Sentiment Analysis using Naïve Bayes Classifier for Faculty Rating System

Pragna Panduranga[1], Sushma C S[1,], Sharanya S[1], Vindhya C R[1], Mr. Shashank D[2]
[1]Student, ISE/The National Institute of Engineering, Mysuru, Karnataka, India
[2]Assistant Professor, ISE/The National Institute of Engineering, Mysuru, Karnataka, India
(E-mail: crvindhya@gmail.com)

Abstract- It is hard to keep track of performance of each faculty in a vast in educational system. Student feedback helps the educational institutions and the faculty to understand if their way of teaching is helping the students and if there is need for new methods. This paper focuses on using sentiment analysis to extract a single word to evaluate the faculty and then use Naïve Bayes Classifier which is a machine learning algorithm to rate the faculty based on few characteristics. The faculty assigns a weightage to each student within an institution for unbiased feedback. Based on feedbacks given by students, the faculty will be given a rating which in turn helps in increasing the standards of education.

Keywords*: Data Mining , Naive Bayes Classifier, Text Mining, Sentiment Analysis, NLTK(Natural Language Toolkit).*

I.   INTRODUCTION

The education system needs to be constantly changing due to the advent of new technologies. So, there is a need for faculty of an educational system to also be on par with the latest teaching methodologies. This model helps in keeping track of performance of each faculty by capturing the emotions of the students in the feedbacks submitted by them using sentiment analysis and Naïve Bayes classifier for processing them. The faculty is then given a rating between 1 to 5 stars with 1 being the lowest and 5 being the highest. Student weight is assigned by the faculty for each student on the basis of few student characteristics and finally a risk class is found for the faculty. Visual Studio is used to develop front end and Microsoft SQL Server is used for communication on server side for storing output in database.

A. Existing Approach

The paper [1] makes use of simple text mining and sentiment analysis to capture the emotions of people regarding a particular politician using Naïve Bayes Classifier.

The paper[2] combines the reviews of people regarding a product and uses sentiment analysis to depict the emotions of people regarding a particular product .

In [3] sentiment analysis is applied on product reviews given by the customers to know the quality of a product. The model used here takes the reviews as input and uses three steps to process them. Data preparation, review analysis and classification of sentiments.

II. OUR APPROACH

The application has front end, backend and server side communication. The student submits feedback through internet. This forms the front end. The backend contains the database where various attributes are maintained.
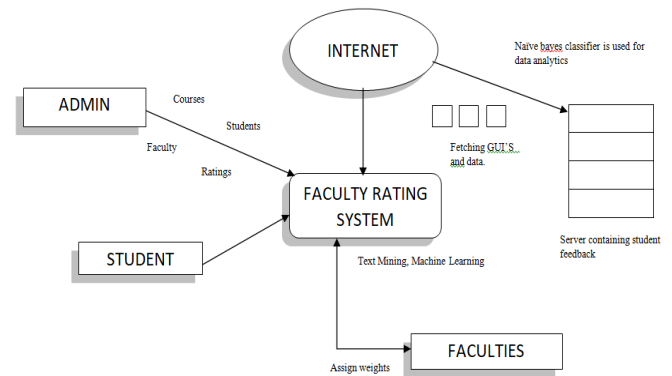

Fig. 1.0 Architecture

The attributes in the database that are maintained are
- Faculty department
- Student name
- Faculty name about whom the feedback is being given
- Student ID and faculty ID
- CGPA of the student

The values of the attributes in the feedback are processed using text mining. The output of this is fed into the Naïve Bayes Classifier as its inputs.

A. Text mining

NLTK (Natural Language Toolkit) module with the Splitter class is made use in text mining. Each sentence that is given as the feedback is split using the Splitter class and is tokenized. All the tokenized words of all the sentences are stored in a list.

Example: The feedback given by a student is as follows:

"He was very rude at first. But now he has improved."

These sentences are tokenized as follows:

['he','was','very','rude','at','first']

['but','now','he','has','improved']

The tokenized sentence is checked for the emotions described in the input. For the above example, the list is as follows,

['very rude' , 'has improved']

To find the best word that describe the qualities of the teacher, two tables are built of the most emotions that are frequently used with their respective values.

