RESEARCH ARTICLE   OPEN ACCESS

Recommendation System for Startups

Shubham Narode[1], Hitesh Pitrubhakta[2], Akash Sonawane[3], Vikram Nikam[4], Shubham Kadam[5]

1,2,3,4,5(Dept of Computer Engineering, SRES Sanjivani college of Engineering, Kopargaon)

Abstract:

Over the past two decades, the Internet has emerged as the mainstream medium for online shopping, social networking, e-mail and more. Corporations also view the Web as a potential business accelerator. They see the huge volume of transactional and interaction data generated by the Internet as research and development that informs the creation of new and more competitive services and products. The proposed system enable startup companies to track Clients interactions across services. This system allows to make open interaction between startup companies and their clients. In this system client ask for the solutions and multiple startup companies are allowed to give solutions. Depending on these solutions client will give feedback to startup companies. This feedbacks are then mined and recommendations are provided to startup companies. Also system allows visitors to view recommendations and solutions. This recommendations are helpful for startups to get more clients and they will get exposure.

Keywords —Feature Selection, Machine Learning, Opinion Mining, Ranking, Review, Startup.

I. INTRODUCTION

Today's people mostly prefers the internet to get the better services. Whenever people looking for any service on internet, there are number of companies available which can provide the same service. It is difficult to see the hundreds of reviews of different companies and compare them manually to make better choice. So people always preferred the popularly known companies in market so that other startup companies which are newly arrived in market didn't get platform to reach up to the more clients. In this paper, we proposed a system where clients can request for different services. When clients are requesting for any service, notification will be send to respective companies and that multiple startup companies will provide the solution for every requested service. After getting services clients must have to give feedback to respective company. Then feedback is get be extracted by using text mining. After that extracted feedback is classified as positive, negative or neutral by using "Naive Bayes Classifier". Each positive, negative or neutral feedback will carry some weight .According

to weight score will be calculated for company. According to score the company which will get highest score that company will be at first rank. The top rank companies will be recommended to customers for their future use.

The role of startups and clients is to register itself to the system, while visitors will be able to view content without registering. Clients will post the requirements and startup will provide solutions. The benefits of the system are all startups will come under one room. Also they will get exposure. Clients will get better solutions by top ranked companies.This paper deals with sentiment analysis hence we are considering clients feedback as their sentiment. Sentiment analysis is the measurement of positive and negative opinion. Using sentiment analysis feedback is evaluated to determine if the feedback is positive, negative or neutral and to what degree.

Sentiment analysis is important as it helps to understand what customer feels about your company and services which you provides.The feedback given by clients may contain very useful information which can be helpful to grow your business. By receiving your customer's

feedbackregularly you can be more proactive regarding changing trends in market. [1]

II. RELATED WORK

Recently there has been many research done on customer reviews for product ranking as well as in many recommendations system. The most related work on ranking is given in [2], which present product ranking model that assigns weights to product review factors to calculate product ranking score. In this paper filtering mechanism was used to preprocessed review. By using Support Vector Machine algorithm. This system proposed model consisting of three stages. The first stage filters out the unrelated comments from sentences. The second stage derives weights for a review based on its helpfulness votes and its date of posting. The third stage calculates product overall ranking score. Ranking score is determined by review contents, relevance of a review to the product quality, helpful votes and total votes from posterior customer posting date and durability of reviews. [3] In machine learning feature selection concept is existing for awarding the rank to each feature .Based on the rank awarded by the feature selection techniques Top 'N' features can be selected.Before applying classification methods preprocessing is required. In which, class imbalance and feature selection are some of the issues need to be considered. There are some literature existed to address these issue. SMOTE is an oversampling technique used to address the class imbalanced problem [4, 5]. Authors of the research article [6] applied the SMOTE with ensembling approaches for increasing the prediction rate of kidney disease data. Filter based feature selection algorithms have been applied by the researcher for the classification of SONAR signal data [4,8].Symmetrical Uncertainty based feature selection method have been applied over some of the medical dataset to derive the best features before applying classification algorithms [9,10]. Feature selection based on correlation coefficient and Symmetrical Uncertainty is proposed by the researchers and applied over various dataset belongs to diverse areas [11]. It is far superior to give feedback to any startup based on provided solution, product qualities, and individual skills very easily and rapidly. Analysing such data and its analysis for their business growth and trends. Automated opinion mining system to help, analyse, and evaluate client's reviews and to provide on click solution of review mined for decision making process. In this extraction of real time datasets is done and on this extracted dataset part of speech tagging is applied to get implicit as well as explicit feature of the solution. After this reviews are classified for further decision making operation.

