Volume 1, Issue 2 (2019)
Article No. 1
PP 1-6

VIVA-TECH IJRI
ANNUAL RESEARCH JOURNAL

VIVA-TECH INTERNATIONAL JOURNAL FOR RESEARCH AND INNOVATION
ANNUAL RESEARCH JOURNAL
ISSN(ONLINE): 2581-7280

SCCAI- A Student Career Counselling Artificial Intelligence

Aditya M. Pujari[1], Rahul M. Dalvi[1], Kaustubh S. Gawde[1], Tatwadarshi P. Nagarhalli[1]
[1](*Computer Engineering Department, VIVA Institute of Technology, India*)

Abstract: *As education is growing day by day, the competition has prompted a need for the student to understand more about the educational field. Many times the counselor isn't available all the time and sometimes due to the lack of proper knowledge about some educational field. Due to this, it creates an issue of misconception of that field. This creates a problem for the student to decide a proper educational trajectory and guidance is not always useful. The proposed paper will overcome all these problem using machine learning algorithm. Various algorithms are being considered and amongst them the best suitable for our project are used here. There are 3 major problems that come across our path and they are solved using Random forest, Linear regression and Searching algorithm using Google API. At first Searching algorithm solves the problem of location by segregating the college's location vice, then Random Forest provides the list of colleges by using stream and range of percentage and finally Linear Regression predicts the current cutoff using previous years' data. Rather than this, the proposed system also provides information regarding all fields of education helping students to understand and know about their field of interest better. The following idea is a total fresh idea with no existing projects of similar kind. This project will help students guide them throughout.*

Keywords *– Machine learning, Random Forest, Linear Regression, K-means, Chatbot.*

1. INTRODUCTION

Artificial Intelligence is also known machine intelligence, is intelligence demonstrated by different machine [14]. Artificial Intelligence is defined as the research of intelligent agents, a device that can learn information from environment and performs action that maximize the chances of successfully achieving its goals. Modern machine capabilities generally classified as artificial include successfully understanding human speech, competing at the highest level in strategic game, autonomously operating cars, and intelligent routing in content delivery networks and military simulations. Artificial intelligence research has been divided into subfields that often fail to communicate with each other. These sub-fields are based on technical consideration, such as particular goals, the use of particular tools, or deep philosophical differences [14]. Artificial Intelligence often revolves around the use of algorithms. An algorithm is a set of unambiguous instructions that a mechanical computer can execute. A complex algorithm is often built on top of other, simpler algorithms. Artificial Intelligence algorithm are capable of learning from data; they can enhance themselves by learning new heuristics, or can themselves write different algorithms. Some of algorithm used Bayesian network, decision trees and nearest-neighbor. This learning of data using different algorithm is known as machine learning.

Machine learning is an interdisciplinary field that uses statistical techniques to give computer systems the ability to learn from data, without being explicitly programmed. Machine learning explores the study and construction of algorithm that can learn for and make prediction on data-such algorithms overcome following strictly static program instructions by making data-driven and make prediction or decisions, through building a model from sample inputs [13]. Machine learning is employed in a range of computing tasks where designing and programming explicit algorithm with good performance is difficult or infeasible. Machine learning is closely related to computational statistics, which also focuses on prediction-making through the use of computers. Within the field of data analytic, machine learning is a method used to devise complex models and algorithms that lend themselves to prediction; in commercial use, this is known as predictive analytics [13]. These analytical models allow researchers, data scientists, engineers, and analysts to "produce reliable, repeatable decisions and results" and uncover "hidden insights" through learning from historical relationships and trends in the data.

Volume 1, Issue 2 (2019)
Article No. 1
PP 1-6

2. RELATED WORK

B. R. Ranoliya et. Al. [] had stated that Artificial Intelligence conversational agents are becoming popular for web services and systems like scientific, entertainment and commercial systems, and academia. But more effective human-computer interaction will take place by querying missing data by the user to provide satisfactory answer. User inquiries are first taken care by AIML check piece to check whether entered inquiry is AIML script or not. AIML is characterized with general inquiries and welcome which is replied by utilizing AIML formats

T. R. V. Anandharajan, et. al. [5] had defined weather prediction based on the previous dataset. They intend to develop an Intelligence Weather predicting module.

S. Kumar, et. al. [8] have focused on use of the Data Mining techniques for predicting rainfall of an area on basis of some dependent feature like precipitation and wet day frequency.

