CNN and RF Classification Method for Course Recommendation to the Students

Ajita Satheesh*, Dr. Aarti Kumar**

* Research Scholar, Rabindranath Tagore University, Bhopal
** Professor, Rabindranath Tagore University, Bhopal

Abstract
Academic information systems are a place of storage for data collection, processing, analysis and reporting of educational information. Improving the quality of education in higher education can be seen from high rates of student success and low failure rates of students. That has become a challenge for Technical education institutions to maintain their reputation and business continuity. Some of the educational organizations have multi-education paths such as Diploma, Engineering and Medicine collages. In such colleges, the behavior of the student in the preparatory year determines which education path the student will join in the future. The improper selection of courses would seriously affect the students' achievements, which enforces students to drop out the improperly selected courses. Therefore, there is an urgent need to develop course recommendation system. In order to solve such an issue, proposed course recommendation system by using Condensed Nearest Neighbour (CNN) with Random Forest (RF) Classification. Here the dataset are collected from school passed out registered for admission to various diploma Courses. The process of data pre-processing takes place by using natural language processing to prepare the data in the correct object oriented format. The process of feature extraction to describe future prediction and recommendation results of object oriented data analysis and interpretation of prediction models. The experimental result shows that the proposed CNN with RF is effective by achieving accuracy of 98.67%. Whereas, the existing K-Nearest Neighbor (KNN) method showed accuracy of 92.52% for course recommendation.

Key words: Condensed Nearest Neighbour , Course Recommendation, Classification, Random Forest, K-Nearest Neighbor.

I. Introduction

Academic information systems are a place of storage for data collection, processing, analysis and reporting of educational information. Besides aiming to collect, store data and process information, it also helps in monitoring, policy formulation, decision making, management, and evaluation. Large amounts of data is stored in academic object oriented databases. This database contains information that is useful if explored more deeply to foster future students so that the risk of failure of studies can be avoided. On the other hand, improving the quality of education in higher education can be seen from high rates of student success and low failure rates of students. In Indonesia based on 2017 statistics, the number of students who experienced study failure was 195,176. It has become a challenge for higher education institutions to maintain their reputation and business continuity. With the rapid growth and spread of information and communication technology (ICT), recommender systems have evolved that have completely reshaped the web experience of users by providing meaningful, effective, and personalized recommendation of products and services to users. Through proper object oriented data modeling and analysis, recommender systems tend to support users in decision making processes by enhancing their ability and quality of thinking. Some of the educational organizations have multi-education paths such as engineering and medicine colleges. In such colleges, the behavior of the student in the preparatory year determines which education path the student will join in the future. The improper selection of courses would seriously affect the students' achievements, which enforces students to drop out the improperly selected courses. Therefore, there is an urgent need to develop an optional course recommendation system. The information provided by existing method was not effective for the students to decide their streams as it was not considered the influencing factors such as the subject interest, background of family and motivation

towards career. Further, the existing systems have not provided the acceptable outcome for higher level of education [10]. In order to solve such an issue, proposed course recommendation system by using CNN with RF.

The paper is organized as follows. the survey of existing techniques based on student performance recommendation system is reviewed in Section 2. The proposed course recommendation system by using CNN with RF for course recommendation system for 10^{th} students is explained in Section 3. The experimental results and discussion is described in Section 4 and the conclusion of proposed method is present in Section 4.

II. Literature Review

The existing methods based on performance prediction system are reviewed in this section. The advantages and limitations of reviewed paper are also explained.

