

Scope Prediction Utilizing Support Vector Machine for Career Opportunities

T. M. Nithya, J. Ramya, L. Amudha

Abstract: Education courses were offered at tutorial establishments round the world. In the present educational society, choosing a career path can be stressful and overwhelming. Selecting the right course for the career is the most crucial decision to make and can be troublesome if not guided properly. The purpose of choosing the appropriate course in any college with employment facilities is the major drawback in the existing system. The conventional method will only give the information about the course. The data is predicted by analyzing the frequent searching pattern of the user by matrix factorization method. To overcome those anomalies, this paper proposes the Course prediction techniques for the student which uses the Base Linear Regression (BLR) Algorithm to predict the course. This algorithm comes under the Support Vector Machine (SVM) Algorithm that will process the data set about the various courses. The student will initially register their details before logging in to the website. This application will filter the data based on their educational qualification when the student login to the homepage. Content Based Filtering (CBF) technique is applied further to filter the data in dataset. This will sort out the information based on the user requirements so that the student can choose the course that they need to study with the accurate predicted data with their future growth in the graphical form.

Keywords: Base Linear Regression Algorithm (BLR), Content Based Filtering (CBF) technique, Support Vector Machine (SVM).

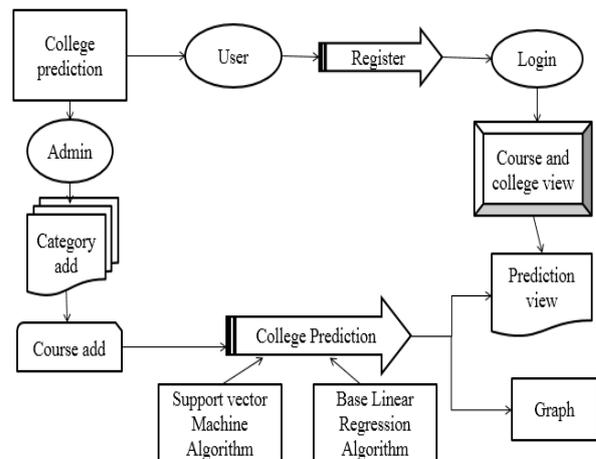
I. INTRODUCTION

Big data analytics is that the typically advanced method of examining massive and varied information sets to uncover info as well as hidden patterns, unknown correlations, market trends and client preferences that may help organizations make informed business decisions. Huge information investigation is that the commonly propelled technique for looking at huge and changed data sets to reveal information just as concealed examples, obscure connections, showcase patterns and customer inclinations that may enable associations to settle on educated business choices. Prescient examination envelops a spread of connected arithmetic systems from information handling, prophetic displaying, and AI that dissect present and chronicled certainties to make forecasts about future or generally obscure occasions. In the present instructive society choosing the course for better vocation is troublesome. Picking the fitting course in any school with business offices is the significant errand for the understudies. This framework predicts the information by

breaking down the incessant looking through example of the client by network factorization technique.

To defeat the inconsistencies of the current framework this paper proposes the Course forecast strategies for the understudy who need to do their higher investigations. The significant strategy used to anticipate the course is the Base Linear Regression (BLR) Algorithm. This calculation goes under the Support Vector Machine (SVM) Algorithm that will procedure the informational collection about the different courses. The help vector machine is a managed learning model to group the preparation information. The Artificial Neural Network is a profound learning approach in concealed information neurons, to conjecture the cost of a stock. A Neural Network comprises of an info layer, concealed layer and yield layer. The loads are relegated for each information layer. The summation capacity is determined for the weighted yield. The yield layer ascertains the sigmoid capacity for the summation esteem. The last yield is the anticipated course.

II. SYSTEM ARCHITECTURE



III. LITERATURE SURVEY

Dimitrios Kotzias, Moshe Lichman and Padhraic Smyth [1], members of IEEE have proposed a paper, "Predicting Consumption Patterns with Repeated and Novel Event". This paper provides the analysis by understanding the user's choice by analyzing the frequent search of the user and then the few items of the user's choice is being suggested to the user. In this paper they have used the technique called Matrix Factorization. This paper had its merits as by providing the accuracy in prediction and the limitations in this paper was that the effect was been mitigated by Empirical Bayes Priority. Qing Li, Yan Chen, Jun Wang, Yuanzhu Chen, Hsinchun Chen [2],

Manuscript published on 30 June 2019.

