

# FEATURE BASED OPINION MINING ON STUDENT FEEDBACK

Veeram Reddy Sai Sushmi

School of computer science and engineering

Lovely Professional University

Gmail: saisushmi2@gmail.com

## I.ABSTRACT:

Opinion mining defines the emotional note behind a text. This is the popular way used to categorize the feedback whether it is positive or negative. Opinion mining is widely known as sentiment analysis. It is used by various organizations for business and other activities. Feedback is evaluation to any product which is very crucial for the growth of it. Large online shopping sites like Flipkart, Amazon seek for reviews of their products. As internet usage is growing day by day a huge amount of opinionated data is available through social media (twitter, blogs). As the importance of reviews or feedback is very essential as the opinion mining is. It is essential to classify the written reviews and categorize them into understandable data whether it is a positive, a negative or a neutral review by machine learning techniques [7].

In this student feedback we have data of teaching, course content, examination, lab work, library facilities, extracurricular. Based on these feature's opinion mining is evaluated and categorized. Before this the data is first pre-processed that is data is being transformed in to an efficient form by making it free from irrelevant elements such as stop words, missing values and special symbols. Only after the data set is accurate and efficient then the opinion mining is done. Thus, the data will be well classified and evaluated by evaluation measures (accuracy, F1 score). In this paper, we will see a detailed approach on feature-based opinion mining on student feedback.

## II.INTRODUCTION:

Opinion mining is also known as sentimental Analysis. It is used to derive an efficient result based on the feedback. It is very essential as it is used to derive results

from a large set of data. Suppose if a customer wants to buy a product, he will check the reviews, instead of checking all the reviews opinion mining can be used to derive a conclusion by analysing large number of reviews. Opinion mining can be evaluated by three levels Document level where the whole document is evaluated and a result is derived, Sentence level in which each sentence is evaluated and a sentence wise result is obtained, Feature level gives result based on each feature. Student Feedback has both numerical data (0, 1, -1) and qualitative data in the form of comments. This data is first processed into consistent data then an accurate set of data is extracted that is training data/feature set. Now, we implement the machine learning models on the feature set and the data is tested [1].

Sentimental Analysis is complicated like in some cases feedback may contain both positive and negative sentences we deal with it by performing sentimental analysis sentence wise. Some sentences may be like "he is punctual" and "he is not punctual" which are almost similar but the meaning has a huge difference these types of sentences are perfectly categorized as positive or negative. Interrogative sentences are accurately classified based on a keyword [2].



Fig1: Opinion mining or sentiment Analysis

### III. LITERATURE REVIEW:

Stamitios-Aggelos et al [3] has proposed data preprocessing in predictive data mining. The quality of the data set plays an efficient role in success of data mining. The data set has to be free from noisy data, missing values, outliers. In case if the data is very large then an efficient set of data is selected by instance selection. Moreover, the data sample plays an important role in the efficiency and performance of a machine learning model. Different learning algorithms (Decision trees, Bayesian, SVM and so on) are implemented and compared their applicability in instance selection, Normalization, Feature selection, Outlier detection and some other parameters.

Vaibhav Jain [4] provided a technique in “Evaluating and summarizing Student’s Feedback Using opinion mining”. As feedback is very important to increase one’s performance and learn accordingly, proposed an online feedback system that processes the student’s feedback and summarize the opinions of student’s by SVM classification algorithm which handle large amounts of data and also avoid overfitting. The categorized data from SVM is passed to generate summary of each teacher for each feature by Performance summary generator. Performance of a teacher is generated by bounded summary algorithm. It also displays the number of positive, negative feedback’s and also summarize the overall feedback’s given by the students.

R. Menaha et al [5] proposed Student Feedback Mining system using sentimental analysis. The data is preprocessed through a series of steps first feedback is collected then text is processed with tokenization and removal of stop words and then clustering is done to divide or cluster the comments and at last these are classified whether the comment is positive or negative by classification using visualization techniques. It mainly concentrated on the qualitative and quantitative feedback to make it more sensible and provide accurate feedback to the teachers to improve their teaching. Explained keyword extraction and classification with algorithms.

T. Nikhil Prakash et al [6] proposed Data Preprocessing in sentiment analysis using twitter data. Unstructured data set is efficiently transformed into a processed data by a sequence the prepared data is cleaned from unwanted data and errors by data cleaning and then data is reduced and is taken as input to Data Mining algorithms. Then, the sentiment score is calculated and are classified as positive, negative, neutral sentiments by Sentiment Classification. Data Cleaning includes “Stemming” through which inflectional words are removed and “Lemmatization” that is morphological analysis of the words. An algorithm is proposed for Sentimental Classification.

