Learning Content Classification and Mapping Content to Synonymous Learners based on 2022 Augmented Verb List of Marzano and Kendall Taxonomy

S. Celine[1]*
Research Scholar
Department of Computer Science
Sacred Heart College, Tirupattur, India

F. Sagayaraj Fransis[3]
Professor, Department of Computer Science
Pondicherry Engineering College
Pondicherry, India

M. Maria Dominic[2]
Assistant Professor
Department of Computer Science
Sacred Heart College, Tirupattur, India

M. Savitha Devi[4]
Assistant Professor, Department of Computer Science
Periyar University Constituent College of Arts and Science
Harur, India

***Abstract*—Finding suitable learning content for learners with different learning styles is a challenging task in the learning process. Hence it is essential to follow some learning taxonomies to deliver learner-centric learner content. Learning taxonomies are used to express various learning practices and learning habits to be followed by the learner for a better learning process. The investigator has already classified the learners based on the 2022 augmented verb list of Marzano and Kendall taxonomy. The main objective of this paper is to minutely classify the tutor-defined learning contents according to the domains as well as the subdomains of the considered taxonomy which is in text format. Providing personalized learning content could help the learners for a better understanding of learning content and their interrelationship which in turn produce better learning outcomes. Mapping the six levels of learning contents into the corresponding learner is a challenging task. Hence the investigator has chosen seven algorithms including Bagging, XG Boost, Support Vector Machine from Machine Learning and four algorithms including Convolutional Neural Network, and Deep Neural Network in Deep Learning algorithm to classify the learning contents. The experimental results indicate that Support Vector Machine performed well in machine learning and Deep Neural Network yields good performance in deep learning in the learning content classification process. These micro contents were organized using a property graph. Further, the micro contents were retrieved from the property graph using SPARQL for mapping the classified contents to the corresponding learners to achieve personalization in the learning process.**

Keywords—Learning taxonomies; marzano and kendall taxonomy; personalization; XG boost; deep neural network; CNN; property graph; action verbs; content classification

I. Introduction

Learning is a process of adapting changes in personal and professional to ameliorate the quality of life. According to Stephen Hawking, Intelligence is the ability to adapt the change. Acquiring intelligence, absorbing, adapting and storing new information in memory is uneven among the learner. Hence it is the need of the hour to identify different learning characteristics of the learner to achieve a better learning outcome. The resource used to provide knowledge to the learner is known as learning content. According to the learner's preference and learning styles, learning content has to be provided to the learners. This process is called personalization in the learning process [1]. Personalized learning must pass some control over the learners, providing some input into how they progress through their learning activities. This can be achieved by adapting learning taxonomies in the learning process. Various taxonomies were developed by researchers in the field of Education and Learning since from the year 1956 [2].

This research work adapted Marzano and Kendall (MK) taxonomy to determine the learning style of the learners. MK taxonomy model provides better knowledge about certain fundamental processes in learning, such as emotion, memory, motivation and metacognition. This model also provides greater precision while creating learning objectives, having a more specific map of the types of knowledge that can be acquired and how they are acquired. Due to this greater precision, it is also possible to evaluate more easily [3]. MK taxonomy has six domains from lower order of thinking skills to higher order of thinking skills.

The investigator prepared the questionnaire based on the 2022 augmented verb list of MK taxonomy to find the learning style of the learner and classified the learners into six domains and 22 sub-domains of the considered taxonomy To classify the learning contents into micro contents the same taxonomy has to be utilized.

Text-based learning content was pre-processed to provide good interpretation and usage. It can also reduce the redundancy in the text content. After the content were pre-processed, it has to be classified based on the considered taxonomy. To accomplish the classification of learning contents into the micro-content process, the investigator has chosen seven algorithms from Machine Learning models such as

*Corresponding Author.

- Naïve Bayes,
- Support Vector Machine (SVM)
- K-Nearest Neighbor (KNN)
- Decision Trees
- Random Forest
- Bagging
- XG Boosting

and four algorithms from Deep Learning models such as

- Deep Neural Network (DNN)
- Recurrent Neural Network (RNN)
- Convolutional Neural Network (CNN)
- Recurrent Convolutional Neural Network (RCNN)

have been considered based on the verb list of six domains and sub-domains of MK's taxonomy.

The classified micro contents were stored in file format. And these micro contents can be represented using Property Graph also called a labelled property graph since it contains nodes(entities), edges(relationships) and properties(attributes). This research work creates ontology for MK Taxonomy to provide learning contents based on the weightage. In ontology, individuals are created for each micro-content with the annotation properties of learning content, keywords and file size. Once all the terms are arranged, the data can be retrieved using the SPARQL query. The representation in the property graph is visualized using the OWLGrEd Visualization tool.

Further, the researcher evaluated the performance of each model and compared them according to precision, recall, F1 score and accuracy. Classified micro contents obtained from the classifiers were mapped to the synonymous learners based on the maximum score on the accuracy of the model.

The rest of the paper is systematized as follows. Section II provides an overview of the related works. Section III provides the design and methodology of the proposed method. Section IV expresses evaluation and results and discussion. Section V illustrates the way to represent the learning content organization using a property graph and the method to extract the contents using SPARQL. Section VI discusses mapping the learning micro-content into the corresponding learner according to six domains and 22 sub-domains of the considered taxonomy. Section VII presents the conclusion and Section VIII illustrates the case study of the proposed method.

II. Related Works

A. Action Verbs

To express the noticeable behaviour of the learner the learning objective must start with action verbs. Action verbs were used to monitor the learner and the throughput of the learning objectives. Choosing the right verb for different types of the learner is an art [4]. The verb list of Marzano and Kendall Taxonomy was first published in 2007 and it needed an up-to-date update to include the later verbs. This is because the recent education system utilizes new vocabularies as per the current technology. The existing action verb list in the taxonomy may not be fulfilling to achieve the throughput of learning objectives. Hence it is the need of the hour to augment the verb list of Marzano and Kendall Taxonomy.

Augmentation is achieved by gathering suitable verbs from sixteen existing taxonomies and open domains. Hence the researcher has made an exhaustive search to update the verb list from 95 to 360 verbs as shown in Table I.

TABLE I. AUGMENTATION OF VERBS IN MK TAXONOMY

Domain / Level	**Sub-domain/Level**	**No. of Existing Verbs**	**No. of Extended Verbs**	**Total Number of Verb List**
Self-System Thinking	Examining Importance	01	15	16
	Examining Efficacy	01	14	15
	Examining Emotions	01	17	18
Metacognition	Examining Motivation	01	14	15
	Specifying Goals	02	13	15
	Process Monitoring	01	13	14
	Monitoring Clarity	01	06	07
	Monitoring Accuracy	02	12	14
Knowledge Utilization	Investigating	07	15	22
	Experimenting	05	14	19
	Problem Solving	06	10	16
	Decision making	04	11	15
Analysis	Specifying	04	11	15
	Generalize	05	06	11
	Analyzing errors	08	11	19
	Classifying	04	09	13
	Matching	08	10	18
Comprehension	Symbolizing	09	08	17
	Integrating	03	17	20
Retrieval	Executing	06	16	22
	Recalling	12	12	24
	Recognizing	04	11	15
	Total No. of Verbs	**95**	**265**	**360**

B. Adaptive Learning Path and Contents

A learning path is a progression of activities and concepts to be chosen by the learner to construct their knowledge or skills in a specific area. Traditional learning system provides the same content and learning path to all learners. But the learner's knowledge, circumstance, and preference are different, and their performance and satisfaction may decrease if they have been given the same content and learning path [5]. Presenting learner-centric learning content can ameliorate the effectiveness and performance of the learning process. To achieve this goal the researcher presented a new model as shown in Fig. 1.