Authors	Methodology	Advantages	Limitations
Ling Huang et.al.[11]	Developed score prediction approach for optional course recommendation via cross-user-domain collaborative filtering.	The predicted scores of all optional courses, the top optional courses with the highest predicted scores without time conflict will be recommended to the student.	It was expected that dramatic decrease of dropout rate could be obtained by the proposed method, although not yet confirmed by the experiments due to the lack of ground-truth dropout information.
Mohamed Ezz et.al.[12]	Established adaptive recommendation system using machine learning algorithms for predicting student's best academic program.	The proposed model recommends the best machine learning algorithm for each faculty department, find the relevant data that are important in the recommendation process and recommend the student with the suitable engineering department.	The developed model can be tested to solve other educational datasets once there is an available one to ensure that the proposed architecture can be generalized to efficiently solve other similar problem.
Hazra Imranet.al .[13]	Developed personalized learning object recommender system.	The developed system supports learners by providing those recommendations about which learning objects within the course are more useful for them.	The developed method need to focus on testing, developing, and implementing the proposed framework, including the process of identifying parameters to be used, testing models, and recommendations for solving problems faced by students.
Youness Madan et.al.[14]	Developed recommendation approach for finding the effective pedagogical content in the platform of e-learning by using reinforcement learning.	The recommendation method is on the basis of social filtering by utilizing the notions for sentimental analysis and employed collaborative filtering to define the effective ways to make the learners understand and recommends the courses in better way by making the profile and social content of learners.	But, the recommendation approach utilized the similar measure which depends on inter-related items among the two users and makes use of overall information of rating.
Vedat Apuk et.al.[15]	Developed educational contents classification method by utilizing K-Nearest Neighbor (KNN) and Long Short Term Memory (LSTM).	The process of text processing takes place to convert the unorganized data into specific format. Then, the features are extracted by	The result of developed method affected because of larger similarity among various transcripts. The transcripts belonging to

		using tokenization, lemmatization, stop words removal, etc to transform into the structured space of features.	various classes during course level included various similarities in keywords and sentences so it was not properly distinguished the class of transcripts.

III. Proposed Methodology

In this research, proposed course recommendation system by using CNN with RF. The bock diagram of proposed CNN with RF method is shown in figure 1.

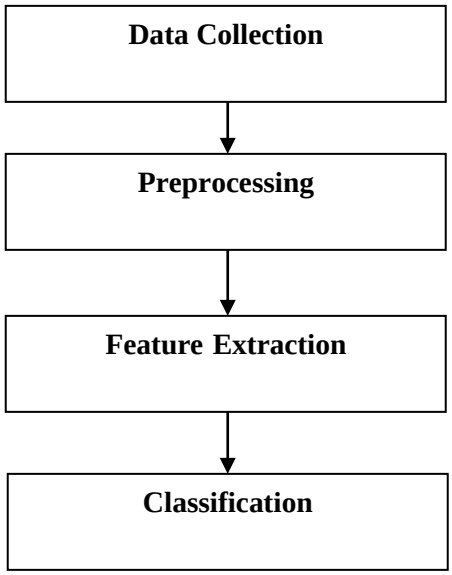

Figure 1. The block diagram of proposed CNN with RF for course recommendation system for students.

A. Data Collection

This students' performance dataset are collected from school passed out students registered for admission to various diploma Courses. The Object Oriented dataset includes students' attributes like academic grades, demographic attributes, social attributes and school related attributes. School reports and questionnaires are used for collecting data from the students.

B. Preprocessing

After collecting the data, data pre-processing is necessary. Steps followed for this purpose in sequence are: Data quality assessment, data cleaning, data transformation and data reduction. Data quality assessment is done in which we look for data relevance, find out data anomalies, data inconsistency and other inherent problems. In data cleaning we use methods like binning, regression and clustering for filling missing data, data correction, removing incorrect and irrelevant data. Data transformation is a process of converting data into proper formats needed for analysis. Aggregation , normalization , feature selection, discretization and Concept hierarchy generation techniques are used for Data transformation. Finally to make analysis easier and more accurate data reduction is needed which involves attribute selection and dimensionality reduction as per requirement of our analysis

C. Feature Extraction

After pre-processing the data, the process of feature extraction -data selection and transformation takes place. A few derived variables were selected and some of the information for the variables was extracted from the database. Descriptive data summarization techniques are used identify data features. Central tendencies and dispersion are used to identify outliers and to learn the behavior of data. Through data cleaning we find out discrepancy, missing values and outliers. Data integration, transformation and reductions are also necessary. Different forms and formats of data are collected and integrated together. Data is transformed into a form suitable for mining. Smoothing, aggregation, generalization normalization and feature construction are also required before passing through appropriate mining model. In order to reduce large volume of data along with keeping the integrity of original data, data reduction technique is used.

