Recommendation System for E-Learning Platforms

G. Senthil Kumar, T.S.Shiny Angel, Anirudh Gangwar, Kaustubh P.M. Saksena

Abstract: ***E- Learning platform provides learners all over the world with information and resources to enhance the education management and delivery. In the current scenario there is a lack of recommendation systems for various online courses on the web. Anyone eager to learn from this huge pool of courses online, they need to get a legitimate course which is best suited for them. The recommendation system which is currently available is not suited for learners. With the rapid increase of E-learning, a new demand of recommendation system has been created which can help a learner in such a way so that they can choose a suitable course for themselves in an easy and accessible manner. In this paper an E-learning recommendation system is constructed to offer better courses to the users. The proposed recommendation system extracts user reviews from various E-learning websites such as Edx, Coursera, Udemy, etc. and suggests the users whether the course is suitable or not. Reviews are collected by using web scraping for an e-learning website. Then sentiment analysis of the collected reviews is performed. This tells the system about the sentiment of a particular review. Afterwards techniques like hybrid SVM and maximum entropy are used to provide a recommendation to the user. Then the users can easily decide which course is better suited for them. So, the tedious task of going through all the courses on the web can be avoided. With the help of this the users can easily map out the best course of their choosing.***

Index Terms: ***Hybrid SVM, Maximum Entropy, Recommendation system, Sentiment Analysis, Web Scraping.***

I. INTRODUCTION

In this modern world of technology students refer to E-Learning platform [1] and Massive open online courses (MOOCS) instead of the conventional learning techniques. Since it has many advantages such as easy access, self-paced learning, cost effective and many more. This has caused an increase in the number of people going online in search of these courses. But the problem arises where the quality of these courses is not up to the standards and it is very difficult for someone new to find out the most suitable course for his/her requirements. Dealing with the enormous amount of data present on the websites, categorizing and analyzing becomes a very tedious task.

Revised Manuscript Received on June 7, 2019

G. Senthil Kumar, Assistant professor, Department of Software Engineering, SRM Institute of Science & Technology, Kattankulathur 603203, Chennai, India

T.S.Shiny Angel, Assistant professor, Department of Software Engineering, SRM Institute of Science & Technology, Kattankulathur 603203, Chennai, India

Anirudh Gangwar, Assistant professor, Department of Software Engineering, SRM Institute of Science & Technology, Kattankulathur 603203, Chennai, India

Kaustubh P.M. Saksena, Assistant professor, Department of Software Engineering, SRM Institute of Science & Technology, Kattankulathur 603203, Chennai, India

This problem can only be solved by a recommendation system [12] which finds a suitable course for learners. Hence a recommendation system is proposed in this paper based on Sentiment analysis [2], [13], [19], [22] in which the sentiment of the reviews is extracted [15] from the user reviews on the E-Learning platforms. Sentiment analysis is basically based on web scraping [3] and classifying of user reviews**.** It helps us to tackle this observed phenomenon. Hence, by creating a list of suitable choice of courses is a thing that is unheard of. Here an analysis is provided based upon the sentiments of the collected reviews. For the purpose of processing the reviews, the immediate step is to change the sentences from the extracted reviews into basic words and then sentence is divided based on the conjunction word into two parts and then check each and every word of both the parts against the established dictionary. After completing this a recommendation [18] is generated. The system uses an incorporated algorithm of Hybrid support Vector Machine [4] to classify reviews in accordance with the distinctive features. In the past technique called the Naïve Bayes [5], [20] classifier is used for calculating the chances of data being negative or positive. Random forest [5] has also been used which is a learning method which targets the storing and enhancing of classification trees. For the purpose of doing sentiment analysis a combination of both Support Vector Machine and Maximum Entropy [6] is used and based on the analysis obtained by the algorithm [21] a recommendation of the courses which suits the needs of the users is generated.

II. LITERATURE SURVEY

Priyadharshini and Lovelin [7] analysed feedbacks using sentiment analysis. They have implemented a product recommendation system for e-commerce websites in which the reviews are classified as positive, negative and neutral. The products which got positive feedbacks earlier were recommended for current users. Analysis takes place based on classification [16]. Alia Karim and Ahmed Bahaaaldeen [8] performed sentiment analysis using NLP Techniques and predicted user satisfaction level on the basis of user reviews. The proposed system works on reviews as a new way to tackle the problem of data insufficiency. The model implements NLP methods in order to foretell user rating from the extracted reviews using various datasets of yelp, IMDB, etc. Eugene Ch'ng, Ziyang Chen 1 and Simon See [9] proposed a dynamic answer to overcome the dialect problem in twitter. Their proposed system processed the data which is streaming dynamically.

