Journal of Engineering Science and Technology
Vol. 16, No. 5 (2021) 3757 – 3771

TOWARDS AN IMPLEMENTATION OF INSTANCE-BASED CLASSIFIERS IN PEDAGOGICAL ENVIRONMENT

IJAZ KHAN[1,*], ABDUL R. AHMAD[1],
NAFAA JABEUR[2], MOHAMMED N. MAHDI[1]

[1]Universiti Tenaga Nasional, Jalan Ikram-Uniten, 43000 Kajang, Selangor, Malaysia
[2]German University of Technology, Athaibah, Muscat, Oman
*Corresponding Author: ijaz@buc.edu.om

Abstract

Monitoring individual student academic performance is indispensable to educational institutes since they are required to provide evidence of their students' academic performance to diverse governmental bureaus. Machine learning classifiers appear productive tools for this purpose; however, instance-based machine learning classifiers have acquired the least consideration. This research measures the suitability of instance-based classifiers, exclusively k-Nearest Neighbours (k-NN) and Locally Weighted Learning (LWL), in the pedagogical environment and proposes solutions to issues related to this class of classifiers. The performance of these classifiers depends upon the number of nearest neighbours (k) and the distance metrics. We performed experiments, with varying values of k and different distance metrics, to evaluate the performance of k-NN and LWL. To authenticate the conclusions drawn from these experiments, we carried out experimental evaluation with 3 more datasets taken from another research. This comparison evidences the suitability of instance-based classifiers, in pedagogical environment, especially LWL which is one of the least addressed classifiers. The comparative analysis highlights the fact that varying value of k and changing the distance metric optimistically affect the classifier's performance. Even though Manhattan distance metric dominates in achieving higher accuracy; however, classifiers may act differently for dissimilar datasets. To resolve this shortfall, we propose a novel framework which carries out extensive experiments with varying value of k and changing distance metrics and conclude a prediction model which emerges appropriate for the provided training dataset. The framework takes training dataset from an instructor and recommends suitable instance-based learning prediction model.

Keywords: Instance-based classifiers, K-nearest neighbours, Lazy classifiers, Locally weighted learning, Machine learning, Student performance prediction.

1. Introduction

Education is a vital necessity for individuals to build up their potential capabilities [1]. Being accountable to different bureaus, the institutes must be capable to make decisions that keep them competent in pursuing an admirable reputation in the pedagogical society [2]. The institutions constantly collect a vast amount of data about their students from various sources. Learning Management Systems (LMSs) are present in most higher education institutions for storing students' academic records [3]. Moreover, due to the integration of web technologies, a huge amount of educational data becomes accessible [4]. The institutions require meaningful resources and accurate practices to examine the flow of ongoing trends to discover the weak spots and subsequently take appropriate actions. The advent of big data analysis has led to an increased interest in the collection and analysis of student data to extract the underlying unseen patterns in the learning activities [5].

Monitoring and tracking individual student performance can emerge as a supportive tool for an instructor to identify the students with unsatisfactory academic progress [6]. Due to increased workload, the adaptation of novel technology is inevitable for the institutions [7]. Machine Learning (ML) is a useful tool to build intelligent classification models to classify the students which helps the institutions to take the appropriate actions [8]. Supervised classifiers are widely used for developing student performance prediction models. Some of the popular supervised classifiers include decision tree, neural networks, naïve bayes, k-Nearest Neighbours (k-NN) and Support Vector Machine (SVM) [9].

Instance-based classifiers, also known as lazy classifiers or memory-based learners, store the training dataset in the memory and find the relevant data in the database to classify an unseen instance [10]. The distance of the unseen instance is computed from each training instance and a predefined number of its neighbours vote to decide the class label of the unseen instance. Therefore, distance metrics and the setting the number of neighbours (k) for voting are the parameters that play vital role in the execution of these classifiers. This set of classifiers is recognized as unstable and an alteration to their parameters can change the overall performance.

Most of the authors [6, 11, 12], especially when they use WEKA, applied k-NN with default settings (k=1; with Euclidean distance metric), few authors [13-16] chose to experiment with varying values of k and different distance metrics. In several cases executing k-NN with particular settings proves best [17]; but yields worst results for a dissimilar dataset in same domain [11]. It signifies that tuning the classifier affects the prediction performance. However, identifying an accurate value of k and apposite distance metric for a given dataset is imprecise. Further, in our knowledge no research work is available which uses LWL for classification in educational domain.

The aim behind this research is not to prove that solely instance-based learning classifiers performs better with the educational dataset, but the major objective is to underline the impact of the proper tuning over the classification performance of these classifiers. Due to having comparable nature, we chose k-NN and LWL from the category of instance-based learning classifiers. The motivation behind using instance-based classifiers is their capability to work well with small datasets with real-value attributes, which resembles the nature of datasets in educational domain.

We compare k-NN with LWL and assess their suitability for the student performance prediction models. We perform experiments to confirm that the performance of instance-based classifiers tends to enhance with varying the value of k and changing the distance metric. Initially, an educational dataset, from a course, has been used to develop student performance prediction models with k-NN and LWL under varying settings of classifiers which lead to several observations. The observations are then tested with 3 other datasets to come up with discrete conclusions. This leads to the fact that, depending upon the dataset complexity; different dataset may work well with different value of k and distance metrics. To address this shortcoming, we propose a novel instance-based classification framework which takes a pre-processed dataset and perform exhaustive experiments, with various combinations of k and different distance metrics, and recommends the most appropriate model for the provided dataset. Following is the list of major research questions we investigate in this research paper.

