

E-Learning Recommender System for Learners: A Machine Learning based Approach

Kamika Chaudhary*, Neena Gupta

Department of Computer Science
Gurukul Kangri Vishwavidyalaya, Haridwar, India
**Corresponding author: kamika.agrohi@gmail.com*

(Received September 15, 2018; Accepted April 4, 2019)

Abstract

Web mining procedure helps the surfers to get the required information but finding the exact information is as good as finding a needle in a haystack. In this work, an intelligent prediction model using Tensor Flow environment for Graphics Processing Unit (GPU) devices has been designed to meet the challenges of speed and accuracy. The proposed approach is isolated into two stages: pre-processing and prediction. In the first phase, the procedure starts via looking through the URLs of various e-learning sites particular to computer science subjects. At that point, the content of looked through URLs are perused and after that from their keywords are produced identified with a particular subject in the wake of playing out the pre-processing of the content. Second phase is prediction that predicts query specific links of e-learning website. The proposed Intelligent E-learning through Web (IEW) has content mining, lexical analysis, classification and machine learning based prediction as its key features. Algorithms like SVM, Naïve Bayes, K-Nearest Neighbor, and Random Forest were tested and it was found that Random Forest gave an accuracy of 98.98%, SVM 42%, KNN 63% and Naïve Bayes 66%. Based on the results IEW uses Random forest for prediction.

Keywords- Web mining, Machine learning, Lexical analysis, Prediction modelling, GPU, Tensor flow.

1. Introduction

The ubiquitous nature of the internet has led to its advancement and growth at a very significant pace. This allows an increase in the popularity of various online web applications such as electronic commerce, distance education, online collaboration, and news broadcasts etc. One of them, which are holding an important place, is online learning. There has been a great development in the area of web-dependent education in upcoming years due to which a significant increase has been noticed in the number of online-based learning systems. This e-learning system exhibits the characteristic of being heterogeneous in nature along with its presence in several media formats (Ghauth & Abdullah, 2010). Due to such a high availability of online learning systems, keyword-based search query method has started losing its relevance. To find the appropriate result for their e-learning related search query has become a cumbersome task for learner (Khribi et al., 2008). This problem leads to the requirement of personalization in internet-based learning environment. Recommendation techniques have been utilized by researchers in order to tackle the information overload problem in web based-learning platforms.

Recommender Systems (RSs) mark their presence as a technology that plays a vital role of rescuer to tackle the information and product overload problem (Adomavicius & Tuzhilin, 2005). Major motive of these systems is to provide the user specific requirements in the form of suggestions by considering their preferences in an explicit or implicit manner. E-learning recommender systems consider the learners as well as their learning activities in order to recommend the learning items that can be an article, web pages or learning web sites etc. Collaborative filtering is a kind of recommendation strategy whose main function is to consider the user with similar choices and then

providing the suggestions based on those similarities. This technique is most popular and widely accepted for recommendation systems in all the domains of either e-learning or e-movies or e-music. Generally, two recommendation techniques collaborative and content based filtering exists for e-learning recommendation system to provide effective recommendations (Bobadilla et al., 2009). Content-based approach considers items or contents as their focal point and examines the properties of these items in predicting the next useful item whereas collaborative filtering finds the similarity among users and their accessing patterns and they recommend those items that are preferred by similar users. These traditional approaches suffer by cold start problem as well as sparsity and scalability issues. These drawbacks have caused the emergence of newly developed techniques for recommending learning resource to the learner by exactly capturing and considering learner requirement of learners in e-learning systems. As each learner have different requirements depending on their learning style, knowledge level or learning goals, an open environment is required in which he can express his specification in natural language without any restrictions.

The paper proposes a unique intelligent web-searching model for the e-learners. Everything is available on the web today; all we require is relevant information to the surfer. Most of the mining techniques depend on the meta-keywords, historical data, cached pages and keywords created from contents of the pages. The result is that the surfer has to find various combinations to get what it wants, as most of the time irrelevant information is fetched in the response of user search query. The proposed work reads the contents of the e-learning pages, carry out lexical analysis of the text, retains the main features, basically the keywords and the Random Forest algorithm is applied to predict the relevant sites. Beside correct information, surfers would like the results to displayed fast, keeping speed in mind the proposed model was designed using Tensor Flow environment for GPU cards (Wang & Alexander, 2016). The proposed model is applicable for heterogeneous systems with a little or no change.