D. Classification

The obtained features are forwarded to the proposed classifier model by using CNN with RF for classification.

Condensed Nearest Neighbor (CNN)

The CNN algorithm was introduced by Hart in the year 1968 to find the subsets of labelled points of data structure which leads to accurate and faster classification. The main limitation of K-Nearest Neighbor (KNN) is that the requirement of larger memory for the storage of sample data. The CNN minimizes the overall number patterns stored in the training set of subsets. The idea behind the patterns during training is that it will be similar and adds extra information and it might get discarded. The CNN classifier is simpler technique for approximating the subsets among the labelled data points. The CNN algorithm is

defined with the T group of labelled data points and $T(t)$ is the label identified by the t nearest neighbor classifier which is trained by T. For example, if the sample S subset is evaluated which is smaller and accurate. The best subset Z is selected from the $2^{|S|}$ of possible subsets. The error measure CNN is defined by using regulation theory as shown equation (1-3)

$$Z^* = Arg \min E(Z) \quad (6)$$

$$E(Z) = \sum_{X \epsilon s} L(x|Z) + \gamma|Z| \quad (7)$$

$$L(x|Z) = \begin{cases} 1, & if\ D(Z_C, x) = (\min \mathrm{D}(\mathrm{Z_j}, \mathrm{x}) \text{ and } \mathrm{class(x)} \neq \mathrm{class(Z)} \\ 0, & otherwise \end{cases} \quad (8)$$

Where, $Z_c \epsilon Z$ is the closest stored pattern from $x \epsilon s$ by utilizing the measure of distance $D(.)$. $L(x|Z)$ is the non-zero when the x ans Z labels do not match.

From the theory of regularization, the solutions for a problem are obtained by combining the smoothness information and data. The initial term measures the miss fitting of data which is because of error obtained during classification by using the rule of nearest neighbor. The next term calculates the size of subset that is stored in such a way it defines the smoothness in the boundary of class. The nearest neighbor classifier divides the space in input as the form of voronoi relation and the boundaries of class are piecewise linear.

Random Forest

RF as characterized by their developer [16], are outfit of trees to such an extent that each tree is built on bootstrapped test of the first training information. To order another item from an information vector, the information vector will be put down every one of the trees in the woods. Each tree gives a vote to demonstrate the tree's choice about the class of the item. The forest picks the order having the most votes over every one of the trees in the forest. Each tree of the forest is developed as follows:

- Let the quantity of models in the first preparing information is N. Draw a bootstrap test of size N from the first training information. This example will be another training dataset for developing the tree. Information which are in the first training information however not in the bootstrap test are called out-of-pack information.
- Let the absolute number of info highlights in the first training information be M. On this bootstrap test information, just m ascribes are picked indiscriminately for each tree where m < M. The credits from this set makes the most ideal split at every hub of the tree. The worth of m ought to be steady during the developing of the forest.

The exactness of the individual trees and the relationship between's the trees in the forest decide the blunder pace of the forest. While expanding the connection builds the forest blunder rate, expanding the precision of the individual tree diminishes the error rate of forest. That is lessening m decreases both the relationship and the strength. Contrasted with a solitary choice tree calculation RF runs productively on huge datasets with a superior exactness. RF can deal with ostensible information and doesn't over-fit. Ultimate choice for characterization of test information is finished by greater part casts a ballot from expectations of the outfit of trees.

