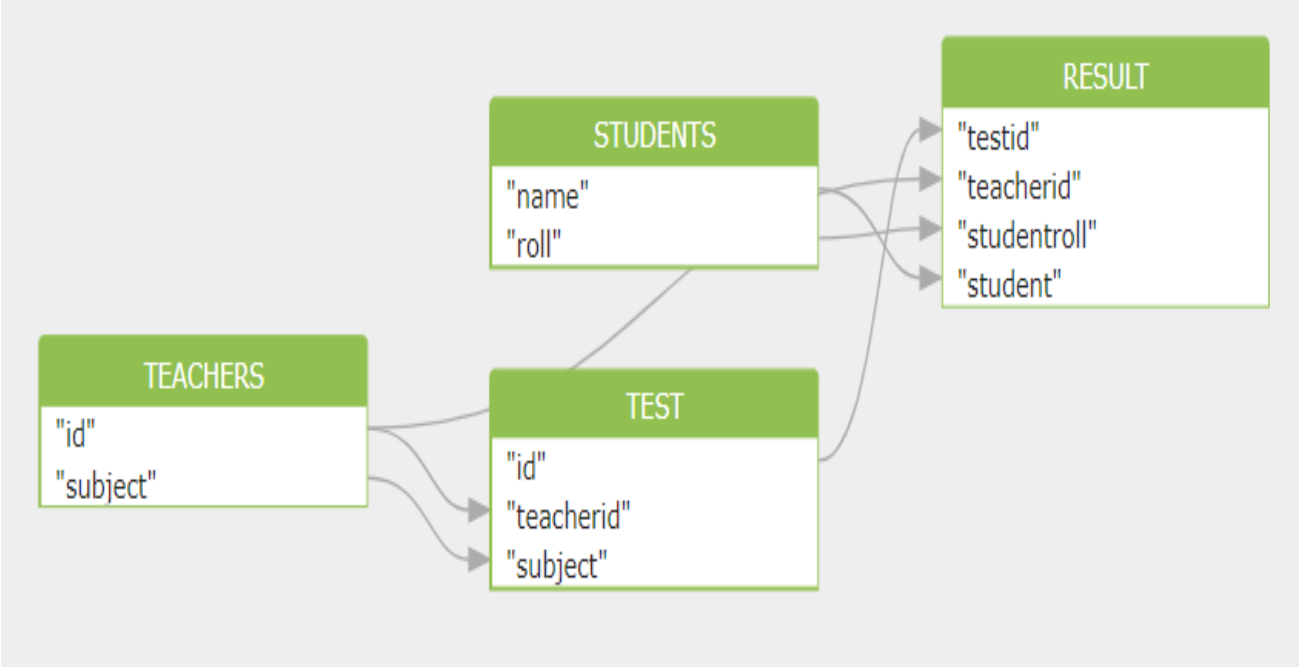

Fig. No.2 : Flow chart representing all tables with their fields

3.6 Database Design

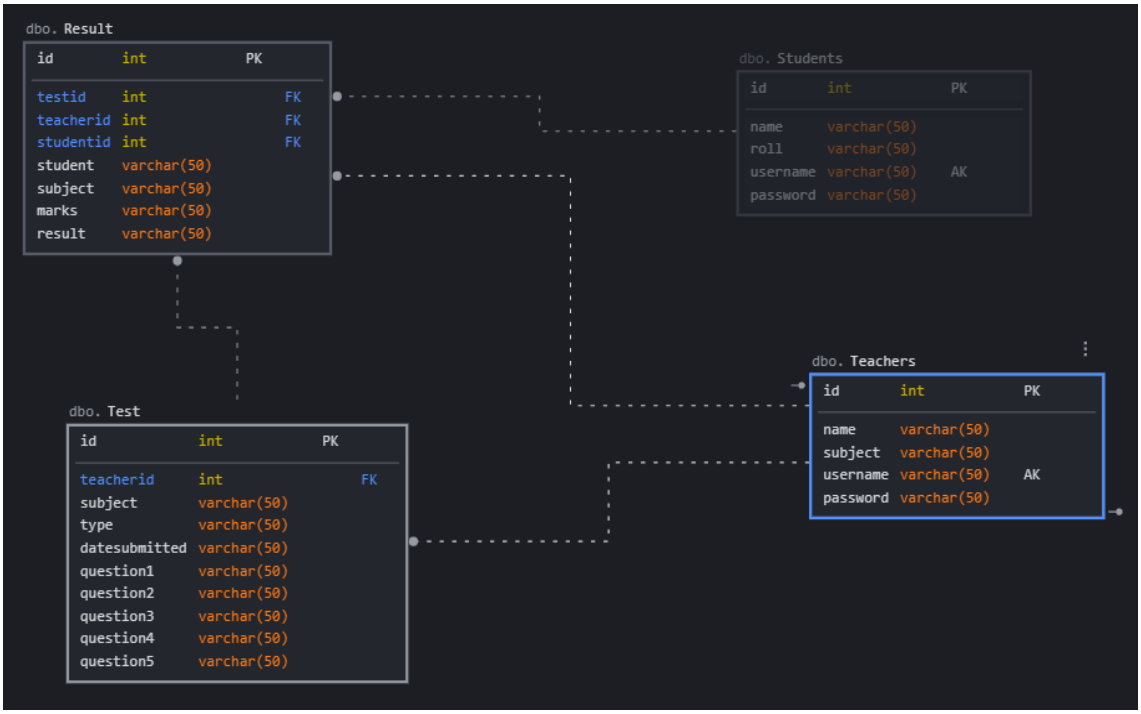

Fig. No. 3 : Database design with all table names

3.7 Table Structure

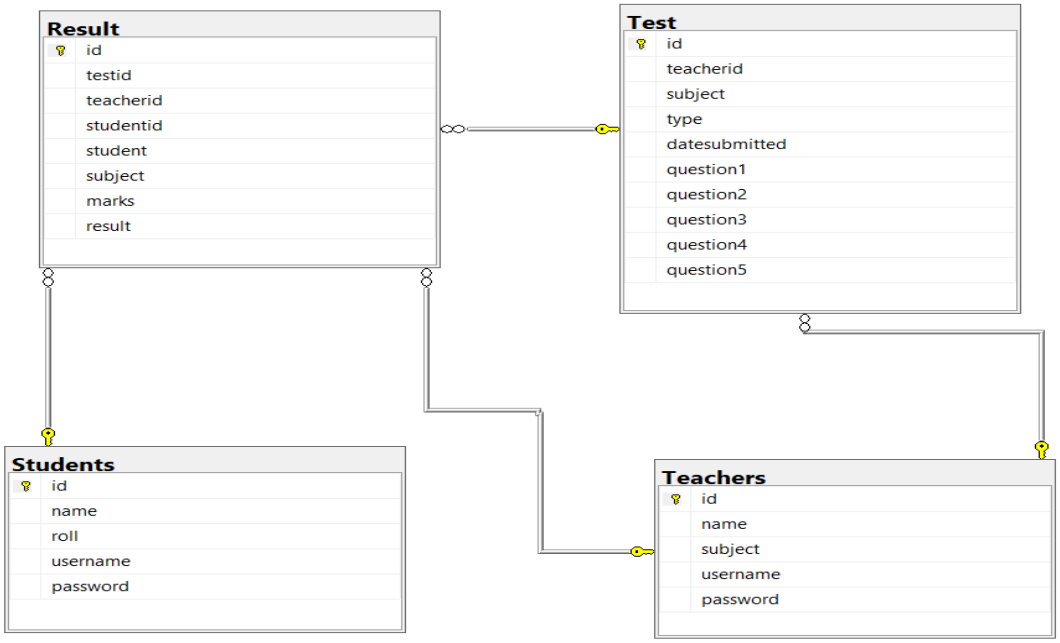

Fig. No. 4 : Table structure

3.8 ER Diagram

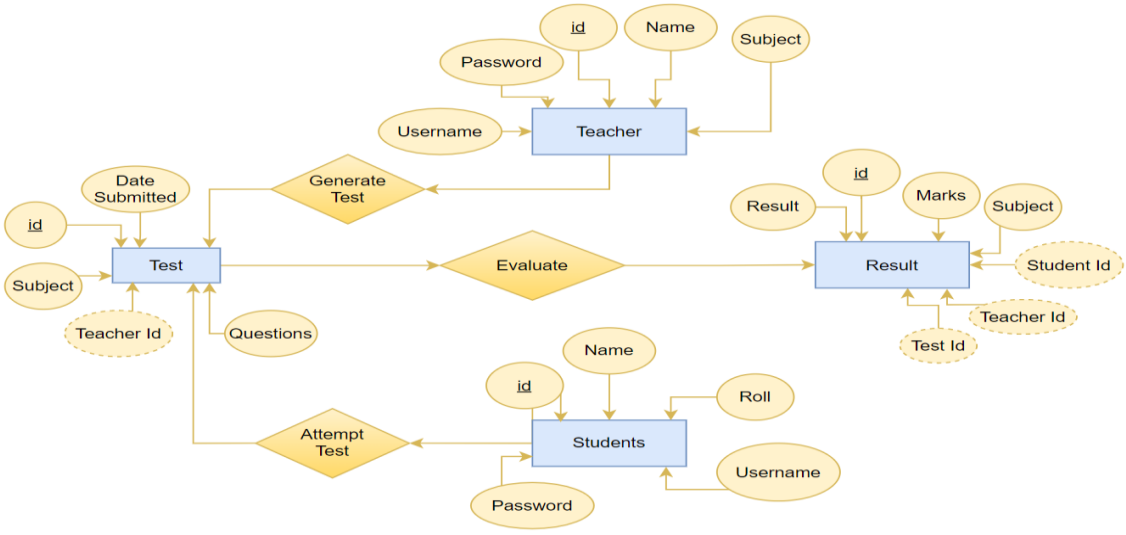

Fig. No. 5 : ER Diagram containing relationship between all the entities

3.9 Assumptions and Dependencies:-

Few assumptions and dependencies that is essential to use the product:

- User should be aware about the operational functioning of the system.
- The Source data provided by the user for the generation of the questions-answer must be in meaningful and usable form.
- User should have enough storage to store the results of students.

3.10 Specific Requirements:

- The project will work only when Python and SQLite is installed on your PC.
- Some libraries use some older version of Python.
- PC with good GPU must be used for faster execution of code.

[Chapter-4] Development, Implementation and Testing

4.1 Introduction to development environment (Front End and Back End)

1. Front End

- HTML:-

  The HyperText Markup Language, or HTML is the standard markup language for documents designed to be displayed in a web browser. It can be assisted by technologies such as Cascading Style Sheets and scripting languages such as JavaScript.

  Web browsers receive HTML documents from a web server or from local storage and render the documents into multimedia web pages. HTML describes the structure of a web page semantically and originally included cues for the appearance of the document. HTML provides a means to create structured documents by denoting structural semantics for text such as headings, paragraphs, lists, links, quotes and other items. HTML elements are delineated by tags, written using angle brackets. Tags such as <img /> and <input /> directly introduce content into the page.

- CSS :-

  Cascading Style Sheets is a style sheet language used for describing the presentation of a document written in a markup language such as HTML. CSS is designed to enable the separation of presentation and content, including layout, colors, and fonts. This separation can improve content accessibility; provide more flexibility and control in the specification of presentation characteristics; enable multiple web pages to share formatting by specifying the relevant CSS in a separate .css file, which reduces complexity and repetition in the structural content; and enable the .css file to be cached to improve the page load speed between the pages that share the file and its formatting.

- Bootstrap :-

  Bootstrap is a free and open-source CSS framework directed at responsive, mobile-first front-end web development. It contains CSS- and JavaScript-based design templates for typography, forms, buttons, navigation, and other interface components. The primary purpose of adding it to a web project is to apply Bootstrap's choices of color, size, font and layout to that project. As such, the primary factor is whether the developers in charge find those choices to their liking. Once added to a project, Bootstrap provides basic style definitions for all HTML elements. The result is a uniform appearance for prose, tables and form elements across web browsers. In addition, developers can take advantage of CSS classes defined in Bootstrap to further customize the appearance of their contents. For example, Bootstrap has provisioned for light- and dark-colored tables, page headings, more prominent pull quotes, and text with a highlight.