S. Prabakaran, et. al. [7] have proposed rainfall prediction on the historical data is trending in research point of view. The existing model use the data mining technique for predicting the state of atmosphere at a given time of a weather variable like rainfall, cloud conditions, temperature etc.

H. L. Siew, et. al. [6] have examined the theory and practice of the regression technique for prediction of stock price by using the transformed data set in ordinal dataset. In this the original pre-transformed data source contain data of heterogeneous data type use for handling of currency value and financial ratio.

Y. Liu, et. al. [2] have explained the problem of traffic congestion is solved by using classification algorithm random forest. The city area is divided in sections and prediction is done of the areas possible of having heavy traffic. This is done by considering the environmental conditions such as Climate, Holiday, Road Condition etc. The results show that the traffic prediction model established by using the random forest classification algorithm has a prediction accuracy of 87.5%.

X. Xun, et. al. [9] have defined the management of land resources, not only solving the existing problems of the land, instead also prediction of the problems of the land and prevention on land misuse are in demand urgently due to the urbanization so that the propose system use Random Forest algorithm for prediction.

A. Ghosh, et. al. [1] hqve defined the problem of urbanization is solved by using Random Forest algorithm along with Landsat archive and ancillary data. It proposes a methodology to map the urban areas with multi-seasonal Landsat data. The Random forest classifier and decision level fusion are applied. The paper gives the general idea about the random forest algorithm of urban landscape.

Y. C. Shiao, et. al. [10] have studied that according to the statistical data, each day there are over one million passengers taking the MRT in Taipei. In this paper, author did a predicting MRT passenger flow with random forest, by using different factors collected from the Taipei Main station as input for training. In this paper, system use only the Taipei main station passenger flow to test the method.

H. Zhang, et. al. [4] have examined each chromosome is made up of a sequence of genes coding. The number of genes of a chromosome is randomly chosen where n is the number of data points, which is randomly selected a given data sets. Canopy is usually employed to capture the number of clusters.

M. Lehsaini, et. al. [3] have proposes a cluster-based routing scheme based on an enhanced version of K-means approach. The improved version of K-means generates balanced clusters in the network, which does not overload one cluster-head over the others unlike LEACH where one of the generated clusters may contain a large number of nodes and another contains a small number of members. This paper proposes a cluster-based routing scheme based on an enhanced version of K-means approach.

S. Ye, et. al. [11] have stated K-means algorithm is a clustering algorithm based on partition. Because of its simplicity and efficiency, it has become one of the most widely used clustering algorithms. The original cuckoo algorithm is influenced by step size A and probability of discovery P, and the step size and discovery probability control the accuracy of CS algorithm global and local search, which has great influence on the optimization effect of algorithm. K-Means algorithm is easy to fall into the local optimum and the Cuckoo search (CS) algorithm is affected by the step size.

Volume 1, Issue 2 (2019)
Article No. 1
PP 1-6

Z. Ya-Ling, et. al. [12] In the literature present an agglomerative fuzzy K-means clustering algorithm for numerical data, an extension to the standard fuzzy K-means algorithm by introducing a penalty term to the objective function to make the clustering process not sensitive to the initial cluster centres. The paper extends the K-means clustering process to calculate a weight for each dimension in each cluster and use the weight values to identify the subsets of important dimensions that categorize different clusters.

3. PROPOSED SYSTEM

The proposed system works on machine learning algorithm, which includes Linear Regression and Random Forest. Dialogflow is a Google-owned developer of human–computer interaction technologies based on natural language conversations. Dialogflow is best known for creating a virtual assistance for Android, iOS, and Windows Phone smartphones that performs tasks and answers users' question in a natural language. Dialogflow has also created a natural language processing engine that incorporates conversation context like dialogue history, location and user preferences. The proposed system also uses Google Dialogflow(API.ai) for Natural Language processing(NLP). Google Assistance is the main framework for the system. The dataset was not available at any database information websites. The dataset was developed manually by using available physical data. Different machine learning algorithm uses different features from dataset.

