Subject Categorization by E-Counselling using Machine Learning Algorithm

Sasi Rekha Sankar[a] Jharna Gupta[b] Riddhi Datta[b] and V.S. Kishore Kumar[b]

[a]*Asst. Professor, Department of, Software Engineering, SRM University, Chennai, India*
E-mail: sasirekha.s@ktr.srmuniv.ac.in
[b]*Student, Department of Software Engineering, SRM University, Chennai, India*
E-mail: jharna71096@gmail.com, toriddhi.datta@gmail.com, kishoresankar100@gmail.com

Abstract: The E-Counsellor Analytical System is designed to act as virtual counsellor who can recommend and suggest students to choose the department and open electives using this system. The main objective of this system is to counsel and recommend them electives based on their ambition, field of interest, hobbies and other parameters so that they can choose the required course offered by the institution hence achieve it. The student have to enter the course completed by them till the previous semester and the field of interest or subject of interest or target they want to achieve by specifying the role they want to play in an organization. The software analyse the input provided by the student user with the syllabus offered by the university and the mapping of a subject to respective careers and propose the appropriate courses they can purse with self-explanatory suggestion as a human counsellor[1] who generally provides subject counselling session.

1. INTRODUCTION

E-Counsellor Analytical System is a web based application that is a virtual counsellor which acts a bridge between a student and the university. Traditionally the university offers department electives and few allied department open electives from which the students can choose their preferred subjects from the given list. The students usually get their suggestion from the class in charge or counsellor; they also go through the subject information from the syllabus provided by the university. In the current scenario the university is adapting to the choice based system in which the students can choose the subjects offered by their core department, allied department and other departments in the university offered by them in that semester. The faculty counsellor generally recommends the students the subjects they can opt from the given department list as they will have profound knowledge in their core areas and superficial knowledge in their allied department subjects, but their familiarity with other department subjects will be very limited. This will restrict the student to explore in other specialized domains which might be very useful in future for his career.

The E-Counsellor Analytical System takes various parameters from the student, university, feedback, marks and individual subject feedback from the previous batch provided by the students who opted it and the faculty who handled the subject; based on which predictive analytics is applied to the factors to generate the tailored made elective subjects options for every individual student.

2.1. Students Analysis Module

The system performs the predictive analytics from various angles and considers the students input and request factor more in making the prediction. The student provides information[5] like ambition, hobbies, Area of Interest, Future Goals, position he like to work in future and so on to suggest him the choices of subjects he can choose from the current semester from his department or other departments.

2.2. University Analysis Module

The main motive is to perform sentiment analysis[5]. The analysis is also performed in another aspect from the information derived from the university. The university provides the subjects already studied by the student and the marks obtained in every individual subject. This factor is used to judge the students learning capacity, their interest on the type of subjects and the scoring capacity of the individual student. Another important aspect this factor can be considered is for checking the subjects already studied by the student doesn't clash with new set of subjects, pre-requisite to be used for selecting advanced subjects.

2.3. Feedback Analysis Module

Another important factor considered is the feedback about the subject received from the previous batch of students and the faculties who handled the subject. This tells the scope of the subject, the mode of teaching, outcomes of the subject and the future implication of the subjects. Also the feedback from the student who studied the subjects, the outcomes achieved by the student, difficulties faced and the subject application in real time. Even the alumni suggestions can be loaded to get the impact the received when applying in the industry at work.

2.4. Visualization

We visualised a data for various factors like CGPA, hobbies, favourite subject, least preferred subject, aim in life, passion, expectations etc. to predict which subject can be suitable for students to take and on basis of that what can be the best career opportunities for a student. (See fig.1)

2.5. Architecture Diagram

So, the architecture is divided into two phases. So, first of all a user feedback will be taken and it will be processed via NLP. After that a faculty survey will be taken and it will also be processed through

NLP. The personal details of every student will be taken and altogether the entire details will be clustered and stored in the database. This is the first phase of the counselling system. This phase is basically the training phase. The second phase is basically the phase in which the training data will be accumulated and it will be used for analyzing the accumulated data and giving the recommendations for the required courses.