* Correspondence Author (s)

Mrs. T. M. Nithya AP/CSE K. Ramakrishnan College of Engineering Trichy

Mrs. J. Ramya AP/CSE K. Ramakrishnan College of Engineering Trichy

Mrs. L. Amudha AP/CSE K. Ramakrishnan College of Engineering Trichy

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an open access article under the CC-BY-NC-ND license <http://creativecommons.org/licenses/by-nc-nd/4.0/>

Scope Prediction Utilizing Support Vector Machine for Career Opportunities

members of IEEE have proposed a paper, "Web Media and Stock Markets". This provided the survey and future directions from a Big Data Perspective. The technique used in this paper was Convolution Neural Network (CNN). It succeeded in the advancements in quantifying the influence of information on financial markets. It had some limitations such it was critical to maintain a financial sentiment dictionary to assist in accurate sentiment analysis for financial documents. Talha Siddique, DiproBarua, ZannatulFerdous, AmitabhaChakrabarthy [3], have proposed a paper, "Automated Farming Prediction". They used a technique as Multiple Linear Regression (MLR) and K-Nearest Neighbor Regression (KNNR). It had merits such as to intend digital cropping prediction and the crop cultivation procedures. It failed to support the farmers who does not have smart phones, it was not applicable to farmers who did not have smart mobile phones. AhelamTikotikar, MallikarjunKodabagi [4], They have proposed a paper, "A Survey on Technique for Prediction of Diseases in Medical Data", In this paper they have provided the analysis and the prediction of diseases for the users to get alert of the disease stages. The technique used in this paper was Data Mining. It helps medical expertise to decide about the disease using computer aided decision support system were its merits but its implementation cost was high. Zhaocong Sun, Jinyi Xia, Chi Zhang [5], have proposed a paper, "Energy Evaluation and Prediction System based on Data Mining". It used the Least Square Support Vector Machine (LSSVM) and Back Propagation (BP) technique. This paper succeeded in prediction of the clean and renewable energy but the machine learning frequency results in various errors of each prediction had failed.

IV. EXISTING SYSTEM

Matrix factorization method is used for finding the accuracy. Matrix factorization item consumption approach is widely used over the past decade for modelling of sparse user dataset. This model over estimates the probabilities of new items for many users for significantly over smoothing on all datasets. The prediction accuracy contains the result of all methods and all datasets corresponding to the optimal probability distribution for each dataset. It gives the idea about implicit user exploration through validation and allowing the learning of exploration preferences directly from the consumption data without external item information.

Disadvantages:

Lack of predictive accuracy at the individual user level. Used only for static distribution dataset. Large amount of data cannot be processed through matrix factorization.

V. PROPOSED SYSTEM

Machine learning approach is used for predicting the dataset. A support vector machine (SVM) is a supervised learning model that divides the data into region separated by linear boundary. Her SVM technique is used for classify the categories and courses. Content based filtering (CBF) is used to filter the range of dataset from undergraduate to postgraduate courses with its duration. Decision tree algorithm is used for analyzing the dataset and predict the scope of the course. Decision tree splits down a dataset into smaller and smaller subsets. It is the combination of mathematical and computational technique. Data comes in

record of $(x, y) = (x_1, x_2, \dots, x_k, y)$ y denotes the growth of each courses whereas x denotes the year. Data can be predicted up to 80% of accuracy.

I. Advantages:

Identification of courses and growth of course. Representation of the scope of the course in bar chart. 80% of accuracy in prediction. It supports huge amount of dataset.

A. Algorithms Used:

Linear Kernel Algorithm:

Input: D {course data};

Output: Label {course Label};

Category's Label {category, Course}

Step1: Pre-processing and data cleansing

Step2: For each instance in D, do

Find feature vector (V)

Step 3: For each V do

Data clustering using Kernal and split data in two halves and classify data using SVM and linear kernel algorithm

Step 4: Determine the total class label

Find

True_positive (TP)

True_negative (TN)

False_positive (FP)

False_negative (FN)

Step 5: Find Performance Parameters

Step 6: Predict Course Class as

if (class=1) Category=Normal State

else_if(class =0) Category=Course State

end for

B. Content Based Filtering

In a content based recommender framework, watchwords are utilized to depict the things and a client profile is worked to show the sort of thing this client likes. As such, these calculations attempt to prescribe things that are like those that a client preferred before (or is looking at in the present). Specifically, different hopeful things are contrasted and things recently evaluated by the client and the best-coordinating things are prescribed. This methodology has its underlying foundations in data recovery and data sifting research. To digest the highlights of the things in the framework, a thing introduction calculation is connected. A broadly utilized calculation is the tf-idf portrayal (additionally called vector space portrayal).

To make a client profile, the framework generally centers around two kinds of data:

1. A model of the client's inclination.

2. A background marked by the client's cooperation with the recommender framework.

Fundamentally, these techniques utilize a thing profile (i.e., a lot of discrete characteristics and highlights) describing the thing inside the framework. The framework makes a substance put together profile of clients based with respect to a weighted vector of thing highlights. The loads signify the significance of each element to the client and can be processed from separately evaluated substance vectors utilizing an assortment of strategies. Straightforward methodologies utilize the normal estimations of the appraised thing vector while other refined strategies use AI systems, for example,

Bayesian Classifiers, group examination, choice trees, and counterfeit neural systems so as to gauge the likelihood that the client is going to like the thing.

