

**AN APPROACH TO GENERATE AUTOMATIC QUESTION AND ANSWER USING
MACHINE LEARNING**

Ifrah Tabassum , M.Tech Scholar, Department of Computer Science & Engineering, Government College of Engineering, Aurangabad 431005 (Maharashtra state, INDIA)

Prof. Prashant .D. Pathak (Assistant Professor), Department of Computer Science & Engineering, Government College of Engineering, Aurangabad 431005 (Maharashtra state, INDIA)

Abstract— Today's time is the era of artificial intelligence and slowly things are based on artificial intelligence, in today's time, many questions arise in the minds of people and they need their answer as soon as possible. And in today's time, many times people find the answers to their questions or the answer to the question on the computer and the computer feeds inside itself already on the basis of the predefined answer, but many times it happens that we want to answer those questions. Which is in front of us in the form of an article, and in the matter of education, the same problem is faced by the children because in today's time when the epidemic has spread and the children are locked in their homes and they cannot go to school, at such a time, we have a There is a need for such a system which reads a sentence itself and prepares its own questions and answers, so that it will be difficult for the child to understand the lesson. Many times it also happens that people are a little educated and when they read a newspaper or article, they do not understand the essence of that news, at such a time if there is such a system that can question and answer a paragraph itself. How good it would be if we prepare the answer, we would have felt the need of such an artificial based system. But the system which is present in today's time does not give much better response.

Keywords: decision tree algorithm, Real News, Fake News, Genuine.

1. INTRODUCTION

In today's time, Kasis does not have enough time to read the entire news or article, he just wants that he can get the answer of the questions related to that article quickly and it is also right, it saves our time. And in today's education system there is a great need for such a system, in today's time the education system is going through its worst phase because all the children are locked in their homes and their studies are getting affected a lot and at such a time they have to be like this. There is a need of system which can solve their problem which can help the children in their studies. My work of a teacher is to get the child to read the chapter and answer the questions and answers related to that chapter. We miss the lack of such an artificial based system at such a time that we can get to know the true purpose of that chapter. It is not that such a system does not exist, but they do not give a perfect question answer because fixed question answers are generated inside them, whereas we need a system that can answer the question intelligently. And we want to make such system. To create such a system, first we prepare the dataset of such articles, questions and answers, and in this dataset, we have included the articles of 10 different fields and related questions and answers from them, once the dataset is ready. Then first of all I have filtered this dataset so that there is no redundant data because redundant data can create problem in generating query answer because for any machine learning algorithm its dataset is its soul and other dataset is a Works as fuel so if this vehicle has to run artificially then your dataset should be as strong. After that we do the classification of the dataset, for the classification we have used the classification algorithm so that we can prepare the training and testing dataset Training dataset means such dataset on whose base we have trined the dataset Training dataset always testing dataset Should be more than this because on the basis of this we trinade our machine so that the machine can take the right decision and make the right prediction. After this we work on Testing dataset Testing dataset is also very important but Testing dataset works as compared to training dataset Testing dataset is real time dataset from which we know whether the system is working properly or not Training and testing We have taken the ratio of dataset as 8:2 and also giving very good result then after that we have done prediction algorithm which is doing query and answer generation based on training and testing dataset. In the auto question and answer system, it is necessary to pay attention to the grammatical errors because to make any sentence it is necessary that attention should

be paid to Noun, Pronoun, Verb and Adjective, because even one mistake in them gives the meaning. It is possible

2. RELATED WORK

Sonam Soni, Praveen Kumar, Amal Saha, (2021)[1] Producing Questions is a significant activity in instructive learning. Questions give different phases of difficulty in the instructive learning strategy. Information building is expensive for PC Assisted assessment when setting exercise questions, instructors use test makers to fabricate question Banks. The paper presents a framework that produces inquiries with answers automatically applying natural language processing and openNMT. Past work question age naturally, makes inquiries from sentences applying grammar and semantic parser. The paper presents another methodology for creating question and answer. And the author has given the accuracy of his algorithm as 87%. But if we analyze this structure carefully, then we find that the correct comparison has not been done at many places.