Given below are the two tables.

Table 1. Values assigned for commonly used words[4].

Word	Value
very rude	1
very poor	2
very bad	3
very strict	4
rude	5
poor	6
bad	7
strict	8
arrogant	9
not nice	10
short tempered	11
unpleasant	12
not good	13
not polite	14
doesn't understand	15
not bad	16
not strict	17
ok	18
indifferent	19
nice	20
polite	21
understands	22
good	23
very polite	24
very nice	25
calm	26

In our example, the list storing the words ['very rude', 'has improved'], the output word would be 'ok'. We get this by taking the average of the values of the words in the list.

Table 2. Values assigned for commonly used words[4].

Word	Value
very poor	1
very bad	2
poor	3
bad	4
doesn't know anything	5
not good	6
not bad	7
must improve	8
doesn't know much	9
ok	10
not interesting	11
average	12
not great	13
not perfect	14
can improve	15
good	16
interesting	17
has improved	18
very good	19
great	20
perfect	21
superb	22
fabulous	23
awesome	24
brilliant	25
excellent	26
best	27
outstanding	28
pleasant	29
friendly	30

B. Naïve Bayes Classifier

Naïve Bayes algorithm is used for classifying a particular input to a particular output known as the risk class. It is a probabilistic model, based on the Bayes theorem which uses conditional probability.

The algorithm [7] for prediction is as follows,

Step 1: Scan the dataset (storage servers), retrieval of required data for mining from database, cloud, excel sheet etc.

Step 2: Calculate the probability of each attribute value.

[n, n_e, e, p]

Here for each attribute we calculate the probability of occurrence using the formula given below. For risk class we should apply the formula.

Step 3: The formula is given by
P(attribute_value(ai)/subject_valuevj)=(n_e + ep)/(n+e)
Where:
n = the number of training examples for which v = vj
n_e= number of examples for which v = vj and a = ai
p = a priori estimate for P(aijvj)
e= the equivalent sample size
Step 4: Multiply the probabilities by p for each class, here we multiple the results of each attribute with p and final results are used for classification.
Step 5: Compare the values and classify the attribute values to one of the predefined set of class. [7]
The list having the final word(s) is given as input to the Naïve Bayes classifier. The value of the input is compared to that of the standard dataset. Then the output results are mapped onto the database and the average rating of the faculty is got.
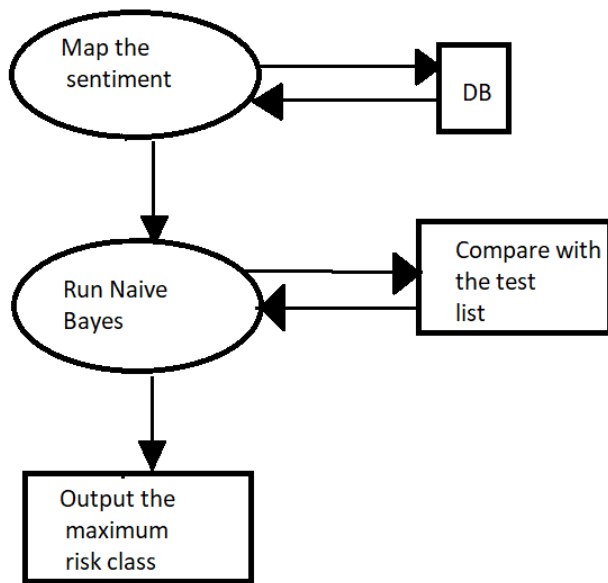
Table 3. Weights assigned for attendance[4].

