# Random Multiple Choice Questions Generation using Nlp

**Ch Hemanth, M Venkat, B Chitti Babu, U Rochasvi, P Daanesh**

***Abstract*: *Students' life is incomplete without exams because exams are those that help students in evaluating themselves and thus proceeding further in studies. So, the starting step in conducting such examinations is creating a question paper. Generating a question paper is still in its traditional way, where lecturers or professors that are the teaching staff are doing it manually and wasting a terrible amount of time in selecting what type of questions are to be generated. It's so difficult to create a question paper as it includes a lot of resource utilization and exhaustion. These tasks can be automated. As we are seeing a lot of development in new, exciting technologies and these technologies can help and can make the process of automation easier. So for automation, we use Machine Learning and Natural Language Processing as this whole task involves using and manipulating textual data. In this solution, we provide our model with a textual paragraph from which the questions are to be selectively generated and we develop the multiple choices using a certain distinctive process for the users.***

***Keywords* : *Answers evaluation, MCQ's generation, Natural Language Processing, NLTK.***

## I. INTRODUCTION

In this competitive world students' future is defined by exams even if we agree with it or not. Exams are the only means that helps students understand the levels of their respective academic skills and knowledge. The examination helps students to truly evaluate themselves and then provides them with an opportunity to work hard and improve their skills. Thus exams help in creating youth who are matured, academically strong with a high level of skill sets who in turn help in developing this country because today's youth are tomorrow's pillars of the society so they need to be shaped in a better way. So, exams are to be carefully designed fairly and transparently. Parents of students take these exams very seriously and also in a prestigious way. Some students prepare for some high-level competitive exams, taking years of training and practice. An institute's reputation and goodwill also depend on these exams. Institutions prepare schedules and prepare their students in hard and tiresome ways so that students succeed in such high-level competitive exams. So, such exams are to be organized effectively and efficiently. For exams like UPSC, JEE mains, and Advanced, the government takes up calibrated measures in organizing these exams. In the case of college exams, respective university management takes up the job of organizing exams. So, we feel like there no problems in organizing these examinations. But we tend to overlook sometimes, here we did it again. We'll explain to you what is that has been overlooked this time.

The very first step is important. With the development of technology organizing an examination has become easy. Steps like allocating students to exam halls and also allocating respective invigilators for the exam hall. Arranging question papers in a better way to respective branches etc can be managed easily with the developed technological resources. But the only process that is still in the traditional way irrespective of the development of technology is question generation. Even now with all sorts of technology available at our hands, the questions are still being created by professors or lecturers manually. So, our focus is on bringing in the use of technology and automating questions generation. The main benefit here is teachers generally consume a lot of amount of valuable time. So what is the point of technology if we can't save valuable time? Therefore we have decided to automate the process of question generation. Here the main aim is to develop a machine learning algorithm that helps in automating questions generation phase efficiently. After developing an algorithm, if it works efficiently, a web-based final solution can be provided for the end-users. We have generally two types of questions. One is subjective questions, these take descriptive texts as answers, these types of questions can be easily generated. Where there are other types of questions which are multiple-choice questions. So such questions are to be created with utmost importance. While generating multiple-choice questions, choices are to be generated in such a way that one of them is the right one and other options are wrong answers but are close to the correct option. If the wrong answers are not relevant to the correct answer then it is easy for students to guess the correct answer. So, ample care has to be taken while generating multiple-choice questions. So the final solution for all these problems is provided for the end-users through a web platform. Web platforms or web-based solutions provide users with a sophisticated way through a browser either on a monitor or a mobile. So, Flask has been chosen to provide this web-based solution and it can be easily deployed in a browser and it is compatible with any type of browser.

**Revised Manuscript Received on June 30, 2020.**

**Ch Hemanth**, department of computer science and engineering, VR Siddhartha engineering college, Vijayawada, Kanuru. Email: hemanthchigurupati25@gmail.com

**M Venkat**, department of computer science and engineering, VR Siddhartha engineering college, Vijayawada, Kanuru. Email: venkatmannam3@gmail.com

**B Chitti Babu**, department of computer science and engineering, VR Siddhartha engineering college, Vijayawada, Kanuru, Email: bhukyachittibabu5@gmail.com

**U Rochasvi**, department of computer science and engineering, VR Siddhartha engineering college, Vijayawada, Kanuru, Email: rochasviuppuleti31@gmail.com

**P Daanesh**, department of computer science and engineering, VR Siddhartha engineering college, Vijayawada, Kanuru, Email: daaneshpotnuri.3690@gmail.com

Flask is a technology that also provides an efficient framework for easy connectivity with the project's backend and frontend. So, here for the frontend HTML, CSS is being used, and coming to backend some algorithms are used which helps us generate a solution. So for this machine learning techniques like Natural Language Processing is used. NLP is an area among the computer science technologies that deals with the textual data and their processing, so that the machine understands and evaluates and processes the human language. NLP provides users with multiple options. So NLP and NLTK generate both multiple-choice questions and subjective questions. Here NLP algorithm makes use of textual data that is being provided to the algorithm, understands it, and proceeds on to create questions. A special type of machine learning algorithm called Gaussian Naïve Bayes is used to generate MCQs and their respective choices.