Miroslav Blšták, Viera Rozinajova (2017) [2], In this paper, we acquaint an intuitive methodology with age of verifiable inquiries from unstructured content. Our proposed system changes input text into organized arrangement of provisions and uses them for question age. Its learning cycle depends on blend of AI methods known as support learning and directed learning. Learning measure begins with beginning arrangement of sets shaped by definitive sentences and doled out questions and it constantly figures out how to change sentences into questions. Cycle is likewise improved by criticism from clients with respect to as of now created questions. We assessed our methodology and the correlation with cutting edge frameworks shows that it is a point of view way for research.

Bidyut Das, Mukta Majumder, Santanu Phadikar & Arif Ahmed Sekh(2021) [3] Learning through the web becomes well known that works with students to learn anything, whenever, anyplace from the web assets. Appraisal is generally significant in any learning framework. An evaluation framework can track down oneself learning holes of students and work on the advancement of learning. The manual inquiry age takes a lot of time and work. In this way, programmed question age from learning assets is the essential assignment of a robotized evaluation framework. This paper presents an overview of programmed question age and appraisal methodologies from literary and pictorial learning assets. The motivation behind this overview is to sum up the best in class strategies for producing questions and assessing their answers naturally, then the author is telling the accuracy of this algorithm to be 92%. That sounds right. But if the algorithm is understood properly, then it turns out that it can be very difficult because of keyword wise searching.

Miroslav Bl and Viera Rozinajov (2016) [4] In this paper, we acquaint an intuitive methodology with age of verifiable inquiries from unstructured content. Our proposed system changes input text into organized arrangement of components and utilizations them for question age. Its learning cycle depends on blend of AI methods known as support learning and administered learning. Learning measure begins with beginning arrangement of sets framed by decisive sentences and allocated questions and it consistently learns instructions to change sentences into questions. Cycle is additionally improved by input from clients with respect to as of now created questions. We assessed our methodology and the correlation with cutting edge frameworks shows that it is a point of view way for research.

Holy Lovenia, Felix Limanta, Agus Gunawan(2018) [5] Potential freedoms for question-answer age have been proposed in the past work, remembering for the field of schooling. The need of inquiries and answers is incited for different purposes, for example self-study, scholarly appraisal, and coursework. In any case, the regular approach to make question-answer sets has been both drawn-out furthermore, tedious. In the current examination, we propose an programmed question age for sentences from text entries in understanding cognizance. We present a rule-based programmed question age for the errand, as well as carry out factual sentence determination and different designs of named substance acknowledgment. Three sorts of WH-questions ("What", "Who", and "Where") can be delivered by our framework. The framework performs well on producing inquiries from basic sentences, however flounders on more mind boggling sentences because of inadequate change rules.

3. METHODOLOGY

• Proposed System

In this system, first we take a text paragraph from the user, then by using the Natural Language Processing Algorithm, we remove whatever problem is there so that the machine does not have any problem in reading the text graph, after that we use the Support Vector Machine Algorithm. By doing this we classify the paragraphs and prepare a trinade and test dataset. In any machine learning algorithm, it is necessary to have the right ratio of trinade and test dataset, that is why we have taken the ratio of trinade and test dataset 8: 2 like this Our model is prepared, after this, as soon as the user will differentiate his text paragraph, the natural processing algorithm first separates the grammatical keywords like Noun, Verb, Adjective, Preposition etc. Then after that using our model in the natural language. Processing algorithm generates whatever possible question answer can be formed, then using prediction algorithm, the algorithm tells what percentage is correct.

4. Natural Language Processing

Natural Language Processing (Natural Language Processing) is that part of computer science and artificial intelligence, within which research is done on how to reconcile humans and computers. It shows the ability of the computer to understand human language. Under this, the human language is automatically understood by the computer with the help of some software. Research is going on in this area for the last 50 years. As computers developed, more in-depth study started here. Example: If you open a website, there is an option of an online assistant working automatically on a website. It works on the basis of natural language processing process.