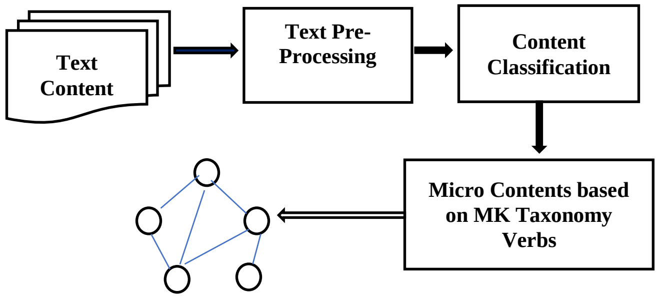


Fig. 1. A Text Content Classification Model.

These micro contents can be represented using Property Graph. Then the micro-content will be retrieved using SPARQL from the property Graph. Further, these micro contents are mapped into the corresponding learners according to six domains and sub-domains of MK taxonomy based on the descending order of the size of the micro contents.

C. Existing Content Classification Approaches

Learning Contents can be in the form of images, text, audio, animations, and video. Text contents can be classified as rule-based, supervised learning-based, and combined classifier-based approaches.

A set of handcrafted rules are utilized in the rule-based approach. In supervised learning text classification approach classification made based on learning past observations. The combined classifier utilized both a machine learning trained base classifier and a rule-based classifier for showing improvement in the throughput [6]. Table II illustrate the various researchers who proposed their model for the classification of questions into Bloom's taxonomy only on a cognitive level. This research work classifies the learning contents into six domains and 22 sub-domains of MK taxonomy.

TABLE II. VARIOUS STUDIES WERE CARRIED OUT TO CLASSIFY THE CONTENTS

S. No.	Name of the Researchers	Model Applied
1	Syahidah Sufi Haris et al [7]	Rule-Based Classification
2	Indika Perera et al [8]	Rule-Based Classification with n-gram Statistical Approach
3	Wen Chih Chang et al [9]	Rule-Based Classification with weighted Technique
4	Anbuselvan Sangodiah et al [10]	Support Vector Machine (SVM)
5	Anwar ali Yahya et al [11]	Support Vector Machine
6	Addin Osman et al [12]	Naive Bayes (NB), SVM, Logistic Regression, and Decision Tree
7	Norazah Yusof et al [13]	Artificial Neural Network
8	Dhuha Abdulhadi Abduljabbar et al [14]	SVM, NB and KNN use a majority voting algorithm.
9	Ali Danesh et al [15]	Combine three classifiers such as NB, KNN and Rocchio
10	Julio Villena Roman et al [16]	K- Nearest Neighbour

III. PROPOSED METHOD: DESIGN AND METHODOLOGY

Learning style is the strategy to accommodate receiving and processing the received information which are two phases of learning [17]. The process of recognizing the behaviour of the learner then spontaneously generates a natural learning path, and tailoring the learning contents to an individual learner is known as adaptation in learning which is the prime need for personalized learning [18]. Learning taxonomy can be employed to understand the learning levels of the learners scientifically. Hence this research focused on classifying the learning contents based on a learning taxonomy for better-personalized learning.

This research focused only on text learning contents. These learning contents were preprocessed and classified based on the augmented verb list of MK Taxonomy into micro contents. Then the suitable micro contents were assigned to the corresponding learner to accomplish the personalized teaching-learning process. Fig. 2 depicts the design architecture for text-based content classification. The design contains two main modules a pre-processing module and a Classification module.

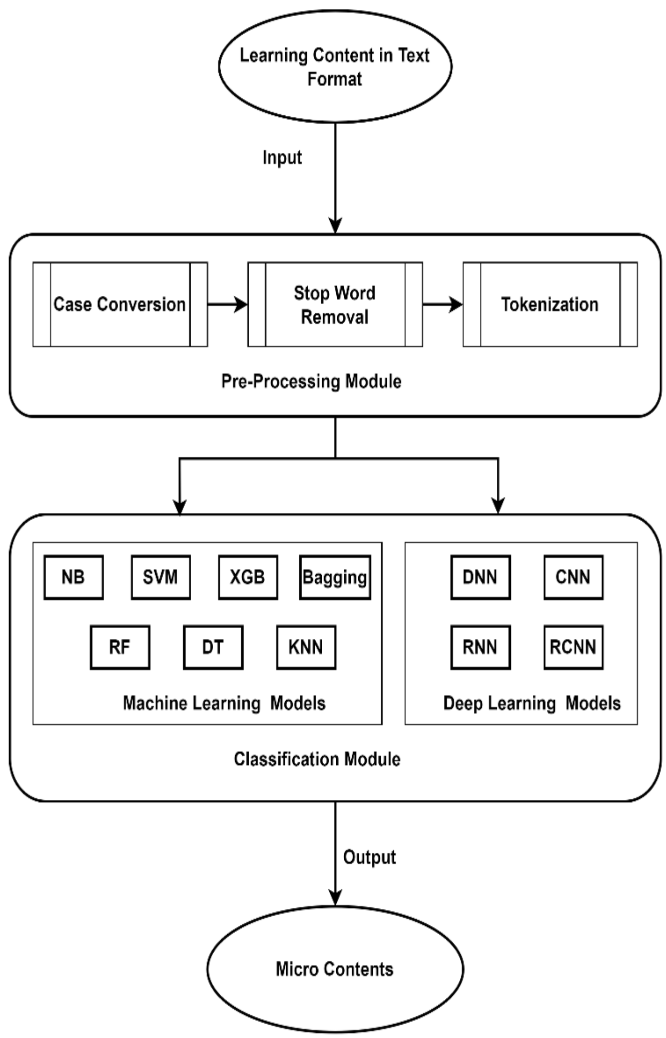


Fig. 2. Design Architecture of Content Classification.

The pre-processing module is the first module of content classification. Text pre-processing is essential for eliminating all the irrelevant objects from the data and making it ready for further processing. This is because raw text data might have insignificant text which makes it difficult to understand and investigate. Hence proper pre-processing must be implemented on raw text data [19]. This research work utilizes three pre-processing techniques as Case Conversion, Stop word removal, and Tokenization.

Case Conversion: Converting all the text content into the lower case is utilized to discard unproductive words [20].

Stop Word Removal: Articles, prepositions, pronouns, and conjunctions in any language are called stop words. "The", "a", "an", "so", and "what" are examples of stop words in English. Removal of such words would help in the size reduction of a dataset and further the training time can also be reduced due to the lesser number of tokens involved in the training [21].

Tokenization: Splitting text contents into smaller units is known as tokenization. The individual units are called tokens. Tokens can be words, phonemes, or maybe full sentences [22]. This research work utilizes sentence tokenization. The learning contents were divided into sentences and considered tokens.

In the e-learning environment, a large volume of learning materials was available in various formats. But it is necessary to provide appropriate learning content to the respective learners according to the six domains and sub-domains of MK taxonomy.

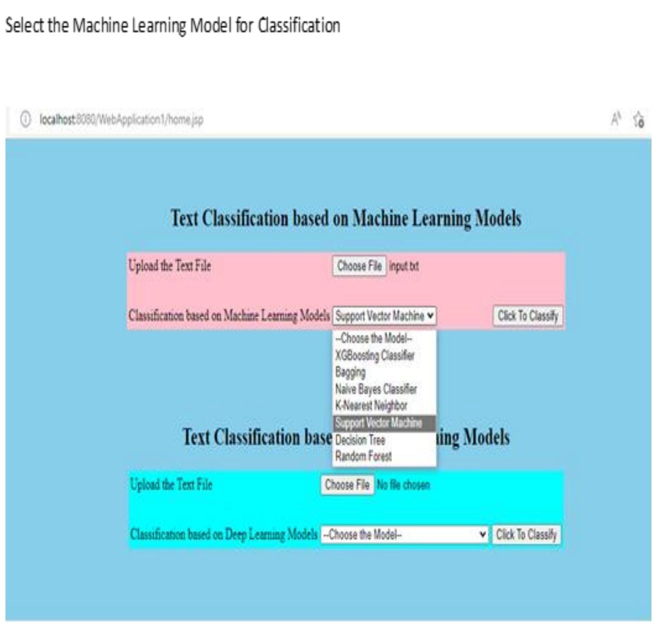


Fig. 3. Text Classification based on Machine Learning Models.