Direct criticism from a client, more often than not as a like or abhorrence catch, can be utilized to allocate higher or lower loads on the significance of specific properties. A key issue with substance-based sifting is whether the framework can take in client inclinations from clients' activities in regards to one substance source and use them crosswise over other substance types. At the point when the framework is constrained to prescribing substance of a similar kind as the client is now utilizing, the incentive from the suggestion framework is altogether not as much as when other substance types from different administrations can be suggested. For instance, prescribing news stories dependent on perusing of news is valuable, yet would be substantially more helpful when music, recordings, items, talks and so forth from various administrations can be suggested. The term recurrence lattice (weight) $TF(d,t)$ measures the related of a term t as for the given archive d . It is characterized as 0 if the archive does not contain the term and non zero generally. The relative term recurrence is estimated utilizing the term recurrence versus the absolute number of event of the considerable number of terms in the archive. The term recurrence is registered

$$TF(d,t) = 0 \text{ if } \text{freq}(d,t) = 0 \quad (1)$$

$$\text{else } 1 + \log(1 + \log(\text{freq}(d,t)))$$

There is other significant measure, called reverse archive recurrence (IDF) in Eq.1 that speaks to the scaling factor, or the significance of a term t and it will be decreased if a term t happens in numerous reports. For instance the term data might be less significant in many research papers. The Formula for IDF(t) is given There is other important measure, called inverse document frequency (IDF) in Eq.1 that represents the scaling factor, or the importance of a term t and it will be reduced if a term t occurs in many documents. For example the term information may be less important in many research papers. The Formula for IDF(t) is given in Eq.2.

$$IDF(t) = \frac{\log 1 + |d|}{|dt|} \quad (2)$$

Where, d is the document collection, and dt is the set of documents containing term t . In complete vector space model, TF and IDF are combined together, which forms the TF-IDF measure is given in Eq.3:

$$TF-IDF(d,t) = TF(d,t) * IDF(t) \quad (3)$$

VI. EXPERIMENTAL RESULTS

A. Admin Module

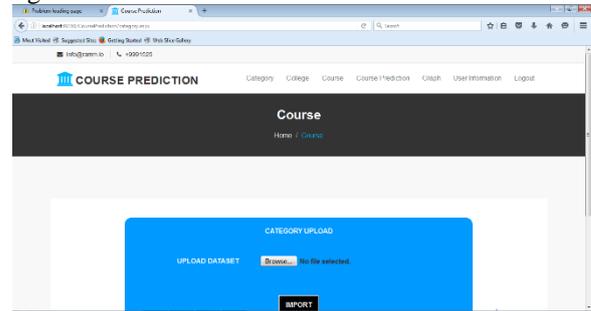
1. Authentication

The module is designed to enter the authorized person where the valid username is to be entered and the correct password should be entered. This will prevent the frequent modification of data by others. This also helps the admin in updating and modifying the data in the database. This will act as the gateway of this application.

2. Category Details

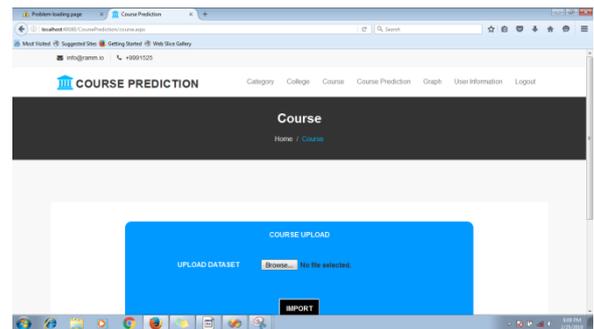
In the category details module, the admin will enter the category of the college like arts, medicine, engineering or

science. This module will have all the needed information about that particular college like their address, contact number, star rating, approval details, courses that offered and the principal details. The admin can also modify the added in information if any changes occurred in any particular colleges or an institution.



3. Course Details

The courses add module is designed to enter the courses that are provided by any institution. The information of the course will have the duration of the course, colleges that are providing that course, predicting the scale to pay for that course and other needed information. This will bring the convenience to sort out the information while the user enters the course that they need to study



4. Analysis Graph

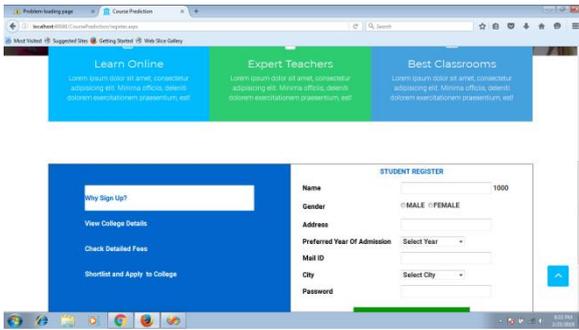
In the graph module the overall results are viewed on the basis of the particular features. This module will bring out the significant feature of the ups and downs of the engineering college and rise of the arts and Science College by the graphical representation.