- How does a change in the value of k and distance/similarity metrics affects the prediction performance of instance-based learning classifiers?
- How useful does LWL appear for classification in pedagogical environment comparing to k-NN?
- How can we automatically identify the most appropriate value of k and distance/similarity metric for the dataset of a course?

This paper is organized as; section 2 provides a literature review of the topic and discusses the prediction models designed with instance-based classifiers. In section 3, describes our dataset, experimental setups, and the evaluation metrics. Section 4 explains the research methodology and explains each step of the research process and section 5 concludes the paper.

2. Literature Review

Humans have the ability to reason and learn from their experience, but computers, being unable to reason, learn with algorithms. Machine Learning (ML) is a branch of artificial intelligence that refer to learning from previous data to improve future performance automatically without any external assistance from humans [18]. Supervised classifiers, a class of machine learning classifiers, constructs a classification model by examines the training dataset and extracts the classification rules [19]. The subsequent step runs the model with a subset of data from the training dataset to determine the performance of the model. Figure 1 depicts the popular supervised classifiers and their brief description.

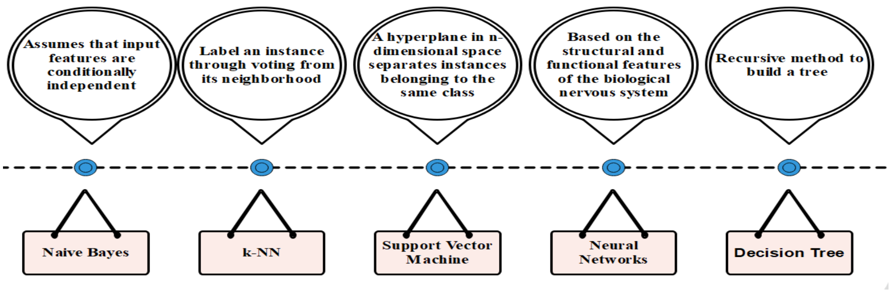


Fig. 1. An illustration of widely used supervised machine learning classifiers.

Instance-based learning classifiers [20] are called as lazy learners because they memorize all the instances from the training set and the classification is based directly on the training examples [21]. The fundamental assumption is that similar instances will have similar classes. These classifiers are considered as unstable due to the fact that tuning their essential parameters can modify their performance behaviour. These classifiers require a reasonable training time [14]. k-NN is one of the simple instance-based learning algorithms [20]. The fundamental principle behind k-NN states that, within a data set, the instances that have similar properties generally exist in close proximity [16]. It plots each instance as a point in multi-dimensional space. The instances are tagged with a classification label, and then the value of the class of an unclassified instance can be determined by observing the class of its nearest neighbours. The value of k determines the number of instances it must consider while classifying an unseen instance. A voting process among the k neighbours decides the class of unclassified instance. Locally Weighted Learning (LWL) performs prediction by using an approximated local model around the point of interest [10]. The instances, in the close neighbourhood to the unclassified point receive a higher weight than instances that are far away.

Numerous prediction models have been proposed under different educational settings to predict the final outcome of the students. Several authors used 1-NN in the pedagogical situations. Rawat and Malhan [17] developed a hybrid classification model with decision tree, naïve bayes, 1-NN and multilayer perceptron. 1-NN achieved higher accuracy than decision tree. In case of Acharya and Sinha [11], 1-NN produces the worst results comparing to other classifiers. In work of Asif et al. [12], 1-NN attains better accuracy than other classifiers for developing model to predict the students' performance at an early stage of the degree program.

Several authors tuned instance-based classifiers to develop prediction models. In work of Kabakchieva [13] 100-NN produced better results than 250-NN, but most of the evaluated classifiers outclassed k-NN. Shehri, et al. [15] observed 9-NN and Manhattan distance metric failed to prevail its solo competitor SVM. Similarly, Ghosh et al. [22] developed prediction model with 50-NN to classify students as either pass or supplementary. Kotsiantis et al. [14] applied 1-NN and 3-NN in their proposed ensemble solutions to identify poor performers.

A number of authors modified instance-based classifiers to enhance the student performance prediction model. Wafi et al. [23] used genetic algorithm for feature selection and proposed modified KNN (M-KNN). The results show M-KNN proves better than the default k-NN in achieving accuracy for same value of k. Ahmad et al. [24] proposed Fast k-NN (FkNN) which merged the concept of moment descriptor with k-NN so instead of making the full search during the distance computation with the whole training set, only a subset of samples is computed. Alfere and Maghari [25] used several forms of k-NN classifiers to forecast the student's final grades.

Similarly, authors in [26-29] used k-NN to develop student performance prediction models. Table 1 provides a brief summary of the work which used k-NN in their prediction models and the essential details of each model. The literature review summarizes the utilization of instance-based classifiers, and LWL has received a negligible concentration in the educational environments. Most of the works provide rare attention to alter classifier's parameters. The major observation reveals the difficulty in setting the correct value for k and choosing the right distance metric. This comes up with the need of a solution to determine most appropriate values of k and distance metric for educational datasets.