The classification module is the second important module of the design. The preprocessed text contents are classified using seven machine learning models Naive Bayes, SVM, Decision Trees, Random Forest, KNN, Bagging, XG Boosting as shown in Fig. 3, and four deep learning models such as DNN, CNN, RNN, RCNN algorithm based on the action verb list of 2022 augmented verb list of MK Taxonomy as shown in Fig. 4.

This study utilizes NetBeans IDE open-source integrated development environment using Java and libraries such as NLKT, pandas, TensorFlow, NumPy, sklearn, text blob, and seaborn for the classification process.

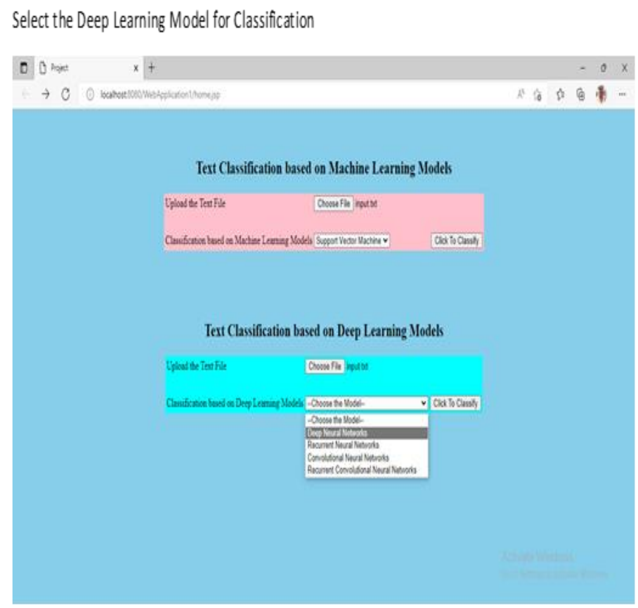


Fig. 4. Text Classification based on Deep Learning Models.

IV. Results and Discussions

For the evaluation process, this study has selected a dataset from the responses received from hundred students for three different learning contents. In a convenient sampling technique analysis on data-set can be carried out either by taking multiple sampling or by repeating the survey. The researcher adapted multiple sampling techniques to produce a reliable result.

A. Evaluation Metrics

Accuracy, precision, recall and f1-score are the measures for evaluation utilized by this study for understanding, measuring relevance and correctness of classification of learning content into micro-contents. Accuracy is used to check the correctness of the model. The exactness of the results is expressed by precision. The completeness of the quality of the results was measured by a recall. F1-score is the weighted average of precision and recall. F1-score is used to evaluate the binary classification system [23].

The evaluation of this study was performed based on the number of keywords classified per domain of MK Taxonomy. A maximum of ten keywords were considered for the classification of learning content into micro-content in each domain of the considered taxonomy.

B. Experiments

The experiments were conducted both on machine learning models and deep learning models and categorized into two.

Experiment 1: Analyze the results of individual Machine learning and deep learning classifier models.

Experiment 2: Results Analysis based on the evaluation metrics.

Experiment 1: Results of Individual Classifier Models

Table III represents the evaluation metric for the XG Boosting classifier in the machine learning model.

TABLE III. EVALUATION METRIC FOR XG BOOSTING CLASSIFIER

Marzano and Kendall Taxonomy Levels	No. of Keywords Classified per level	Accuracy	Precision	Recall	F1-Score
	1	68	0.69	0.7	0.69
Retrieval	2	69	0.7	0.68	0.69
Comprehension	3	71	0.64	0.72	0.68
Analysis	4	78	0.79	0.79	0.79
Knowledge Utilization	5	74	0.83	0.64	0.72
Meta Cognition	6	83	0.81	0.84	0.82
Self -System Thinking	7	82	0.84	0.75	0.79
	8	88	0.9	0.86	0.88
	9	88	0.9	0.85	0.88
	10	92	0.93	0.86	0.9
	Avg.	**80**	**0.8**	**0.77**	**0.78**

The overall accuracy of this classifier is 80%. The percentage of all the measures will be incremented if the number of keywords is increased in each level of MK Taxonomy.

Table IV represents the evaluation metric for the Bagging classifier in the machine learning model.

TABLE IV. EVALUATION METRIC FOR BAGGING CLASSIFIER

Marzano and Kendall Taxonomy Levels	No. of Keywords Classified per level	Accuracy	Precision	Recall	F1-Score
	1	65	0.57	0.74	0.65
Retrieval	2	59	0.60	0.56	0.58
Comprehension	3	59	0.62	0.54	0.58
Analysis	4	56	0.54	0.57	0.55
Knowledge Utilization	5	59	0.63	0.54	0.58
Meta Cognition	6	66	0.68	0.62	0.65
Self -System Thinking	7	53	0.55	0.46	0.50
	8	75	0.77	0.67	0.72
	9	81	0.79	0.82	0.80
	10	77	0.74	0.77	0.76
	Avg.	**65**	**0.65**	**0.63**	**0.64**

TABLE V. EVALUATION METRIC FOR NAÏVE BAYES CLASSIFIER

Marzano and Kendall Taxonomy Levels	No. of Keywords Classified per level	Accuracy	Precision	Recall	F1-Score
	1	71	0.81	0.65	0.63
Retrieval	2	74	0.82	0.65	0.72
Comprehension	3	75	0.67	0.78	0.73
Analysis	4	78	0.86	0.77	0.72
Knowledge Utilization	5	83	0.89	0.75	0.81
Meta Cognition	6	84	0.93	0.69	0.82
Self -System Thinking	7	86	0.85	0.92	0.80
	8	91	0.94	0.93	0.88
	9	91	0.92	0.90	0.93
	10	93	0.89	0.97	0.91
	Avg.	**83**	**0.86**	**0.80**	**0.80**

Table V represents the evaluation metric for the Naïve Bayes classifier in the machine learning model. This classifier achieved a considerable score in precision measurement. This shows the exactness of the results.

The evaluation metric for the SVM classifier is illustrated in Table VI. This study observed that the SVM classifier successfully classifies the content with much accuracy since the overall accuracy of the SVM classifier is 86%.

Tables VII, VIII and IX represent the evaluation metrics for KNN, Decision Trees and Random Forest classifiers.

Table X depicts the evaluation metrics for all the seven models in machine learning models.