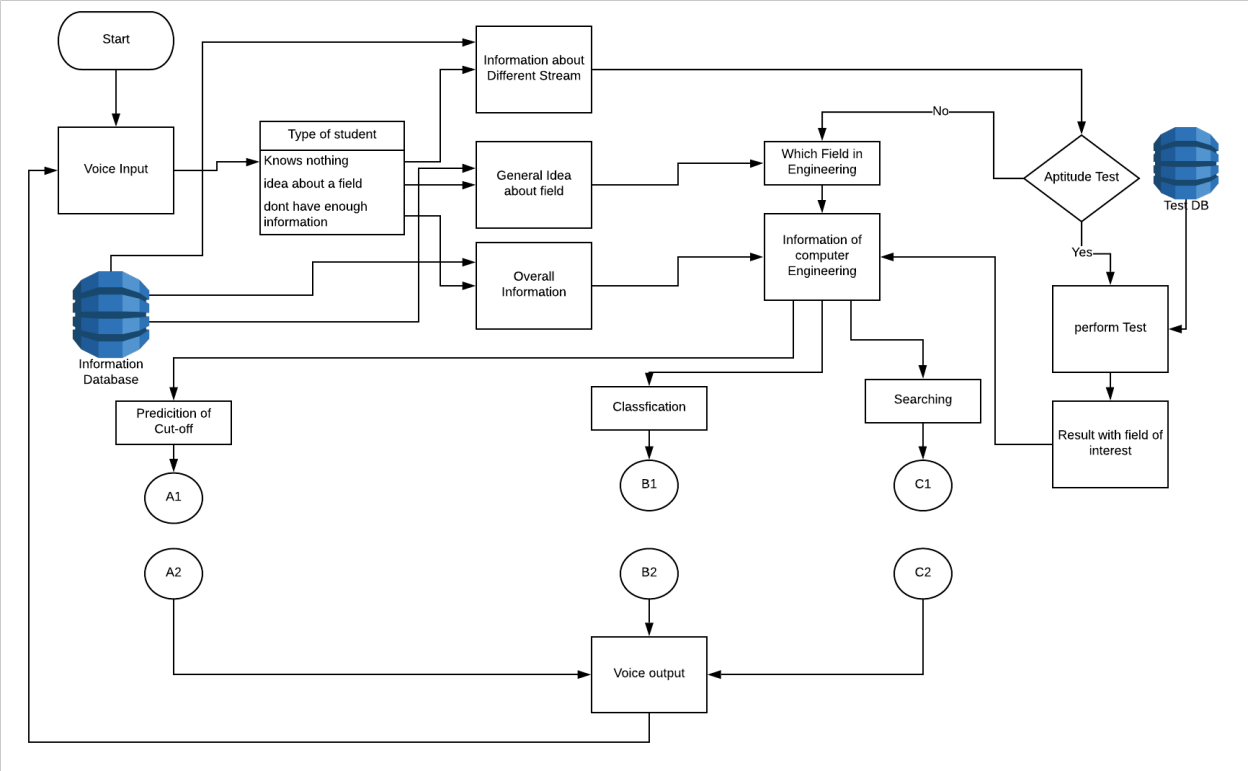

Figure 3.1 System Flow Diagram

Figure 3.1 shows the flow of overall proposed system. As the user gives command “Talk to hey SCCAI” the bot receives the activation signal and gets activated. Then the proposed system requests for input from user about how familiar he/she is regarding educational fields and also their knowledge about those fields. Later analyzing the given information, it decides a path of how to approach the user and solve user’s problem. The problems of the users are divided mainly into 3 types, they are: Information about different streams, general idea about the fields, overall information. When the student chooses the first option he/she is directly directed to the aptitude test, when he/she opts for the second option they get a question which field they want to know about and then they are provided with the desired information they required. And finally when they opt the third option they are directly provided with the information. All this information is stored in a data base which is updated regularly, in all three problems the solution is fetched from the database. Also the aptitude test contains the questions that are randomly selected amongst the questions stored in the data base. This test gives the system the required field of interest of the user. Now all this information is provided to achieve various answers to the questions like college name, cut-off list of college, and desired college with its location. All this is done with voice command in a simple question answer way so that all group age people can use this.
people can use this.