- JavaScript :-

  It’s an object-oriented computer programming language commonly used to create interactive effects within web browsers. JavaScript is a high-level, often just-in-time compiled language that conforms to the ECMAScript standard. It has dynamic typing, prototype-based object-orientation, and first-class functions. It is multi-paradigm, supporting event-driven, functional, and imperative programming styles. It has application programming interfaces (APIs) for working with text, dates, regular expressions, standard data structures, and the Document Object Model (DOM).

  The ECMAScript standard does not include any input/output (I/O), such as networking, storage, or graphics facilities. In practice, the web browser or other runtime system provides JavaScript APIs for I/O.

- jQuery:-

  jQuery is an open-sourced JavaScript library that simplifies creation and navigation of web applications. jQuery's syntax is designed to make it easier to navigate a document, select DOM elements, create animations, handle events, and develop Ajax applications. jQuery also provides capabilities for developers to create plug-ins on top of the JavaScript library. This enables developers to create abstractions for low-level interaction and animation, advanced effects and high-level, themeable widgets. The modular approach to the jQuery library allows the creation of powerful dynamic web pages and Web applications.

2. Back End

   - Flask :-

     It is a micro web framework written in Python. It is classified as a microframework because it does not require particular tools or libraries. It has no database abstraction layer, form validation, or any other components where pre-existing third-party libraries provide common functions. However, Flask supports extensions that can add application features as if they were implemented in Flask itself. Extensions exist for object-relational mappers, form validation, upload handling, various open authentication technologies and several common framework related tools. Flask has become popular among Python enthusiasts. As of October 2020, it has second most stars on GitHub among Python web-development frameworks, only slightly behind Django, and was voted the most popular web framework in the Python Developers Survey 2018.Flask has become popular among Python enthusiasts. As of October 2020, it has second most stars on GitHub among Python web-development frameworks, only slightly behind Django, and was voted the most popular web framework in the Python Developers Survey 2018.

- Python :-

  It is an interpreted high-level general-purpose programming language. Python is dynamically-typed and garbage-collected. It supports multiple programming paradigms, including structured (particularly, procedural), object-oriented and functional programming. It is often described as a "batteries included" language due to its comprehensive standard library.

  Guido van Rossum began working on Python in the late 1980s, as a successor to the ABC programming language, and first released it in 1991 as Python 0.9.0. Python 2.0 was released in 2000 and introduced new features, such as list comprehensions and a cycle-detecting garbage collection system (in addition to reference counting). Python 3.0 was released in 2008 and was a major revision of the language that is not completely backward-compatible. Python 2 was discontinued with version 2.7.18 in 2020.