The paper is organized in different sections; section 2 presents existing research done in the area. Section 3 presents proposed approach for preprocessing and prediction phase. Section 4 shows an experimental evaluation and then section 5 analyses various algorithms and their accuracy. Section 6 concludes the work.

2. Related Work

In the last few years, growth in the area of e learning has acquired a faster pace that generally leads to making it difficult for learners to find their most suitable learning resource easily. There arises a need for improving the searching efficiency, which is fulfilled by proposing the recommender system for e learning. These systems function by providing automatic suggestions for the most suitable learning resources for learners. Due to the high importance of e learning in the area of education, a lot of research has been conducted by utilizing a number of techniques.

Herath and Jayaratne (2017), proposed a web mining based recommendation system, in which a number of parameters are considered for giving suggestions of resources to learners such as navigational pattern, web content, profile and past and present performance of learner. In this, authors have used content and collaborative filtering technique in order to cluster learner and web content group that ultimately leads to providing suggestions.

Salehi et al. (2013) presented a solution for the following two problems sparsity and cold start because as the field of recommendation grows it has been noticed that the recommendations to any e-learning systems suffer from these two major challenges. Here a recommendation system

utilizing both implicit and explicit attribute has been proposed by authors. Nearest-neighbor collaborative filtering technique has been applied for generating recommendations both implicit and explicit attribute along with learner preference tree for the explicit and genetic algorithm for implicit attributes. A number of performance measures like precision, recall and mean absolute error have been applied and it has been found that the accuracy of the system surpasses its ancestors.

In the next work of Choudhary et al. (2017), authors have utilized particle swarm optimization technique for proposing a Multi-Criteria Recommender System (MCRS) in which the focus is on considering several criteria for rating any item in order to get user real preferences. It mainly targets providing recommendations that are more accurate and hence preference is given to multiple criteria for rating. In order to make system performance better different similarity measures such as Jaccard and Dice was utilized and the performance of the system reveals its superiority on traditional systems.

Fraihat and Shambour (2015) explored the semantic relationship between learning objects and learner's requirement. Here a query keyword based extension algorithm has been proposed and utilized for developing semantic based framework for recommendation.

Tam et al. (2014), introduced a more complete recommendation framework incorporating a trio of semantic analysis, concept clustering, and learning path optimization. In this, a pre-processed concept information reaches to rule based evolutionary algorithm after passing through concept clustering algorithm with the aim of generating optimized learning paths of learners. The implementation of the overall process had performed and a hill climber system utilized for optimization and the performance of the system compared with benchmark shortest-distance optimizer. The concept utilized and implemented found to be effective by authors.

Labib et al. (2017) presented a cross model that brings together learning style framework dimensions and learning style along with appropriate learner's characteristics in the form of learner's attributes ontology. This process is advantageous in providing suitable content as well as content reusability. Bernard et al. (2017) defined how the precision and accuracy in automatically finding out the learning style required an improvement place to hold. For this, authors proposed an approach by utilizing intelligence algorithm named as the artificial neural network, genetic algorithm, ant colony system, and particle swarm optimization and resulted in improving the precision with a noticeable difference.

Fok et al. (2018) described another model for prediction of student future degree program on the basis of his present performance in the academic and non-academic field. For achieving this tensor flow, based deep learning environment is used and there is a significant improvement in prediction accuracy. Wan and Niu (2018) defined how diversity and adaptability are the two important factors that need to be achieved by learning recommendation system. For achieving these two parameters a distributed self-organization based recommendation technique has been given. Application on self-organizing maps on recommendation system helps in achieving better performance and reduced excessive recommendations.

Moubayed et al. (2018) discussed about a k-nearest neighbor based classification method and based on this proposed a Really Simple Syndication (RSS) reader website in which the dataset is pre-processed in the first phase and then after developing classifier recommendations provided to readers in real time. The proposed approach found to be useful when less knowledge about data

distribution is present. Dwivedi et al. (2018), paper handled the challenges faced by the recommender system of curriculum sequencing. A variable length genetic algorithm had been used to propose a system that resulted in providing optimal learning paths by considering both the learner’s needs and his preferences.

