Automatic Question Paper Generator-Bloom's Taxonomy

Uma.D
Department of CSE
Professor
PES University
Bangalore.India
umaprabha@pes.edu

Jyosna.S
Department of CSE
Student
PES University
Bangalore.India
hema.m456@gmail.com

Megha.M
Department of CSE
Student
PES University
Karnataka.India
canonmegha@gmail.com

Radhika. Sadanand
Department of CSE
Student
PES University
Karnataka.India
radhikasadanand55@gmail.com

Roshan Poojary
Department of CSE
Student
PES University
Karnataka.India
roshanpoojary0@gmail.com

Abstract—**Assessments or exams generally play an essential role in education and as well as in a student's life. It is also considered as the primary indicator of students' learning process. Facing exams is also considered as academic readiness and also learning progress. One of the main forms of the assessment used is exams in learning.**

Based on the classification process, a question paper can be redesigned by examiners. This study follows a rule-based approach to achieve multiclass classification. Questions are collected from University Question Papers, Online sources and Textbooks. These questions are related to various subjects under Computer Science like Operating Systems, Databases, Machine Learning, and Networking etc. These questions are pre-processed. To make it more dynamic and self-learning, we replaced the rule-based classifier with an SVM (Support Vector Machine) Classifier.

Then the method of weighting the classes accordingly is used to correctly assign the categories to the questions. This weighting technique is developed using Bloom's Taxonomy verbs. Once the questions are classified, based on user-specified input for bloom's levels expected in the question paper the questions are grouped and a question paper is generated. Results show that the hybrid classifier performs better than the SVM classifier alone.

Keywords—Bloom's Taxonomy, Verbs, question, question paper.

I. INTRODUCTION

In the present world of education, the assessment process is a dominant and vital venture for educational establishments and universities for student performance evaluation. It is a demanding, time consuming and very tedious job for the tutors to come up with worthy questions.

There is much taxonomy that has been developed in order to define questions so that each question has significance for its presence. It might be conversational, accuracy-based, social or knowledge according to one Taxonomy whereas another research says one can classify questions into easy, hard and moderate but there was a need of in-depth classification of the cognitive levels of assessment.

Then in 1956 an American educational psychologist Benjamin Samuel Bloom came up with a Rubric that systematically classified the cognitive levels. This was later named as Bloom's Taxonomy. This Taxonomy has three ranked models which are cognitive, psychomotor and affective domains.

The cognitive domain is a knowledge-based model which is divided into six levels of objectives. These six levels are named as knowledge/remember, analysis/analyze, application/apply, comprehension/understand, evaluation/evaluate and synthesis/create. The affective domain is an emotion-based model which is split into five levels. These five levels are responding, receiving, valuing, characterizing and organizing.

The psychomotor domain is an action-based model which is divided into seven levels. These seven levels are perception, guided response, set, mechanism, origination, complex overt response and adaptation. Since the main focus here is on examination for educational establishments, the first cognitive (knowledge based) model is considered to be a better fit for this kind of research.

Asking and setting proper assessment questions in order to attain the expected result from the course is a difficult work for the examiner to deal with. So, the research on the whole focuses on categorizing the queries from all the cognitive level domains that follow Bloom's Taxonomy with better model accuracy and on which technical skills are used to equally cover the questions from all the domain levels in bloom's taxonomy.

The rest of the paper is organised as follows. In Section II we brief about the literature survey we have done via various research papers in the field of Machine learning and various other techniques used in classification of questions. In Section III we give a brief about Bloom's taxonomy. In Section IV we give information about the data, techniques that have used. Section V we discuss about the methodology used, Section VI experimentation which tell the performance of our system. Section VII we give information of the results obtained. Section VIII is conclusion and future work.

II. Literature Survey

In this section we going to brief about the papers that where researched. These paper mainly are from fields where ML algorithms and other techniques are used for classification.

In [1], the authors have presented a system which based on a set of rules categorize questions into distinct cognitive categories of Bloom's nomenclature. The rule-based approach makes use of NLP methods to recognize key verbs that aids in the classification of the given input question. Followed by this they use category weighting given by experts to classify the questions into a single class.