Table 1. Summary of the prediction models developed with instance-based classifiers.

Ref.	Dataset Size	Instance-Based Classifierused	Performance	No. of Classes
[23]	480	1-NN to 7-NN	Modified 4-NN accuracy 0.826 4-NN accuracy: 0.736	3
[13]	10330	100-NN 250-NN	100-N (Precision): 0.57 (for 10-fold cross validation) and 0.578 (for percentage split) 250-NN (Precision):0.592 (for 10-fold cross validation) and 0.565 (for percentage split|)	5
[17]	27	1-NN	0.89 accuracy	3
[11]	309 (training) 104(testing)	1-NN	F-Measure: 0.58 to 0.72 for various experiments.	7
[22]	Not specified	50-NN	Does not provide details	2
[14]	1047	1-NN and 3-NN	1-NN performs better than Naïve Bayes. 3-NN performs better than most of the classifiers. Decision tree C4.5 overcome 3-NN in some experiments.	2
[12]	347	1-NN	Accuracy: 66.15% and 74.04 percent for different datasets.	5
[15]	395	19-NN 100-NN	0.612 correlation coefficient.	Not reported

3. Material and Methods

3.1. Experimental setup

We performed our experiments in WEKA (Waikato Environment for Knowledge Analysis) [30]. Classifier's training is performed using 10-fold cross-validation [31] in which the training dataset is split into 10 identical length intervals. In each cycle, the nine intervals are used for learning purpose and the tenth for testing the classifier's performance. It is an iterative process, and, in each iteration, a new interval is chosen for testing.

3.2. Performance evaluation metrics

Confusion matrix [32] visualizes the performance of a classification model (Table 2).

True Positive (TP) is the cell where the class names match the subscript (e.g., TP for class A is AA; False Negative (FN) for a class consists of the entire row of actual outcomes except the TP cell (e.g., TP of class A is the sum of AB, AC, and AD; False Positive (FP) for a class consists of the entire column of predicted outcomes except the TP cell (e.g., FP for class A is the sum of BA, CA, and DA; True Negative (TN) for a class is sum of all cells of confusion matrix except the row and column which contains the TP for that class (e.g., TN for class A is the sum of all cells except TP, FP, and FN).

Table 2. The view of confusion matrix for 4 classes

		Predicted Outcomes			
	Class	***A***	***B***	***C***	***D***
Actual Outcomes	***A***	A_A	A_B	A_C	A_D
	B	B_A	B_B	B_C	B_D
	C	C_A	C_B	C_C	C_D
	D	D_A	D_B	D_C	D_D

We used accuracy and f-Measure for evaluation. Accuracy encompasses the performance of the model into a single value. Recall is a measure of all positive instances and the number of instances the model predicted correctly. Precision shows, out of all the positive instances that model predicted correctly, how many are actually positive. F-Measure takes both precision and recall into account and calculates their weighted average.

$$Accuracy = ((TP + TN)\ ) \ / \ ((TP + FN + FP + TN)) \tag{1}$$

$$Recall = TP \ / \ ((TP + FN)) \tag{2}$$

$$Precision = \ TP \ / \ ((TP + FP)) \tag{3}$$

$$F - Measure = 2 \times \ ((Precision \ \times Recall)/(Presicion + Recall)) \tag{4}$$

4. Research Methodology

Figure 2 illustrates our 4-phase research methodology consisting of analogous kind of operations in each phase. The first phase handles the description and pre-processing of dataset. The primary operation in this phase includes, decreasing imbalance ratio among the classes. The experimental evaluation phase applies instance-based learning classifiers, with different settings in each experiment, over the dataset and record its behaviour. This phase ends with a list of conclusions from the experiments. To validate the conclusions, in comparative evaluation phase, instance-based classifiers are applied over several datasets taken from other research. The final phase proposes a novel framework which encompasses the major contributions of the research.

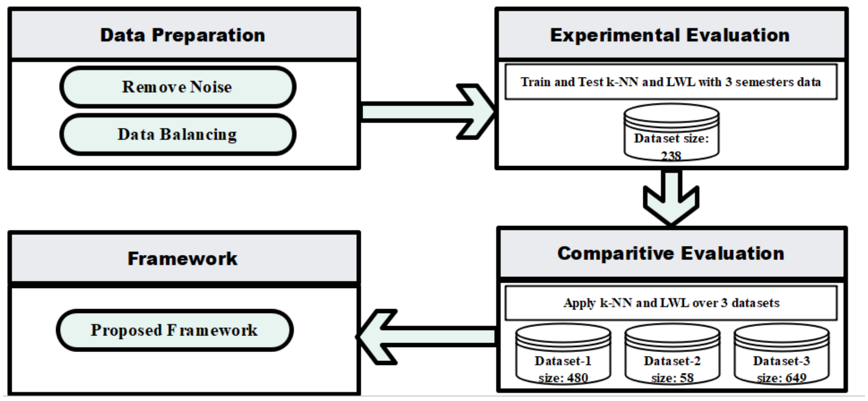


Fig. 2. The 4-phase methodology used in this research.