The authors have presented a graphics processing unit-based project which implements compute unified device architecture. In order to disperse the analysis and processing of text data to imagine the data dynamically. Prema Monish, Santoshi Kumari and Narendra Babu C [10] developed a technique to find out the topics which are interesting in nature from the discussion on twitter. They found the topics of unknown nature, expectations of people based upon empirical results and some of their concerns related to unknown topics. The proposed system based upon linear discriminant analysis to find out the topics from the enormous number of tweets that are present in relation to two major political supremoes of India. The system was based on SparkR to provide better and efficient results. Rita Guimaraes, Demóstenes Z. Rodríguez, Renata L. Rosa and Graça Bressan [11] analysed social networks by finding adverbs in a sentence. A solution was developed to make a low complexity feedback system. They presented a model that analyses the sentiment of the comments extracted from social media. They have proposed an algorithm which finds out the polarity of the adverb in the comment. The model has low amount of complexity and hence is quick and reliable.

III. RECOMMENDATION SYSTEM FOR E-LEARNING PLATFORM

The proposed system is for learners who want to know which course is best suited for them. It extracts reviews dynamically in order automate the data extracting process. It works on reviews which are extracted from users as a new way to tackle the problem of data insufficiency. This new way is getting reviews from E-Learning websites using web scraping technologies using python-based tools so that datasets obtained are genuine in nature and are not manually fed by users to the algorithm for conducting sentiment analysis.

For the purpose of doing the sentiment analysis a combined version of both support vector machine and maximum entropy is used and based on the analysis a recommendation is produced. For performing analysis of sentiments, manipulating the data is needed in a processed manner. For manipulating the dataset, a support module was developed in the early stage of the project. The support module was integrated and tested along with the following parts:

A. Data Extracting Modules

The data extracting module collects the data for which the sentiment analysis will be carried out. It basically has two sub modules one for live data gathering and another existing data that has already been gathered.

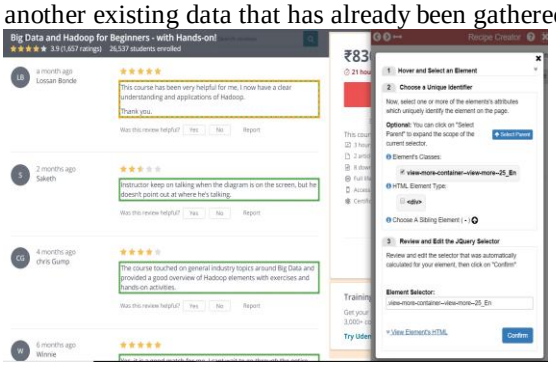


Fig. 1. Review extraction process.

Fig. 1 shows the user reviews from a course of the website such as Udemy is extracted by using the Data Miner tool.

B. Data Filtering Modules

In this module the process of tokenization takes place in which the sentences are broken into words, phrases, symbols and other elements. The proposed system discards punctuation marks in this process (stop words). Also feature selection is carried out in which a feature is selected which contribute most to the prediction variable.

C. Association Rules Modules

Here classification of reviews takes place. The reviews get classified as negative, positive, neutral and unpredicted based upon the sentiment observed from the extracted reviews. The algorithm is based on the concepts of hybrid SVM which is integrated with maximum entropy.

a. Maximum Entropy Technique

Maximum entropy technique [6] also known as MaxEnt focuses upon the features of dataset which is being considered without making any independent assumptions. In this technique there is no restriction upon the number of features and hence one can add as much features as one wants without overlapping the features. The most unique thing about the maximum entropy technique is that it favors the model which is most uniform in nature i.e. it fulfills a given limitation. These models which are feature based are useful to estimate any probability [17] distribution. The below equation depicts the technique:

$$p^{*}(y|x)=\frac{\exp(\sum_i \lambda_i f_i(x,y))}{\sum_y \exp(\sum_i \lambda_i f_i(x,y))}$$

In the above given equation x represents class/category, y represents review, λ represents weight vector and each fi(x, y) represents feature. The more the amount of wait, the stronger the measure for mentioned classes.