Padmapani P et al [8] explained steps for feature-based opinion mining by focusing on three tasks that is identifying the features and classifying them and finally summarizing. Explained different approaches for feature-based opinion mining for identifying polarity. Measured the performance of the algorithms by Recall, Precision, Accuracy.

### IV. PROPOSED METHODOLOGY:

Feature based opinion mining on student feedback. We will see this in detail.

#### 1.Data Set:

I collected the dataset from the internet from Kaggle. It consists of 6 categories they are teaching, course content, examination, labwork, library, extracurricular and 184 rows.

#### 2. Data Cleaning:

Cleaning the incomplete, noisy, inaccurate data from the dataset is called Data Cleaning [9].

##### a. Missing values:

In the data we may have some missing rows. This data can be handled by

- i. Manually filling the missing data
- ii. Filling the missing values by mean, median and mode
- iii. Deleting the rows containing missing values

b. Noisy data:

Meaningless and faulty data is called Noisy data. This data can be removed by Clustering that is grouping the data into clusters so that the irrelevant data fall outside the clusters.

After the dataset is free from noisy data and missing values, opinion mining is done on the cleaned dataset so that it gives accurate results.



Fig2: Data Cleaning

3.Opinion mining:

Opinion mining or sentimental analysis is identifying and categorizing the sentences based on the opinion behind the sentence, whether the sentence is positive, negative or neutral.

a. Tokenization:

Splitting the sentences into words. It can be splitting them into words, sometimes phrases also.

Ex: Explanation of the topic is excellent!

Explanation

of

the

topic

is

excellent

!

b. Removing the special characters:

Removing characters like, #, %, &. As these doesn't add any meaning to the sentence these are removed.

c. Removing stop words:

Stop words are unwanted words which doesn't contribute any value in determining the opinion behind the sentence are removed. Stop words are is, am, the....

d. opinion mining:

Categorising each sentence into positive, negative, neutral based on the sentiment. If the sentence is positive +1, negative -1, neutral 0.

I used TextBlob a python library for processing the textual data by calculating polarity and subjectivity of a sentence. Polarity gives the measure how positive or how negative a sentence is, it lies in the range of [-1,1]. Subjectivity is the opinion of the sentences lying in the range of [0,1].

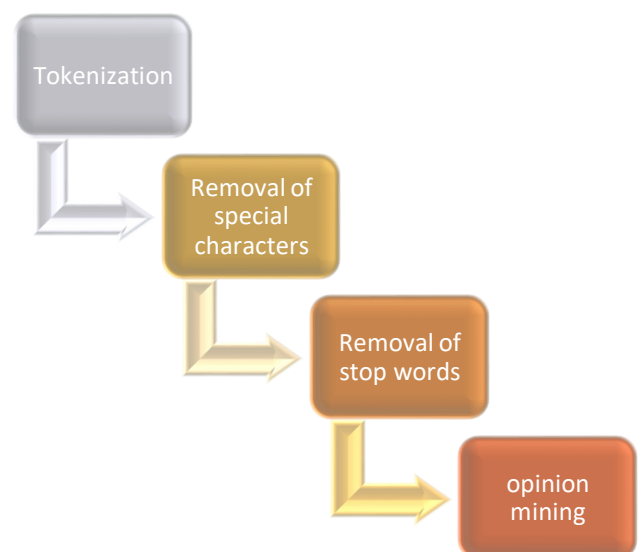


Fig3: Flow of the project

#### 4. Applying SVM for classification:

Support vector machine (SVM) is especially used for classifying the data into two classes. It is about finding the best hyperplane for classification by maximizing the width of the margin. In this case the student feedback is divided into positive feedback and negative feedback. SVM is best suitable as it can efficiently handle high-dimensional data. SVM is used for polarity classification which can efficiently differentiate between the positive and the negative comments. The risk of overfitting is very less in case of SVM. It also gives better results that is the best hyperplane compared to other algorithms or classifiers [10].

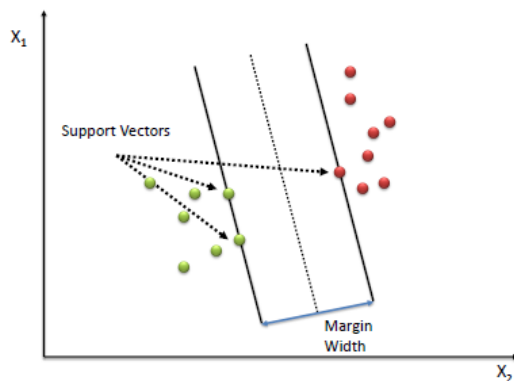


Fig4: Support vector machine

#### V. RESULT AND DISCUSSION:

We use some evaluation measures to find how accurately our algorithm is predicting.