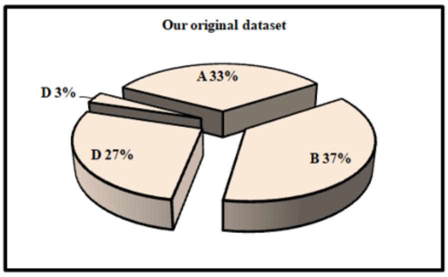
4.1. Data preparation

The dataset is taken from Buraimi University College (BUC), Oman. It contains 191 instances which are the total number of students enrolled in a core course during the previous 3 semesters. Table 3 provides the features used in our dataset and their description. The dataset is prepared in Microsoft Excel and converted to ARFF format.

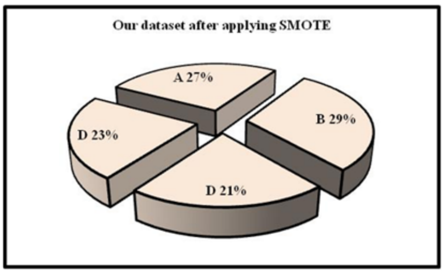
Table 3. The list of features in the dataset with description

Feature	Description	Values
Marks_MidT	Grades in Midterm Exam	Continuous (Percent)
Prev_Sem_GPA	Grade Point Average (GPA) of the student in last semester.	Continuous (0-4)
CGPA	Current Cumulative Grade Points Average (CGPA) of student	Continuous (0-4)
Marks_Pre_Req	Grades in the prerequisite subject	Continuous (Percent)
Final_Grade	Prediction Class	A, B, C, D

Figure 3(a) illustrates non-uniform class distribution in our dataset. Although there is no unified rule for class balancing, however, the preceding literature shows that tuning class distribution can improve classifier performance [33]. SMOTE (Synthetic Minority Over-sampling Technique) [34] oversamples the minority class with random under sampling of the majority class. It forms new instances for minority class by interpolating among several minority class instances that recline together [35]. Figure 3(b) shows the class distribution after sampling with SMOTE. The new file contains 238 instances. SMOTE appends new instances with the original dataset, we use “randomize” option in WEKA to shuffle the instances.



(a) Non-uniform class distribution in our dataset



(b) Class distribution after sampling with SMOTE

Fig. 3. Class distribution before and after applying SMOTE.

4.2. Experimental evaluation

4.2.1. Nearest neighbours (*k*)

The instance-based classifiers hold the capability to act differently if their essential parameter i.e., value of neighbours (*k*) and distance metric are attuned appropriately. No specific rule is available to select the number of neighbours; hence, heuristic approaches based on human judgement are applied to choose the value of k. Several research, especially conducted in WEKA, perform experiments with default settings (*k*=1, with Euclidean metric). One of the approaches is to keep k as the square root of the number of instances in the dataset. In our evaluation, we performed experiments with k=1 (being minimum possible and default value), 15 (the nearest square root to the total number of instances), 23 (Square root of roughly twice the number of instances), 11 and 19 (values above and below 15). To break the ties during voting process, the value of k is recommended to be an odd positive integer.

4.2.2. Distance/similarity metrics

The distance summarizes the relative difference between two rows of data that describe an instance. Once the training instances are plot in an n-dimensional space, the distance metric determines the member of the training set closest to unknown test instance. The k nearest neighbours votes and the unseen instance is labelled with the class receiving maximum votes. Minkowski distance or Lp norm distance is a generalized distance metric. Manhattan, Euclidean and Chebyshev are derived from Minkowski by substituting 'p' by 1, 2 and infinity, respectively. We chose Euclidean, Eq. (7), Manhattan, Eq. (6), and Chebyshev, (Eq. (8) distance metrics in experiments. Distance metric and similarity metrics are used interchangeably in rest of this paper.

$$(\sum_{i=1}^{n} | x_i - y_i|^p)^{1/p} \tag{5}$$

$$\sum_{i=1}^{n} |x_i - y_i| \tag{6}$$

$$\sqrt{\sum_{i=1}^{n} (x_i - y_i)^2} \tag{7}$$

$$\lim_{p\to\infty} (\sum_{i=1}^{n} | x_i - y_i|^p)^{1/p} \tag{8}$$

4.2.3. Experiments with our dataset

Figure 4 illustrates the fluctuation in the accuracy with varying value of k and different distance/similarity metrics for k-NN (left) and LWL (right). It shows, the classifiers achieve the least accuracy with k=1 and the accuracy enhance but fluctuates above k=1. Manhattan distance metric dominates in achieving higher accuracy; however, it is imprecise to identify an appropriate value for k. It is also observed that Chebyshev distance metric fails to prevail any of the classifier. Due to least difference in the achieved accuracies, it is right to say that both k-NN and LWL appears suitable for classification. Figure 5 demonstrates the fmeasure values of both the classifiers at varying value of k and different distance metrics. It displays similar trends in the fmeasure values of the classifiers.