The main reason of this project is to make the process of creating a question paper. The main advantage involves the prevention of leakage of paper and it also reduces the scope of malpractice because no two students have the same set of questions. It also saves lots of time. It also has some disadvantages which are security, in this project, there is an absence of authentication of students in this process. Another disadvantage is sometimes the choices of MCQs are too similar. These can be overcome through future work.

## II. RELATED WORK

Vamsi Krishna Yepuri, et. al [1] designed a system called "Examination Management Automation System" which is supported by android and web browsers. It is created to manage and handle various operations during examinations like the timetable, seating arrangement, and invigilators allocations to the respective exam halls. These tasks are generally performed by various institutions to date consuming a lot of time and manpower. So this automated system is proposed by them which solves the above-stated problems, saving a lot of amount of time and energy. This model has a high-end architecture that provides the user with a simple interface so that the user can enter their inputs from a webpage or an android application. This model mainly consists of 3 main tiers are Presentation tier, Business tier, Data tier. These tiers have their individual and specific functionalities that ultimately reduce the stress of staff and students. Some of the complexity is reduced and a lot of time is saved. This project helped us understand that a lot of tasks are still being manually done and automation is needed. This also helped us in understanding the basic concept of automation, how it is done, and how it is implemented. Priti Mishra, et. al [2] proposed a system called “An automated examination system using cloud computing technology” which also focuses mainly on removing all the manual paperwork. The main aim is centralization, this centralized system provides a solution so that all examination activities can be effectively and efficiently performed. The main examination activities that are focused on are Exam form filling, ATKT form filling, and other types of forms that are involved during the examination process. These people observed that are a lot of forms that are filled by students before writing an important exam, these processes automation is a major thing since it consumes a lot of manpower and time. So they proposed a centralized CLOUD COMPUTING solution. Because of this solution provided by this team, there is no more standing in queues for form filling. As it is based on cloud computing technology and cloud computing is booming in the IT industry nowadays this proposed solution is very beneficial. Prof. Sejal Dmello, et. al [3] proposed an " Examination cell Automation System" whose main focus is on automating the examination activities. These papers are chosen as references to get a brief knowledge and idea about automation and how it is being implemented. This proposed system mainly replaces the manual calculations and all the enrolment activities which are performed by students before writing an exam or before registering for an exam. These activities are replaced by this automation system. This project also provides a centralized system so that all the manual work of teaching staff and students can be saved and be able to save valuable time. This proposed system consists of 3 main actors they are Admin, Staff, Student. Admin manages everything that is update, create, delete, and maintaining the database. These sorts of work are done by admin. Student and staff perform their respective activities but they are only limited to certain activities not much more because that is the role of admin.

Bondre Rutuja Avinash, et. al [4] have designed an "Automated Examination Support System" which replaces their current examination system which is being followed in their institute with a new system. They felt the traditional system is outdated so they designed this new system. Basic technologies like PHP and HTML are being used to develop this system. This system stores details of students and provides all functionalities required for both staff and students. It is based on CGPI/SGPI. A "UserCakeManagement System” is used to manage data and information of teachers and students in an efficient manner.

These papers gave us a basic idea of automation and also made us understand how the examination system is still manual in a lot of activities. After referencing these papers we noticed that a small area in the examination system which is yet to be automated that is the process of creating a question paper. So we took up this idea and from the help of this research paper, the automation of questions generated is done.

## III. PROPOSED METHOD

The proposed system is a simple yet effective one. This system mainly has 3 actors. That is Admin, Staff, and Students. The admin takes care of the update and deletion and creation of the database. He also maintains the website and prevents it from any more troubles or attacks. He also takes care of adding or removing students' details as per the need. The only role of staff is submitting the text from which the questions are to be generated. This text can be directly fed up to the algorithm or it can be provided to the admin who in turn directly feeds this textual data to the algorithm. Thus textual data given by staff or admin is used by the algorithm and generates questions. Students can only view the exam portal whatever they see is the examination portal alone which is a web application that displays the questions. Students’ only role is to select the subject and select the type of question paper they need to answer that is objective question paper or subjective question paper. After selecting they will be provided with the question paper where they read the paper and submit their response in the given time.