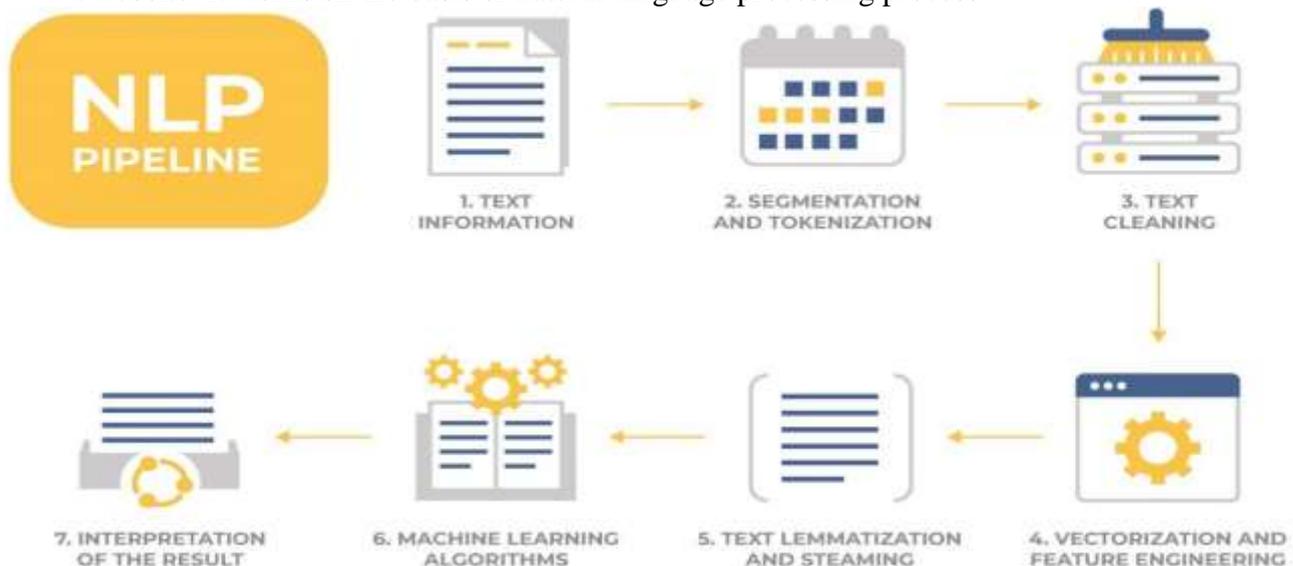


Figure 1: Natural Language Processing

Proposed Algorithm

1. Read the paragraph text in p.
2. Split the string in sentences s[] string array.
3. After search grammatical tokens in each sentences and store in t[i][j]
4. Clean the array t[i][j].
5. Apply the Support Vector Machine Algorithm and classify the paragraph and the select the features Of text on the basis of grammatical token and store it in F[]
6. Extract the feature of each row
 - For k_x in s[]
 - If $t[i][j] = s[i]$
 - If $t[i][j]$ in tokens
 - Collect in $F[] = t[i][j]$
7. Trained the dataset p[] and create the model $m[x][y]$

8. Test the paragraph on the basis of LSTM and get prediction score.
9. if score=0 then
 No Question Answer Generate
 Else if score>0 and score<=10
 Print Question Answer Will be print