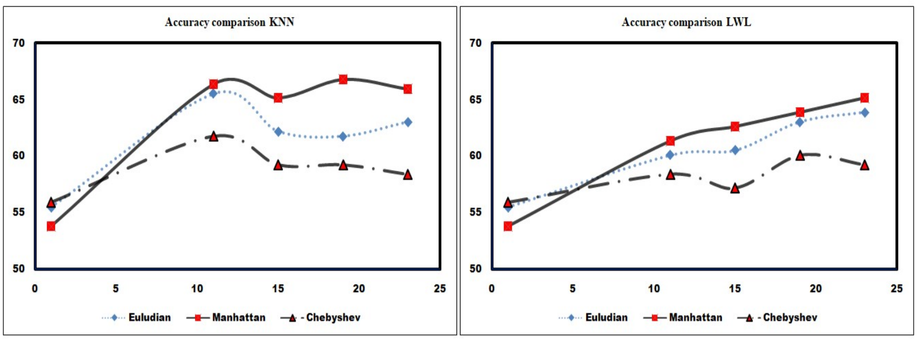


Fig. 4. Comparing the accuracy metric of k-NN and LWL with various settings.

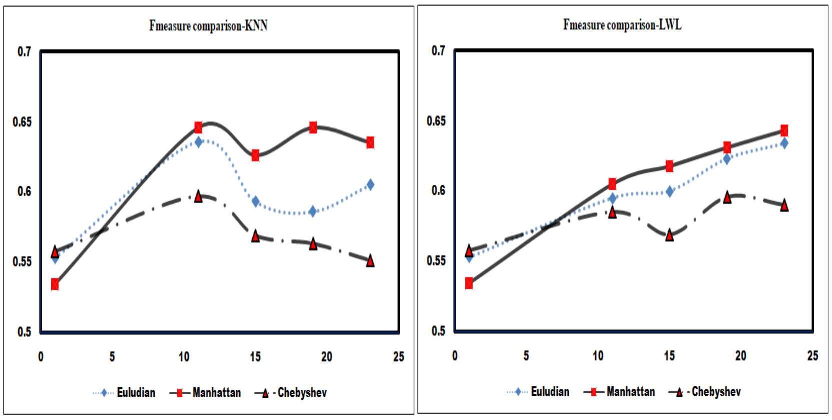


Fig. 5 Comparing the fmeasure metric of k-NN and LWL with various settings.

This section concludes that both k-NN and LWL appear efficient in pedagogical environment. Further, the classifiers achieve least accuracy and fmeasure at *k*=1 and the overall performance enhances with an increase in k. The Manhattan distance metric proves better comparing to Euclidean and Chebyshev.

4.3. Comparative evaluation

To validate the conclusions of previous phase, we perform experiments with 3 datasets taken from other research in the education domain. The dataset 1 taken from [36] contains 480 instances, distributed in 3 classes. We applied k-NN and LWL over the pre-processed dataset 1 to observe the behavior of the classifiers under different settings. We used value of k as 1, 15, 21, 27 and 33 through our heuristic approach (section 4.2.1). The second dataset, dataset 2, taken from [6] contains 58 instances from an introductory programming course with 11 features. Similarly, dataset from [37], dataset 3, consists of 649 instances distributed into 5 classes and 31 features. The feature selection algorithms are applied over each dataset to extract most essential features.

Figures 6, 7, and 8 display the outcome of the experimental evaluation for the classifiers with dataset 1, dataset 2 and dataset 3, respectively. Since the class distribution is uniform, therefore, we chose to compare accuracy metric. The prime observation reveals both classifiers achieving distinct accuracy at varying value of k and changing distance metric. A closer look at dataset 1 and dataset 3 reveals the classifiers are achieving lowest accuracy at k=1. However, classification with Chebyshev behaves slightly different as its accuracy drops down than its accuracy at *k*=1. Dataset 2 does not follow the trends observed in other datasets and the accuracy goes up and down at several values of k.

Figure 9 illustrates the net rise in the accuracy of the classifiers with different distance metrics. Net rise is the difference between the highest and lowest accuracy achieved by a classifier for a distance metric in the whole dataset. For instance, in our dataset, the minimum and maximum accuracy achieved for k-NN with Manhattan was 66.8 (at *k*=19) and 53.8 (at *k*=1) respectively. The figure 9 demonstrates classification with Manhattan distance yielding better results for most of the datasets. Both Euclidean and Chebyshev shows similar up and downs. This concludes that proper tuning may put gigantic affirmative effects over classifier.

Both the classifiers produced satisfactory results, however, choosing appropriate value for k and distance metric is still an ambiguous issue.

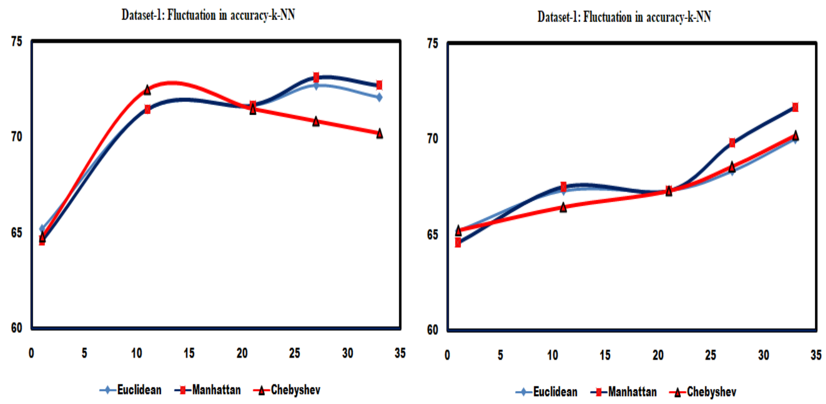


Fig. 6. Comparing accuracy of k-NN and LWL for dataset 1.