IV. Experimental Result and Discussion

In this section, the experimental results of proposed CNN with RF for course recommendation system for students are described. The validation of the proposed recommendation method carried out with the students' performance dataset are collected from the schools in Madhya Pradesh against the existing approaches described. The proposed method is applied on a computer with 8GB RAM with 2.2 GHz using Python 3.7.3. The performance metrics and performance analysis for the proposed CNN with RF for course recommendation system for students against the existing approaches are explained as follows:

Performance metrics

To evaluate the performance of the proposed classification method by using CNN with RF for course recommendation system for students. The proposed work considered 70% of data for training and 30% of data for testing the proposed method. The proposed CNN with RF for course recommendation system for students is compared with the existing techniques and it is estimated by using the various parameters that are used to check the property of model. The performance metrics considered in the CNN with RF is explained as follows:

- **Accuracy:** Accuracy is the ratio of a number of correct predictions to the overall predictions and it is used for evaluating the classification of models, which is defined in equation (16).

$$Accuracy = \frac{Number\ of\ correct\ predictions}{overall\ predictions} \quad (16)$$

- **Precision:** Precision is defined as the ratio of truly predicted positive observation to the overall predicted observation for positives. The precision is explained in equation (17).

$$Precision = \frac{TP}{TP+FP} \quad (17)$$

Where TP is True positive and FP is False Positive cases.

- **Recall**: Recall is defined as the ratio of truly predicted as the fault-modules which is explained in equation (18).

$$Recall = \frac{TP}{TP+FN} \quad (18)$$

Where TP is True positive and FP is False Positive cases.

- **F-score:** F-score calculates the accuracy of the model and it is the combination of precision and recall, which defined in equation (19)

$$F - score = \frac{TP}{TP+1/2(FP+FN)} \quad (19)$$

Where TP is True positive, FP is False Positive and FN is False Negative cases.

Quantitative Analysis

The proposed quantitative analysis of proposed classification method by using CNN with RF for course recommendation system for students is explained in this section. The values obtained for the proposed classification method by using CNN with RF for course recommendation system for students is shown in table 1.The table 1 includes the evaluation results for pedagogical content classification in terms of accuracy, precision, recall and f-measure.

Metrics	Proposed CNN with RF
Accuracy	98.67
Precision	94.32
Recall	95.54
F-score	96.29

Table 1. The quantitative analysis of proposed classification method by using CNN with RF for course recommendation system for students

The table 1 shows the quantitative analysis of proposed classification method by using CNN with RF for course recommendation system for students. The performance is evaluated in terms of accuracy, precision, recall and f-score. The accuracy is a number of correct predictions to the overall predictions which is utilized to evaluate the model of classification. The proposed classification method by using CNN with RF for course recommendation system for students method achieved accuracy of 98.67%, precision of 94.32%, recall of 95.54% and f-score of 96.29%. The CNN classifier is simple process and constructs the subsets of example that classifies the original object oriented data correctly. The RF improves the performance of model in sequence classification process. So, the classification method showed effective performance. The graphical representation of quantitative

analysis of proposed method is shown in figure 2.

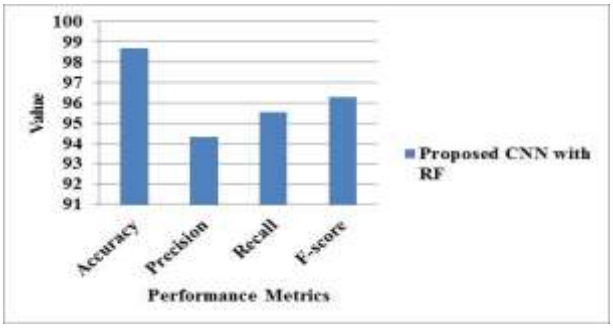

Figure 2. The quantitative analysis graphical representation of proposed classification method by using CNN with RF for course recommendation system for students.

Metrics	RNN	DT	CNN	Proposed CNN with RF
Accuracy	89.56	90.23	93.45	98.67
Precision	88.25	91.0	92.43	94.32
Recall	90.23	91.25	92.39	95.54
F-score	92.56	93.21	94.35	99.98

Table 2. The quantitative analysis of proposed classification method by using CNN with RF for course recommendation system for students with the existing methods.