In [2], the idea is specifically and mainly centred on classifying the test questions accordingly into its learning objectives in Bloom's taxonomy. Similarity based algorithms like wordnet and cosine algorithms are utilized to produce a set of rules that recognizes the category to which the question belongs to and the weight that are given for each question based on Bloom's taxonomy. The scores of each of this method is combined and a final score for each category is determined where the input gets classified into highest scored category. This method has achieved an accuracy of 70 percent.

In [3], in order to classify the questions automatically, this research suggests a way by drawing out TF-POSIDF and word2Vec as two different features. The calculation of term frequency-inverse document frequency is one feature based on parts of speech, sequence of allocation a significant weight for essential and important words in the input. The process of classification was enhanced using the prior-trained word2vec as the second feature. Then, the blend of these features was sent into three different classifiers; Logistic Regression, K-Nearest Neighbour, and Support Vector Machine for the classification process. The proposed method has achieved an accuracy of 71-83 percent.

In [4], The paper focuses on two descriptions of studies to identify the learning complexity of given cognitive questions automatically. The former approach used is labelled Latent Dirichlet Allocation (LDA). This approach represents texts as arbitrary mixtures over quiescent topics, where everything could be represented by a diffusion of words in the corpus. The second approach uses the Bidirectional Encoder Representations from Transformers famously known as BERT framework that aids the multi-class text classification under deep learning. The algorithm uses prior-trained deep multi-directional presentations from the unprocessed text by jointly setting up either on right and left contexts. This paper has achieved an accuracy of 83 percent.

In [5], This paper focuses on automatic question tagging with the knowledge units. The researchers propose two models for automatic tagging one is position based attention model and a keyword-based model, both these models focus on the idea that answers help in deducing the knowledge units more effectively as in the case of multiple-choice questions. Both these models use the Bi-LSTM to get contextual information from the inputs. The position-based attention model has achieved an accuracy of 73 percent and the keyword- based model has an accuracy of 80 percent.

In [6], the paper discusses a genetic algorithm for classifying the email as spam or non-spam. They extract the keywords from the body of the email. Based on the database corpus they have defined different categories under the class based on which for a particular email the weight of the class is calculated by taking the average of the categories in the class. These weights are then converted into bit strings which are the chromosomes that are fed into the genetic algorithm which performs crossover and the results are generated. The results when observed show improvement as the database corpus size increases and the proposed methodology show an accuracy of 81.7 percent.

In [7], This paper proposes an improvised version of the decision tree classification algorithm by using genetic programming and genetic algorithm. Genetic programming is used for the feature construction and the Genetic algorithm for the feature selection. This Paper uses GP individuals consisting of a number of separate trees/ automatically defined function (ads') to construct the features for the decision tree algorithm. A Genetic Algorithm is then used to select over the original and constructed features for a final hybrid decision tree classifier system. This hybrid classifier is tested with various datasets. Results show that this system was able to outperform the normal decision tree algorithm on eight out of ten datasets when tested and compared.

III. Bloom's Taxonomy

This study needs a detailed understanding of the Bloom's Taxonomy levels. Below shown figure depicts the hierarchical model of the cognitive model of Bloom's Taxonomy starting from remember (being the least) to create (being the highest). The figure also shows what each of these levels signifies.

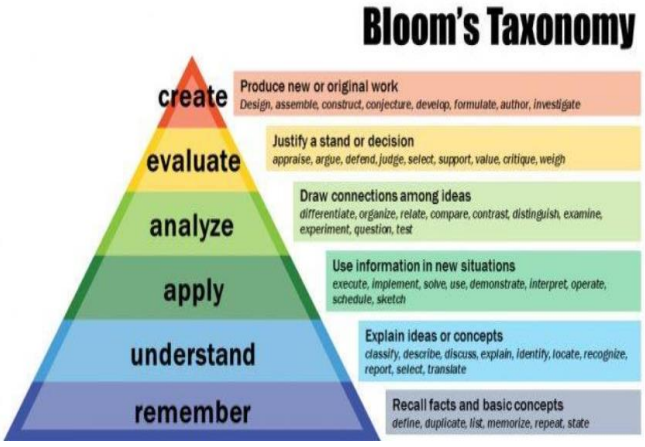

Figure 1: Blooms Taxonomy levels and their significance

IV. PREPARATION OF DATASET AND TRAINING

A. Data Creaction

As seen above, the main requirement with respect to data is to have a proper dataset which consists of a set of questions that are labelled according to the abovementioned Blooms taxonomy levels. The questions are collected from University Question Papers, Online sources and Textbooks. These questions are related to various subjects under Computer Science like Operating Systems, Databases, Machine Learning, Networking etc.,

Here the data set has two columns; one is questions columns and another one is its respective classified classification

column. So when we look at the very first question, the question starts with or has a keyword which is ‘what’, which has been classified under knowledge domain level, just like this remaining keywords like recite, define, name etc., are classified into knowledge domain level. Just like this the classification goes on with different key words in different questions.