Knowledge discovery from textual database refers to the process and extracting interesting pattern from unstructured text document. This is called as text mining or knowledge discovery. This system will lead to increase the stored data tremendously day by day. This is unstructured form so we cannot extract the needed information .Some text mining techniques are used to extract useful info from text documents.

III. PROPOSED SYSTEM

To overcome the limitations of earlier work, proposed system is developed. Uptil now so many research is done on product ranking system on the basis of product rating, time of review only. However proposed system decide ranking based on the review which includes quality and cost of services.In earlier work all review carries equal weights.However proposed system assigns weight according to type of review whether it is positive negative or neutral. Ranking will be decided on the basis of total score.Input for proposed system is client's requirement solution for specified requirement and client's feedback. Final output is ranking of startup companies. Following figure shows different components of recommendation System.
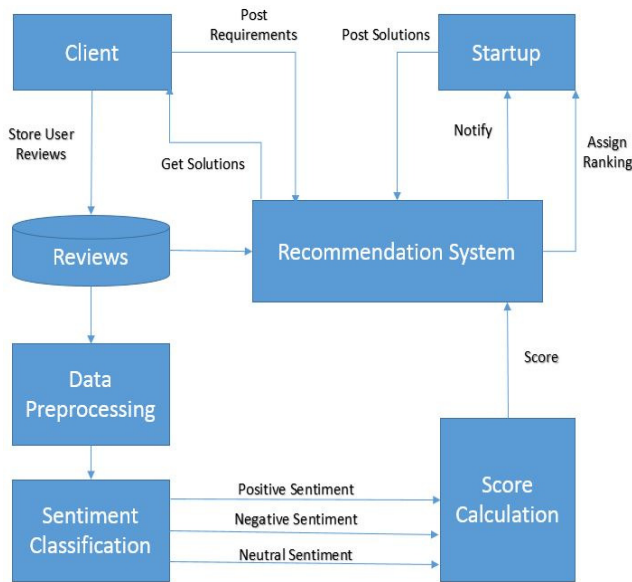
Figure I.   ARCHITECTURE OF SYSTEM

Working of System

Basically proposed system is consist of three main modules namely user module, admin module, ranking module.

A) User module

Clients and Startups are included under user module. Both clients as well as startups must have to register first. After registration verification is done by admin. Whenever any client will post their requirement on the system. The system will send notification to respective startup companies. As soon as notification will be received, multiple startups will provide solution for specific requirement. After getting solution client has to provide feedback to company. Client feedback is the input for ranking module.

B) Admin Module

Administrator is the owner of system. Roles of administrator are as follows:
- Manage Accounts: All accounts of user i.e. startups and clients are managed by admin.
- Verification of users.
- It provide feedback facility to clients.

C) Ranking Module

This is the main module of proposed system. In this module there are following subtasks.

1. Retrieval of reviews:
All the reviews given by clients are stored into database. All these reviews are retrieved from database and preprocessing is done on that reviews.

2. Preprocessing of data:
Preprocessing is the process of formatting the retrieved data in specific format so that it can be useful for next process. It involves three phases filtration, tokenization and removal of stop words.

i. Filtration: Many times people uses repeated letters in the words to express their feelings. For example peoplewrites the word "hiii" instead of writing like "hi". To avoid such type of repetition of letters, filtering is done and all these words are removed.

ii. Removal of stop word:The words like "a","the","an" are not more useful so I data preprocessing such words are removed.Special characters like [],{},(),/ also removed.

3. After preprocessing sentiment classification is done. Sentiment classification is done on preprocessed data. Using Naïve Bayes classifier preprocessed data is classified into different sentiments such as positive, negative or neutral sentiment.

For example:

Consider classifier classifies the review as positive or negative. Consider training set has 4 reviews.

TABLE I.   SAMPLE DATASET

Text/Review	Category
Performance is poor	Negative
The company is good	Positive
The solution is satisfactory	Neutral
The solution is better	Positive

As Naïve Bayes is probabilistic classifier we should calculate probability that reviews is positive or negative. Mathematically it can be given as

P(Positive/Review)-The probability of category of review is positive and the review is "company is good."

Bayes Theorem:

Bayes theorem works on conditional probabilities.Mathematically it is given as,

$$P(A|B) = P(B|A)*P(A)/P(B)$$

In this case
P(Positive | The company is good)=(P(The company is good |positive)*P(Positive))/P(The company is good)

As we are just finding out larger probability category. We can remove divisor which is same for both. Compare P(The company is good |Positive)*P(Positive) with P(The company is good |Negative) To obtain P(The company is good |Positive),count how many times the appears in positive category divide it by total.

