Improving feedback analysis: Deep learning approach to college customer satisfaction assessments

Priyank Singhal¹ | Bulbul Chaudhary² | Vikas³

¹Teerthanker Mahaveer University, Moradabad, Uttar Pradesh, India, Associate Professor, Computing Sciences and I.T.
²ATLAS SkillTech University, Mumbai, India, Pro-Vice Chancellor, School of Design & Innovation.
³IIMT University, Meerut, Uttar Pradesh, India, Assistant Professor, Computer Science and Applications.

Abstract Establishing consumers' views via text-based feedback in a questionnaire is crucial for organizations, include education, since it gives a summary of significant areas that help administrators plan, regulations, and decision making. Through surveys, academic organizations have gathered huge quantities of textual data all over the years. For the organization, it is still difficult to analyse the vast quantities of unstructured feedback from customers to understand their concerns and opinions generally. In this study, we propose deep learning (DL) based technique called topic modelling that utilizing Naive Bayesian (NB) to automatically summarize text and retrieve ideas from this raw data. Additionally, it discusses the text mining procedure used to extract relevant information from the vast volume of text-based data. The most significant issues obtained through feedback from customers were subsequently identified. The findings showed particular issues for workplaces, including environment, staffing, IT infrastructure, and customer feedback system. The feedbacks also prominently highlight difficulties with the attitude of student assistance and security staff as well as the library's management and operations.

Keywords: deep learning, Naive Bayesian, topic modelling, customer satisfaction

1. Introduction

Higher education institutions must prioritize student happiness. Maintaining a healthy learning environment, enhancing educational results, and drawing in new students depend on assessing and comprehending the satisfaction levels of students and other stakeholders within the college ecosystem (Qu tieshat et al 2020). To learn more about student happiness, universities have traditionally used surveys, ways to provide feedback and assessments. But presumptions, small sample numbers, and laborious data processing procedures often plague these approaches. As machine learning technology develops, there is a chance to use it to increase and simplify the assessment of college student happiness.

A critical component of every organization, including schools and universities, is customer satisfaction measurement. Machine learning algorithms have seen a considerable change in recent years in analysing consumer feedback and gauging client happiness (Dash et al 2021).

Machine learning, an aspect of machine learning, concentrates on training neural networks with numerous layers to discover intricate patterns and representations from big datasets. In a number of fields, including voice recognition, picture recognition, and natural language processing, this method has shown astounding performance. We can extract valuable insights from enormous volumes of data by using machine learning methods to the assessment of college customer satisfaction. This would enable schools to make data-driven choices to enhance students’ experiences and institutional performance (Asghar et al 2021).

This strategy may assist universities in identifying locations where patrons are expressing unfavourable opinions, enabling administrators to take remedial action to raise patron contentment. The modelling of a topic is another kind of machine learning that is used for evaluating customer happiness. In topic modelling, algorithms find recurring themes or subjects in a vast dataset. This method enables universities to pinpoint the most important aspects that impact student happiness and concentrate their efforts there (Capuano et al 2021).

The main goal of this study is to create a machine-learning model capable of reliably predicting and evaluating college consumer happiness. To find trends, correlations, and predictors of satisfaction levels, the model will be generated on the data that have been gathered. We may learn more about the important areas where schools can concentrate their efforts to increase customer satisfaction by examining the model's findings (Özkan et al 2020).
In this regard, machine learning algorithms provide a potential method for analyzing customer feedback and measuring customer satisfaction in schools and universities. Colleges may improve the entire customer experience by using these strategies to acquire insightful information about consumer sentiment and pinpoint areas that need improvement (Ibrahim et al. 2019).

The data gathered from this study (Ho et al. 2021). Suggest that higher education institutions should place a high priority on supporting and enhancing initiatives made by instructors, modulating assessment procedures to enhance the workload, suitability, and impartiality for unforeseen modifications using ERL during a crisis, developing contingency strategies and other complementary educational activities or assets to compensate for the inferiority, or learning deficiencies in ERL.