SYSTEM ARCHITECTURE

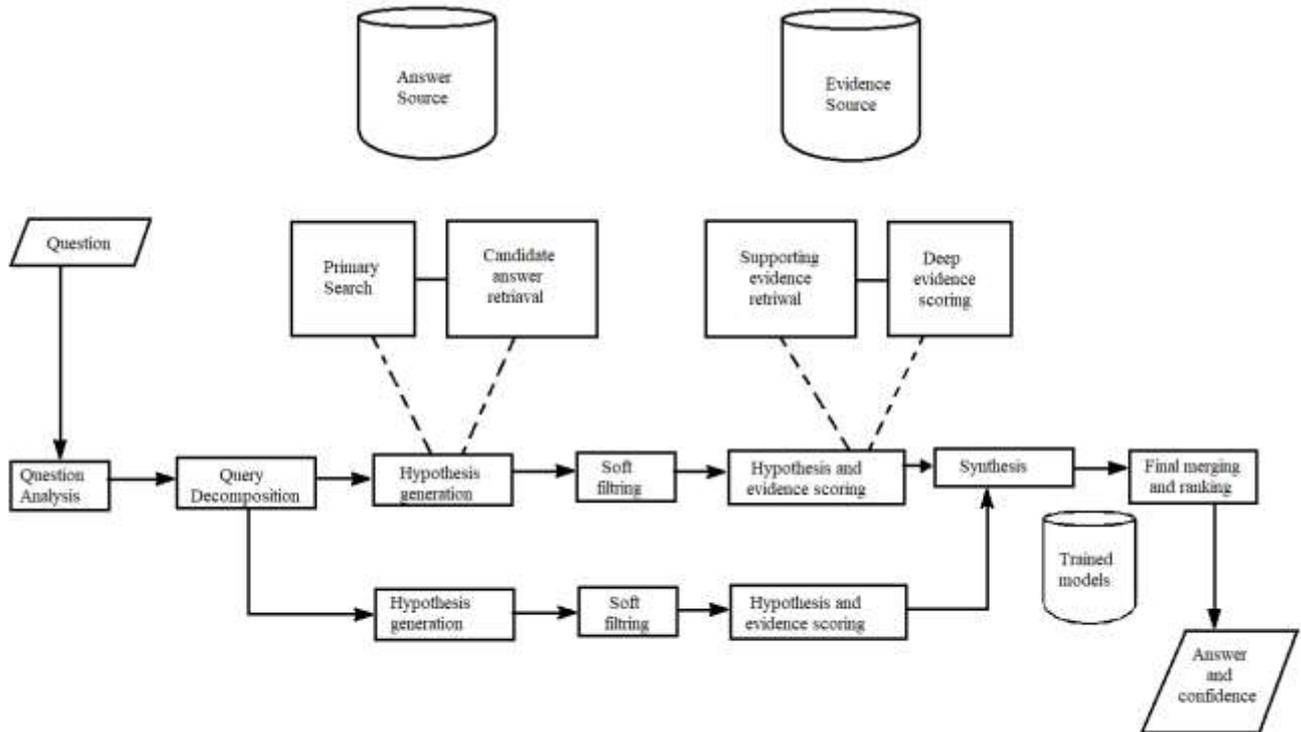


Figure 2: Architecture diagram

5. RESULTS

In this paper first we have prepared a dataset of 100 text paragraphs then separated its grammatical tokens using Natural Language Processing Algorithm and then performed feature extraction using Trined & Test dataset Prepar Kia Support Vector Machine Algorithm and Then by using the LSTM algorithm, a list of questions and answers has been prepared. We have given equal priority to dataset and NLP algorithm in algorithms so that there is no gap in the result and to a large extent we have been successful in this.

So far, the existing algorithm, it generates only question answers related to a particular field, but there is no such restriction in our algorithm, the accuracy of our algorithm is 89% which is quite good.