- SQLite :-

  It is a relational database management system contained in a C library. SQLite generally follows PostgreSQL syntax. SQLite uses a dynamically and weakly typed SQL syntax that does not guarantee the domain integrity. This means that one can, for example, insert a string into a column defined as an integer. SQLite will attempt to convert data between formats where appropriate, the string "123" into an integer in this case, but does not guarantee such conversions and will store the data as-is if such a conversion is not possible.

4.2 Other supporting languages or tools

1. NumPy:-

   NumPy is a library for the Python programming language, adding support for large, multi-dimensional arrays and matrices, along with a large collection of high-level mathematical functions to operate on these arrays. The ancestor of NumPy, Numeric, was originally created by Jim Hugunin with contributions from several other developers. In 2005, Travis Oliphant created NumPy by incorporating features of the competing Numarray into Numeric, with extensive modifications. NumPy is open-source software and has many contributors. NumPy is a NumFOCUS fiscally sponsored project.

2. Pandas :-

   Pandas is a software library written for the Python programming language for data manipulation and analysis. It is free software released under the three-clause BSD license. The name is derived from the term "panel data", an econometrics term for data sets that include observations over multiple time periods for the same individuals. Its name is a play on the phrase "Python data

analysis" itself. Wes McKinney started building what would become pandas at AQR Capital while he was a researcher there from 2007 to 2010.

3. Scikit – Learn :-

   Scikit-learn is a free software machine learning library for the Python programming language. It features various classification, regression and clustering algorithms including support vector machines. Scikit-learn is largely written in Python, and uses NumPy extensively for high-performance linear algebra and array operations. Furthermore, some core algorithms are written in Cython to improve performance. Support vector machines are implemented by a Cython wrapper around LIBSVM; logistic regression and linear support vector machines by a similar wrapper around LIBLINEAR. In such cases, extending these methods with Python may not be possible.

4.3 Implementation of problem

4.3.1 Pseudo code of Tokenization

```
Define : List of unwanted_character
Define : String_article
For i=1 to number_of_character in String Article
        If (String_article[i]==“unwanted character”)
                Remove String_article[i];
        End If
End For
String_split (String_article)
```

Fig. No. 6 : This represents tokenization of string

4.3.2 Pseudo code of Keyword extraction

```
sort sentences by weignt
while (desiredSumLength is not met and there are unused sentences)
for (all sentences x )
        if (sentence x not already in summary)
                if (segment of sentece x has the lowest or equally low use)
```

set sum_sentence = x

break out of for loop

end if

end if

end for

add sum_sentence to the summary

record sum_sentence as having been used

increment sum_sentence's segment use

increment currentSumLength

end while

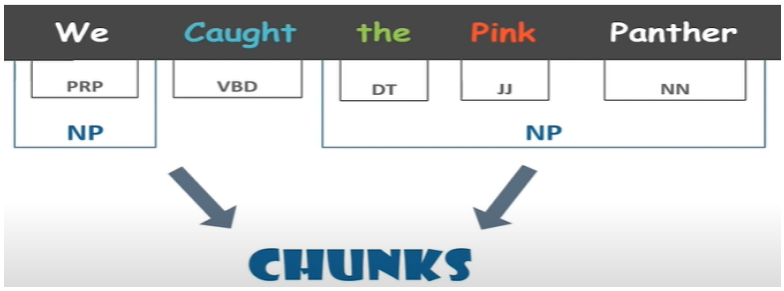

Fig. No. 7 : Chunking of a sentence

4.3.3 Pseudo Code of Part of speech tagging

Algorithm Noun<n>

for each word in the list n >= 1do

if the word=noun then mark it up with <n>

else mark it up with <a>

return <n>

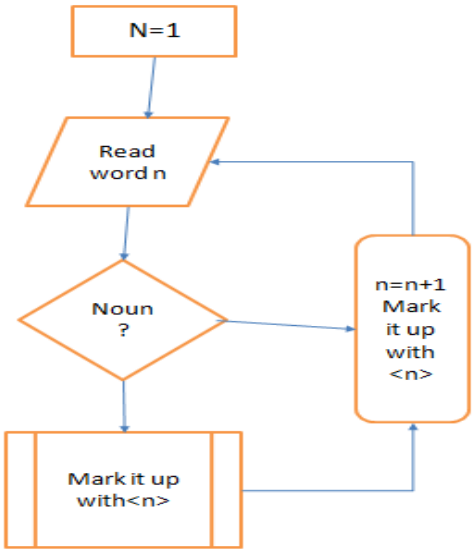

Fig. No. 8 : Flow chart of part of speech tagging (Pseudocode)

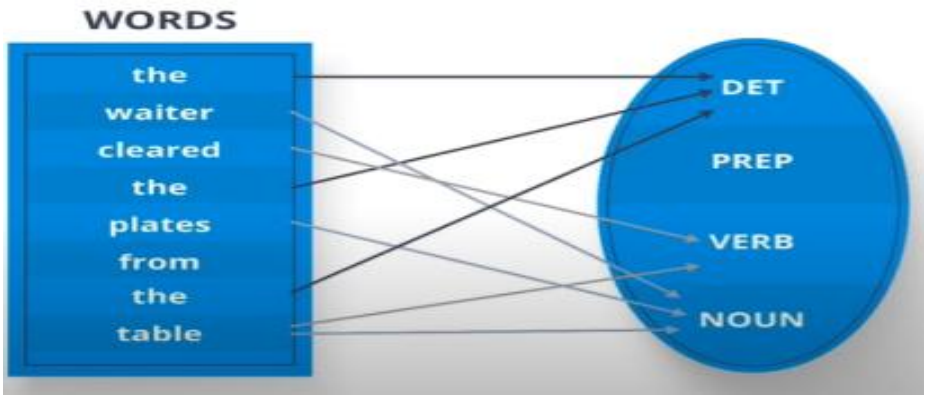

Fig. No. 9 : Pictorial representation of part of speech tagging

[Chapter - 5] Results and Discussions

5.1 User interface representation

User Interface Module	Is the module user directly interacts with.
NLP Module	This module is acting as backbone for project. Keyword extraction is done using various processes like tokenization, morphology, part of speech tagging, distributional similarity , chunking , name entity recognition.
Database Module	All students' records will be saved in database.

Table No. 2 : Different Modules Used in Project

4. User login
   - Teacher Login:-

     Teacher can generate the test by providing some data. Test will be available to every student and after the student completes the test, teacher can either see the results on the website or can download the CSV file.
   - Student Login:-

     Each student will receive subjective or objective test with random questions and after they submit the test, result is evaluated and sent to teacher's portal.
5. Online generated test questions
   - Objective type questions:-

     Five objective questions will be generated.
   - Subjective type question:-

     Two objective questions will be generated.
6. Result declaration
   - CSV file format:-

     Teacher can download the result of all the students.
   - Online Portal:-

     Teacher can view the results online.