The table 2 shows the quantitative analysis of proposed classification method by using CNN with RF for course recommendation system for students with the existing methods. The performance is evaluated in terms of accuracy, precision, recall and f-score. The existing Recurrent Neural

Network(RNN) method for mobile application showed accuracy of 89.56%, precision of 88.25%, recall of 90.23% and f-score of 92.56%. The existing Condensed Nearest Neighbour (CNN) method for mobile application showed accuracy of 93.45%, precision of 92.43%, recall of 92.39% and f-score of 94.35%. The existing Decision Tree (DT) method for mobile application showed accuracy of 90.25%, precision of 91.0%, recall of 91.25% and f-score of 93.21%. The proposed classification method by using CNN with RF for course recommendation system for students achieved accuracy of 98.67%, precision of 94.32%, recall of 95.54% and f-score of 99.98%. The RF classifier is simple process and constructs the subsets of example that classifies the original data correctly. The graphical representation of quantitative analysis of proposed method with existing methods is shown in figure 3.

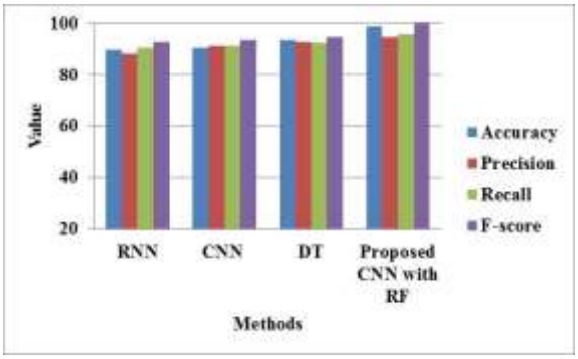

Figure 3. The graphical representation of quantitative analysis of proposed classification method by using CNN with RF in course recommendation system for students with the existing methods.

Comparative Analysis

The comparative analysis of the proposed classification method by using CNN with RF for course recommendation system for students is carried out and the values are tabulated as described in table 3. The existing techniques such as [14] and [15] are compared with the proposed approach. Table 3 shows the comparison results of the proposed method with existing methods.

Table 3. The comparative analysis of proposed classification method by using CNN with RF for course recommendation system with the existing methods

Table 3 shows the comparison results of the proposed method with existing methods in terms of performance metrics such as accuracy, precision, recall and f-score. The proposed classification method by using CNN with RF is compared with existing methods such as [14] and [15] for the course recommendation system for students. The existing [14] method showed accuracy of 97.0%, precision of 89.0%, recall of 86.0% and f-score of 87.0%. Similarly, the existing [15] method showed accuracy of 92.52%, precision of 92.63%, recall of 92.52% and f-score of 92.533%. Whereas, the proposed CNN with RF method showed accuracy of 98.67%, precision of 94.32%, recall of 95.54% and f-score of 96.29%. In existing method the transcripts belonged to different classes at the course level had many similarities in context of sentences and keywords so the model could not properly distinguish in which class the transcripts belonged. In proposed CNN with RF method showed effective performance by combining CNN with RF . Similarly, The CNN classifier is simple process and constructs the subsets of example that classifies the original data correctly. The RF improves the performance of model in sequence classification process. The graphical representation for the comparison of proposed CNN with RF method for course recommendation system for students is shown in figure 4.

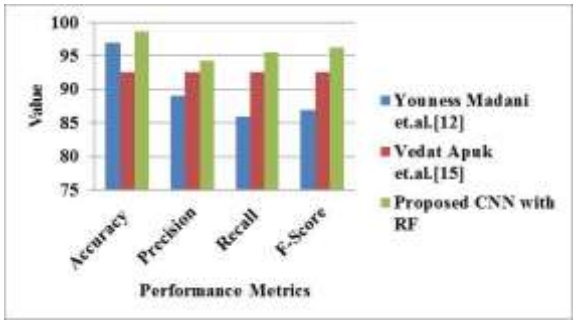

Figure 4. The comparative analysis graphical representation of proposed CNN with RF method for course recommendation system for students.