3. Proposed Approach

The proposed approach composed of two phases: pre-processing and prediction. In the first phase, the process begins by searching the URL of different e-learning websites specific to computer science subjects. The contents of searched URL’s is read and then from their keywords are generated related to the specific topic after performing the pre-processing of the content. Retained keywords along with their associated URL represented by using notations u_1 , u_2 , u_3 , and so on are used to form a classified dataset.

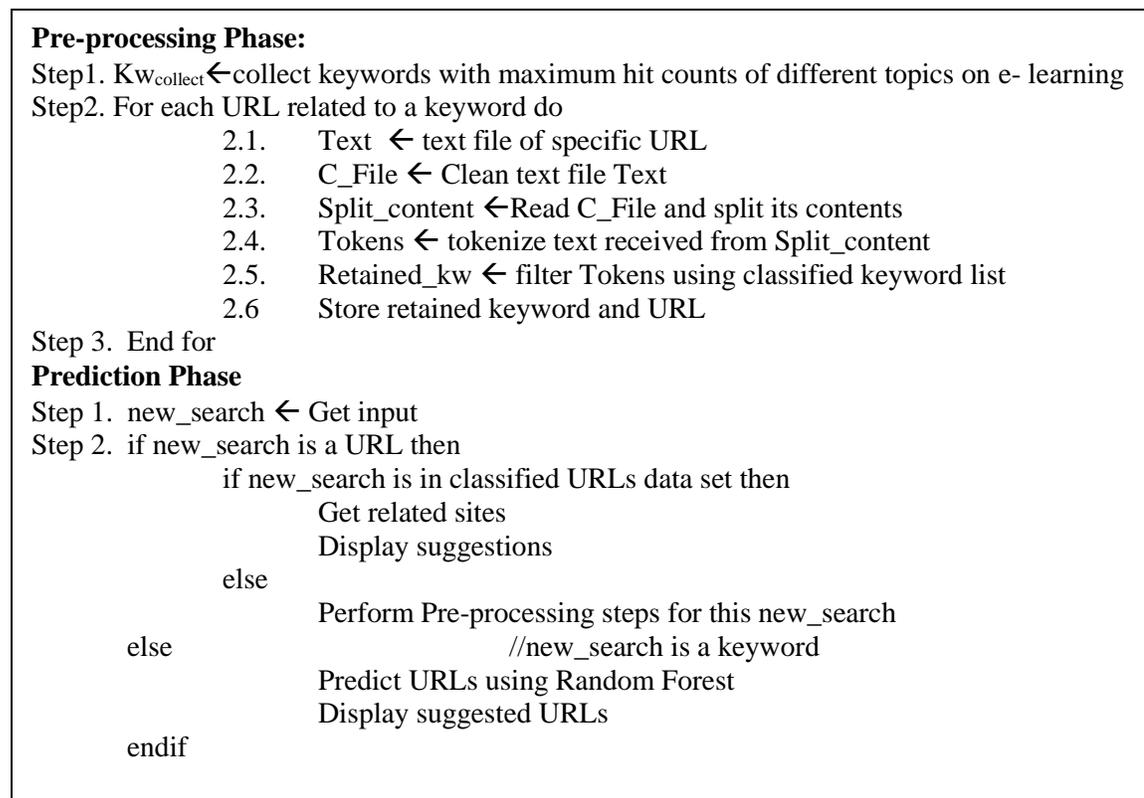


Figure1. Proposed algorithm

Second phase starts with a new search (*new_search*) of the user. This *new_search* could be a URL or a keyword if it is URL then search in the classified dataset if it is not in URL’s list then perform pre-processing steps mentioned above. In addition, if it is found in the dataset then simply display the suggestions. Last possibility of the *new_search* is to be a keyword. In this case, prediction algorithm is applied and gives suggested URLs. These suggestions are the recommended result of Random Forest Algorithm. A description of the proposed algorithm is presented in Figure 1 and

flow charts of the overall process of pre-processing and prediction phase is given in Figure 2, and 3 respectively.