Career planning includes the concept of machine learning algorithm and artificial intelligence which helps students in determining most suitable subject and even career option for them. It is a technique of determining individual's ability through their interest, hobbies, character attributes, passions etc. which gives easy solutions for career selection problem. Various algorithms can be considered for such prediction including Naïve Bayes, SVM (Support Vector Machine), CBR etc. which are making the prediction work much simpler and automatic

by filling some of the details like Name, Gender, Age, Grades, Hobbies, Skills on basis of which classification can be done for various students in different career and subjects selection according to their interest. The idea of SVM[2][3] is to find such linear separator . It categorizes data both theoretically and empirically. The theoretical analysis includes that SVM[6] can categorize text on basis of particular properties:

	Candidate 1	Candidate 2	Candidate 3	Candidate 4	Candidate 5	Candidate 6	Candidate 7
Current CGPA	9.2	7.8	7.9	8.35	8	9.4	4.5
Year	3	3	3	3	3	3	3
Semester	Even	Even	Even	Even	Even	Even	Even
Favorite Subject (In college/ Overall)	Data Structures And Algorithm	Networks	Python, Web programming, SW Testing	Software engineering, Software Testing	AI	Data Structures and Algorithm	Programming in C
	NA	NA	Physics	Mathematics	Mathematics	Mathematics	Biology
Least Favorite Subject	Java	Java	Biology, Electrical Eng, Chemical Eng, SW Design, SW Architecture	Biology	Physics	Biology	Web programming
Hobbies	Finance of real world	Reading Fiction	Swimming	Reading Stories, Hearing songs	Driving, Music	Reading	
Aim	Richest man	IAS	Technical	Quality Analyst	Data Analyst	Make tons of money	Be a developer
Favorite Domain	Finance	History	Web Programming	Event Management	Data Science	Tech	Web development
Expected Salary After Graduation	₹1; 80000 shares of company	₹75,000	₹5L to ₹8L	₹4.8L	₹10L	₹8-9 L	₹3L+
Pass time	Playing poker	Dancing, Singing, Painting	Gaming	Social Media	Music	Music, Books	Video in YouTube
Passion	Problem solving in C	Writing stories	Politics	Interior Design	Game Dev	Politics	Food
Hate Most	Change	Unfinished work	NA	Politicians, Racism		Sexists	Teachers
Favorite Color	Black	Pink/Blue/Black	NA	Blue	Black	Blue	NA
Status After 10 years	• India • Own company • 2-3 blonde assistants • Business class tickets	• IAS	• Abroad • Technical Head in Fb/ Google/ Twitter	• India With Satisfied bank account	• At Google	• Travelling the world	• India • Consultant

Figure 1: Survey of some students taken manually
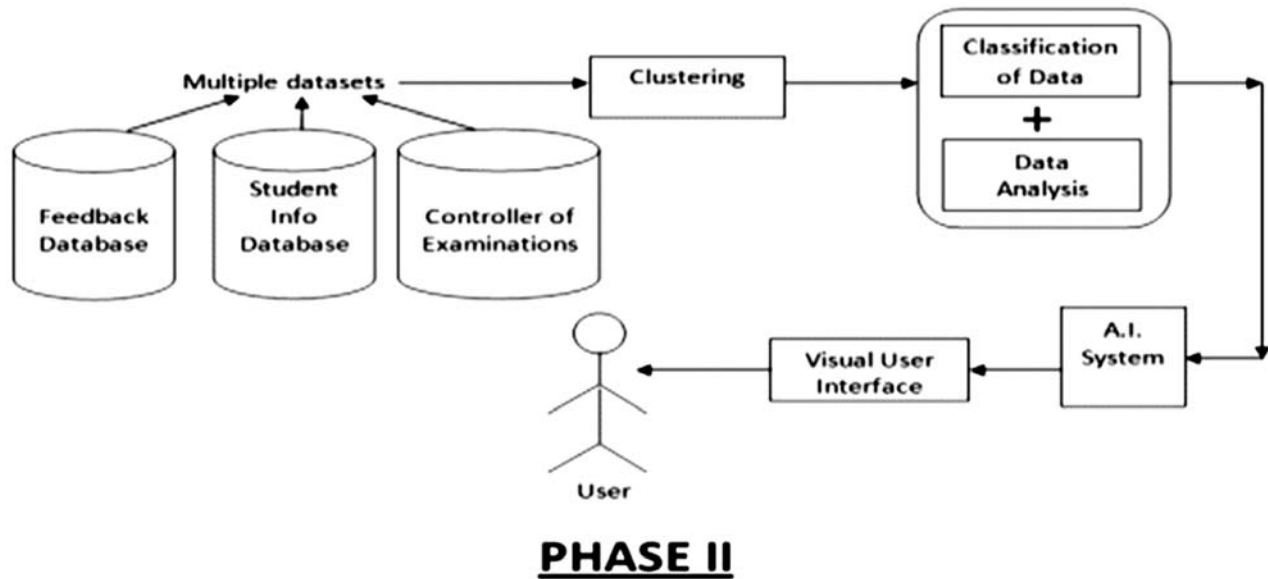