5.3 Project screenshots

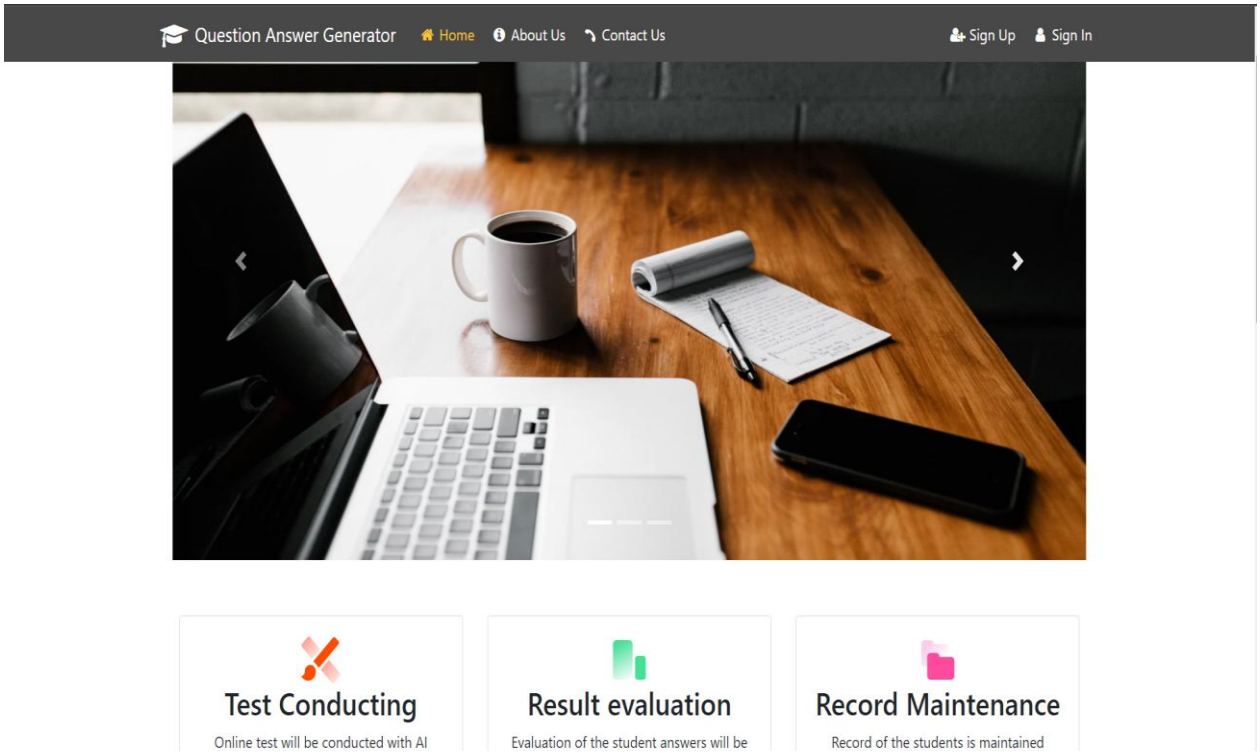

Fig. No. 10 : Home Page

This page depicts the home page of our website called Question answer generator. We provide this page with the option of login Sign in and sign-up option for teacher and student.

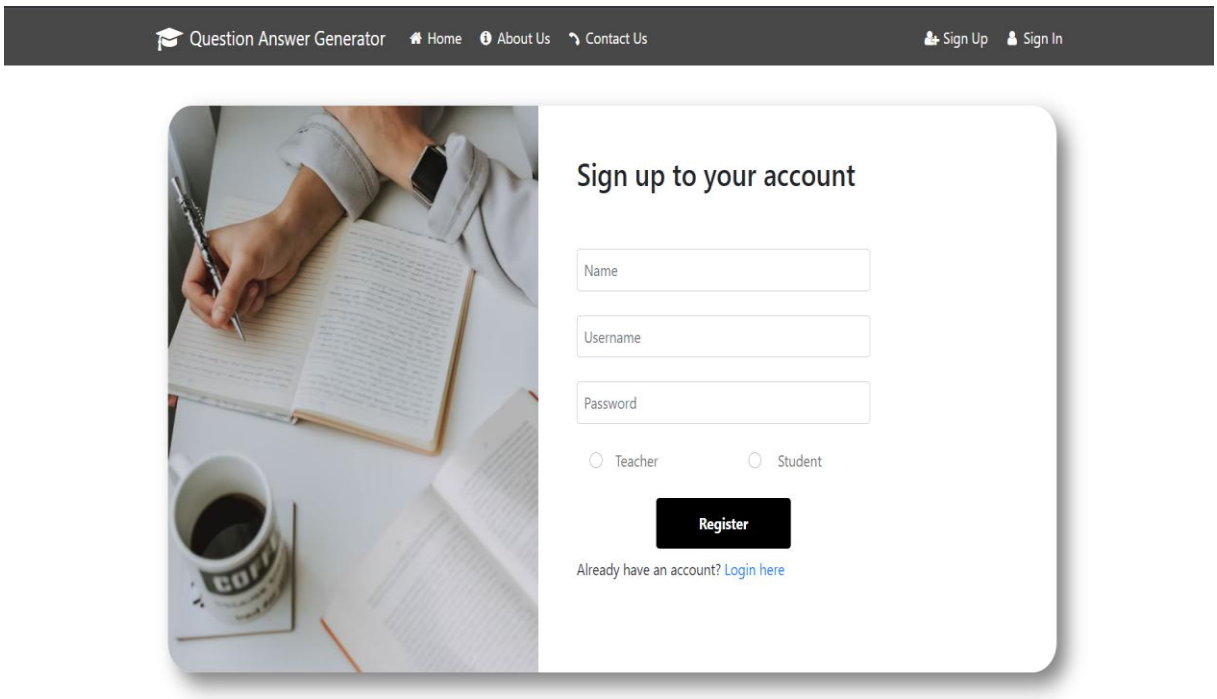

Fig. No.11 : Sign up Page

This webpage shows the registration page as a student or teacher with their credentials.