Being Naïve:

All words in sentences are independent of each other so we are considering individual words rather than sentence Example: "The company is good" is same as "The good company is"Therefore,

P(The good company is)=P(The)*P(good)*P(company)*P(is)
It works well for less data.
P(The company is good |positive) =P(The|Positive)*P(Company}positive)*P(is}posittive)*(Good|Positive)

Calculating probabilities:

The final step is to calculate all probability & find out larger one.
Posterior Probability = (likelihood probability *class prior probability)/predictor prior probability
On the basis of probability score is calculated and rank will be decided.

IV. RESULT

We have executed the application with dummy datasets and obtained results. Following pictures gives evidence of testing.
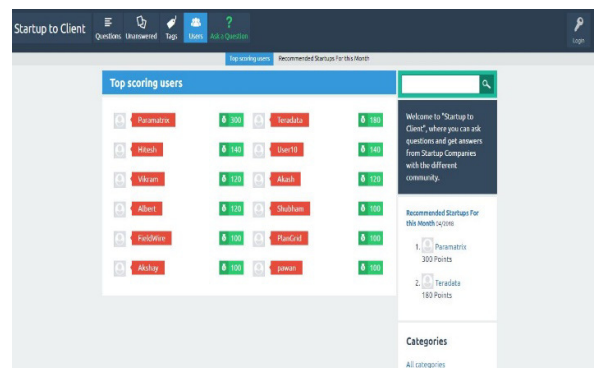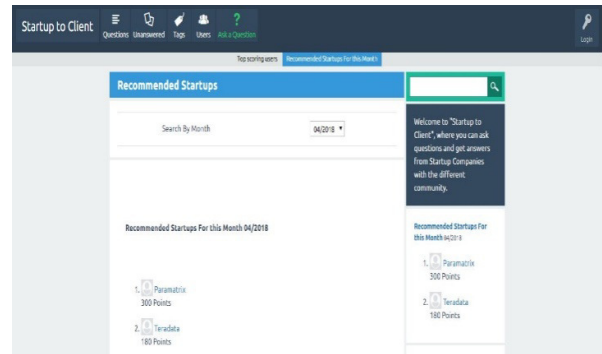
Naïve Bayes algorithm is tested on different positive and negative datasets and results are recorded.

V. CONCLUSIONS

The proposed system as the name indicates, Recommendation System for Startups provides an attractiveuser interface. This system give solution to track user interaction across services. The system design is flexible enough to do analysis and translate it into code for the development. This system implements data mining algorithms to generate recommendations to startups. Truly this system will able to provide exposureto startup.

REFERENCES

1. Lun-Wei Ku, Yong-Sheng Lo, Hsin-Hsi Chen, Using Polarity Scores of Words for Sentence-level Opinion Extraction. Proceedings of NTCIR-6 Workshop Meeting, May 15-18, 2007, Tokyo, Japan

2. Kunpeng Zhang, Yu Cheng, Wei-kengLiao, Alok Choudhary, Mining Millions of Reviews: A Technique to Rank Products Based on Importance of Reviews ,2011, Association for the Advancement of Articial Intelligence

3. M.A.Jawale, Dr.D.N.Kyatanavar, A.B.Pawar , Implementation of Automated Sentiment Discovery System ",IEEE International Conference on recent advances and innovations in engineering(ICRAIE-2014).

4. Potharaju, S.P. and Sreedevi, M., 2016. An Improved prediction of Kidney disease using SMOTE. Indian Journal of Science and Technology, 9(31).

5. Chawla, Nitesh V., Kevin W. Bowyer, Lawrence O. Hall, and W. Philip Kegelmeyer. "SMOTE: synthetic minority over-sampling technique." Journal of artificial intelligence research 16 (2002): 321-357.

6. Potharaju, S.P. and Sreedevi, M., 2016. Ensembled Rule Based Classification Algorithms for predicting Imbalanced Kidney Disease Data. Journal of Engineering Science and Technology Review, 9(5), pp.201-207.

7. Kumar, Vipin, and Sonajharia Minz. "Feature selection." SmartCR 4, no. 3 (2014): 211-229.

8. Potharaju, S.P. and Sreedevi, M., 2018. A Novel Subset Feature Selection Framework for Increasing the Classification Performance of SONAR Targets. Procedia Computer Science, 125, pp.902-909.

9. Novaković, Jasmina. "Toward optimal feature selection using ranking methods and classification algorithms." Yugoslav Journal of Operations Research 21, no. 1 (2016).

10. Potharaju, S.P. and Sreedevi, M., 2017. A Novel M-Cluster of Feature Selection Approach Based on Symmetrical Uncertainty for Increasing Classification Accuracy of Medical Datasets. Journal of Engineering Science & Technology Review, 10(6).

11. Potharaju, S.P. and Sreedevi, M., 2017. A Novel Clustering Based Candidate Feature Selection Framework Using Correlation Coefficient for Improving Classification Performance. Journal of Engineering Science and Technology Review, 10(6), pp.38-43.