The dataset consists of 1100 labelled Questions.

Name the factors that affect the performance of the network?	Knowledge
Define the terms Unicasting, Multicasting and Broadcasting	Knowledge
Describe at one disadvantage of a peer to peer network.	Comprehension
How can you manage a network using a router?	Application
One way of securing a network is through the use of passwords. What can be considered as good passwords?	Analysis
What happens when you use cables longer than the prescribed length?	Knowledge
You need to connect two computers for file sharing. Is it possible to do this without using a hub or router?	Evaluation
When you move the NIC cards from one PC to another PC, does the MAC address gets transferred as well?	Analysis
How are IP addresses arranged and displayed?	Knowledge
When troubleshooting computer network problems, what common hardware-related problems can occur?	Analysis
How does dynamic host configuration protocol aid in network administration?	Application
Explain profile in terms of networking concept?	Comprehension
How do bridges pass spanning tree information between themselves?	Application
Describe a recent short term stressful situation and how you managed it.	Comprehension

Figure 2: Glance of the dataset being used for the research

B. Training the classifier

The dataset discussed above is used to train the classifier .As discussed above in this study we are using the SVM classifier along with a weighting algorithm which will be discussed in the later section . The reason for choosing the SVM classifier is the machine learning technique is a highly preferred Technique for Classification as it produces significant accuracy with less computation power. So the above dataset is used to train the SVM classifier. Later it is also used by the weighting algorithm in order to calculate the probability of the input question belonging to certain categories.

V. PROPOSED METHODOLOGY

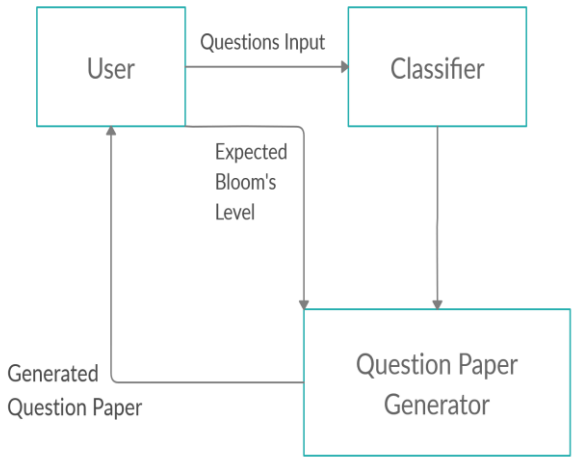

Figure 3: Proposed model

As seen in the above diagram the steps in the proposed methodology are as follows:

A. Collection of input from the user

The first step of the proposed methodology is to collect user input. The user is given with three option in the user interface as shown in the figure below. First option is single question prediction which the user can choose if he wants to known the predicted class of one question. So here the input is taken as a pdf document and the predicted class is displayed. Second option is multiple questions where the user can give a question paper in a csv format and the predicted classes for those question are displayed Then the third option is generating a question where the user can give the question bank in csv format. Later the user is prompted to give the percentages for each Bloom’s level and the number of questions that needs to be there in the final question paper. Taking these input our system generates a question paper in word format.

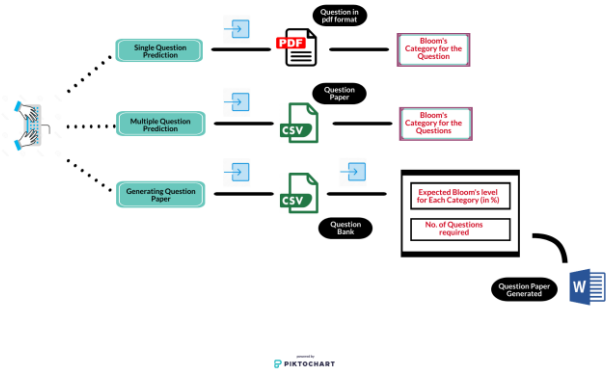

Figure 4: User Interface Diagram

```
for each question in question bank:
    list=SVM_classsification(question)
    weighted_list=weight_algo(question,list)
    find the max (weighted_list)
    classify that question to the max weight category
    write the question into the that particular category dataset
```