An algorithm for text analysis of consumer feedback was developed in study (Soriano and Palaoag 2018) using machine learning techniques like topic modelling. This went on to detail the methods used throughout the text to perform mining to extract meaningful data from the customer survey feedback of Bicol State College of Applied Sciences and Technology (BICAST), one of the SUCs in the Philippines. Additionally, automatic text synthesis and theme extract from the information sources based on texts were completed. The major issues that emerged from the input were noted. Both inputs for policymaking and helpful insights for management analysis are provided by this information.

The study of the author (Shambour 2021) suggested a machine learning-based approach for multi-criteria recommender systems. An algorithm describes the irregular and complicated user-item correlations. To improve the sources and learn the delicate user-item connections, an extensive auto encoder-based multicriteria suggestion algorithm was generated. It enables the coding of more combined definitions as data representations in the more advanced layers, in addition to the multicriteria preferences of customers.

The study (Almuqren et al. 2019) uses an analysis of sentiment which utilizes a corpus of Arabic tweets, to determine customer happiness for Saudi telecom organizations. According to the study’s findings, the bidirectional-GRU with method of attention performed better in the telecommunications sector and enabled very accurate customer satisfaction detection in that area.

The study (Ligiarta and Ruldeviyani 2022) used Support Vector Machine (SVM) model established to predict the sentiment of the obtained data, which might be favourable or adverse. According to the model training results, this model has an accuracy rate of 92.5%. Using sentiment analysis on BCA Mobile, Livin’ by Mandiri, BRI Mobile (Brimo), and BNI Mobile, the study examines consumer happiness.

The purpose of the research (Nwachukwu 2023) was to analyse the impact of artificial intelligence (AI) advertisements on consumer satisfaction. When analyzing the data, the Pearson Moment regression coefficient and the SPSS spreadsheet were used. The important rising correlation parameters with subsequent orders and customer referrals prove that artificial customization has a considerable optimistic link with customer happiness.

The study of the research (Aslam et al. 2019) investigates the relationship between customer satisfaction and automated teller machine (ATM) service quality. Investigative factor analysis (EFA), confirmatory factor analysis (CFA), and structural equation modelling (SEM) were implemented to discover the aspects of the quality of ATM services and their associations with customer satisfaction and regard for others. In this research outlines the problems that need to be fixed to elevate the quality of ATM service and offers specific recommendations for bank management to enhance customer satisfaction with ATMs.

They suggested a method for simulating consumer happiness using data from internet reviews in (Bi et al. 2019). The process begins by applying naive bayesian algorithm allocation to extract customer satisfaction dimensions (CSDs) from online reviews. An SVM was used to calculate the sentiment orientations of the obtained CSDs. (SVM). Then, a collective neural network-based typical (ENNM) was presented to quantify the impacts of consumer opinions toward various CSDs on customer happiness, considering complicated interactions between various CSDs and customer satisfaction. Finally, a case study was carried out to demonstrate the viability of the suggested approach.

Authors of study (Zhou et al. 2020) analyzed many online user-generated assessments of products inside a product ecosystem to offer a machine-learning method for customer requirements analysis. They implemented a rule-based appraisal technique to forecast the reviews’ sentiment and sentiment intensity ratings. Finally, focusing on the dissatisfaction-satisfaction ratio from the perspective analysis, they categorized customer needs relating to numerous themes using an analytical Kano model.

Aim: In this paper, they offer a topic modeling approach that uses Naive Bayesian (NB) algorithms to summarize text and recover concepts from this unprocessed information. The text mining process used to sift through a large amount of text-based data and find pertinent information is also covered.

2. Materials and Methods

To examine client problems gleaned from their text-based comments and recommendations from the feedback form, an analytical approach was used in this study. Additionally, this study made use of text-mining techniques. Data mining includes text mining, which aims to automatically uncover intriguing and complex configurations from unorganized textual
data. Figure 1 depicts the text extraction procedure flow, which starts with data collection from different sources, followed by pre-processing, text mining method applied on the cleaned data analysis, evaluation, and interpretation of the results.

![Text data mining process flow](https://www.malque.pub/ojs/index.php/msj)

**Figure 1** Text data mining process flow.

2.1. **Extraction of Text**

The informational resource of the state college's customer satisfaction assessment system was used to obtain data for the 2021–2022 academic year for this research. The school's various constituents, which include students, staff, and customers from the outside, are conducting the purpose of this study. Their criticism, improvement requests, and opinions were removed from the database.