TABLE VI. EVALUATION METRIC FOR SVM CLASSIFIER

Marzano and Kendall Taxonomy Levels	No. of Keywords Classified per level	Accuracy	Precision	Recall	F1-Score
	1	79	0.76	0.80	0.78
Retrieval	2	76	0.77	0.73	0.75
Comprehension	3	74	0.71	0.76	0.74
Analysis	4	85	0.84	0.86	0.85
Knowledge Utilization	5	84	0.87	0.76	0.81
Meta Cognition	6	86	0.83	0.91	0.87
Self -System Thinking	7	91	0.92	0.91	0.91
	8	95	0.95	0.95	0.95
	9	93	0.92	0.95	0.93
	10	97	0.96	0.98	0.97
	Avg.	**86**	**0.85**	**0.86**	**0.86**

TABLE VII. EVALUATION METRIC FOR KNN CLASSIFIER

Marzano and Kendall Taxonomy Levels	No. of Keywords Classified per level	Accuracy	Precision	Recall	F1-Score
	1	56	0.50	0.61	0.55
Retrieval	2	57	0.56	0.57	0.57
Comprehension	3	56	0.53	0.58	0.56
Analysis	4	58	0.59	0.56	0.57
Knowledge Utilization	5	66	0.69	0.60	0.64
Meta Cognition	6	53	0.58	0.45	0.51
Self -System Thinking	7	73	0.75	0.69	0.72
	8	83	0.84	0.81	0.82
	9	75	0.77	0.72	0.74
	10	85	0.85	0.84	0.84
	Avg.	**67**	**0.67**	**0.64**	**0.65**

TABLE VIII. EVALUATION METRIC FOR DECISION TREE CLASSIFIER

Marzano and Kendall Taxonomy Levels	No. of Keywords Classified per level	Accuracy	Precision	Recall	F1-Score
	1	42	0.42	0.42	0.42
Retrieval	2	55	0.51	0.56	0.53
Comprehension	3	43	0.46	0.42	0.44
Analysis	4	55	0.50	0.56	0.53
Knowledge Utilization	5	49	0.50	0.47	0.44
Meta Cognition	6	64	0.66	0.73	0.53
Self -System Thinking	7	58	0.60	0.59	0.48
	8	68	0.66	0.72	0.69
	9	56	0.53	0.55	0.59
	10	67	0.68	0.66	0.69
	Avg.	**56**	**0.55**	**0.57**	**0.53**

TABLE IX. EVALUATION METRIC FOR RANDOM FOREST CLASSIFIER

Marzano and Kendall Taxonomy Levels	No. of Keywords Classified per level	Accuracy	Precision	Recall	F1-Score
	1	66	0.56	0.69	0.62
Retrieval	2	74	0.67	0.78	0.72
Comprehension	3	67	0.67	0.67	0.67
Analysis	4	75	0.71	0.78	0.74
Knowledge Utilization	5	74	0.78	0.68	0.73
Meta Cognition	6	86	0.74	0.92	0.82
Self -System Thinking	7	80	0.81	0.79	0.80
	8	90	0.90	0.89	0.90
	9	85	0.80	0.89	0.84
	10	92	0.90	0.93	0.91
	Avg.	**79**	**0.75**	**0.80**	**0.78**

TABLE X. CONSOLIDATION OF THE EVALUATION METRICS FOR MACHINE LEARNING MODELS

Machine Learning Models	Accuracy (%)	Precision	Recall	F1-Score
XG Boosting	80	0.80	0.77	0.78
Bagging	65	0.65	0.63	0.64
Naïve Bayes (NB)	83	0.86	0.80	0.80
K-Nearest Neighbor	67	0.67	0.64	0.65
Support Vector Machine (SVM)	86	0.85	0.86	0.86
Decision Tree (DT)	56	0.55	0.57	0.53

The evaluation metric for Deep Neural Network is illustrated in Table XI. It is observed that the DNN classifier successfully classified the contents because the accuracy of the classifier is 83% which is high score than the remaining classifiers considered in this study.

The evaluation metrics for CNN and RNN were represented in Tables XII and XIII.

TABLE XI. EVALUATION METRIC FOR DEEP NEURAL NETWORK (DNN) CLASSIFIER IN DEEP LEARNING MODEL

Marzano and Kendall Taxonomy Levels	No. of Keywords Classified per level	Accuracy	Precision	Recall	F1-Score
	1	76	0.67	0.80	0.73
Retrieval	2	72	0.82	0.63	0.71
Comprehension	3	74	0.76	0.69	0.72
Analysis	4	76	0.65	0.86	0.74
Knowledge Utilization	5	78	0.84	0.75	0.79
Meta Cognition	6	85	0.82	0.87	0.84
Self -System Thinking	7	87	0.86	0.90	0.88
	8	93	0.95	0.91	0.93
	9	92	0.91	0.92	0.92
	10	96	0.98	0.92	0.95
	Avg.	**83**	**0.83**	**0.83**	**0.82**

TABLE XII. EVALUATION METRIC FOR RECURRENT NEURAL NETWORK (RNN) CLASSIFIER

Marzano and Kendall Taxonomy Levels	No. of Keywords Classified per level	Accuracy	Precision	Recall	F1-Score
	1	71	0.72	0.68	0.71
Retrieval	2	68	0.76	0.62	0.69
Comprehension	3	61	0.67	0.58	0.62
Analysis	4	68	0.68	0.67	0.68
Knowledge Utilization	5	74	0.75	0.73	0.74
Meta Cognition	6	79	0.82	0.74	0.78
Self -System Thinking	7	83	0.83	0.83	0.83
	8	86	0.81	0.91	0.86
	9	91	0.92	0.91	0.91
	10	92	0.91	0.94	0.93
	Avg.	**78**	**0.79**	**0.76**	**0.78**

TABLE XIII. EVALUATION METRIC FOR CONVOLUTIONAL NEURAL NETWORK (CNN) CLASSIFIER

Marzano and Kendall Taxonomy Levels	No. of Keywords Classified per level	Accuracy	Precision	Recall	F1-Score
	1	71	0.63	0.74	0.68
Retrieval	2	66	0.74	0.54	0.62
Comprehension	3	68	0.49	0.76	0.60
Analysis	4	64	0.60	0.70	0.64
Knowledge Utilization	5	68	0.79	0.57	0.66
Meta Cognition	6	74	0.73	0.76	0.74
Self -System Thinking	7	79	0.83	0.74	0.78
	8	87	0.86	0.88	0.87
	9	87	0.95	0.78	0.86
	10	93	0.93	0.93	0.93
	Avg.	**76**	**0.76**	**0.74**	**0.74**

The evaluation metric for the Recurrent Convolutional Deep Neural Network (RCNN) Classifier is illustrated in Table XIV. The combination of RNN and CNN is known as an RCNN classifier. The performance metrics of this classifier range from 75% to 78%.

Table XV provides the consolidation of the evaluation metrics for the four deep learning models.

Experiment 2: Overall Result Analysis based on the measures for evaluation.

The overall score obtained by all the classifiers using both machine learning and deep learning models were illustrated in Table XVI.

TABLE XIV. EVALUATION METRIC FOR RECURRENT CONVOLUTIONAL NEURAL NETWORK (RCNN) CLASSIFIER

Marzano and Kendall Taxonomy Levels	No. of Keywords Classified per level	Accuracy	Precision	Recall	F1-Score
	1	72	0.81	0.73	0.58
Retrieval	2	74	0.64	0.83	0.71
Comprehension	3	69	0.77	0.64	0.61
Analysis	4	69	0.84	0.57	0.64
Knowledge Utilization	5	66	0.45	0.78	0.62
Meta Cognition	6	69	0.55	0.69	0.83
Self -System Thinking	7	85	0.87	0.77	0.87
	8	86	0.85	0.79	0.91
	9	95	0.98	0.89	0.96
	10	87	0.93	0.84	0.82
	Avg.	**78**	**0.77**	**0.75**	**0.76**

TABLE XV. CONSOLIDATION OF THE EVALUATION METRICS FOR DEEP LEARNING MODELS

Deep Learning Models	Accuracy (%)	Precision	Recall	F1-Score
DNN	83	0.83	0.83	0.82
RNN	78	0.79	0.76	0.78
CNN	76	0.76	0.74	0.74
RCNN	78	0.77	0.75	0.76

TABLE XVI. CONSOLIDATION OF THE EVALUATION METRICS FOR ALL THE MODELS USED

ML and DL Models	Accuracy %	Precision	Recall	F1-Score
XGB	80	0.8	0.77	0.78
Bagging	65	0.65	0.63	0.64
NB	83	0.86	0.8	0.8
KNN	67	0.67	0.64	0.65
SVM	86	0.85	0.86	0.86
DT	56	0.55	0.57	0.53
RF	79	0.75	0.8	0.78
DNN	83	0.83	0.83	0.82
RNN	78	0.79	0.76	0.78
CNN	76	0.76	0.74	0.74
RCNN	78	0.77	0.75	0.76

Table XVII shows the Accuracy measure values obtained for each classifier and arranged in descending order based on the percentage of Accuracy.