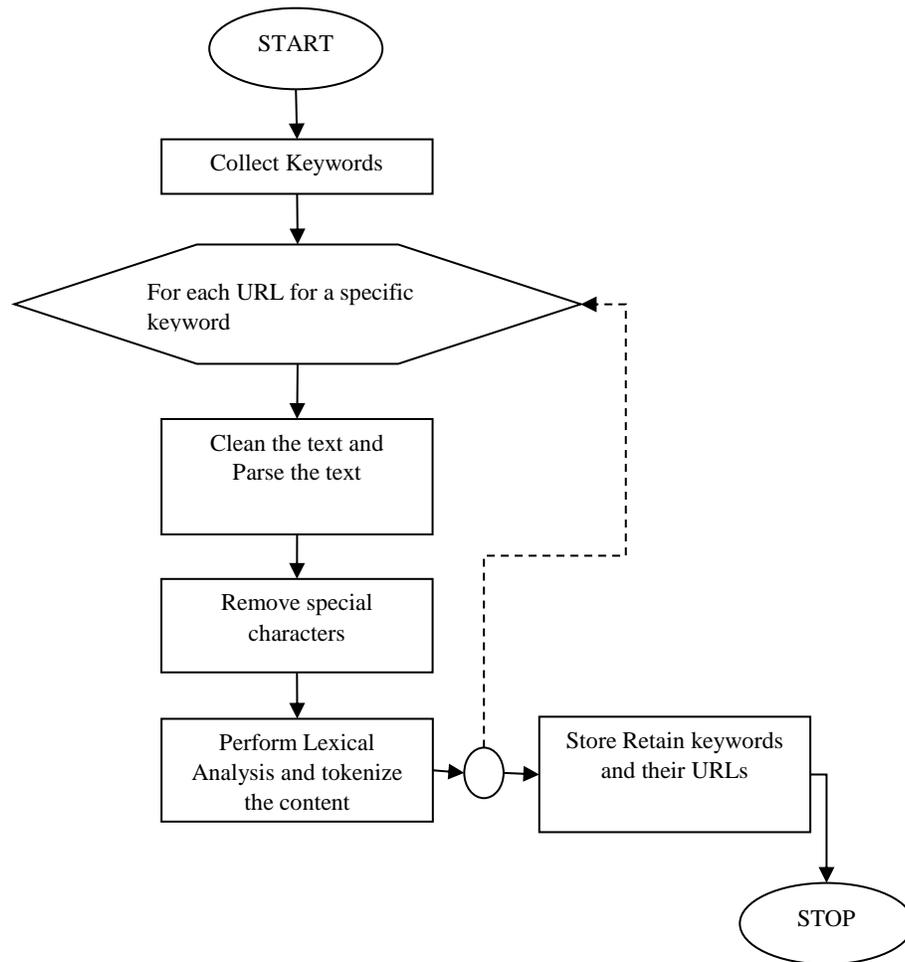


Figure 2. Pre-processing phase

Both the phases, pre-processing and prediction applied to the dataset collected for the domain of online education specific to computer science. The application of both the phases resulted into an intelligent e-learning system for the web by suggesting the recommendations to user in the form of URL's. This recommendation makes it easier for the user to go to his page of interest and eradicating the process of wandering from one web page to another. This saves his/her time also resulted into an increase in confidence of user on the web page.