This technique is solely based upon feature which is determined with the help of weight vectors. It is considered that features are strong if the weight vector is high. For the purpose of classification, a standard classifier of maximum entropy is used. Maximum entropy is way better that Naïve Bayes technique as it handles the overlapping of features in a better and efficient way.

b. Support Vector Machine (SVM)

Support Vector Machine (SVM) [4] is another classification technique of supervised nature which provides with better results in consideration with classification of data which is of textual manner. It is even better than Naïve Bayes and maximum entropy. What it does is it founds a plane in order to divide the training sample into two halves that is negative and positive. This technique has exceptional performance over the other two methods. Many types of SVMs are available that makes it more adaptable to the real-world scenarios. In the proposed system a Hybrid SVM technique is used that successfully distinguishes between the reviews and classifies them as categories that are positive, negative, neutral and unpredicted. Since there are four classes it requires a multi- dimensional approach which is only possible with the case of hybrid support vector machine.

To increase the efficiency of the algorithm the combined version of support vector machine with maximum entropy is implemented.

D. Sentiment Classification Modules

Now the recommendation of course takes place based upon the positive or negative polarity of the reviews that has been obtained in the previous step using the classification algorithm SVM and maximum entropy.

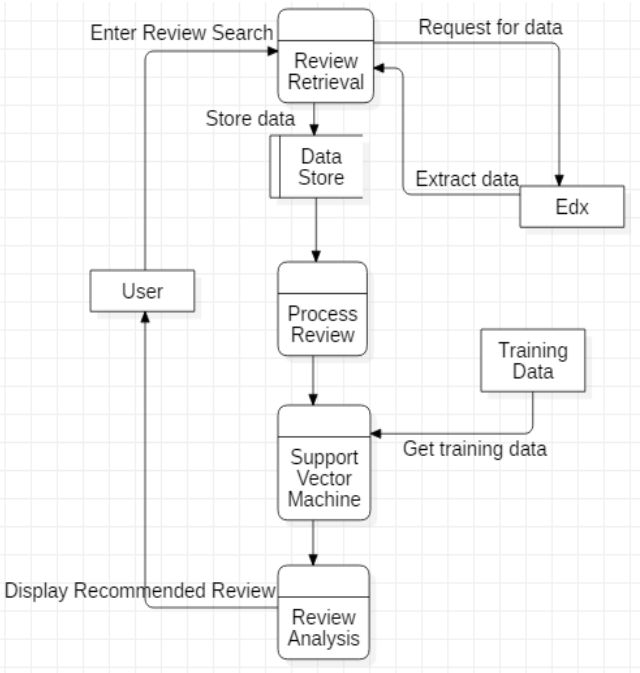


Fig. 2. Framework of the designed system

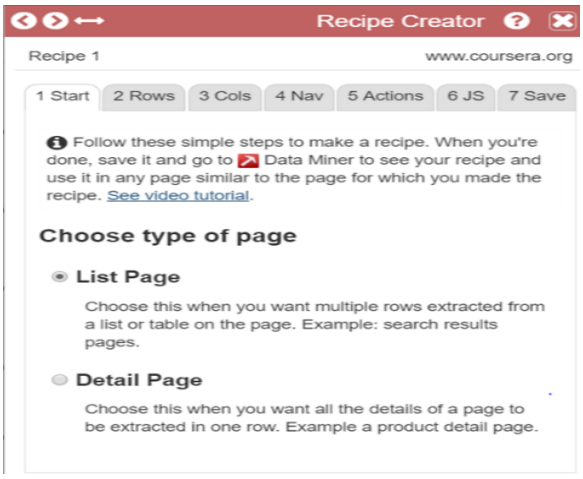
In the above Fig. 2 the process of our recommendation system has been explained diagrammatically including all the four modules. Hence finally generating a recommendation based upon the reviews extracted.

E. ALGORITHM

1. Start
2. Read a complete review.
3. Check for conjunction.
4. if Conjunction found
    a. split into before and after based on conjunction
    b. check whether words before and after is positive/negative
        i. Check for negation words
        ii. if Negation found
        iii. check whether next word is positive/negative
        iv. if (negation and positive) then set as negative
        v. if (negation and negative) then set as positive
        vi. ELSE End
    c. End if
5. if both sides have different polarity set as neutral.
6. if both sides are positive set as positive
7. if both sides are negative set as negative
8. else if Conjunction not found
    a. Check for negation words
    b. if Negation found
    c. check whether next word is positive/negative
        i. if negation + positive set as negative
        ii. if negation + negative set as positive
    d. if Negation found End
    e. if No Negation found
        i. check whether word is positive/negative
    f. if No Negation found End
    g. Check for negation words End
9. if No Conjunction Found End
10. END else.
11. Stop.