According to the results, the SVM classifier performed well toward the correctness of classification and the accuracy is measured as 86 per cent. Further, the F1-score, the weighted average of precision and recall is also 86% as shown in Table XVIII. The analysis based on accuracy is depicted in Fig. 5.

TABLE XVII. EXPERIMENTAL RESULTS AS PER ACCURACY

Machine Learning and Deep Learning Models	Accuracy %
Support Vector Machine (SVM)	**86**
Naïve Bayes (NB)	83
Deep Neural Networks	83
XG Boosting	80
Random Forest (RF)	79
Recurrent Neural Networks (RNN)	78
Recurrent Convolutional Neural Networks (RCNN)	78
Convolutional Neural Networks (CNN)	76
K-Nearest Neighbor	67
Bagging	65
Decision Tree (DT)	56

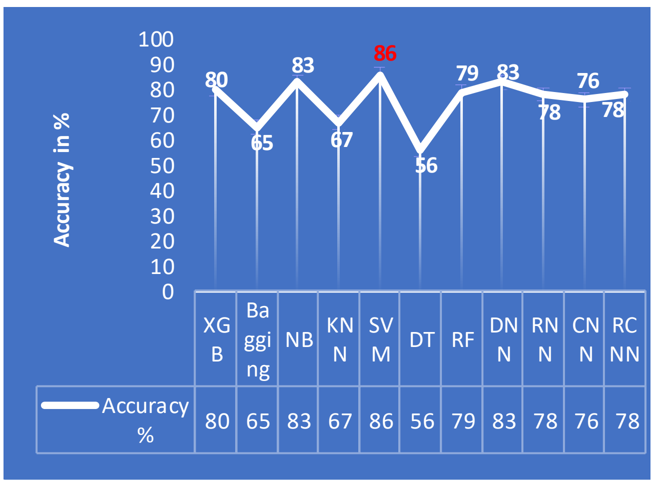


Fig. 5. Analysis based on Accuracy.

TABLE XVIII. EXPERIMENTAL RESULTS AS PER F1-SCORE

Machine Learning and Deep Learning Models	F1-Score
Support Vector Machine (SVM)	0.86
Deep Neural Networks	0.82
Naïve Bayes (NB)	0.8
XG Boosting	0.78
Random Forest (RF)	0.78
Recurrent Neural Networks (RNN)	0.78
Recurrent Convolutional Neural Networks (RCNN)	0.76
Convolutional Neural Networks (CNN)	0.74
K-Nearest Neighbor	0.65
Bagging	0.64
Decision Tree (DT)	0.53

As per the analysis, it is observed that the Naïve Bayes classifier achieved a considerable value in precision. The higher precision indicates that, less false positive measure. It shows the exactness of the classification of learning contents. Table XIX represents the experiment results based on the precision measure.

TABLE XIX. EXPERIMENTAL RESULTS AS PER PRECISION

Machine Learning and Deep Learning Models	Precision
Naïve Bayes (NB)	**0.86**
Support Vector Machine (SVM)	0.85
Deep Neural Networks	0.83
XG Boosting	0.8
Recurrent Neural Networks (RNN)	0.79
Recurrent Convolutional Neural Networks (RCNN)	0.77
Convolutional Neural Networks (CNN)	0.76
Random Forest (RF)	0.75
K-Nearest Neighbour	0.67
Bagging	0.65
Decision Tree (DT)	0.55

The completeness of the quality of the results was measured by a recall. SVM classifier again occupies the top place among other classifiers for the completeness of the classification of learning content according to keywords of MK Taxonomy. Table XX illustrate the experiment results based on recall measure.

TABLE XX. EXPERIMENTAL RESULTS AS PER RECALL

Machine Learning and Deep Learning Models	Recall
Support Vector Machine (SVM)	**0.86**
Deep Neural Networks	0.83
Naïve Bayes (NB)	0.8
Random Forest (RF)	0.8
XG Boosting	0.77
Recurrent Neural Networks (RNN)	0.76
Recurrent Convolutional Neural Networks (RCNN)	0.75
Convolutional Neural Networks (CNN)	0.74
K-Nearest Neighbor	0.64
Bagging	0.63
Decision Tree (DT)	0.57

According to the above analysis, this study concludes that the SVM classifier model provides more accuracy. Hence the micro-contents classified by utilizing the SVM classifier are considered for mapping to the synonymous learner based on the verb list of MK Taxonomy.

V. Learning Content Organization and Retrieval

Information can be represented in the form of tables, charts and graphs to support organizing, analyzing and fetching them when required. Graphs are shines well in representing the connections and relationships among diverse data.

A graph in which the connections between nodes are used to represent the relations along with name and some properties is called Property Graphs. Nodes, labels, relationships, and properties are the components of a property graph. Relationships and connecting nodes of data are capable of storing properties. Appropriate and easy-to-recognize labels are utilized by the property graph for modelling data and its connections. This structured form of data can be easily understood by laymen [24].

This research work utilizes the property graph to represent the classified micro contents according to MK taxonomy. Visualization of six domains and sub-domains and the corresponding micro-content of MK Taxonomy using OWLGrEd is depicted in Fig. 6. OWLGrEd is an editor for OWL to represent graphical notation in an ontology. Ontology becomes common in the fields of artificial intelligence and machine learning, where knowledge plays a vital role. Ontology creates a generic vocabulary which can be shared with researchers and different stack holders. It consists of machine-interoperable definitions of the domain concepts and the relationship between them. It enables the researcher to retrieve data based on knowledge which is known as knowledge-based retrieval. Ontology shares the knowledge to understand the structure of information which can be reused from the domain knowledge. This feature motivates to development of learning content to enhance the learner to gain knowledge of the subject based on their interest and learning ability.