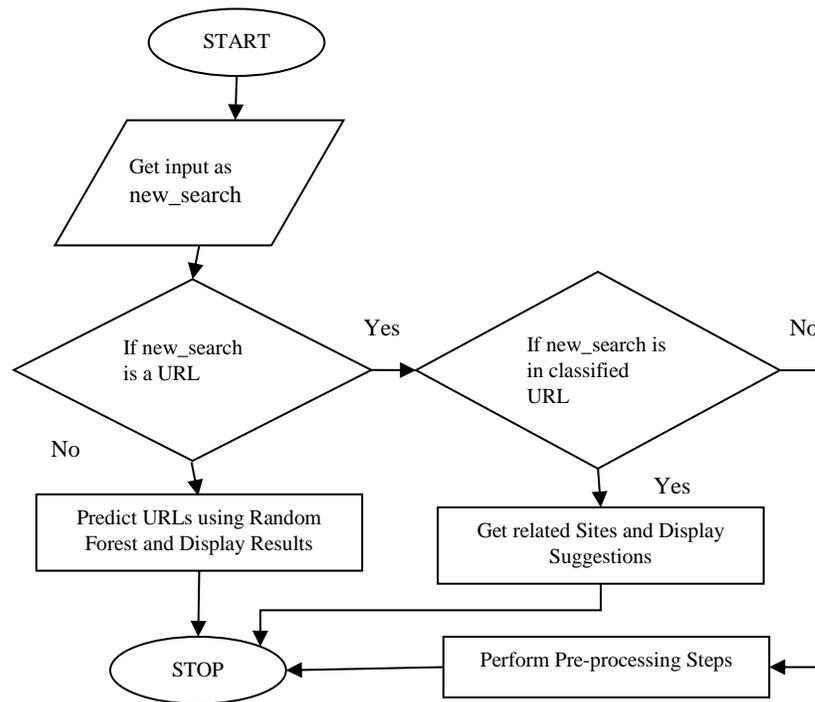


Figure 3. Prediction phase

4. Experimental Evaluation

For performing the experiment, a fifth-generation Intel core i5 system with 16 GB RAM used. Python 3.6 under anaconda cloud environment works as a programming language. A classified dataset consisting of keywords and URL's is shown in Figure 4. This dataset is a matrix of 0's and 1's, where 'zero' represents the absence of a particular keyword on a particular URL and 'one' represents its presence. Now this dataset divided in the ratio of 70%:30% for the purpose of training and testing respectively. Testing dataset mainly used for serving the validation requirement of the proposed work.

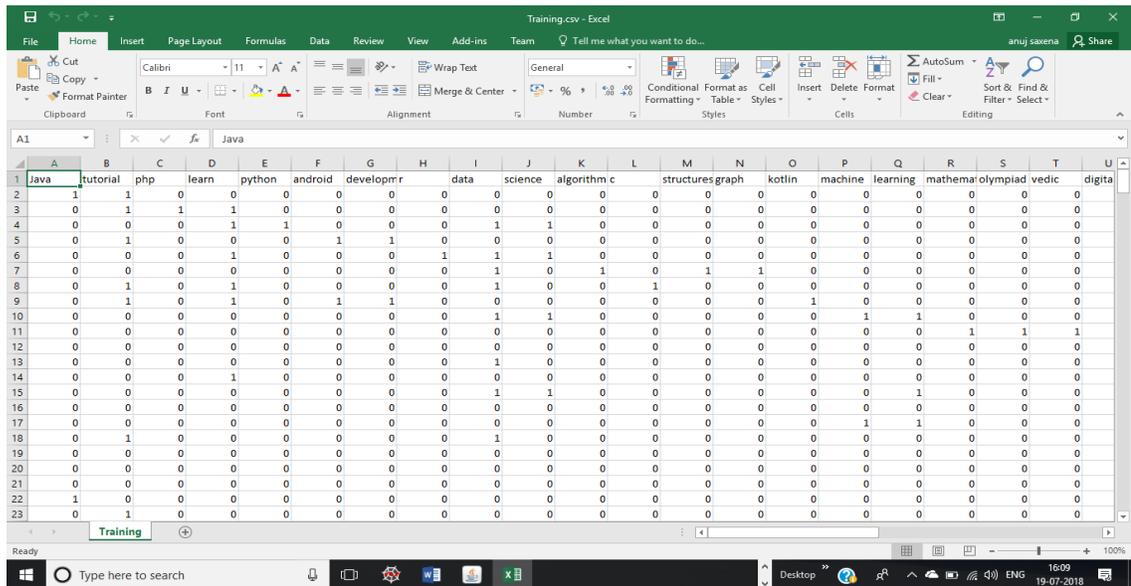


Figure 4. Classified dataset created

Once dataset created and bifurcated into training and testing part then a user- based interface is developed, in which a user puts up his query in the form of keywords. The application of Random Forest on the classified dataset is performed and it works on the weights of the data so it will calculate the frequency of keywords in the URL's by using its standards mean, median and mode.