The metrics that I used in my project to measure the accuracy of the algorithm are:

Accuracy:

It is the ratio of number of correct predictions to the total number of instances.

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN}$$

Recall:

$$\text{Recall} = \frac{TP}{TP + FN}$$

Precision:

$$\text{Precision} = \frac{TP}{TP + FP}$$

F1 Score:

$$\text{F1 Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

To calculate the TP (true positive), TN (true negative), FP (false positive), FN (false negative) we have confusion matrix:

|        |       | Predicted           |                     |
|--------|-------|---------------------|---------------------|
|        |       | Negative            | Positive            |
| Actual | False | True Negative (TN)  | False Positive (FP) |
|        | True  | False Negative (FN) | True Positive (TP)  |

```
(185, 2) (185,)
```

```
Accuracy score 0.7321428571428571
```

```
confusion matrix [[ 0  0]
```

```
[ 0  0]
```

```
[ 4 11 4]]
```

```
classification report
```

|              | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| -1           | 0.00      | 0.00   | 0.00     | 0       |
| 0            | 0.00      | 0.00   | 0.00     | 0       |
| 1            | 1.00      | 0.73   | 0.85     | 56      |
| accuracy     |           |        | 0.73     | 56      |
| macro avg    | 0.33      | 0.24   | 0.28     | 56      |
| weighted avg | 1.00      | 0.73   | 0.85     | 56      |

Fig4: Values obtained after using evaluation measure of teaching feature

I got different values of accuracies for each feature that the algorithm will predict the results.

There are also some issues while performing opinion mining:

- Interrogative Sentences: They do not have any proper opinion as positive or negative.
- Sarcastic Sentences: destroy the opinion behind a sentence.

- Some sentences without sentiment words: It is difficult to know the opinion and classify the sentence.

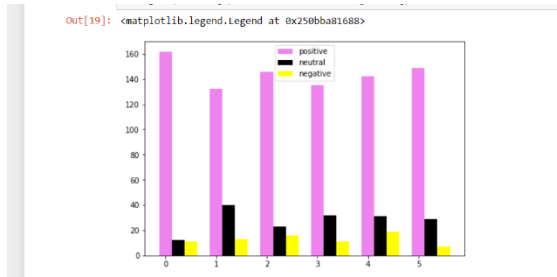


Fig5: Bar plot representation of each feature

Fig5 is the result that i have got by opinion mining each aspect that is teaching, Course content, Examination, Labwork, library, Extracurricular. I calculated and represented how much percentage of each content is positive, negative, neutral and visualized it.

Comparison with other literatures:

| Title   | Approach                   | Result   |
|---|----------------------------|--|
| Data preprocessing in predictive data mining [3]                      | SVM, Bayesian network      | Importance in Processing the data efficiently  |
| Evaluating and summarizing students feedback using opinion mining [4] | SVM, Summary generator     | Proposed a performance summary generator   |
| Student feedback mining system using sentiment analysis [5]           | Clustering, Classification | Accurately classified the sentiments of the sentences as positive and negative           |
| Data preprocessing in sentiment analysis using twitter data [6]       | Stemming Lemmatization     | Accurately predicted the sentiment of twitter data and data cleaning is efficiently done |

A peer review of feature-based opinion mining and summarization [8]

Lexicon based Approach, Statistical Approach, Intelligent feature selection approach

A step by step approach in feature-based opinion mining and efficiently classified and summarized

## VI. CONCLUSION:

As feedback is most important in evaluating and for improvisation of a person or a product. Manually evaluating each and every review or feedback is impossible. In this a system for opinion mining is used to do opinion mining to a high-dimensional data accurately. The important aspect of this process is data cleaning. Only if the data is efficient and free from irrelevant data then only it is possible to predict accurately using a machine learning algorithm. If the data is irrelevant the data should be cleaned that is removing of missing values, outliers to make it relevant is explained in detail. A step by step approach to opinion mining is explained in detail. First Tokenization, removal of stop words, removal of special characters only after that the opinion mining is done that is each sentence is labelled as positive or negative. I used SVM for classifying the data. I also used some metrics like precision, recall, F1 score, accuracy for measuring how good the algorithm performs in predicting the opinion of the sentence. I visually represented the data in the form of graph. I represented how much percentage of the feedback's in each feature are negative and positive. Successfully predicted each feedback and evaluated the examination, lab work, Teaching, course content, extracurricular, library features.