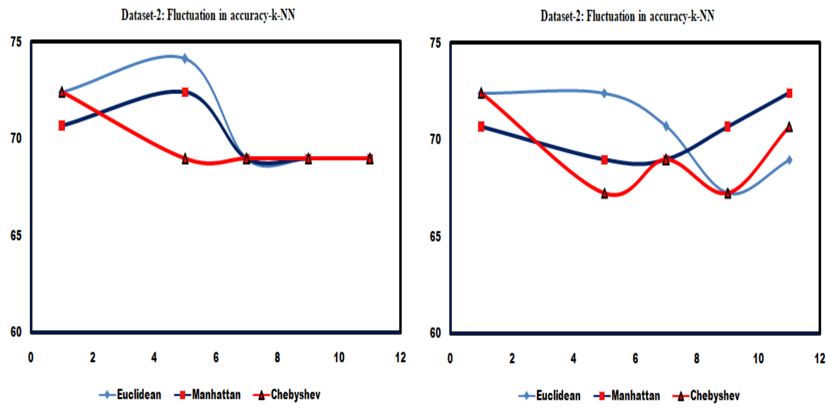


Fig. 7. Comparing accuracy of k-NN and LWL for dataset 2.

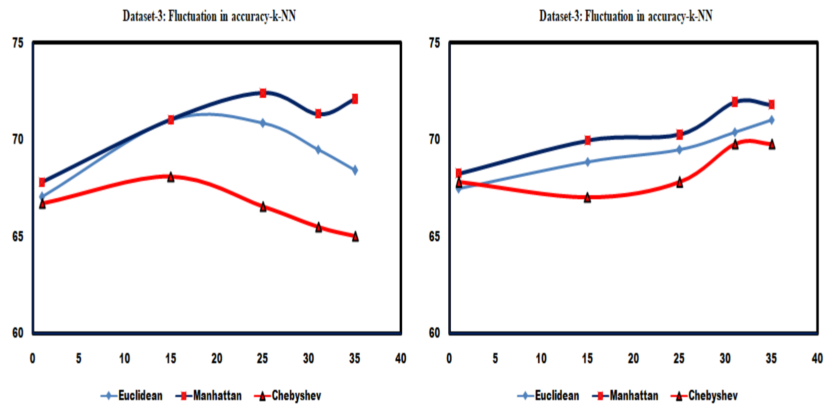


Fig. 8. Comparing accuracy of k-NN and LWL for dataset 3.

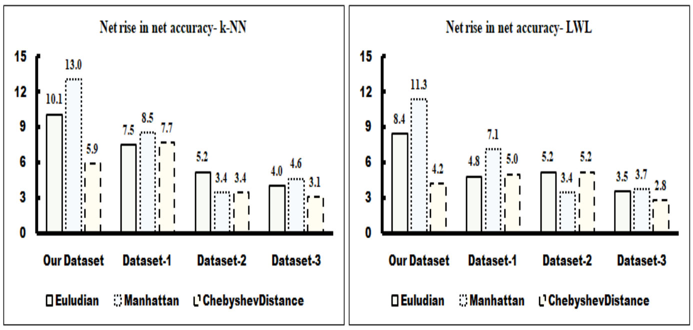


Fig. 9. Illustrating the net rise in the accuracy of classifiers under different settings.

4.4. Proposed framework

The core objective of this framework is to execute exhaustive experiments with varying value of k and changing distance measure and conclude a prediction model which emerges appropriate for the provided training dataset. The framework takes a pre-processed dataset as input and provides a single prediction model as output. Figure 10 demonstrates the framework.

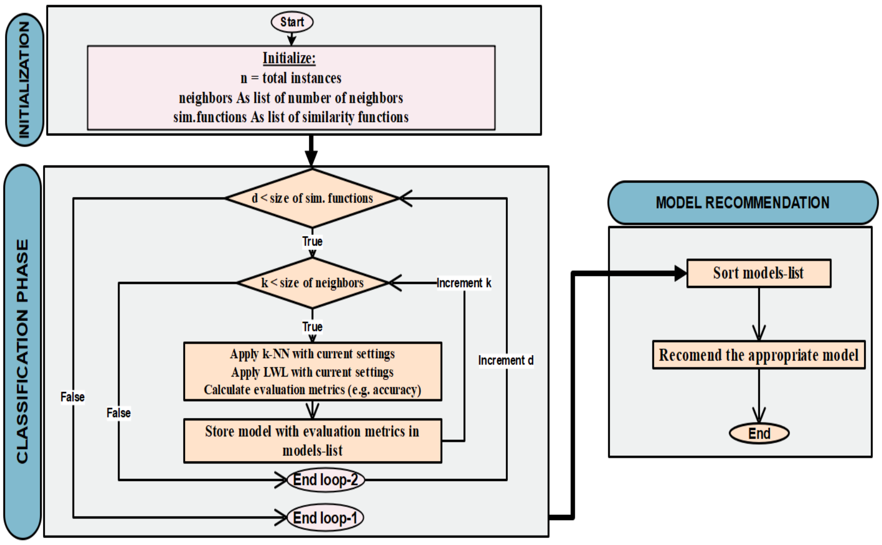


Fig. 1. The proposed framework.