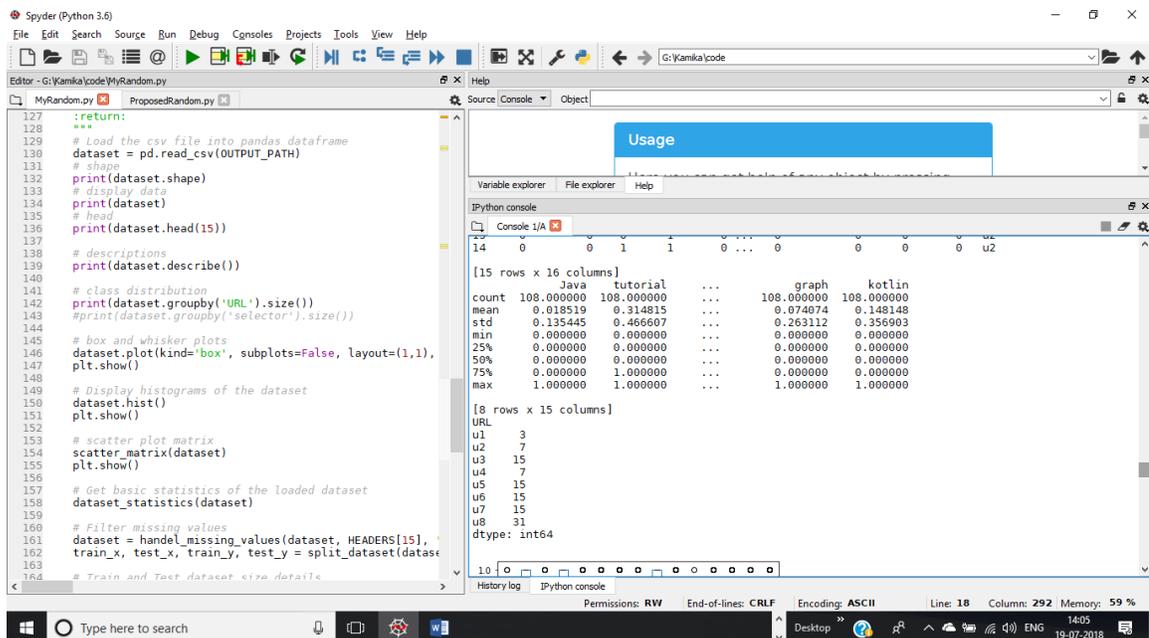


Figure 5. URL classification considered for testing

Figure 5 shows the classification of URL's considered for testing. As it can be seen from the snapshot that three datasets belong to class U1, seven belong to U2 and so on.

A relation between loss and accuracy is depicted in Figure 6. Both loss and accuracy share an inverse relationship with each other as with the decrease in the loss, accuracy is increasing in every run of the proposed algorithm.

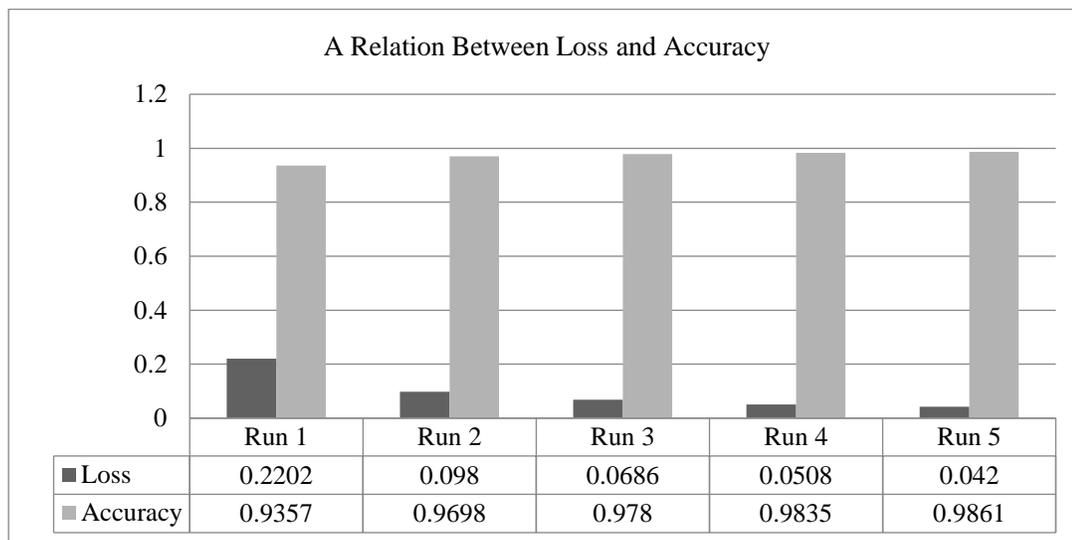


Figure 6. Relationship between loss and accuracy

5. Performance Analysis

The criterion of performance evaluation of the proposed approach is based on several statistical measures such as:

True Positive (TP) is the number of truly recognized class examples which are correctly predicted URL's that is actual URL is one (True) and predicted URL is one (True).