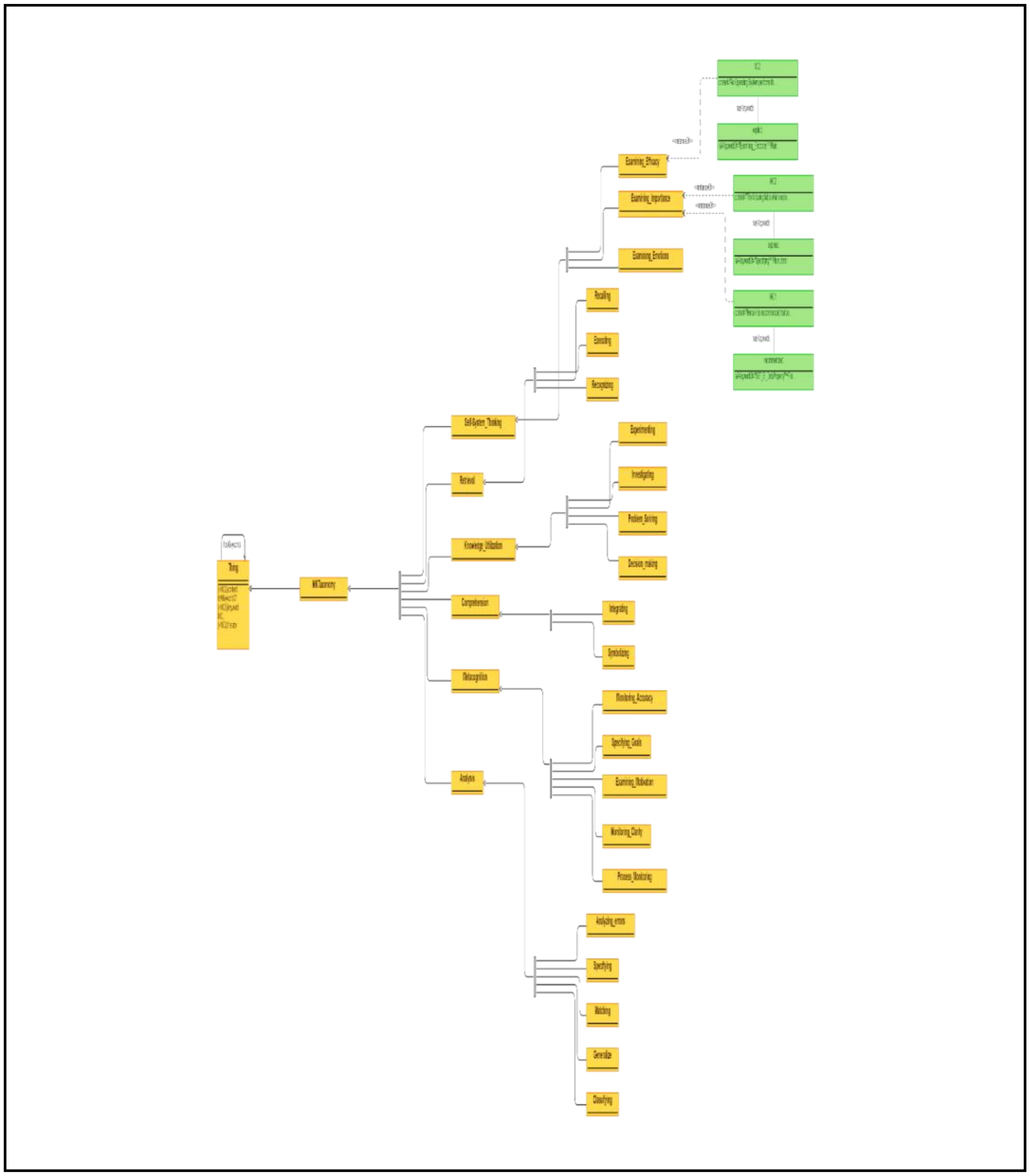


Fig. 6. Visualization of Six Domains and Sub-Domains and the Corresponding Micro-Content of MK Taxonomy using OWL.

Ontology is a collection of classes, properties, instances and axioms. Classes are also known as the concepts of the domain, properties define the relationship between the concepts, instances are the individuals of each class, and axioms denote the restrictions. Ontology can be defined as, a formal explicit specification of a shared conceptualization'. The key terms of a domain are identified and arranged hierarchically and the relationships between the terms are established before developing ontologies.

This research work creates ontology for MK Taxonomy to provide learning contents based on the weightage. The levels and sublevels of the considered taxonomy are arranged as classes hierarchically to frame ontology using the Protégé ontology development tool.

The level/domain of learning is identified through the keywords used in the learning content. Each level of MK taxonomy contains a different set of keywords to group the learning content. The keywords are listed as individuals and the relationship between the classes and keywords is established.

The learning contents were partitioned into micro contents to improve the learning ability of the learner. In ontology, individuals are created for each micro-content with the annotation properties of learning content, keywords and file size.

Micro-content (MC) can be represented as

$$MC_{ij} = \{K_{ij}, C_{ij}, FS(C_{ij})\} \quad (1)$$

Where

i represents domains of MK Taxonomy,

j represents sub-domains of MK Taxonomy,

K is a Keywords,

C is a Learning Content,

FS is the File Size of the learning content.

In this study, MC_{11} represents a micro-content in the sub-domain Recognizing in the domain Retrieval. Each micro-content is defined with these annotation properties to retrieve the content based on the file size given in Fig. 7.

Each micro-content is related to the type of class and object property it belongs. Variable content is created to hold the value of micro-content. Once all the terms are arranged, the data can be retrieved using the SPARQL query.

The SPARQL query to retrieve the micro-content based on the file size in descending order is given below and the result is shown in Fig. 8.

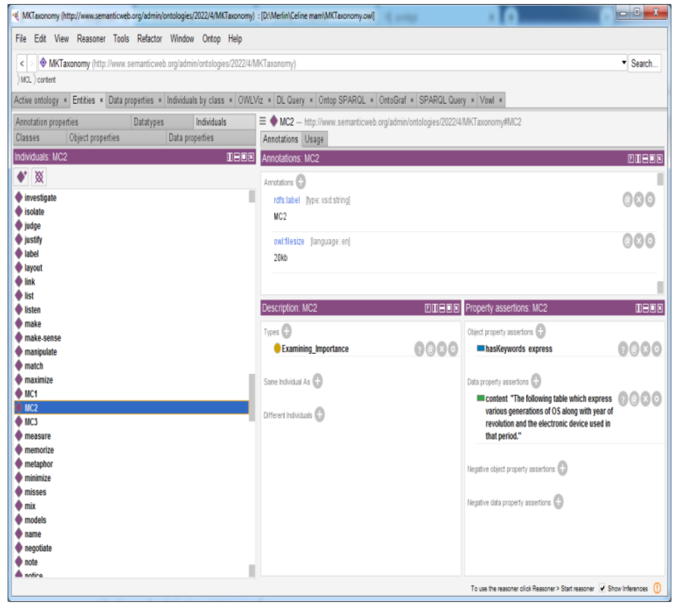


Fig. 7. Creation of Individuals for Micro-contents.

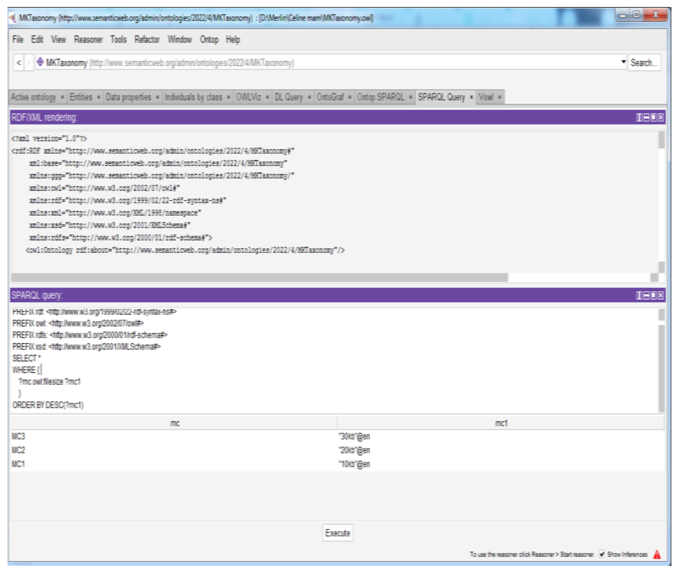


Fig. 8. Retrieval of the Micro-Contents using SPARQL.