2.2. **Pre-processing of Text**

Pre-processing is a crucial and significant step in the text mining process previously using any text mining techniques. In this step, the corpus was cleaned to reduce, if not eliminate, imposed terms and defects in consistency present in the text. For the purpose of creating a usable domain model, data cleansing is vitally essential. The content was normalized by removing words that ended in punctuation, preceding and utilizing trailing whitespaces, and other formatting elements. The text was also changed to be lowercase, subsequent whitespaces were compressed, and stemmed words were added.

Data purification was first carried out manually. Misspelled words were fixed, joyful expression indicators were eliminated, "ok" and "ty" phrases were dropped, and terms from the regional dialect were transformed into formal English. The cleansing procedure was then finished using OpenRefine, an open-source program for data purification and layouts modification.

2.3. **Text-mining Algorithm**

Naive Bayes is a probabilistic classification algorithm that is commonly used in text mining for tasks such as document classification, sentiment analysis, spam filtering, and more. The algorithm is based on Bayes’ theorem, which is a way of calculating conditional probabilities. In text mining, Naive Bayes works by analysing the frequency of words in a document or a set of documents and using this information to predict the class of a new document.

The basic theory behind the technique is that documents comprise various hidden themes, with the possibility of the characters given to each subject. These subject probabilities provide a succinct summary of a text.

Naive Bayes is a prominent text mining technique since it is quite simple and effective, particularly when working with huge datasets. However, in certain circumstances, its exactitude may be constrained by its feature independence assumption. Additionally, the algorithm's accuracy depends on how well the feature extraction procedure was done.

3. **Discussion and Result**

3.1. **Techniques for Data Processing**

The statistics were obtained for the academic year 2021–2022 from the state college customer satisfaction survey (CSS) database.

The 500 entries made up the majority of the dataset. Pre-processing procedures were applied to the extracted data to eliminate noise and inefficiencies. The elimination of the preceding and following whitespaces, punctuation, symbols,
unimportant but often reaffirmed words, and stop words is one of them. Other methods include grouping terms and word stemming. The dataset is now ready for further processing. After maintenance, the corpus was reduced to 475 records.

3.2. Topic Models

To produce topics or themes that appeared in the customer feedback, the Mallet with naive bayesian algorithm was utilized.

The original values for the parameters were 1500, 20, and 20, respectively, for the number of iterations, themes, and top words. It was discovered after studying the subject composition breakdown that similar terms fell under several themes. This indicates that the context is too wide and should be made more focused by fewer themes and more top words. As a result, the top 8 words were increased, and the number of themes was reduced from 5 to 15. In Figure 2, the list of produced subjects is shown.

![Figure 2 list of generated topics.](image1)

Each line in the picture has three elements: the subject number, the topic weight, and the phrases that follow, which are the words that are related to that topic and are used the most often.

The initial component on each line of the figure is the subject number, the subsequent component is the topic's weight, and the phrases that execute are the terms that are used the most often and come under that topic.

3.3. Procedure for Validation and Assessment

Authorities in the relevant fields are consulted as part of the validation process for the created subject models. The Director of Student Development Services, the Institutional Planning Officer, and the Vice President for the College of Education made up the group of specialists. They were asked to go through the vocabulary associated with each subject and determine what each set's overarching theme was. Table 1 summarizes three themes from the dataset that were randomly selected, along with the top terms and their weight.

The professionals have categorized the subject models taken from the comments after carefully examine groupings of words from each topic. Figure 3 contains a list of the recognized key topics. The experts have categorized the subject models taken from the comments after carefully examining groupings of words from each topic. Figure 3 contains a list of the recognized key topics.

![Figure 3 Labeled Topic Models](image2)

On the contrary, the evaluation procedure involved using the trained algorithm to deduce new collections of documents and manually analyzing the topic assignments when the model properly predicted the subject. The outcome of the prediction conduct applied to the newly acquired document work is shown in figure 4.

https://www.malque.pub/ojs/index.php/msj
The initial document work, mostly defined by the office setting, demonstrates that the model successfully anticipated the themes of future documents. Additionally, the third and fourth documents discuss customer service excellence.

### 3.4. Primary Client Concerns

After analyzing the feedback from the database of the customer satisfaction survey system, the top customer complaints were identified and arranged by composition, as shown in Table 2.