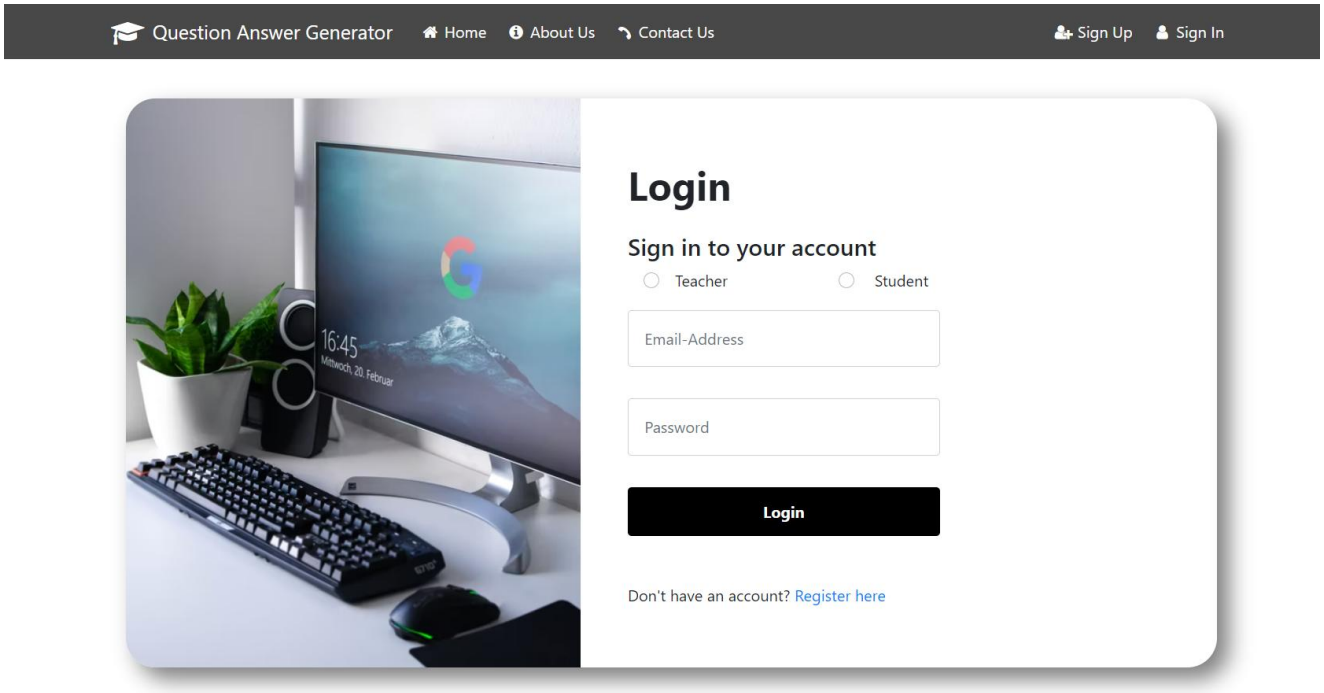

Fig. No. 12 : Login Page

This webpage depicts the login page as a student or teacher.

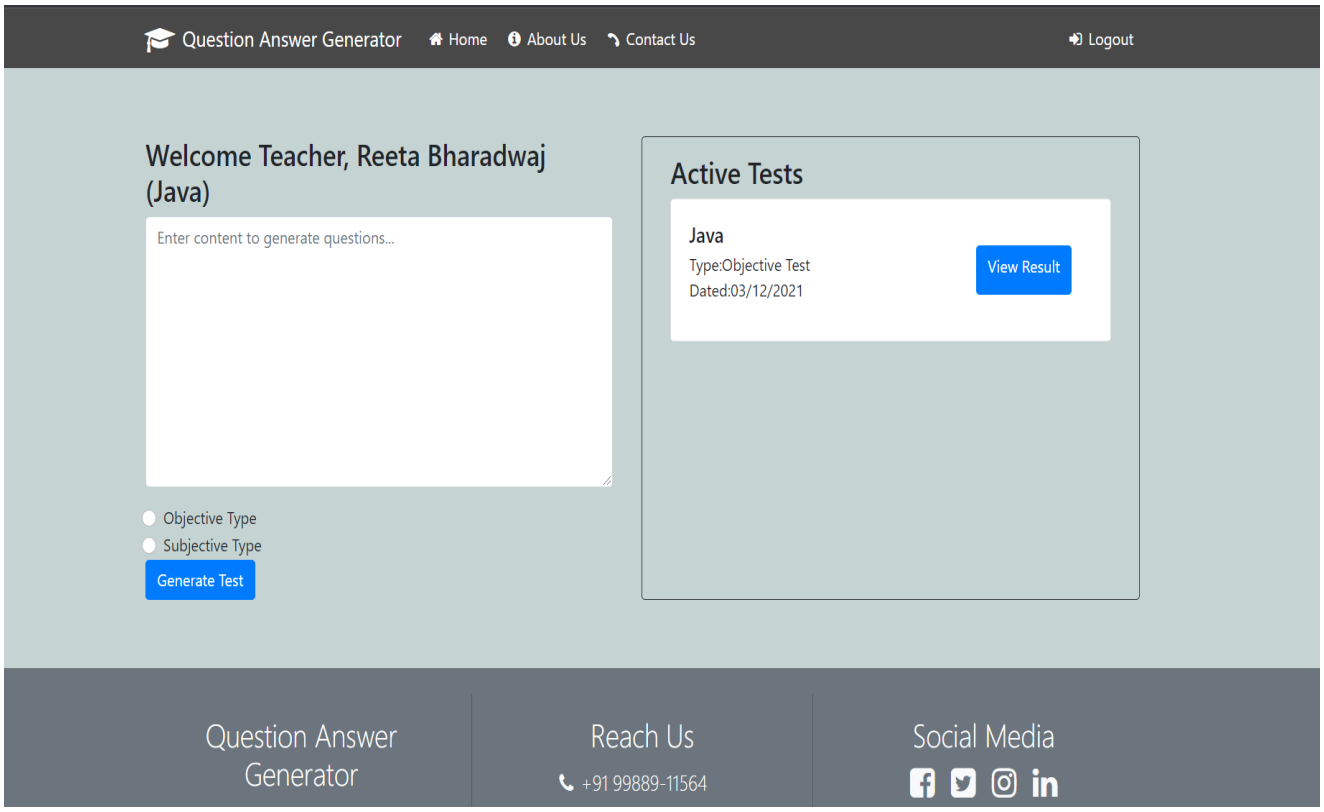

Fig. No. 13 : Teacher Dashboard

This Picture depicts the teacher Dashboard where teacher can add content and can easily generate objective and subjective question by choosing one of the options. Generate test button help in generating test. Active tests section represents all the tests that the teacher has generated.

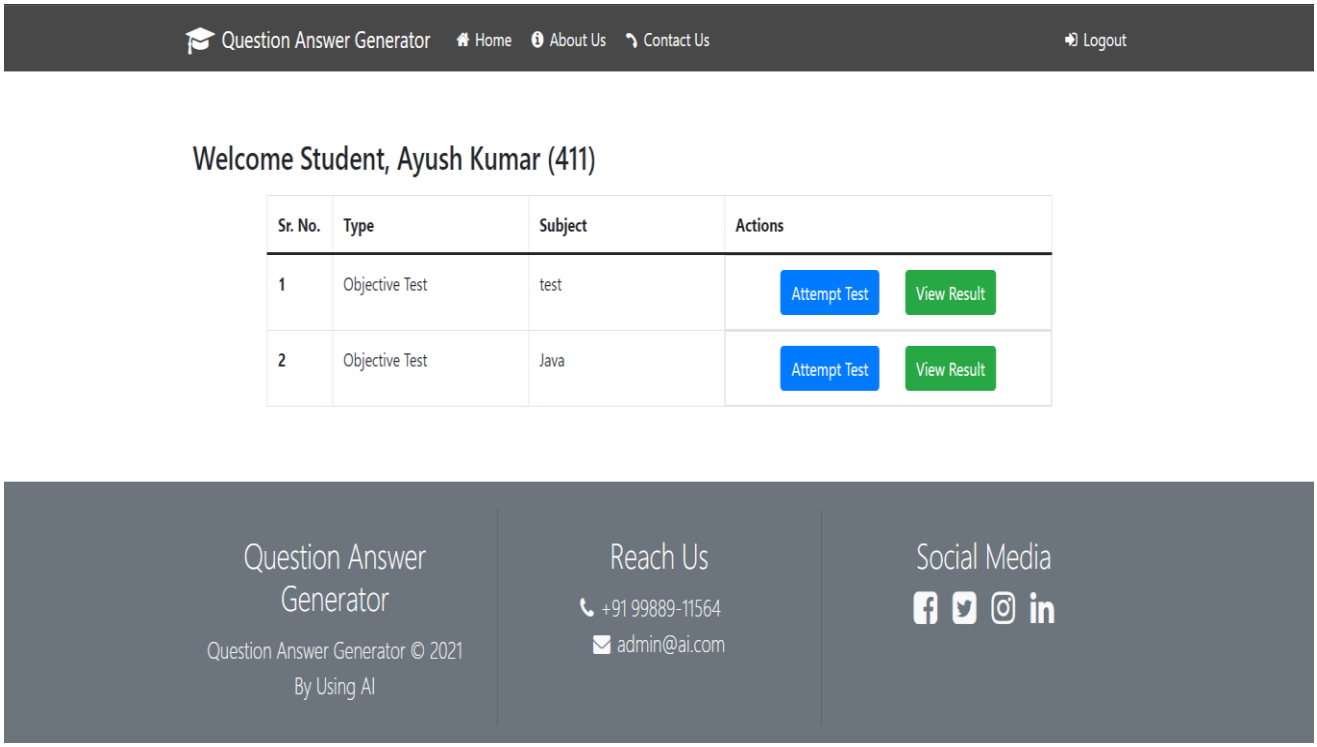

Fig. No. 14 : Student Dashboard

This picture represents student dashboard where student can attempt test and view his/her result.