B. User Module

1. Registration Details

In the register module the end user will register basic needed information about them like name course like to study, address, phone number and the date of birth. This will ensure the admin for the authentication and the future clarification. After registering the user will have valid username, qualification and the password for the login purpose. This will help to filter the course for the particular user.

Scope Prediction Utilizing Support Vector Machine for Career Opportunities



2. Prefiltering

The login module will be the gateway to the end user where the username, qualification and password are entered. While entering the qualification of the user filtering process is executed for displaying the needed information of the particular student. Content Based Filtering technique is used to filter the needed information based on their qualification of the course. By this the accurate information is displayed to the end user.

3. Course and College View

The course and college view module is designed to display the needed course that the user needed to study and the colleges that are offer that course. This information is extracted by the filtering technique that will avoid the course conflict that are related to that particular course. This module displays the college name, address, contact number and the other description. It also views the course details like duration for the course and the other details.

4. Prediction View

In this module the predicted college that the user needed to study is viewed. This will ensure the data about the college details. This will give the accurate information about the course information and the platform of the course.

VII. CONCLUSION

Our paper will allow the students to take a correct decision after their schooling, it provides a great vision and the predicted data will make the users to choose their career for them to enrich in their life. The future works is done by developing it to a prediction base to a large country by including all the college and course database. And it can also allow the parents to know about the courses that their children choose.

REFERENCES

1. DimitriosKotzias, Moshe Lichman and Padhraic Smyth. "Predicting Consumption Patterns with Repeated and Novel Event".
2. Qing Li, Yan Chen, Jun Wang, Yuanzhu Chen, Hsinchun Chen, members of IEEE, "Web Media and Stock Markets".
3. Talha Siddique, DiproBarua, ZannatulFerdous, AmitabhaChakrabarthy, "Automated Farming Prediction".
4. AhelamTikotikar, MallikarjunKodabagi, "A Survey on Technique for Prediction of Diseases in Medical Data".
5. Zhaocong Sun, Jinyi Xia, Chi Zhang, "Energy Evaluation and Prediction System based on Data Mining".
6. Tong Wang, HuijunGao, JianbinQiu, "A Combined Adaptive Neural Network and Nonlinear Model Predictive Control for Multirate Networked Industrial Process Control"16 April 2015.
7. Min Chen, YixueHao, KaiHwang, LuWang, Lin Wang, "Disease Prediction by Machine Learning Over Big Data from Healthcare Communities",26 April 2017.
8. Xinyi Xu, XingHe, QianAi, RobertCaimingQiu, "A Correlation Analysis Method for Power Systems Based on Random Matrix Theory", 29 December 2015.

9. Kai Lin,Jiming Luo, Long Hu, Ahmed Ghoneim,M. Shamim Hossain, "Localization Based on Social Big Data Analysis in the Vehicular Networks", 19 December 2016.
10. Sicheng Zhao,Hongxun Yao,Yue Gao,Rongrong Ji,Guiguang Ding, "Continuous Probability Distribution Prediction of Image Emotions via Multitask Shared Sparse Regression", 13 October 2016.
11. Huisu Jang,Jaewook Lee, "An Empirical Study on Modeling and Prediction of Bitcoin Prices with Bayesian Neural Networks Based on Block Chain Information", 04 December 2017.
12. Xiaojun Chang,Zhigang Ma,Yi Yang,Zhiqiang Zeng,Alexander G. Hauptmann, "Bi-Level Semantic Representation Analysis for Multimedia Event Detection", 28 March 2016.

Authors Profile



Mrs. T. M. Nithya M.E., Currently pursuing PhD (Part time) in Anna university chennai,working as an Assistant Professor in the Department of CSE,K.Ramakrishnan College of Engineering. Area of interests includes Software Testing and Artificial intelligence.



Mrs. J. Ramya M.E.M.B.A., Currently working as an Assistant Professor in the Department of CSE,K.Ramakrishnan College of Engineering. Area of interests includes Neural networks and data mining.



Mrs. L. Amudha M.E., Currently pursuing PhD (Part time) in Anna university chennai,working as an Assistant Professor in the Department of CSE,K.Ramakrishnan College of Engineering. Area of interests includes Machine learning and Internet of Things.