SPARQL query to Retrieve the micro-content

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
SELECT *
WHERE {
 ?mc owl:filesize ?mc1
 }
ORDER BY DESC(?mc1)
```

VI. Mapping the Microlearning Content to the Synonymous Learners

The classified learning contents which are retrieved from the graph were mapped to the corresponding leaner to achieve personalization in the learning process. Each micro-content with learning content, keywords and file size defined with annotation properties was used to retrieve the content. Further, these retrieved micro-contents were arranged in descending order based on the size of the files. Based on the score obtained by the learner, they were classified by the researcher. The micro-content classified based on the keyword under the sub-domain Recognize in domain Retrieval is MC_{11}.

This study proposed a novel method to perform the mapping process. The learners' characteristics were obtained by the response received from them through the tool questionnaire according to the 2022 verb list of MK Taxonomy. Questions were rationalized to 50 according to six levels of MK Taxonomy as 8, 8, 10, 8, 8, 8 which can be considered as weightage (w) for each domain as shown in Fig. 9. Eight questions in D1 in turn sub-dived into 3, 3 and 2.

Dataset has been constructed from the response fetched from the hundred learners. The correct response was represented as 1 and the incorrect response was represented as 0. Further, the total score against each domain was calculated as illustrated in Table XXI. This provides a way to quantify each type of learning style in the learner.

Based on the score (SC) obtained by the learner out of each domain and sub-domains of MK Taxonomy, the number of micro-contents (NMC) retrieved from a graph as per each domain, and file size are the parameters for providing micro contents to the synonymous learner. Equation (2) is utilized for mapping the micro-contents to the corresponding learner.

$$K_{ij} = SC_{ij} / w(SD_{ij}) * NMC_{ij} \tag{2}$$

Where i represent six domains of MK Taxonomy,

j represents sub-domains of MK Taxonomy,

K - Number of micro-contents to be provided to the learner,

SC - Score obtained by the learner,

SD – Sub-Domains of MK Taxonomy,

w – Weightage assigned to SDs as shown in Fig. 9,

NMC – Number of Micro-Contents.

The Pseudo code for the mapping process is illustrated below.

Start the process.
If (i = 1) then j = 1 to 3
{Calculate NMCs for D1
$$K_{ij} = SC_{ij} / w(SD_{ij}) * NMC_{ij}$$
}
If (i = 2) then j = 1 to 2
{Calculate NMCs for D2
$$K_{ij} = SC_{ij} / w(SD_{ij}) * NMC_{ij}$$
}
If (i = 3) then j = 1 to 5
{Calculate NMCs for D3
$$K_{ij} = SC_{ij} / w(SD_{ij}) * NMC_{ij}$$
}
If (i = 4) then j = 1 to 4
{ Calculate NMCs for D4
$$K_{ij} = SC_{ij} / w(SD_{ij}) * NMC_{ij}$$
}
If (i = 5) then j = 1 to 5
{ Calculate NMCs for D5
$$K_{ij} = SC_{ij} / w(SD_{ij}) * NMC_{ij}$$
}
If (i = 6) then j = 1 to 3
{ Calculate N MCs for D6
}
Stop the process.

The MCs were arranged in descending order based on the file size. Hence as per the above calculation, the MCs were mapped to the synonymous learners to achieve personalization in the learning process according to MK Taxonomy.

VII. Conclusion

The main objective of this paper is to specifically classify the learning contents based on the specific characteristics of the learner and according to the domains as well as the subdomains of the considered taxonomy. The learning contents in text format were represented in a property graph and retrieval of the same is achieved to fulfil the personalization process in the learner-centric environment. The learners were classified according to MK Taxonomy. Hence the classified learning contents were assigned to the synonymous learners to achieve personalization in the learning process.

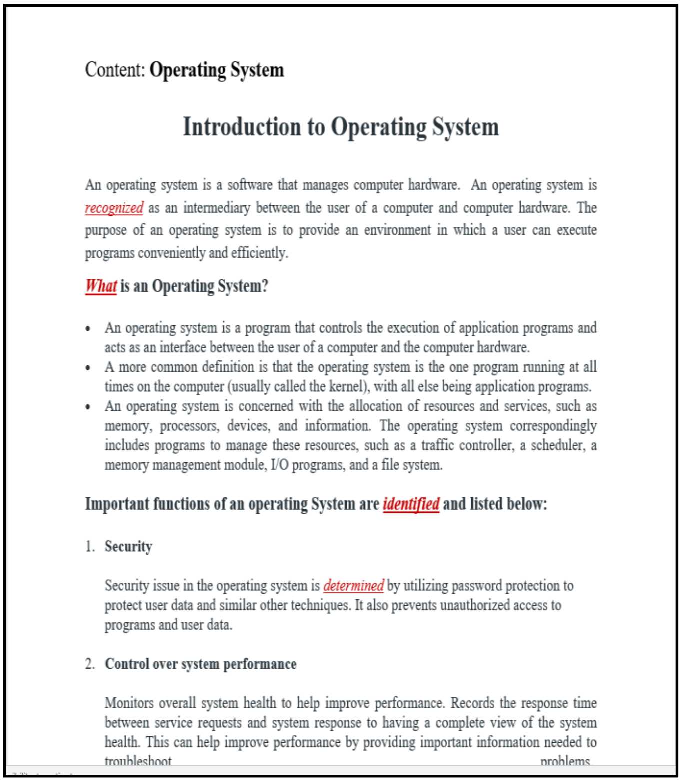
Many researchers classified the learners based on Bloom's Taxonomy's cognitive level. But this research work proposed a novel contribution towards the classification of learning contents into micro contents according to the six domains and 22 sub-domains of MK Taxonomy and represents them using a property graph. Further, these micro contents were retrieved from the graph and mapped to the corresponding learners who were classified according to MK Taxonomy. Hence the learner-centric learning contents were provided to the learners for better learning outcomes.

VIII. CASE STUDY

Learning Contents classification can be carried out by the following steps. Fig. 9 shows the Screenshot of the learning content.

Input: Subject: Operating Systems-Tutor defined Text Contents.

An Operating System is recognized as an intermediate between the user of the computer and computer hardware. Important functions of an operating system are identified and listed below.


Content: **Operating System**

Introduction to Operating System

An operating system is a software that manages computer hardware. An operating system is *recognized* as an intermediary between the user of a computer and computer hardware. The purpose of an operating system is to provide an environment in which a user can execute programs conveniently and efficiently.

What **is an Operating System?**

- An operating system is a program that controls the execution of application programs and acts as an interface between the user of a computer and the computer hardware.
- A more common definition is that the operating system is the one program running at all times on the computer (usually called the kernel), with all else being application programs.
- An operating system is concerned with the allocation of resources and services, such as memory, processors, devices, and information. The operating system correspondingly includes programs to manage these resources, such as a traffic controller, a scheduler, a memory management module, I/O programs, and a file system.

Important functions of an operating System are *identified* and listed below:

1. **Security**

   Security issue in the operating system is *determined* by utilizing password protection to protect user data and similar other techniques. It also prevents unauthorized access to programs and user data.

2. **Control over system performance**

   Monitors overall system health to help improve performance. Records the response time between service requests and system response to having a complete view of the system health. This can help improve performance by providing important information needed to troubleshoot problems.

Fig. 9. Screenshot of Learning Contents.

Step 1: Text content Pre-Processing

Step 1.1: Case Conversion – convert into a lower case

an operating system is recognized as an intermediate between the user of the computer and computer hardware. important functions of an operating system are identified and listed below.

Step1.2: Stop word Removal

the operating system recognized intermediate user computer hardware. Important function operating system identifies list.

Step 1.3: Tokenization (Sentence Tokenization)

Token 1: operating system recognizes intermediate user computer hardware **Token 2:** important function operating system identify the list.

Step 2: Verbs are Extracted from the tokens.

Verb list: recognize, identify, list

Step 3: Classification based on the verb list according to MK Taxonomy domains and sub-domains using seven ML models and four DL models as shown in Fig. 2. As per the performance metrics, the SVM model is used to classify this study. Keywords or the action verbs in MK Taxonomy were utilized for the classification of tokens into micro contents.