In initialization phase, the instructor provides a pre-processed dataset, and the framework sets the total training instances, lists the distance metrics, and recommends values for k. The value of k can be selected with an adequate heuristic judgement approach. Apart from our approach (section 4.2.1) another approach may be "all the odd positive integers starting from 1 to the square root of twice the total instances in the dataset".

In the classification phase, the first loop keeps the distance/similarity metric constant and the second loop results in the development of as many prediction models as the number of predefined values for k. In each of the iterations of the first loop, the distance metric is changed, and the second loop continues with the production of models. The accuracy of each developed model is stored for evaluation purpose. This phase ends with a list of developed prediction models and their accuracy measures.

In final phase, the framework sorts the list of developed models and recommends the model with highest accuracy. As an output, the framework recommends the most appropriate model from the entire set of models tested.

In this paper, we keep accuracy as the comparison metrics, however, the framework is flexible to select appropriate model based on other evaluation metrics. Accuracy is useful if the dataset has uniform class distribution. Sensitivity and specificity are useful metrics if the identification of satisfactory students or the identification of weak students is of prime importance, respectively. Similarly, the same framework can be used for LWL.

5. Conclusion

The aims of this research are to evaluate the suitability of instance-based learning classifiers in the pedagogical environment and suggest solutions to issues related to this unstable class of classifiers. We applied k-NN and LWL classifiers over a dataset of 238 instances to evaluate the performance for varying value of k and distance/similarity measuring metrics. The conclusions from this evaluation are then further validated with comparative evaluation with 3 datasets taken from another research.

The oscillation, observed in the accuracy, draws attention to the fact that varying value of k and changing the distance metric positively affects the classifier's performance. However, it is observed that choosing Manhattan distance metric tends to prevail Euclidean and Chebyshev for most of the datasets. Thus, it seems right to say that most of the authors, who applied k-NN, could improve their results by properly tuning the classifiers.

This research validates the aptness of instance-based classifiers, in pedagogical environment, particularly LWL which is one of the least used classifiers in educational domain. Since classifiers may act in a different way for other datasets, therefore, we propose a novel framework which recommends suitable instance-based learning prediction models for dataset of different courses.

References

1. Riza, L.S.; Firdaus, D.S.; Junaeti, E.; Hidayat, T.; Abdullah, A.G.; Nandiyanto, A.B.D.; and Abdullah, C.U. (2018). A concept and implementation of instructional interactive multimedia for deaf students based on inquiry-based learning model. *Journal of Engineering Science and Technology* (*JESTEC*), 13(7) 2016-2035.
2. Musso, M.; and Cascallar, E. (2009). New approaches for improved quality in educational assessments: Using automated predictive systems in reading and mathematics. *Problems of Education in the 21st Century*, 17, 134-151.

3. Stockless, A. (2018). Acceptance of learning management system: The case of secondary school teachers. *Education and Information Technologies*, 23(3) 1101-1121.
4. Bharara, S.; Sabitha, S.; and Bansal, A. (2018). Application of learning analytics using clustering data mining for students' disposition analysis. *Education and Information Technologies*, 23(2) 957-984.
5. Greller, W.; and Drachsler, H. (2012). Translating learning into numbers: A generic framework for learning analytics. *Educational Technology and Society*, 15(3), 42-57.
6. Khan, I.; Al Sadiri, A.; Ahmad, A.R.; and Jabeur, N. (2019). Tracking student performance in introductory programming by means of machine learning. *Proceeding of the 4th MEC International Conference on Big Data and Smart City*. Muscat, Oman.
7. Chatterjee, S.; and Bhattacharjee, K.K. (2020). Adoption of artificial intelligence in higher education: A quantitative analysis using structural equation modelling. *Education and Information Technologies*, 25, 3443-3463.
8. Abdalla, O.A.; Aborokbah, M.M.; and ELfaki, A.O. (2019). Developing classification model to investigate the problem of computing students studies length. *International Journal of Machine Learning and Computing*, 9(2), 174-180.
9. Shahiri, A.M.; Husain, W.; and Rashid, N.A. (2015). A review on predicting student's performance using data mining techniques. *Procedia Computer Science*, 72, 414-422.
10. Atkeson, C.G.; Moore, A.W.; and Schaal, S. (1997). Locally weighted learning. *Artificial Intelligence Review*, 11, 11-73.
11. Acharya, A.; and Sinha, D. (2014). Early prediction of students performance using machine learning techniques. *International Journal of Computer Applications*, 107(1), 37-43.
12. Asif, R.; Merceron, A.; and Pathan, M.K. (2015). Predicting student academic performance at degree level: A case study. *International Journal of Intelligent Systems and Applications*, 7(1) 49-61.
13. Kabakchieva, D. (2013). Predicting student performance by using data mining methods for classification. *Cybernetics and Information Technologies*, 13(1) 61-72.
14. Kotsiantis, S.; Patriarcheas, K.; and Xenos, M. (2010). A combinational incremental ensemble of classifiers as a technique for predicting students' performance in distance education. *Knowledge-Based Systems*, 23(6) 529-535.
15. Al-Shehri, H.; Al-Qarni, A.; Al-Saati, L.; Batoaq, A.; Badukhen, H.; Alrashed, S.; Alhiyafi, J.; and Olatunji, S.O. (2017). Student performance prediction using support vector machine and k-nearest neighbor. *Proceeding of the 30th Canadian Conference on Electrical and Computer Engineering*. Windsor, ON, Canada.
16. Kotsiantis, S.; Pierrakeas, C.; and Pintelas, P. (2004). Predicting students' performance in distance learning using machine learning techniques. *Applied Artificial Intelligence*, 18(5) 411-426.