Attendance (in percentage)	Weight
< 75	0.0
75-80	0.06
80-85	0.12
85-90	0.18
90-95	0.24
95-100	0.30

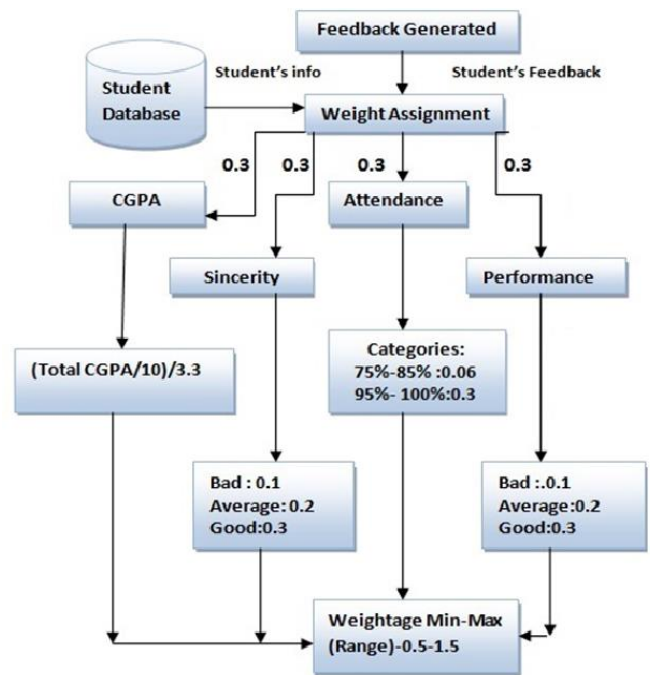Fig 1.2 Naïve Bayes classifier[4].

Fig 1.3 Weightge Assignments[4].

III. CONCLUSION AND FUTURE WORK

Our system uses the data mining technology to evaluate the faculties of an educational institution. The feedback given by the students is the best way to rate the faculties by using the text mining and Naïve Bayes techniques. The main advantage of our system is that it uses the multi point system for rating the faculty. The student gives a feedback in the subjective manner which uses the basic text mining approach. The student weight given to the students helps in increasing the accuracy of the rating. As of future work, we can extend the rating to the supporting staff and other degrees such as BCA, MCA, MBA, M. Tech etc.

C. Updating the student's weight

The student weight is required because sometimes, a student who has a grudge or dislikes the faculty may give negative feedbacks deliberately. To prevent this, the faculty will assign each student a weightage based on certain characteristics. They are attendance, sincerity, behavior and CGPA of student. Each of these characteristics can be given value up to 0.3. The weightage of student feedback can range from 0.5 to 1.5. This is combined along with the rating obtained by applying Naïve Bayes Classifier to get the final risk class. This process is repeated for every student within an institution. The weight assignment for table is as shown in the table 3. The similar process is repeated for other characteristics.

REFERENCES

[1] Jyoti Jain, Asst. Prof. Archana Shinde & Prachi Panchal, Sentiment Analysis using Machine Learning. Department of Information Technology, Sinhgad Academy of Engineering, Maharashtra, India, 2016

[2]Cane W.K. Leung & Stephen C.F. Chan. Sentimental Analysis of Product Reviews. The Hong Kong Polytechnic University, Hong Kong SAR, 2008

[3]Yustinus Eko Soelistio and Martinus Raditia Sigit Surendra. Simple Text Mining For Sentiment Analysis of Political Figure Using Naive Bayes Classier Method. CoRR, 2015

[4]Krishnaveni K S Rohit R Pai Vignesh Iyer. Faculty Rating System Based on Student Feedbacks Using Sentimental Analysis. Department of Computer Science and Engineering, Amrita School of Engineering, Amritapuri, Amrita Vishwa Vidyapeetham, Amrita University, India, 2017.

[5]G.F. Luger. Artificial Intelligence Structures And Strategies For Complex

Problem Solving, Pearson Education Inc., 182-185., 2009.

[6]Yamanishi, K., & Li, H. Mining open answers in questionnaire data. IEEE

Intelligent Systems, 17(5),pp. 58-63, 2002.

[7]Mythri G, Bhat Geetalaxmi Jayram. "Feature based Encryption for Data Privacy and Access Control for Medical application" , 2017 International Conference on Current Trends in Computer, Electrical, Electronics and Communication (CTCEEC), 2017

.