V. Conclusion

In this proposed course recommendation system by using CNN with RF Classification, dataset of the students' performances in object oriented form are collected from different schools of Madhya Pradesh. After collecting the data, the process of data

Methods	Accuracy (%)	Precision (%)	Recall (%)	F-Score (%)
Youness Mad ani et.al.[14]	97.0	89.0	86.0	87.0
Vedat Apuk et.al.[15]	92.52	92.63	92.52	92.53
Proposed CNN with RF	98.67	94.32	95.54	96.29

pre-processing takes place by using natural language processing to prepare the data in the correct format. Then, the process of feature extraction to describe future prediction and recommendation results of data analysis and interpretation of prediction models. The obtained features are forwarded to the proposed classifier model by using CNN with RF for classification. The CNN classifier is simple process and constructs the subsets of example that classifies the original data correctly. The RF improves the performance of model in sequence classification process. The experimental result shows that the proposed CNN with RF is effective by achieving accuracy of 98.67%. Whereas, the existing KNN method showed accuracy of 92.52% for course recommendation.

VI. **References**

[1] IEEE Transactions on Information Theory, May 1968, pp.515-516.

[2] Kurniadi, D., Abdurachman, E., Warnars, H.L.H.S. and Suparta, W., 2019, December. A proposed framework in an intelligent recommender system for the college student. In *Journal of Physics: Conference Series* (Vol. 1402, No. 6, p. 066100). IOP Publishing.

[3] Gray, C.C. and Perkins, D., 2019. Utilizing early engagement and machine learning to predict student outcomes. *Computers & Education*, *131*, pp.22-32.

[4] Pliakos, K., Joo, S.H., Park, J.Y., Cornillie, F., Vens, C. and Van den Noortgate, W., 2019. Integrating machine learning into item response theory for addressing the cold start problem in adaptive learning systems. *Computers & Education*, *137*, pp.91-103.

[5] Ezz, M. and Elshenawy, A., 2019. Adaptive recommendation system using machine learning algorithms for predicting student's best academic program. *Education and Information Technologies*, pp.1-14.

[6] Farzan, R. and Brusilovsky, P., 2011. Encouraging user participation in a course recommender system: An impact on user behavior. *Computers in Human Behavior*, *27*(1), pp.276-284.

[7] Al-Badarenah, A. and Alsakran, J., 2016. An automated recommender system for course selection. *International Journal of Advanced Computer Science and Applications*, *7*(3), pp.166-175.

[8] Aher, S.B. and Lobo, L.M.R.J., 2012. Course recommender system in E-learning. *International Journal of Computer Science and Communication*, *3*(1), pp.159-164.

[9] Bakhshinategh, B., Spanakis, G., Zaïane, O.R. and ElAtia, S., 2017, April. A Course Recommender System based on Graduating Attributes. In *CSEDU (1)* (pp. 347-354).

[10] Morsomme, R. and Alferez, S.V., 2019. Content-Based Course Recommender System for Liberal Arts Education. *International Educational Data Mining Society*.

[11] Huang, L., Wang, C.D., Chao, H.Y., Lai, J.H. and Philip, S.Y., 2019. A score prediction approach for optional course recommendation via cross-user-domain collaborative filtering. *IEEE Access*, *7*, pp.19550-19563.

[12] Ezz, M. and Elshenawy, A., 2019. Adaptive recommendation system using machine learning algorithms for predicting student's best academic program. *Education and Information Technologies*, pp.1-14.

[13] Imran, H., Belghis-Zadeh, M., Chang, T.W. and Graf, S., 2016. PLORS: a personalized learning object recommender system. *Vietnam Journal of Computer Science*, *3*(1), pp.3-13.

[14] Madani, Y., Ezzikouri, H., Erritali, M. and Hssina, B., 2019. Finding optimal pedagogical content in an adaptive e-learning platform using a new recommendation approach and reinforcement learning. *Journal of Ambient Intelligence and Humanized Computing*, pp.1-16.

[15] Apuk, V. and Nuçi, K.P., 2021. Classification of Pedagogical content using conventional machine learning and deep learning model. *arXiv preprint arXiv:2101.07321*.