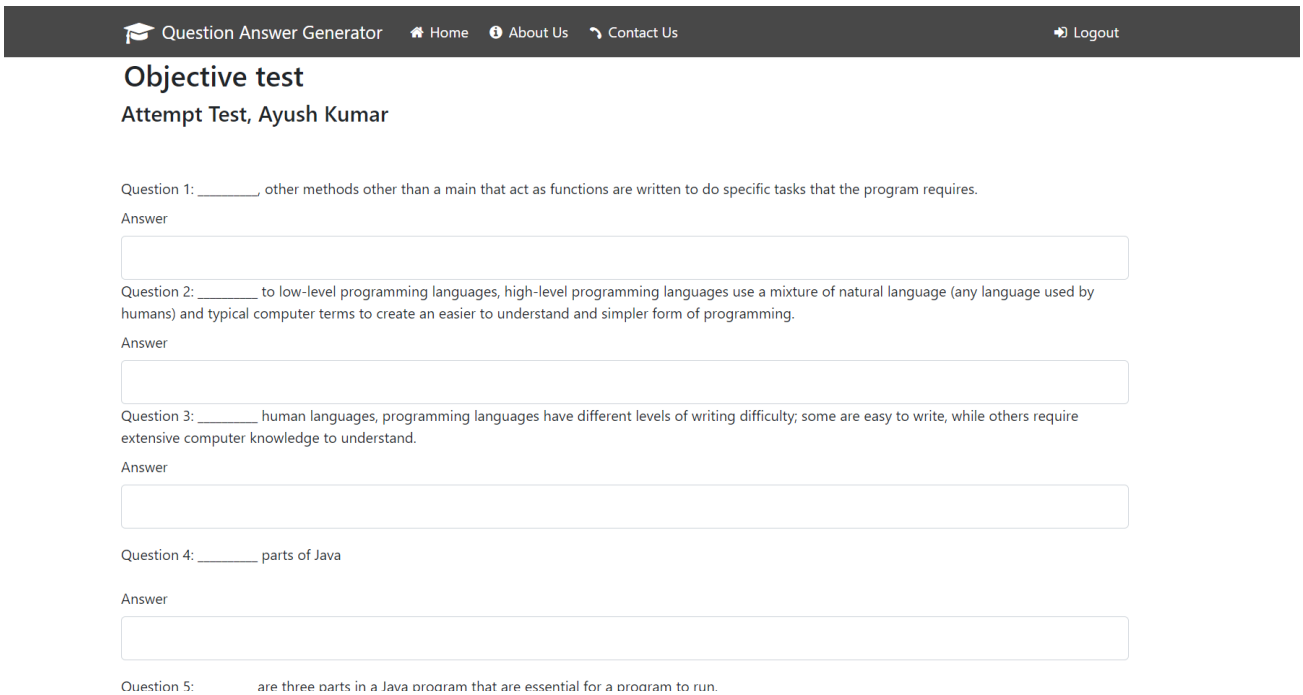

Fig. No. 15 : Objective Test Page

This page represents the objective questions generated from the text. It generates five questions and takes only one word answer.

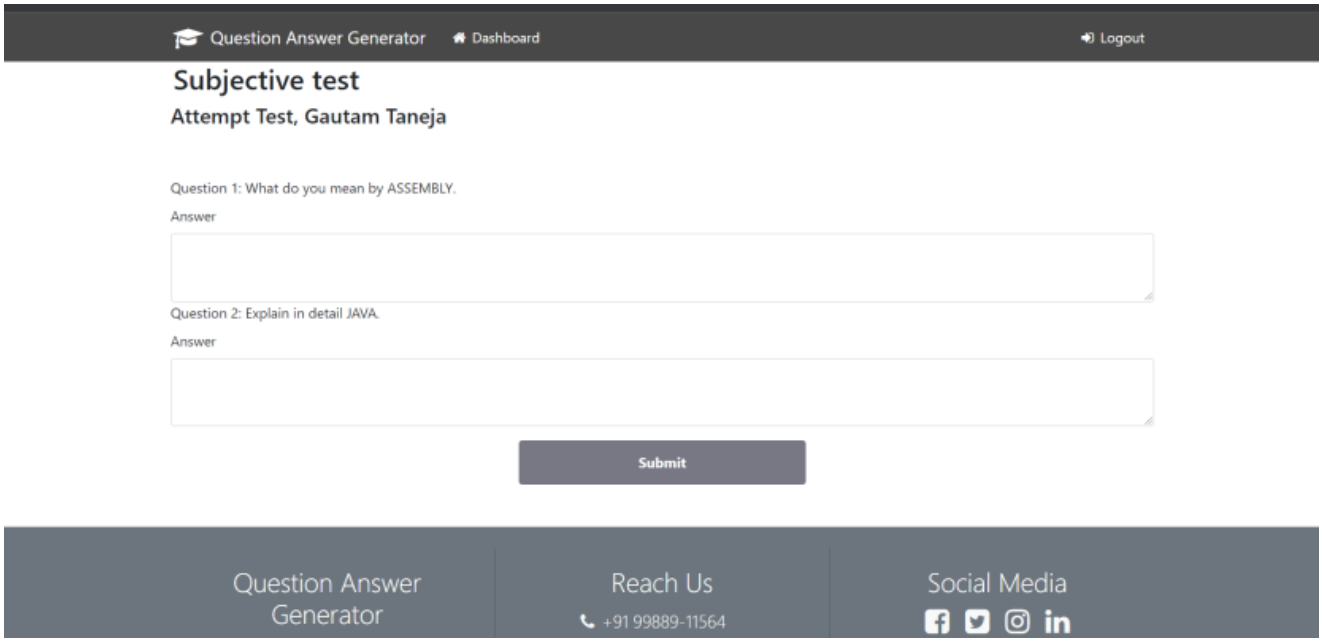

Fig. No. 16 : Subject Test Page

This page represents the subjective questions generated from the text. It generates two questions and takes the whole descriptive answer.