Action verbs 'recognize' and 'identify' the sub-domain Recognizing in domain Retrieval. Hence the corresponding MCs were assigned to that sub-domain.

Output:

MC 1: operating system recognizes intermediary user computer hardware.

MC 2: Important functions operating System identify list.

Step 4: These MCs were represented in the property graph as illustrated in Fig. 8 and retrieved using SPARQL.

Step 5: Mapping the MCs to the synonymous learners.

The total number of MCs in Sub-domain1 Recognizing in domain Retrieval were 02. These two MCs were to be mapped to the learners who were already classified under the same sub-domain as shown in Fig. 9 and the score obtained by the learners as shown in Table XXI were applied in the equation (2).

The score obtained by learner 1 in SD1 (SC) = 02

$$NMC = 02$$

$w(SD_{11}) = 03$

By utilizing equation (2) $K_{11} = 02/03*02 = 1.33 \approxeq 02$

Result:

Hence two MCs were provided to the learner in Sub-domain1 Recognizing in domain Retrieval according to MK Taxonomy in a personalized manner.

TABLE XXI. SCORE OBTAINED IN LEVEL 1 (RETRIEVAL) FOR FIFTEEN LEARNERS

Learner ID	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Score for Recognition	Score for Recalling	Score for Executing	Score for Leve 1-Retrieval
L1	1	0	1	1	0	0	0	1	2	1	1	4
L2	0	1	0	0	1	0	1	0	1	1	1	3
L3	1	0	0	1	0	1	0	0	1	2	0	3
L4	1	1	0	1	1	1	0	1	2	3	1	6
L5	0	0	1	1	0	0	1	1	1	1	2	4
L6	1	1	1	0	1	1	1	1	3	2	2	7
L7	1	0	0	0	0	1	1	1	1	1	2	4
L8	0	0	1	0	1	1	1	0	1	2	1	4
L9	1	0	0	0	1	0	1	1	1	1	2	4
L10	1	0	1	1	1	1	1	0	2	3	1	6
L11	1	0	1	0	1	0	0	1	2	1	1	4
L12	0	1	1	1	0	0	0	1	2	1	1	4
L13	1	1	0	1	1	0	0	1	2	2	1	5
L14	0	0	0	1	0	0	1	1	0	1	2	3
L15	1	0	1	1	0	1	1	0	2	2	1	5

REFERENCES

[1] Sinem Aslana , Zehra Cataltepeb , Itai Dinerc , Onur Dundara , Asli A. Esmea , Ron Ferensc , Gila Kamhic , Ece Oktaya , Canan Soysala , Murat Yenera , " Learner Engagement Measurement and Classification in 1:1 Learning ", Department of Computer Engineering, Istanbul Technical University, Istanbul, Turkey.

[2] Cox, R.C. & Wildeman, C.E. (Eds.) (1970), " Taxonomy of Educational Objectives: Cognitive Domain; An Annotated Bibliography", Pittsburgh, PA: Learning and Research Centre.

[3] Marzano, R.J. (2001), " Designing a New Taxonomy of Educational Objectives", Thousand Oaks, CA: Corwin Press.

[4] https://study.com/academy/lesson/using-action-verbs-for-learning-objectives.html.

[5] Ahmad kardan, Maryam Bahojb Imani, Molood Ale Ebrahim, " A Novel Adaptive Learning Path Method", The 4th International Conference on e-Learning and e-Teaching, ICELET , February 2013.

[6] https://monkeylearn.com/text-classification/.

[7] Syahidah Sufi Haris and Nazlia Omar, " Determining Cognitive Category of Programming Question with Rule-based Approach", International Journal of Information Processing and Management, 4(3), 86-95, 2013.

[8] K. Jayakodi, M. Bandara, I. Perera, and D. Meedeniya. WordNet and cosine similarity based classifier of exam questions using bloom's taxonomy. International Journal of Emerging Technologies in Learning, 11(4):142–149, 2016.

[9] Wen Chih Chang and Ming Shun Chung. Automatic applying Bloom's taxonomy to classify and analysis the cognition level of english question items. 2009 Joint Conferences on Pervasive Computing, JCPC 2009, pages 727–733, 2009.

[10] Anbuselvan Sangodiah, Rohiza Ahmad, Wan Fatimah, and Wan Ahmad. A Review in Feature Extraction Approach in Question Classification Using Support Vector Machine. 2014 IEEE International Conference on Control System, Computing and Engineering, (November):536–541, 2014.

[11] AA Yahya and A Osman. Automatic classification of questions into Bloom's cognitive levels using support vector machines. In The International Arab Conference on December 2011, 2011.

[12] Anwar Ali Yahaya Addin Osman. Classifications Of Exam Questions Using Linguistically- Motivated Features : A Case Study Based On Bloom ' S Taxonomy Research Questions Research Aim. In The Sixth International Arab Conference on Quality Assurance in Higher Education, volume 2016, Saudi Arabia, 2016.

[13] Norazah Yusof and Chai Jing Hui. Determination of Bloom's Cognitive Level of Question Items using Artificial Neural Network. In 10th International Conference on Intelligent Systems Design and Applications (ISDA), pages 866–870, 2010.

[14] Dhuha Abdulhadi Abduljabbar and Nazlia Omar. Exam questions classification based on Bloom's taxonomy cognitive level using classifiers combination. Journal of Theoretical and Applied Information Technology, 78(3):447–455, 2015.

[15] Ali Danesh, Behzad Moshiri, and Omid Fatemi. Improve text classification accuracy based on classifier fusion methods. 2007 10th International Conference on Information Fusion, pages 1–6, 2007.

[16] Julio Villena Roman, Sonia Collada P ´ erez, Sara Lana Serrano, ´ and Jose Carlos Gonz ´ alez Crist ´ obal. Hybrid Approach Com- ´ bining Machine Learning and a Rule-Based Expert System for Text Categorization. In Proceedings of the Twenty-Fourth International Florida Artificial Intelligence Research Society Conference — Twenty-Fourth International Florida Artificial Intelligence Research Society Conference — 18/05/2011 - 20/05/2011 — Palm Beach, Florida, EEUU, pages 323–328, 2011.

[17] Maria Dominic, Sagayaraj Francis, "An Adaptable E-Learning Architecture Based on Learners' Profiling", Published Online March 2015 in MECS (http://www.mecs-press.org/) DOI: 10.5815/ijmecs.2015.03.04 Copyright © 2015 MECS I.J. Modern Education and Computer Science, 2015, 3, 26-31.

[18] Bloom ,B. S., Engelhart .M. D., Furst E.J., Hill, W.H., & Krathwohl, D.R. (Eds.)(1956), " Taxonomy of Educational Objectives. The Classification of Educational Goals", Handbook I: Cognitive Domain. New York: David McKay Company, Inc.

[19] https://www.analyticsvidhya.com/blog/2021/08/why-must-text-data-be-pre-processed/.

[20] https://www.pluralsight.com/guides/importance-of-text-pre-processing.

[21] https://towardsdatascience.com/text-pre-processing-stop-words-removal-using-different-libraries-f20bac19929a.

[22] https://www.lexalytics.com/blog/text-analytics-functions-explained/.

[23] https://deepai.org/machine-learning-glossary-and-terms/f-score.

[24] https://www.dataversity.net/property-graphs-vs-knowledge-graphs/.

AUTHORS' PROFILE

First Author: Mrs S. Celine is currently working as an Assistant Professor in the Department of Computer Science, Government of Arts College for Men, Krishnagiri, Tamil Nadu. She has published four papers in International Journals. Her area of research is e-Learning using Deep Learning.

Second Author: Dr M. Maria Dominic is currently working as an Assistant Professor in Computer Science, at Sacred Heart College. He has published one book and more than 30 Research Articles in International Journals. His area of interest is AI, Machine Learning, and Deep Learning.

Third Author: Dr. F. Sagayaraj Francis is an Associate Professor in the Department of Computer Science and Engineering, Pondicherry Engineering College, Pondicherry, India. He has published more than 65 Research Articles in International Journals. He specializes in Data Management, Data Modeling, Information Systems and e-Learning.

Fourth Author: Dr M. Savitha Devi is currently working as an Assistant Professor and Head, Department of Computer Science, Periyar University Constituent College of Arts & Science, Harur, Tamil Nadu. She has published 30 research articles in various National and International Journals. She has published 2 books. Her area of research is Network Security.