<table>
<thead>
<tr>
<th># doc</th>
<th>New Text Data</th>
<th>Customer Feedback</th>
<th>Office Environment</th>
<th>Customer Service Quality</th>
<th>Office Staffing</th>
<th>Security Personnel and Student Assistant</th>
<th>Attitude</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>maintain office cleanliness</td>
<td>0%</td>
<td>98%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>lack of employee</td>
<td>0%</td>
<td>26%</td>
<td>0%</td>
<td>77%</td>
<td>0%</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>improve building facilities</td>
<td>0%</td>
<td>73%</td>
<td>0%</td>
<td>9%</td>
<td>0%</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>continue what you are doing, excellent service</td>
<td>0%</td>
<td>0%</td>
<td>99%</td>
<td>0%</td>
<td>0%</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>very accommodating</td>
<td>0%</td>
<td>2%</td>
<td>86%</td>
<td>14%</td>
<td>0%</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>some monitor are not working well</td>
<td>5%</td>
<td>15%</td>
<td>3%</td>
<td>9%</td>
<td>4%</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>install air-condition</td>
<td>0%</td>
<td>93%</td>
<td>0%</td>
<td>3%</td>
<td>0%</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Add staff</td>
<td>0%</td>
<td>2%</td>
<td>0%</td>
<td>98%</td>
<td>0%</td>
<td></td>
</tr>
</tbody>
</table>

Figure 4: The distribution of topics in a new document using probability.

Table 1: Top 5 Words from 3 Topics.

<table>
<thead>
<tr>
<th>Topic 3</th>
<th>Topic 1</th>
<th>Topic 0</th>
</tr>
</thead>
<tbody>
<tr>
<td>regular</td>
<td>6.38</td>
<td>4.32</td>
</tr>
<tr>
<td>staff</td>
<td>59.45</td>
<td>31.27</td>
</tr>
<tr>
<td>personnel</td>
<td>32.18</td>
<td>41.13</td>
</tr>
<tr>
<td>approachable</td>
<td>41.23</td>
<td>19.85</td>
</tr>
<tr>
<td>employee</td>
<td>11.19</td>
<td>8.15</td>
</tr>
</tbody>
</table>

Table 2: Top 5 Customer Concerns.

<table>
<thead>
<tr>
<th>Rank</th>
<th>Main Topic/Theme</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Office IT System</td>
<td>0.789</td>
</tr>
<tr>
<td>2</td>
<td>Office Environment</td>
<td>0.547</td>
</tr>
<tr>
<td>3</td>
<td>Office Staffing</td>
<td>0.836</td>
</tr>
<tr>
<td>4</td>
<td>Customer Service Quality</td>
<td>0.213</td>
</tr>
<tr>
<td>5</td>
<td>Customer Feedback System</td>
<td>0.452</td>
</tr>
</tbody>
</table>

According to the outcome, respondents’ top worry is the level of customer service. Given that the study is focused on customer pleasure, this is to be anticipated. Additional office-related issues, including employees, the atmosphere, a mechanism for collecting client feedback, and an IT system, were subsequently found. Additionally, difficulties with the attitude of the security staff and student assistants, as well as library administration and management, are mentioned in the feedback.

Following that, these were given to the administrators as inputs for decision- and policy-making. This is because one of the recommendations in the ISO audit findings for the school was the reorganization and clustering of customer feedback to see their overall perception and to prioritize and concentrate on the necessary actions.

### 5. Conclusion

In this work, the researcher described how an algorithm based on machine learning was implemented to assess and identify the top issues from customer feedback that was left unstructured. According to the research, topic modelling using a Naïve Bayesian algorithm successfully located the main topics in the client assessments of the state college. The findings, which are thought to be helpful in making judgments and establishing school rules, were presented to school leaders. It also addressed the remarks and suggestions regarding the state college’s use of consumer feedback in the ISO audit results. To acquire a comprehensive understanding of the customer perspectives and issues, the researcher advises using the same...
methodology for additional text-based surveys that the institution conducts. The school could also consider using an online survey method to get more people to provide honest feedback.

Ethical considerations
Not applicable.

Declaration of interest
The authors declare no conflicts of interest.

Funding
This research did not receive any financial support.

Reference