17. Rawat, K.S.; and Malhan, I.V. (2019). A hybrid classification method based on machine learning classifiers to predict performance in educational data mining. *Proceedings of the Second International Conference on Communication, Computing and Networking*. Singapore.
18. Das, K.; and Behera, R.N. (2017). A survey on machine learning: Concept, algorithms and applications. *International Journal of Innovative Research in Computer and Communication Engineering*, 5(2) 1301-1309.
19. Sivashankari, R.; and Valarmathi, B. (2018). An empirical machine learning approach to extract and rank multi-word product names using ICIW approach. *Journal of Engineering Science and Technology*, 13(1) 167-186.
20. Aha, D.W.; Kibler, D.; and Albert, M.K. (1991). Instance-based learning algorithms. *Machine learning*, 6, 37-66.
21. Cunningham, P.; and S.J. Delany (2020). *k-Nearest neighbour classifiers:* (*with Python examples*)*.* arXiv preprint arXiv:2004.04523.
22. Ghosh, C.; Saha, S.; Saha, S.; Ghosh, N.; Singha, K.; Banerjee, A.; and Majumder, S. (2020). Machine learning based supplementary prediction system using k nearest neighbour algorithm. *Proceeding of the Second International Conference on Industry Interactive Innovations in Science, Engineering and Technology*. Kolkata, India.
23. Wafi, M.; Faruq, U.; and Supianto, A.A. (2019). Automatic feature selection for modified k-nearest neighbor to predict student's academic performance. *Proceeding of the Third International Conference on Sustainable Information Engineering and Technology*. Lombok, Indonesia.
24. Ahmed, S.T.; Al-Hamdani, R.; and Croock, M.S. (2020). Enhancement of student performance prediction using modified K-nearest neighbor. *Telecommunication, Computing, Electronics and Control*, 18(4), 1777-1783.
25. Alfere, S.S.; and Maghari, A.Y. (2018). Prediction of student's performance using modified KNN classifiers. *Proceeding of the First International Conference on Engineering and Future Technology*. Gaza, Palestine. 143-150.
26. Mayilvaganan, M.; and Kalpanadevi, D. (2014). Comparison of classification techniques for predicting the performance of students academic environment. *Proceeding of the International Conference on Communication and Network Technologies*. Sivakasi, India.
27. Minaei-Bidgoli, B.; Kashy, D.A.; Kortemeyer, G.; and Punch, W.F. (2003). Predicting student performance: An application of data mining methods with an educational web-based system. *Proceeding of the 33rd Annual Frontiers in Education*. Westminster, CO, USA.
28. Gray, G.; McGuinness, C.; and Owende, P. (2014). An application of classification models to predict learner progression in tertiary education. *Proceeding of the 4th International Advance Computing Conference*. Gurgaon, India.
29. Imdad, U.; Ahmad, W.; Asif, M.; and Ishtiaq, A. (2017). Classification of students results using KNN and ANN. *Proceeding of the 13th International Conference on Emerging Technologies*. Islamabad, Pakistan.

30. Hall, M.; Frank, E.; Holmes, G.; Pfahringer, B.; Reutemann, P.; and Witten, I.H. (2009). The WEKA data mining software: An update. *Special Interest Group on Knowledge Discovery in Data Explorations Newsletter*, 11(1) 10-18.
31. Olson, D.L.; and Delen, D. (2008). *Advanced data mining techniques*. Berlin: Springer-Verlag Berlin Heidelberg.
32. Tharwat, A. (2018). Classification assessment methods. *Applied Computing and Informatics*, 17(1), 168-192.
33. Tan, P.N.; Steinbach, M.; and Kumar, V. (2006). *Introduction to data mining* (1st ed.). United Kingdom: Pearson Education.
34. Chawla, N.V.; Bowyer, K.W.; Hall, L.O.; and Kegelmeyer, W.P. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 16, 321-357.
35. Batista, G.E.A.P.A.; Prati, R.C.; and Monard, M.C. (2004). A study of the behavior of several methods for balancing machine learning training data. *Special Interest Group on Knowledge Discovery in Data Explorations Newsletter*, 6(1), 20-29.
36. Amrieh, E.A.; Hamtini, T.; and Aljarah, I. (2016). Mining educational data to predict student's academic performance using ensemble methods. *International Journal of Database Theory and Application*, 9(8), 119-136.
37. Cortez, P.; and Silva, A.M.G. (2008). Using data mining to predict secondary school student performance. *Proceeding of the 5th Future Business Technology Conference*. Porto, Portugal.