IV. EXPERIMENTAL RESULTS AND RESULTS

A website with a user interface is created. A dynamic dataset which consists of the reviews extracted from e learning platforms is obtained using a web scraping technique. It consists of 1000 reviews from multiple users across 4 courses. As many kinds of datasets are available in the selected domain, one has to find out the dataset that has not been used anywhere and confines to the various boundaries which are considered here. Hence, a dataset is created by web scraping the reviews from the website itself with help of a web scrapping tools which are based on python in order to maintain the authenticity of the dataset. The obtained dataset is dynamic in nature and any change that occurs in the review pool of the course will be reflected in the dataset also. The user extracted reviews have been obtained through web scraping the websites such as Udemy, Coursera, etc. as shown in Fig 3. Then sentiment analysis of these reviews is carried out which classifies them into four categories namely, positive, negative, neutral and unpredictable as shown in Table 1.



Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

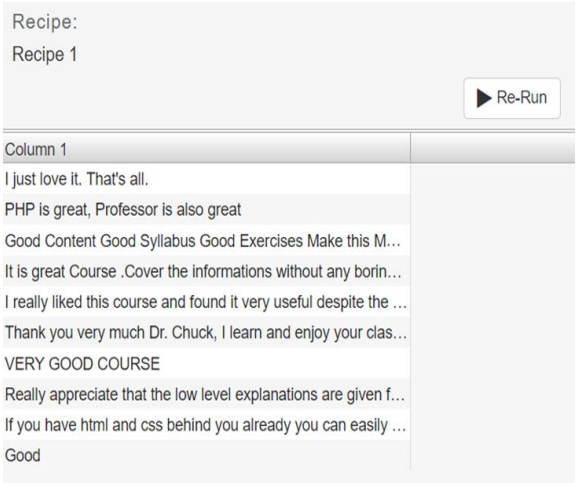


Fig. 3. Data Miner Tool

In the above Fig. 3 the reviews collected by the data miner tool are collected and stored in an excel format for addition to the database. This database acts as a base of the proposed recommendation system.

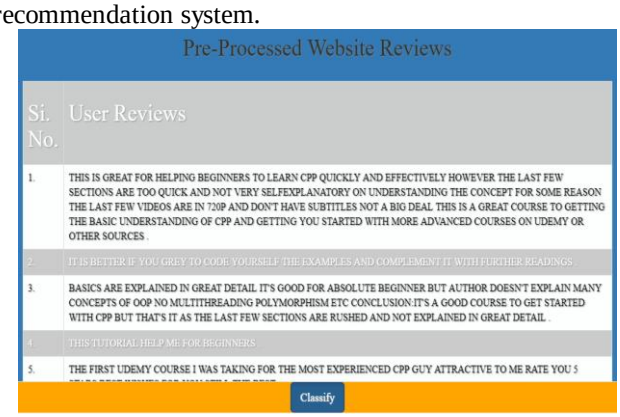


Fig. 4. Review classification UI

Fig. 4 shows the extracted user reviews from the E-learning platforms have been displayed which are classified into four categories based upon sentiment of the reviews.

Fig. 5. Recommendation obtained from the carried-out classification

Table I shows recommendations generated by analyzing the sentiments of the user reviews obtained from the user.

Table I. Comments and sentiment comparison from the extracted reviews

Review	Sentiment analysis
IT WAS A NICE COURSE AND LEARNED SOME NEW THINGS EVERYTHING FROM THE BASICS WAS COVERED	Positive
GOOD EFFORT BUT HE SPEAKS MEANINGLESS SOMETIMES AND IRRITATING JOKES THAT HE MAKES	Negative
THE INTRO AND THE PROCEDURES ARE PRETTY EASY TO FOLLO HOWEVER CONFUSION IS PRESENT	Neutral
WROTE MY FIRST CPP PROGRAM	Unpredictable