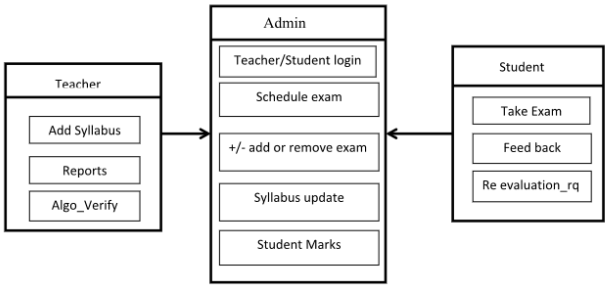


**Fig. 1. Proposed system architecture**

Fig. 1 also explains the actors present in our model and their detailed functionalities. Jumping into the technical functionalities of the project the question generation especially multiple-choice questions generation can be done by divide and conquer approaches. The plain text is fed to the algorithm for generating MCQ's. The algorithm first identifies the key elements from the given text. These keywords can be used as answers to the generated questions. A later question is created or generated by inserting a blank space by replacing the keywords that are the answer. Now the sentences that are generated with a blank space between them are to be converted into proper questions that are in the format of questions for example sentences starting with What, How, etc. Now we have the question and its corresponding answer, but we also need to generate other incorrect answers for MCQ's. So for that, we need to generate some incorrect answers. So a machine learning model is trained with an existing SQUAD(standard questions and answers dataset). After training the algorithm it generates incorrect answers which are closer or relevant to the correct answer.

## IV. CONCLUSION AND FUTURE WORK

This provides a flexible and easy solution for especially teachers because they need not put in a lot of effort and time. So this helps teachers in focusing on what has to be taught to the future generation and how to shape them in a good way because that's how a good society will be created. But this idea can be further developed like here we just automated the task of generating questions so this can be extended so that it evaluates the user's response and generate results. Even though automation is done, some vulnerabilities are to be overcome. There are mainly security vulnerabilities. So in the future, it can be worked through and security can be implemented. In the future, this project can be further developed and can also be put in use in schools as well as small colleges where this automation of question paper generation is missing.

## REFERENCES

1. Vamsi Krishna Yepuri, Gopi Chand Pamu, Naveen Kodali, Pradyumna LV, "Examination Automation System", International Research Journal of Engineering and Technology, Volume 5 Issue 4 April 2018.
2. Priti Mishra, Muhib Lambay, Namrata Kadam, Pooja Tiwari, Krinjal Kadu, "An Automated Examination System using Cloud Computing Technology" International Research Journal Of Advance Research, Ideas And Innovation In Technology, Volume 4 Issue 2, 2018.
3. Prof. Sejal Dmello, Ashraf Pinjari, Harishchandra Damre, Aishwarya Sankhe, Vrushali Ambre" IOSR Journal of Engineering, Volume 5, March 2018.
4. Bondre Rutuje Avinash, Durga Varsha Vijayakumar, Mohite Pradnesh, Parkar Vishal V, "Automated Examination Support System", International Journal of Current Engineering and Technology 2015.
5. Dayand G Savakar, Ravi Hosur, "Automation of Examination System", International Journal of Science and Research, Volume 4 Issue 11, November 2015.
6. J. Cui, F. Wen, and X. Tang (2008), "Real Time Google and Live Image Search Re-Ranking", Proc. 16th ACM Int'l Conf. Multimedia.
7. D. Tao, X. Tang, X. Li, and X. Wu (2006), "Asymmetric Bagging and Random Subspace for Support Vector Machines-Based Relevance Feedback in Image Retrieval," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 28, no. 7, pp. 1088-1099.
8. A.W.M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain (2000),"Content-Based Image Retrieval," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 22, no. 12, pp. 1349-1380.

## AUTHORS PROFILE

**Ch Hemanth** pursuing 4/4 B. Tech in department of CSE at VR Siddhartha engineering college, Vijayawada, AP. His area of interest includes Data Science and Machine Learning. He is also planning to do master's in business analytics

**M Venkat** pursuing 4/4 B. Tech in department of CSE at VR Siddhartha engineering college, Vijayawada, AP. His area of interest includes AI and Machine Learning.

**B Chitti Babu** pursuing B. Tech in department of CSE at VR Siddhartha engineering college, Vijayawada, AP. His area of interest includes Data Structures and DBMS.

**U Rochasvi** pursuing B. Tech in department of CSE at VR Siddhartha engineering college, Vijayawada, AP. Her area of interest includes Data Science and Networking.

**P Daanesh** pursuing B. Tech in department of CSE at VR Siddhartha engineering college, Vijayawada, AP. His area of interest includes Cyber Security. This project is done under the guidance of Ms. V. Deepa, Assistant Professor, VRSEC.