Figure 5: Hybrid Classifier Algorithm

B. Classification

In the second step of the methodology the user questions in the provided question bank must undergo classification so that they can be classified into one of the Bloom’s levels. So here the classification takes place in two steps:

1B. SVM Classification

The initial step of classification takes place using the machine learning techniques that is the SVM classifier. As mentioned earlier the SVM is a highly preferred Machine Learning Technique for Classification as it produces significant accuracy with less computation power. The main objective of this algorithm is to find a hyperplane that distinctly classifies the data points, in an N-dimensional space where N is the number of features. Here the data points play an important role in determining the position and the orientation of the hyperplane so it classifies the datapoints of one class from the data points of the other class. Now in this step most of the questions belong to more than two level/categories of the Bloom’s taxonomy. So, in order to classify them into one proper level we do the second part of classification.

2B. Weighting Algorithm

In order to classify the questions further into to one of the categories of the SVM classification we introduce a weight algorithm. The weight algorithm identifies the keyword that is present in the input question then calculates the probability of it belonging to the classes of the SVM classification that is the confidence level of that question belonging to any one of the categories. By this the algorithm gives a score which ranges from 0 to 1 for the categories. Then we classify the question into the category which has more weightage (higher confidence level). Algorithm for the classifier is shown in the figure

C. Question Paper Generation

From the previous phase, we have questions classified into their respective Bloom's categories (exactly one). Now, a question paper has to be generated from a given question bank. Also, the user-specified Bloom's levels (in percentage) that are expected in the question paper must also be taken into consideration while choosing the questions from the question bank.

STEP 1: Once a question bank is selected for generating a question paper, it's passed to the hybrid classifier and the questions are grouped separately based on the predicted Bloom's classes. The same is shown below.

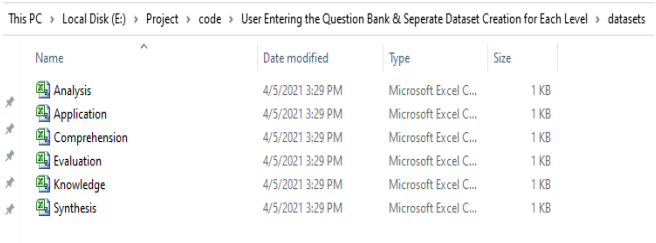

Figure 6: Dataset for each level containing classified questions

STEP 2: Now that we have questions categorized, we consider the percentages of Bloom's levels expected by the user and pick the questions accordingly from each dataset and put them in the final question paper. The total number of questions expected by the user is also considered. All of this is shown in the below figure.

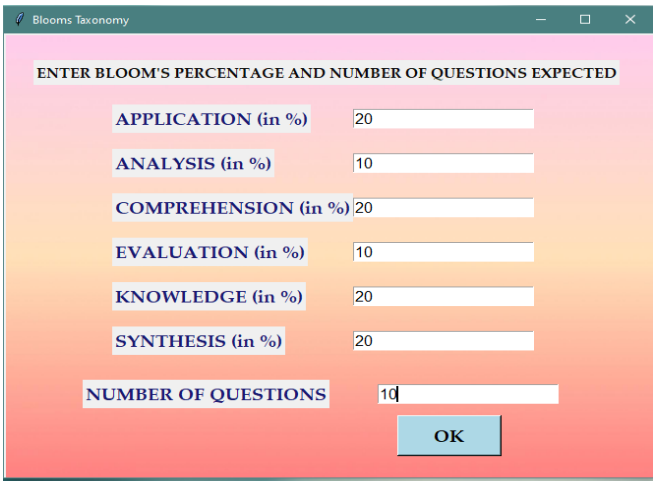

Figure 7: Inputting user specified percentages and number of questions

VI. Experimentation

A. Data Information

The dataset consists of questions that belong to all six categories of Bloom's Taxonomy. The same is shown in the below figure.

```
[nltk_data] Downloading package punkt to
[nltk_data]     C:\Users\mamatha\AppData\Roaming\nltk_data...
[nltk_data]   Package punkt is already up-to-date!
Comprehension    201
Knowledge        191
Analysis         184
Application      179
Synthesis        170
Evaluation       164
Name: Category, dtype: int64
```

Figure 8: Data Information

B. Performance in comparison to other ML techniques

For classification of the questions into Bloom's levels we considered to train different ML techniques with our training dataset to see their performance. So the machine learning considerations are SVM (Support Vector Machine), Linear regression, Multinomial NB and Random forest classifier. After training and testing we obtained the below results.