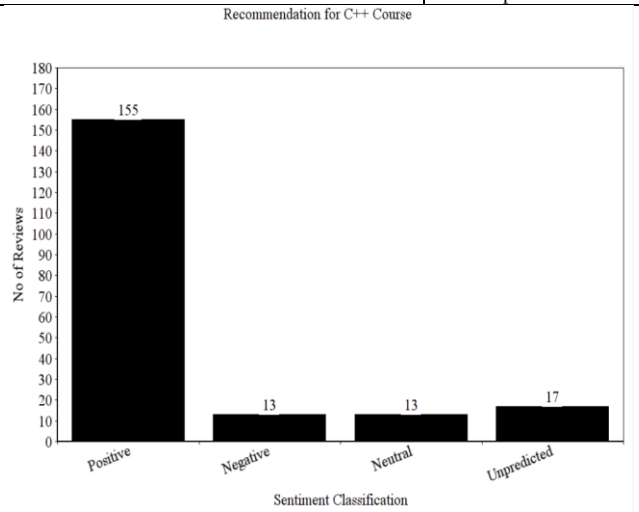


Fig. 6. Graphical representation of the sentiment classification for C++ course

Fig. 6 shows the analysis of the proposed recommendation system of the number of positive, negative, neutral and unpredicted reviews is shown for a course of C++ on Udemy. Using the sentiment classification of the reviews the user gets a general idea of the course based upon the number of reviews classified as positive, negative, neutral or unpredictable classes and recommends the user whether to approach the course or not. More positive reviews or neutral review are present than the course is recommended and if more negative reviews or unpredictable reviews are present than the course is not recommended.

V. CONCLUSION AND FUTURE WORK

A Recommendation system is proposed in this paper which analyses the sentiment of user reviews from E-learning websites such as Edx and tell the users which of the following E-learning courses from the websites are good or bad. The application was developed with web server platform. PHP programming language was used for development. SQL Server was used for database. The proposed application is useful for people who want to know which is the best course for a given topic. This application performs sentiment analysis of the extracted reviews with the help of Hybrid SVM and Maximum Entropy. A recommendation is provided to the users which tells them whether the course is good or bad. For extracting reviews, a web scraping tool called Data Miner is used. Overall this system was found very useful and efficient by the users. The more the number of reviews present, better are the chances for providing an accurate recommendation. In other words, a greater number of reviews will increase the accuracy of the system. For a limited number of reviews analysis can be done but sureness is not guaranteed and it will also impact the accuracy of the system.

With the use of REST APIs, the reviews extraction process can be carried out automatically and any change in the number of reviews will hence be dynamic which further can be analysed in real time [14] and hence provide the user with a recommendation which is based on absolute live data.