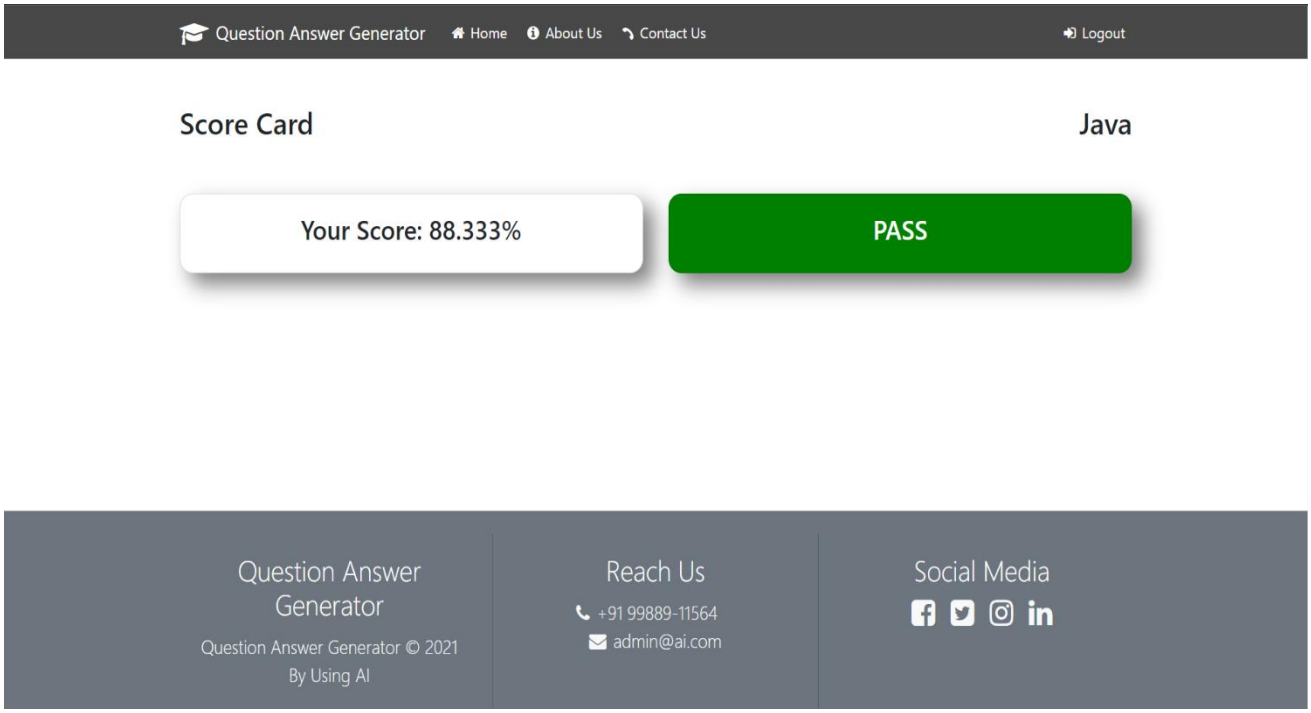

Fig. No. 17 : Result Page

Scorecard of student is displayed and whether he/she passes or fails is also shown. The result can also be downloaded by teachers portal.

5.4 Comparative Analysis

1. Duration for question generation (Lesser is better)

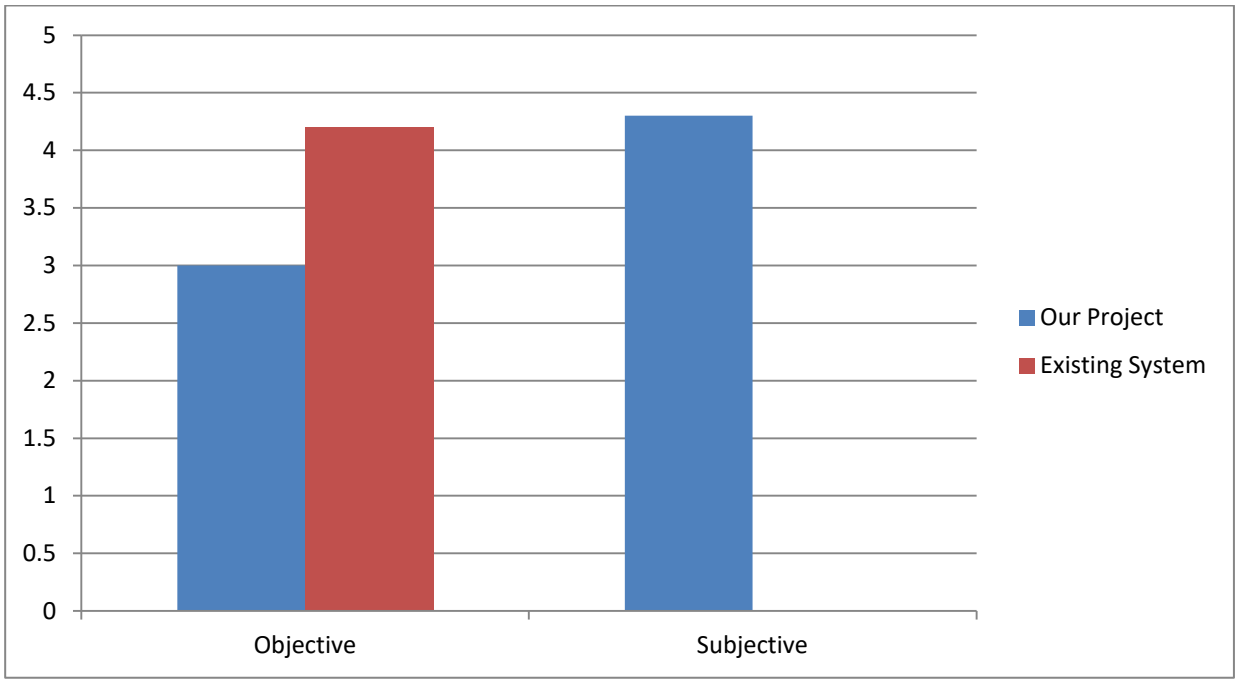

Fig. No. 18 : Comparative Analysis Bar Graph (Duration of question answer generation)