Algorithm	Decision Tree	RNN	LSTM	My Algo(NLP+SVM+LSTM)
Accuracy	84	79	86	89

Table 1: comparative Study Table

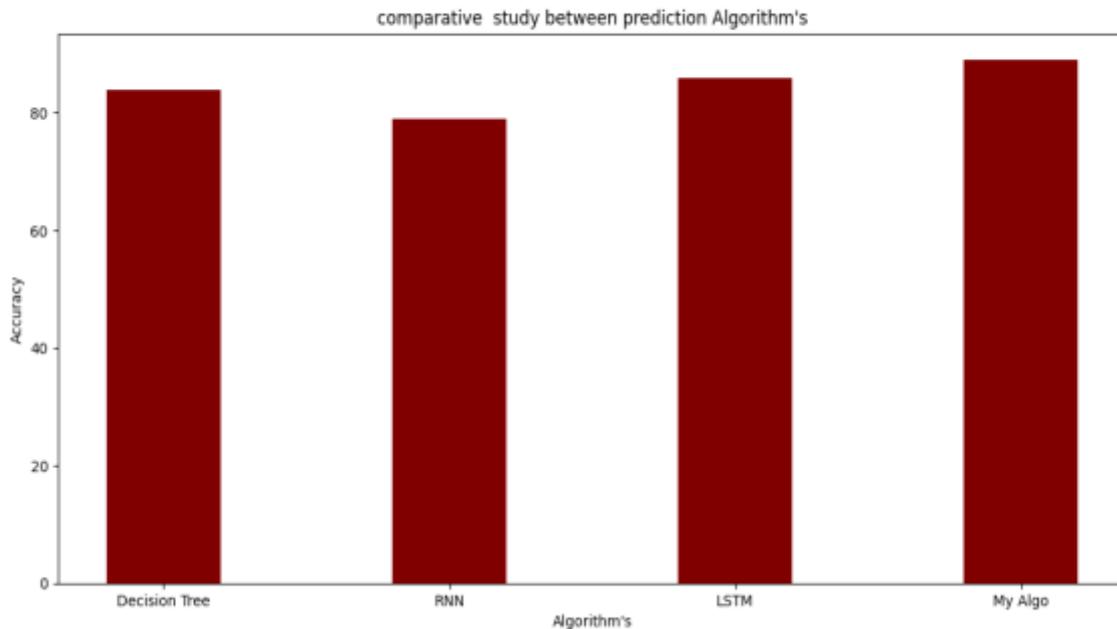


Figure 2. shows the accuracy for Automatic Question and answer generation

6. CONCLUSION & FUTURE SCOPE

In today's time, according to the way people have work time, then there are many requirements of such algorithms and we have developed our algorithm keeping this in mind and the algorithm has also worked as per our expectation and The accuracy of our algorithm is 89%, which is giving a very good result according to this type of algorithm, but we will try to improve it further, now we have got the result only on the basis of text paragraph but we are using audio and video. Will work even with it, because in today's time people do not even want to write, they do their work even after watching videos and listening to audio, so it is necessary that there should be such an algorithm which can do this work and can generate related question answers.

REFERENCE

- [1] Agarwal, M., Shah, R., and Mannem, P. 2011. Automatic question generation using discourse cues. In Proceedings of the 6th Workshop on Innovative Use of NLP for Building Educational Applications, Association for Computational Linguistics.
- [2] Babko-Malaya, O. 2005. Propbank annotation guidelines. URL: <http://verbs.colorado.edu>
- [3] Blanco, E., and Moldovan, D. 2011. Some issues on detecting negation from text. In FLAIRS Conference.
- [4] Boyer, K. E., and Piwek, P., editors. 2010. In Proceedings of QG2010: The Third Workshop on Question Generation. Pittsburgh: questiongeneration.org
- [5] Carpenter, S. 2012. Testing enhances the transfer of learning. In Current directions in psychological science, 21(5), 279-283.
- [6] Charniak, E., and Elsnar, M. 2009.
- [7] EM works for pronoun anaphora resolution. In Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics. Association for Computational Linguistics.
- [8] Collobert, R., Weston, J., Bottou, L., Karlen, M., Kavukcuoglu, K., & Kuksa, P. 2011. Natural language processing (almost) from scratch. The Journal of Machine Learning Research, 12, 2493-2537.
- [9] Curto, S., Mendes, A., and Coheur, L. 2012.
- [10] Question generation based on lexico-syntactic patterns learned from the web. Dialogue & Discourse, 3(2), 147-175.
- [11] Heilman, M., and Smith, N. 2009.
- [12] Question generation via overgenerating transformations and ranking. Technical Report CMU-LTI-09-013, Language Technologies Institute, Carnegie-Mellon University.
- [13] Heilman, M., and Smith, N. 2010a. Good question! statistical ranking for question generation. In Proceedings of NAACL/HLT 2010. Association for Computational Linguistics.
- [14] Wolfe, J. 1976. Automatic question generation from text-an aid to independent study. In Proceedings of ACM SIGCSE-SIGCUE.
- [15] Yao, X., and Zhang, Y. 2010. Question generation with minimal recursion semantics. In Proceedings of QG2010: The Third Workshop on Question Generation.