Volume 1, Issue 2 (2019)
Article No. 1
PP 1-6

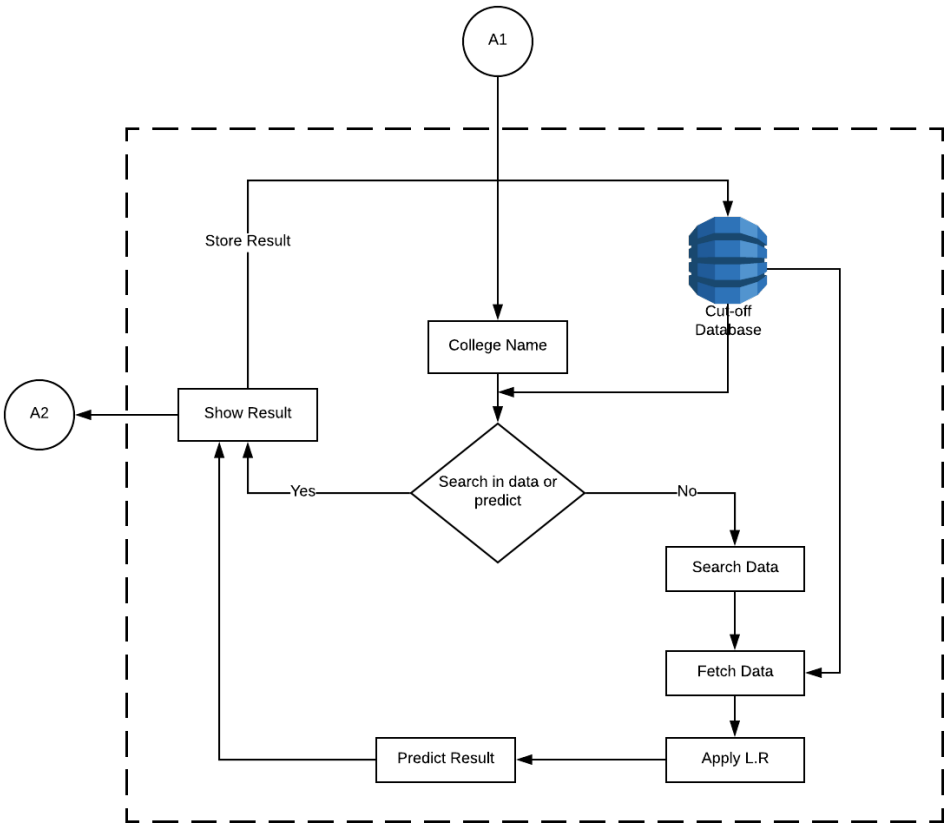

Figure 3.2 Linear Regression

After finishing the informative part, the next step is the working of algorithms to solve various problems and provide the required information to the user. A1 here is the user request to perform Linear Regression algorithm to provide the desired list of college cut-off. Figure 3.2 shows the flow of linear regression. Here the input is taken from the data base where the data of previous all cut-off is stored..