True Negative (TN) is the number of correctly recognized examples which are not a part of class that is actual class is zero (False) and predicted class is zero (False).

False Positive (FP) is the cases when the actual class of the URL point was zero (False) and the predicted is one (True).

False Negative (FN) is incorrectly identified URL's examples that is when the actual class of the URL point was one (True) and the predicted is zero (False).

These four parameters (TP, FP, TN and FN) are utilized to evaluate the accuracy of the proposed algorithm by utilizing the following formula (Anand & Bansal, 2016):

$$Accuracy = (TP + TN) / (TP + FP + TN + FN).$$

The accuracy of the proposed algorithm is calculated and compared with the accuracy of another existing algorithm such as k-Nearest Neighbor (kNN), Naive Bayes (NB) and Support Vector Machine (SVM) applied on same data set (Sadhasivam and Kalivaradhan, 2019). For instance, for the proposed algorithm a keyword “java free online” has been put by the user as a query and in return, he is getting the link of an e-learning site containing the java content. The correctly predicted URL according to user input query is <https://www.tutorialspoint.com/java/index.htm>. This shows the prediction system is performing correctly and recommending the links of keyword related website to the user.

The plotted graph in Figure 7 shows the maximum accuracy in the case of proposed random forest algorithm, which is 98.98%. This percentage of accuracy is much higher than other existing algorithms.

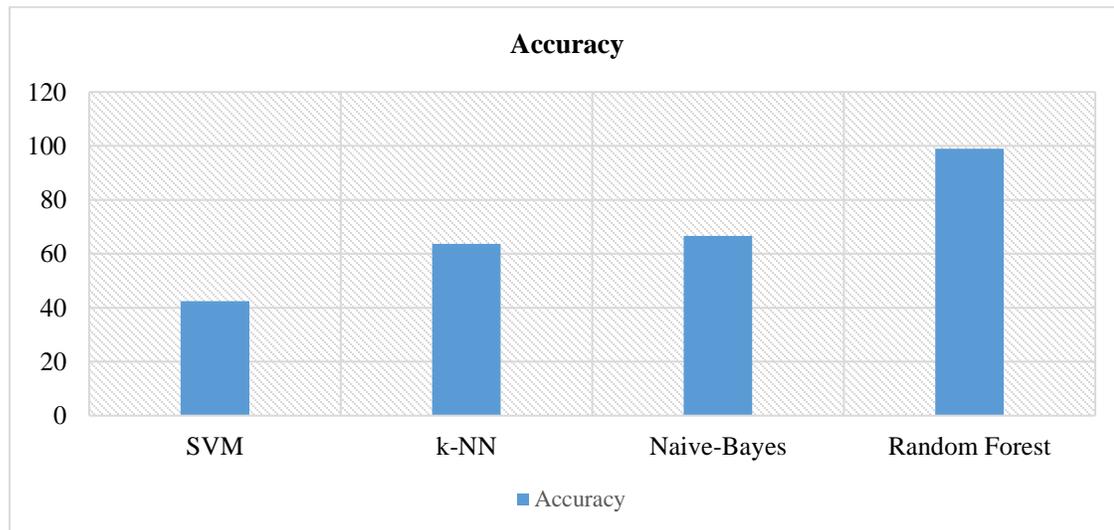


Figure 7. Comparison of various machine learning algorithms accuracy

6. Conclusion

E learning has experienced quick improvement lately. With the development of numerous web based learning frameworks, countless learning assets have been produced which are exceptionally heterogeneous and in different media designs. It is troublesome for students to find the most proper assets utilizing a watchword look strategy. Researchers utilize recommendation techniques to resolve information overload in the new learning environment. The proposed approach has content mining, lexical analysis, classification and machine learning based prediction as its key features. It is a two-phase approach. In the first phase, data is pre-processed and the second phase is concerned with prediction. Proposed approach compared with various machine learning based algorithms and it is found that Random forest algorithm gives the highest accuracy for prediction.