2. Quality of questions/answer : (Higher is better)

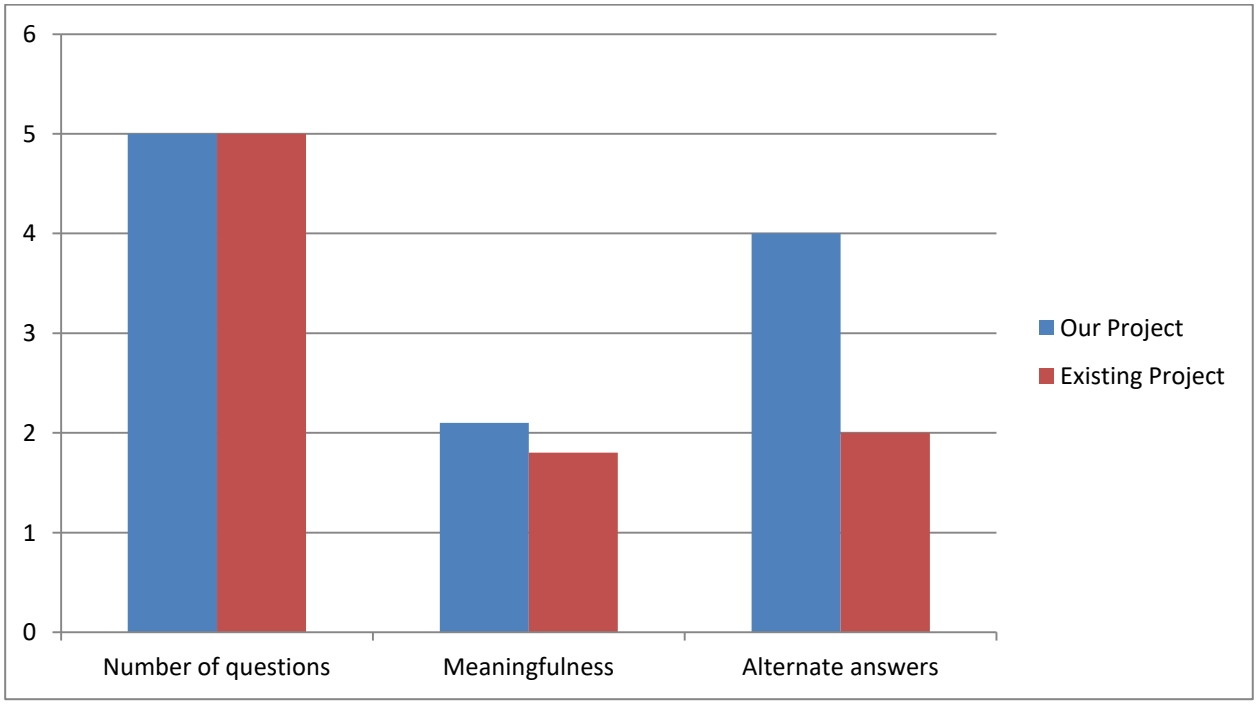

Fig. No. 19 : Comparative Analysis Bar Graph (Quality of questions/answer)

3. User Interface : (Higher is better)

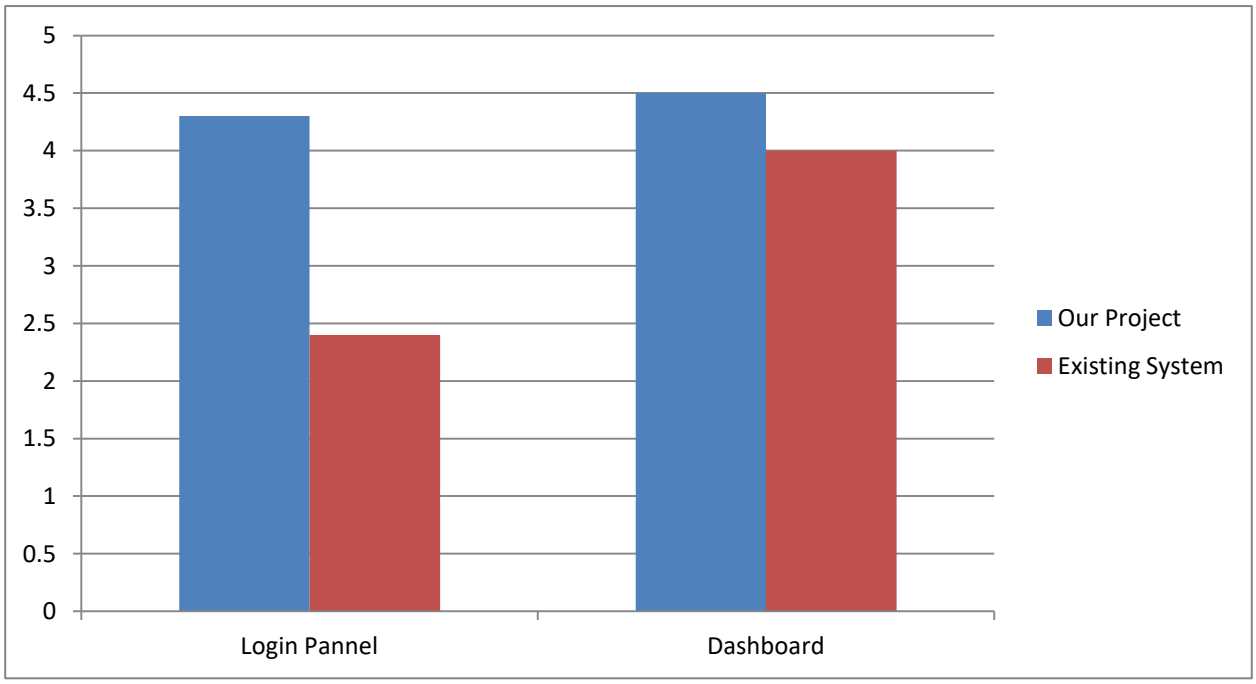

Fig. No. 20 : Comparative Analysis Bar Graph (User Interface)

[Chapter - 6] Conclusion and Future Scope

6.1 Conclusion

Our project is only a humble venture to satisfy the needs of educational institutes to manage their Test-taking task conveniently. We created the both subjective and Objective questions answers: The "Objective &Subjective Question" section presents the different types of objective & Subjective questions generation with NLP. Our approach composes of two steps selecting relevant and informative sentences and identifying keywords from the selected sentences and generate Objective & Subjective Questions. Implement automatic computation of results: Our project calculate the score of the student and display in their respective student portal.

To test and compare the proposed system with existing system: The comparison we made between proposed system with existing system is in our system we provide only the data from which question we have to extract but in existing system teacher have to create the question by itself. Incorporate the transparency in test system: With our system we ensures that the each and every student gets the different set of question paper and they are not perform illegal activity during the exam.

Several user-friendly coding has also been adopted. This package shall prove to be a powerful package in satisfying all the requirements of the educational institutes. The manual question generation takes much time and labour. Therefore, automatic question generation from learning resources is the primary task of an automated assessment system. This project presents a survey of automatic question generation and assessment strategies from textual learning resources. The purpose of this project is to summarize the state-of-the-art techniques for generating questions and evaluating their answers automatically.

6.2 Future Scope

Nowadays the fastest developing field in NLP is one of the quickest developing fields from recent years. By the evolution of the period, its significance will increase day by day because online data is present in an enormous amount due to digitalization. Extracting useful information from text has a long way to go. By developing the benefit of connecting content mining to other fields such as machine learning, perception, normal dialect preparing, it could be conceivable to sketch more effective and helpful content mining frameworks. NLP is very useful for the industry to utilize and develop a way of learning that can't be devoured by people. In this project, we attempted to introduce the application of content mining i.e. the extraction of keywords from text with the use of ensembles approach, NLP, instruments, and applications. Few improvements that can be made are:

- Difficulty level of the question can be improved in the advance version of this project in future.
- User interface can be enhanced.
- Grammatically errors can be improved.
- Timer can be added for the tests.

References

[1] Indrashis Das, Bharat Sharma, Manjusha Pandey and Siddhart S. Rautaray"An Examination System Automation Using Natural Language Processing", Fourth International Conference on Communication and Electronics Systems, 978-1-7281-1261-9,2019

[2] Cristley, "Natural Language Processing (NLP) Tutorial with Python & NLTK", freeCodeCamp.org, 2018 (https://www.youtube.com/watch?v=X2vAabgKiuM)

[3] Catherine Henry, "Natural Language Processing (NLP): Automatic generation of questions and answers from Wikipedia", Clearwater DevCon - Clearwater Analytics,2017 (https://www.youtube.com/watch?v=3vRZ6dBPL2A)

[4] Zolzaya Luvsandorj , "Introduction to NLP : Preprocessing text in Python", Towards Data Science, 2020 (https://towardsdatascience.com/introduction-to-nlp-part-1-preprocessing-text-in-python-8f007d44ca96)

[5] B. Kaur, and S. Jain "Keyword extraction using machine learning approaches", IEEE 3rd International Conference on Advances in Computing, Communication & Automation (ICACCA) (Fall), 2017