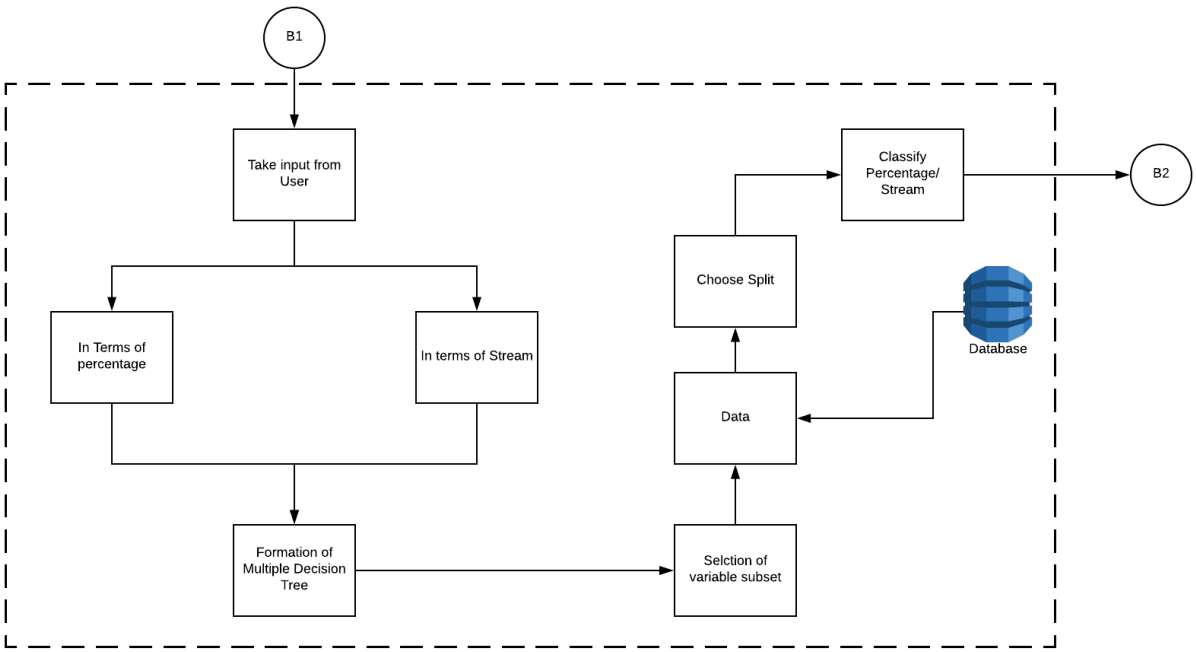

Figure 3.3 Random Forest

Volume 1, Issue 2 (2019)
Article No. 1
PP 1-6

Figure 3.3 shows random forest algorithm. Random forest is used here to classify colleges according to range of percentage and stream. B1 is where the user provides input and B2 is the required output here. The user will provide a range of percentage and according to that classification algorithm will be undertaken.

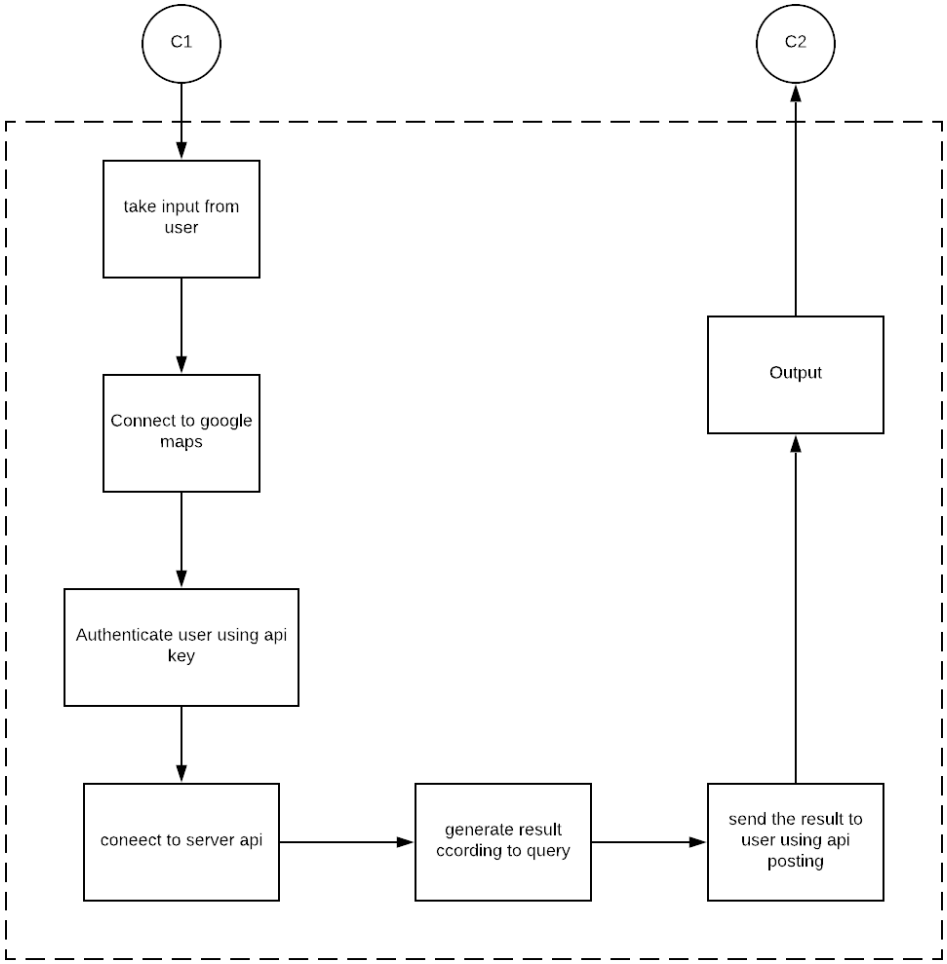

Figure 4.6 Search API

Figure 4.6 shows the flow of Google maps search API. User can provide location according to their needs. Here C1 is where the user request for the information and C2 is where the output is generated.

4. RESULTS AND ANALYSIS

The system is unique in nature and there is no such existing system in market. The system is created using google assistant as framework. The natural language processing is done by google API. The system voice output time is quick and have moderately high accuracy. The proposed system will use various machine algorithm.