ML ALGORITHM	ACCURACY
SVM ACCURACY	68.86446
Logistic ACCURACY	58.36441
MultinomialNB ACCURACY	56.36433
Random Forest ACCURACY	48.16674

Figure 9: Accuracy of different ML techniques

VII. Results

So as we see in the above image, the SVM classifier performance is better compared to other techniques so we used the SVM classifier which resulted with multiple categories being predicted for each question. In order to solve this ambiguity the primary classifier was blended with a weighting algorithm based model classifier which reduced the prediction to exactly one. The hybrid model which is SVM combined with weighting technique performs better than other ML algorithm which predict multiple classes.

In research done so far the focus has been given more on developing a set of rules for classification. Even compared to the research done that includes use of ML techniques SVM classifier performs better than all others. Some researches have also implemented weighting technique but the results were not as expected. The weighting algorithm implemented in this work gives weight by focusing more on the keywords of Bloom's levels. Considering all these facts we can say that our approach works better than the already developed ones.

VIII. Conclusion and future work

The hybrid, as in SVM along with the weighting algorithm-based classifier performed better than other ML algorithm which predict multiple classes. This performs better and is also more dynamic than rule-based classifier that we tried to use during the first phase of the project.

As for future work, collecting more keywords belonging to the six different learning levels as much as possible. Adding more rules to the second classifier can lead to the better model. A hybrid of SVM, weighting and rule-based classifier is a good option to go with so that accuracy can be increased by including more rules and including more number of keywords.

Acknowledgment

We would like to express our deep and sincere gratitude to our Chairperson, Shylaja.S. Sharath, Computer Science Department, PES University for giving the opportunity to do this research on sentiment analysis. We would also like to express our heartfelt thanks for the Computer science Department in our college for providing with all the resources throughout the project. This project would have been impossible without the college's guidance and support throughout the project.

Refrerences

[1] Nazlia Omar, Syahidah Sufi Haris, Rosilah Hassan, Haslina Arshad, Masura Rahmat, Noor Faridatul Ainun Zainal, and Rozli Zulkifli. "Automated Analysis of Exam Questions According to Bloom's Taxonomy" Procedia - Social and Behavioral Sciences, vol. 59, 2012. doi:10.1016/j.sbspro.2012.09.278

[2] Jayakodi, Kithsiri & Bandara, Madhushi & Perera, Indika & Meedeniya, Dulani. (2016). WordNet and Cosine Similarity based Classifier of Exam Questions using Bloom's Taxonomy. International Journal of Emerging Technologies in Learning (iJET). 11. 142. 10.3991/ijet.v11i04.5654..

[3] Mohammed M, Omar N (2020) Question classification based on Bloom's taxonomy cognitive domain using modified TF-IDF and word2vec. PLoS ONE 15(3): e0230442. https://doi.org/10.1371/journal.pone.0230442

[4] Das, Syaamantak & Mandal, Das & Basu, Anupam. (2020). Identification of Cognitive Learning Complexity of Assessment Questions Using Multi-class Text Classification. Contemporary Educational Technology. 12. ep275. 10.30935/cedtech/8341

[5] Bo Sun, Yunzong Zhu, Yongkang Xiao, Rong Xiao and Yungang Wei. Automatic Question Tagging with Deep Neural Networks. : DOI 10.1109/TLT.2018.2808187, IEEE .vol 12 issue 1

[6] Mandeep Choudhary , V. S. Dhaka Jaipur National University, Jaipur, India. Automatic e-mails Classification Using genetic Algorithm. Mandeep Choudhary et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 6 (6) , 2015, 5097-5103

[7] Matthew G. Smith and Larry Bull. Feature Construction and Selection Using Genetic Programming and a Genetic Algorithm. C. Ryan et al. (Eds.): EuroGP 2003, LNCS 2610, pp. 229–237, 2003. © Springer-Verlag Berlin Heidelberg 2003