Conflict of Interest

The authors declare that there is no conflict of interest to declare for this publication.

Acknowledgment

The authors would like to thank anonymous reviewers for their constructive comments to improve the quality of this paper.

References

- Adomavicius, G., & Tuzhilin, A. (2005). Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge & Data Engineering*, 17(6), 734-749.
- Anand, A., & Bansal, G. (2016). Predicting customer's satisfaction (dissatisfaction) using logistic regression. *International Journal of Mathematical, Engineering and Management Sciences*, 1(2), 77-88.
- Bernard, J., Chang, T. W., Popescu, E., & Graf, S. (2017). Learning style identifier: improving the precision of learning style identification through computational intelligence algorithms. *Expert Systems with Applications*, 75, 94–108.
- Bobadilla, J., Serradilla, F., & Hernando, A. (2009). Collaborative filtering adapted to recommender systems of e-learning. *Knowledge-Based Systems*, 22(4), 261–265.
- Choudhary, P., Kant, V., & Dwivedi, P. (2017, February). A particle swarm optimization approach to multi-criteria recommender system utilizing effective similarity measures. In *Proceedings of the 9th International Conference, 2017, on Machine Learning and Computing*. ACM.
- Dwivedi, P., Kant, V., & Bharadwaj, K.K. (2018). Learning path recommendation based on modified variable length genetic algorithm. *Education and Information Technologies*, 23(2),
- Fok, W.W., He, Y.S., Yeung, H.A., Law, K.Y., Cheung, K. H., Ai, Y.Y., & Ho, P. (2018, May). Prediction model for students' future development by deep learning and tensorflow artificial intelligence engine. In *4th International Conference, 2018, on Information Management (ICIM)*. IEEE.
- Fraihat, S., & Shambour, Q. (2015). A framework of semantic recommender system for e-learning. *Journal of Software*, 10(3), 317-330.
- Ghauth, K.I., & Abdullah, N.A. (2010). Measuring learner's performance in e-learning recommender systems. *Australasian Journal of Educational Technology*, 26(6), 764-774.
- Herath, D., & Jayaratne, L. (2017, September). A personalized web content recommendation system for E-learners in E-learning environment. In *National Information Technology Conference, 2017, (NITC)*. IEEE.
- Khribi, M.K., Jemni, M., & Nasraoui, O. (2008, July). Automatic recommendations for e-learning personalization based on web usage mining techniques and information retrieval. In *Eighth International Conference, 2008, on Advanced Learning Technologies*. IEEE.
- Labib, A.E., Canós, J.H., & Penadés, M.C. (2017). On the way to learning style models integration: a Learner's Characteristics Ontology. *Computers in Human Behavior*, 73, 433–445.
- Moubayed, A., Injadat, M., Nassif, A.B., Lutfiyya, H., & Shami, A. (2018). E-learning: challenges and research opportunities using machine learning data analytics. *IEEE Access*, 6, 39117–39138.

- Sadhasivam, J., & Kalivaradhan, R.B. (2019). Sentiment analysis of Amazon products using ensemble machine learning algorithm. *International Journal of Mathematical, Engineering and Management Sciences*, 4(2), 508–520.
- Salehi, M., Kamalabadi, I.N., & Ghouschi, M.B.G. (2013). An effective recommendation framework for personal learning environments using a learner preference tree and a GA. *IEEE Transactions on Learning Technologies*, 6(4), 350-363.
- Tam, V., Lam, E.Y., & Fung, S.T. (2014). A new framework of concept clustering and learning path optimization to develop the next-generation e-learning systems. *Journal of Computers in Education*, 1(4), 335–352.
- Wan, S., & Niu, Z. (2018). An e-learning recommendation approach based on the self-organization of learning resource. *Knowledge-Based Systems*, 160, 71–87.
- Wang, L., & Alexander, C.A. (2016). Additive manufacturing and big data. *International Journal of Mathematical, Engineering and Management Sciences*, 1(3), 107–121.