PHASE I
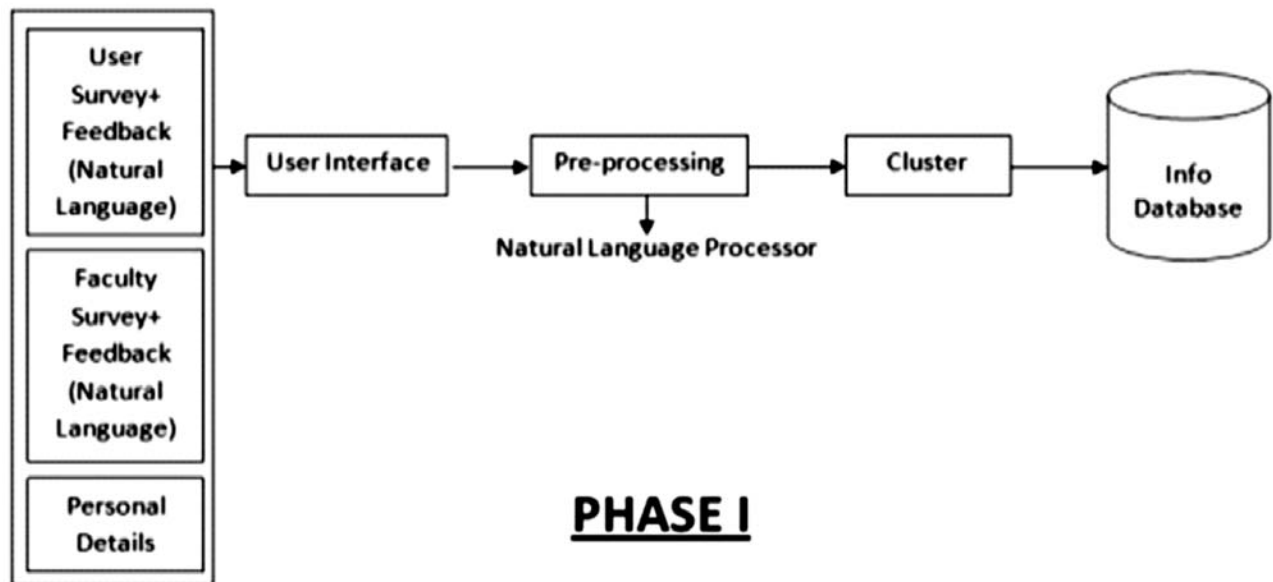
Figure 2: Architecture diagram for *e*-counseling system

1. High dimensional feature spaces 2.some of the irrelevant features and 3. Sparse instance vectors. SVM has ability to generalise the dimensional features, thus eliminating the requirement of feature selection and hence making text categorization much simpler. It does not require any specific tuning of parameters as it consist of some of the trained data sets which are automatic.

In SVM we can have pool based active learning where instead of choosing some random set of data a particular pool of data is chosen which is unlabelled beforehand and can request the labels for them afterwards. With SVM we have opportunity to choose which instance to choose next.

Recently, probabilistic framework was inferred into SVM. In this framework, we seek to obtain the evidence gradient of the training dataset and the likelihood of the model as a negative log-likelihood and predict the class given an unknown input. This approach has been used on regression as well as classification problems and has shown promising performance and results. By inferring SVM into the probabilistic framework, it is able to automatically tune the hyper parameters to their near optimal values. This can also be thought of as model selection.

2.6. Advantages

Also this website will increase the efficiency of the existing Manual Systems like:

1. Better suggestion based multiple factors.

2. Understanding about various subjects from other department

3. Better horizon for future career based on the subjects chosen

4. Broad scope and choice to be adapted based on interest

5. Reducing the mentoring time.

6. Avoiding Mistakes Due To Human Error (Accurate).

2. CONCLUSION

E-Counsellor Analytical System is a web application. The key concept is to provide analytical suggestion to the student in choosing the elective based on various factors. The student users with minimum knowledge about subjects are able to get feedback from the system easily which is most appropriate for the individual student.

REFERENCES

[1]   "Automated Career Counselling System for Students using CBR and J48"Maha Nawaz, Anum Adnan, Unsa Tariq, Jannat Fatima Salman, Rabia Asjad, and Maria Tamoor Journal of Applied Environmental and Biological Sciences 2014

[2]   "Text Categorization with Support Vector Machines: Learning with Many Relevant Features" Thorsten Joachims

[3]   "Support Vector Machine Active Learning with Applications to Text Classification" Simon Tong

[4]   "Sentiment Analysis of Twitter Data" , Apoorv Agarwal Boyi Xie Ilia Vovsha Owen Rambow Rebecca Passonneau.

[5]   https://en.wikipedia.org/wiki/Sentiment_analysis

[6]   "Data Classification usinng Support Vector Machine ",Durgesh K. Srivastava, Lekha Bhambhu.