As opinion mining has become essential in predicting the opinion from the feedback, there are some libraries like text blob which are used to measure the polarity

and subjectivity of each sentence. Supervised machine learning approach perform accurately in case of high dimensional data and also provides accurate results compared to other approaches. In this paper I gave a detailed and a step by step approach to measure and classify the sentences based on the opinion as positive and negative and also evaluated the accuracy of the algorithms used. I conclude that supervised machine learning algorithms perform well in classifying the data and also opinion mining accurately predicts the sentiment behind a sentence.

## VII. REFERENCES:

- [1][https://www.leadingindia.ai/downloads/projects/ED/ed\\_3.pdf](https://www.leadingindia.ai/downloads/projects/ED/ed_3.pdf)
- [2][https://www.researchgate.net/publication/334490227\\_Aspect\\_Based\\_Opinion\\_Mining\\_On\\_Student%27s\\_Feedback\\_For\\_Faculty\\_Teaching\\_Performance\\_Evaluation/fulltext/5d2dc75ba6fdcc2462e4ac5c/334490227\\_Aspect\\_Based\\_Opinion\\_Mining\\_On\\_Student%27s\\_Feedback\\_For\\_Faculty\\_Teaching\\_Performance\\_Evaluation.pdf?origin=publication\\_detail](https://www.researchgate.net/publication/334490227_Aspect_Based_Opinion_Mining_On_Student%27s_Feedback_For_Faculty_Teaching_Performance_Evaluation/fulltext/5d2dc75ba6fdcc2462e4ac5c/334490227_Aspect_Based_Opinion_Mining_On_Student%27s_Feedback_For_Faculty_Teaching_Performance_Evaluation.pdf?origin=publication_detail)
- [3]Data preprocessing in predictive data mining [Stamitios-Aggelos N. Alexandropoulos](#) <sup>(a1)</sup>, [Sotiris B. Kotsiantis](#) <sup>(a1)</sup> and [Michael N. Vrahatis](#) <sup>(a1)</sup>
- [4] Evaluating and Summarizing Student's Feedback Using Opinion Mining, Vaibhav Jain Institute of Engineering & Technology, Devi Ahilya Vishwavidyalaya Indore, India
- [5]<http://ijcat.com/archieve/volume6/issue1/ijcatr06011009>
- [6] INTERNATIONAL EDUCATIONAL APPLIED RESEARCH JOURNAL (IEARJ) Volume 03, Issue 07, July 2019 E-ISSN: 2456-6713 89 | DATA PREPROCESSING IN SENTIMENT ANALYSIS USING TWITTER DATA T. Nikhil Prakash1, Dr. A. Aloysius2
- [7]<https://ieeexplore.ieee.org/abstract/document/6812968>
- [8][https://www.researchgate.net/publication/261842757\\_A\\_Peer\\_Review\\_of\\_Feature\\_Based\\_Opinion\\_Mining\\_and\\_Summarization](https://www.researchgate.net/publication/261842757_A_Peer_Review_of_Feature_Based_Opinion_Mining_and_Summarization)
- [9]  
[https://en.wikipedia.org/wiki/Data\\_cleansing](https://en.wikipedia.org/wiki/Data_cleansing)
- [10]  
[https://www.researchgate.net/publication/290788060\\_A\\_Survey\\_on\\_Sentiment\\_Analysis\\_Algorithms\\_for\\_Opinion\\_Mining](https://www.researchgate.net/publication/290788060_A_Survey_on_Sentiment_Analysis_Algorithms_for_Opinion_Mining)
- [11][https://www.researchgate.net/publication/290788060\\_A\\_Survey\\_on\\_Sentiment\\_Analysis\\_Algorithms\\_for\\_Opinion\\_Mining](https://www.researchgate.net/publication/290788060_A_Survey_on_Sentiment_Analysis_Algorithms_for_Opinion_Mining)
- [12]<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.449.3593&rep=rep1&type=pdf>
- [13][https://www.researchgate.net/publication/258704904\\_Feature\\_Level\\_Opinion\\_Mining\\_of\\_Educational\\_Student\\_Feedback\\_Data\\_using\\_Sequential\\_Pattern\\_Mining\\_and\\_Association\\_Rule\\_Mining](https://www.researchgate.net/publication/258704904_Feature_Level_Opinion_Mining_of_Educational_Student_Feedback_Data_using_Sequential_Pattern_Mining_and_Association_Rule_Mining)
- [14] <https://www.geeksforgeeks.org/data-preprocessing-in-data-mining/>
- [15]<https://searchbusinessanalytics.techtarget.com/definition/opinion-mining-sentiment-mining>