5. CONCLUSIONS

Education system is growing very rapidly and with this rapid growth the competition to be the best amongst all has also increased. The system is unique and has never been implemented before. Here we have defined five main features of the system. The first feature helps the user to gain all possible information regarding education and education system. The second feature is a psychometric test that helps the user to know where he stands or can see a better self in the future. The third feature provides the location of colleges. The fourth provides the cutoff of various colleges with respect to cast as well and the fifth provides the list of colleges to the user. Clubbed together these five features act as one single artificial brain that helps as a counselor to the user.

The proposed system uses various machine learning algorithms to solve various problems. The random

forest helps in providing solution to predict the college list that the user wants provided his percentage and stream. The linear regression helps with the solution of college cutoff and google maps API helps with the location of the college the user wants to know about. The communicative part of the system is taken care by the dialogue flow, where it helps with natural language processing.

REFERENCES

[1] A. Ghosh, R. Sharma, P.K. Joshi, "Random forest classification of urban landscape using Landsat archive and ancillary data: Combining seasonal maps with decision level fusion", Applied Geography Journal, 2014, pp. 31-41.

[2] Y. Liu, H. Wu, "Prediction of Road Traffic Congestion Based on Random Forest", 10th International Symposium on Computational Intelligence and Design, 2017, pp. 361-364.

[3] M. Lehsaini, M.B. Benmahdi, "An improved K-means Cluster-based Routing Scheme for Wireless Sensor Networks", IEEE, 2018.

[4] H. Zhang, Z. Zhou, "A Novel clustering algorithm combining Niche genetic algorithm with canopy and K-means", International Conference on artificial Intelligence and Big Data, 2018, pp. 26-32.

[5] T.R.V. Anandharajan, G.A. Hariharan, K. K. Vignajeth, R. Jitendiran, "Weather Monitoring Using Artificial Intelligence", International Conference on Computational Intelligence and Networks, 2016.

[6] H. L. Siew, M.J, Nordin, "Regression Techniques for the Prediction of Stock Price Trend", International Conference on Statistics in Science, Business and Engineering (ICSSBE), 2012, pp. 1-5.

[7] S. Prabakaran, P. N. Kumar, P. S. M. Tarun, "Rainfall Prediction Using Modified Linear Regression", ARPN Journal of Engineering and Applied Sciences, 2017, pp. 3715-3718

[8] S. Kumar, M. Anamika Upadhyay, C. Gola, "Rainfall prediction based on 100 years of Meteorological data", IEEE, 2017, pp. 162-166.

[9] X. Xun, L. Mo, Y. Yu, "Discovery and Prediction of the Unused Land for Construction Based on Random Forest", Fifth International Conference on Agro-Geoinformatics, 2016.

[10] Y. C. Shiao, L. Liu, Q. Zhao, R. C. Chen, "Predicting Passenger Flow using Different Influence Factors for Taipei MRT System", IEEE 8th International Conference on Awareness Science and Technology (iCAST), 2017.

[11] S. Ye, X. Huang, Y. Teng, Y. Li, "K-Means Clustering Algorithm Based on Improved Cuckoo Search Algorithm and Its Application", IEEE 8th International Conference on Awareness Science and Technology, 2018, pp. 447-451.

[12] Z. Ya-Ling, W. Ya-nan, Y. Lil, "An Improved Sampling K-means Clustering Algorithm Based on MapReduce", IEEE 3rd International Conference on Big Data Analysis,2017.

[13] https://en.wikipedia.org/wiki/Machine_learning , Last Accessed on 05th Sept. 2018.

[14] https://en.wikipedia.org/wiki/Artificial_intelligence , Last Accessed on 05th.Sept. 2018.

[15] https://en.wikipedia.org/wiki/Linear_regression , Last Accessed on 04th Sept. 2018.

[16] https://en.wikipedia.org/wiki/Random_forest , Last Accessed on 05th Sept. 2018.

[17] https://en.wikipedia.org/wiki/K-means_clustering , Last Accessed on 05th Sept. 2018.

[18] B. R. Ranoliya, N. Raghuwanshi, S. Singh, "Chatbot for University Related FAQs", International Conference on Advances in Computing, Communications and Informatics (ICACCI), 2017.

[19] R. Ravi, "Intelligent Chatbot for Easy Web-Analytics Insights", International Conference on Advances in Computing, Communications and Informatics (ICACCI), 2017.