REFERENCES

1. D. Bentaa, G. Bologaa , I. Dzitaca,b, “E-learning Platforms in Higher Education. Case Study”, 2nd International Conference on Information Technology and Quantitative Management, ITQM 2014.
2. Mohey El-Din, Doaa. (2016). “Analyzing Scientific Papers Based on Sentiment Analysis (First Draft)”. 10.13140/RG.2.1.2222.6328.
3. Lalita Junjoewong, SupatsaraSangnapachai, ThanwadeeSunetnanta, “ProCircle: A promotion platform using crowd sourcing and web data scraping technique” , 2018 Seventh ICT International Student Project Conference (ICT-ISPC).
4. Durgesh K. Srivastava, LekhaBhambhu, “Data Classification Using Support Vector Machine”, Journal of Theoretical and Applied Information Technology © 2005 - 2009 JATIT.
5. Wint Nyein Chan 1, a, Thandar Thein 2,b, “A Comparative Study of Machine Learning Techniques for Real-time Multi-tier Sentiment Analysis”, 1st IEEE International Conference on Knowledge Innovation and Invention 2018, ISBN: 978-1-5386-5267.
6. Wang Yaqi1, Xu Yonghai1, Wang Jinhao2, “The Maximum Entropy Probability Distribution and the Improvement of Its Solving Method for Negative Sequence Current of a Traction Substation”, 2016 China International Conference on Electricity Distribution (CICED 2016) Xi’an, 10-13 Aug, 2016.
7. R. Lydia Priyadharsini, M. LovelinPonnFelciah,“Recommendation System in E-Commerce using Sentiment Analysis”, International Journal of Engineering Trends and Technology (IJETT), Volume 49 Number 7 July 2017.
8. Alia Karim Abdul Hassan, Ahmed BahaaaldeenAbdulwahhab, “Reviews Sentiment analysis for collaborative recommender system”, Kurdistan Journal of Applied Research (KJAR), Volume 2 | Issue 3 | August 2017.
9. Eugene Ch'ng 1, Ziyang Chen 1 and Simon See, “Real-Time Sentiment Analysis of Saudi Dialect Tweets Using SPARK”, 2016 IEEE International Conference on Big Data (Big Data), 978-1-4673-9005-7/16/2016 IEEE.
10. PremaMonish, Santoshi Kumari, Narendra Babu C, “Automated Topic Modeling and Sentiment Analysis of Tweets on SparkR”, 9th ICCCNT 2018 July 10-12, 2018, IISC, Bengaluru, India, IEEE – 43488.
11. Rita Guimarães, Demóstenes Z. Rodríguez, Renata L. Rosa, GraçaBressan, “Recommendation System using Sentiment Analysis considering the Polarity of the Adverb”, 2016 IEEE International Symposium on Consumer Electronics, 978-1-5090-1549-8/16
12. Kavinkumar.V, Rachamalla Rahul Reddy, Rohit Balasubramanian, Sridhar.M, Sridharan.K, Dr. D.Venkataraman, “A Hybrid Approach for Recommendation.
13. System with Added Feedback Component”, 2015 International Conference on Advances in Computing, Communications and Informatics (ICACCI), 978-1-4799-8792-4/15/2015 IEEE.
14. Xian Fan, Xiaoge Li, Feihong Du, Xin Li, Mian Wei, “Apply Word Vectors for Sentiment Analysis of APP Reviews” , The 2016 3rd International Conference on Systems and Informatics (ICSAI 2016), 978-1-5090-5521-0/16/2016 IEEE.
15. Eugene Ch'ng, Ziyang Chen and Simon See, “Real-Time GPU-Accelerated Social Media Sentiment Processing and Visualization”, 2017 IEEE, 978-1-5386-4028-9/17/2017 IEEE.
16. R. L. Rosa, D. Z. Rodriguez and G. Bressan, "Music recommendation system based on user's sentiments extracted from social networks”, in IEEE Transactions on Consumer Electronics, vol. 61, no. 3, pp. 359-367, Aug. 2015.
17. DeukHee Park, HyeaKyeong Kim, ll Young Choi, Jae Kyeong Kim, “A literature review and classification of recommender systems research”, Expert Systems with Applications, Vol. 39, No. 11, pp 1005910072,2012.
18. Zhuo Chen, Yong Feng, Heng Li, "A Novel Top-K Automobiles Probabilistic Recommendation Model Using User Preference and User Community", Proc. IEEE 11th International Conference on eBusiness Engineering (ICEBE), pp. 105111, 2014.
19. Z. Wang, L. Sun, W. Zhu, S. Yang, H. Li, and D. Wu, “Joint Social and Content Recommendation for User-Generated Videos in Online Social Network”, IEEE Trans. Multimedia, vol. 15, no. 3, pp. 698-709,2013.
20. Payal S Gadhave,Mahesh D Nirmal,“Twitter Based Sentiment Analysis using Hadoop”, Sixth Post Graduate Conference for Computer Engineering (cPGCON 2017) Procedia International Journal on Emerging Trends in Technology (IJETT).
21. Vaibhavi.N ,Patodkar,Shaikh I.R. “ Sentimental Analysis on Twitter Data using Naïve Bayes”, Sixth Post Graduate Conference for Computer Engineering(cPGCON 2017) Procedia International Journal on Emerging Trends in Technology(IJETT).
22. M. LovelinPonnFelciah and R. Anbuselvi, “A comparative study of machine learning algorithms on opinion classification” International Journal of Applied Engineering Research, ISSN 0973-4562 Vol. 10 No.85 ,2015.
23. A. Bhardwaj, Y. Narayan, V. Anarj, Pawnan, M.Dutta, " Sentiment analysis for Indian stock market predictor using Sensex and nifty", in proceeding 4th international conference on eco-friendly computing and communication systems, pp: 85-91, 2015.

AUTHORS PROFILE

First Author :G.Senthil Kumar holds M.Tech. Degree in Information Technology from SRM Institute of Science and Technology and is an Assistant professor in the Department of Software Engineering, SRM Institute of Science and Technology. His main area of interest includes Web services, software Engineering, machine learning, web Intelligence and Agile Software process. He has published several papers in well-known peer-reviewed journals**..**

Second Author Dr.T.S.Shiny angel holds PhD. Degree in computer science and engineering from SRM Institute of Science and Technology and is an Assistant professor in the Department of Software Engineering, SRM Institute of Science and Technology. His main area of interest includes software Engineering, machine learning, and Agile Software process. She has published several papers in well-known peer-reviewed journals..

Anirudh Gangwar

B.Tech (4th Year), Software Engineering, Student at SRM Institute of Science and Technology, Kattankulathur, Chennai

Kaustubh P.M. Saksena

B.Tech (4th Year), Software Engineering, Student at SRM Institute of Science and Technology, Kattankulathur, Chennai

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication