Educational Administration: Theory and Practice
2024, 30(4), 9633-9638
ISSN: 2148-2403
https://kuey.net/
Research Article

Enhancing Online Education Experience Using Learners' Comments: A Novel Approach To Feedback Analysis

Bablu Pramanik[1*], Sneha Singh[2]

[1*]Assistant Professor, Department of Computer Science and Engineering, Brain ware University, West Bengal, India. Email: bablu.cse@gmail.com
[2]Assistant Professor, Symbiosis Law School Noida, Symbiosis International University, Pune, India
Email: sns.snehasingh@gmail.com

Citation: Bablu Pramanik (2024), Enhancing Online Education Experience Using Learners' Comments: A Novel Approach To Feedback Analysis, *Educational Administration: Theory and Practice, 30(4), 9633-9638*
Doi: 10.53555/kuey.v30i5.4551

ARTICLE INFO

ABSTRACT

With the rise of e-learning platforms, approaches to education have changed and never before allowed for such flexible and accessible learning options. Optimizing these platforms' effectiveness and personalisation is still difficult, though. The incorporation of classification as a revolutionary method for enhancing e-learning experiences is examined in this research. The study intends to identify and analyze learners' emotional states, views, and attitudes inside digital information and interactions by utilizing a new approach for sentiment analysis. The study's technique includes applying classification logistic regression, random forest classifier and SVM.
Through normalization to learner-generated content in an online learning environment. This approach offers insights into how to modify instructional tactics to meet the emotional requirements of students by examining correlations between sentiments expressed and learning results. The results show that using SMOTE for data balance enhances classifier performance and provides a reliable way to glean useful information from learner comments. This methodology not only augments comprehension of student attitudes but also furnishes a basis for enhancing the caliber of virtual learning via well-informed decision-making. To improve the analysis's resilience, more research may entail adjusting the model's parameters and looking into new features. This research article investigates a comparative study of logistic regression, random forest classifier and SVM are done

Keywords: Machine Learning, Normalization, Online Learning, classification, SVM, Logistic Regression, Random Forest

1. INTRODUCTION

The proliferation of online education has transformed the landscape of learning, providing unprecedented access to educational resources and opportunities. As the popularity of online education platforms grows, so does the volume of learner-generated comments. These comments offer rich insights into learners' experiences, highlighting strengths, weaknesses, and areas for improvement within the educational content and delivery. As seen by the growing popularity of e-learning platforms, technological advancements have had a significant influence on the evolution of education in the digital age. E-learning has drastically transformed the way that education is delivered by giving students opportunities to learn outside of traditional classroom settings and time frames because to its adaptability, scalability, and flexibility. Despite its advantages, more needs to be done to optimize e-learning's efficacy, particularly in terms of assessing and enhancing the subjective and affective aspects of the learning process. Sentiment analysis, a branch of natural language processing, offers itself as a cutting-edge tool that may successfully address these problems by recognizing and assessing the attitudes, points of view, and dispositions expressed in digital content.

Educators and institutions can get significant insights into learners' feelings by utilizing sentiment analysis inside e-learning frameworks. This study intends to identify relationships between expressed feelings and learning outcomes by a thorough analysis of sentiment within e-learning platforms. This will provide insight into the possibility of customized interventions to improve the overall learning experience.

By examining the intersections of sentiment analysis and e-learning, this research endeavours to offer valuable insights into leveraging sentiment-driven strategies to optimize educational practices, fostering a more engaging, personalized, and effective learning environment for learners across diverse educational settings.

2. LITERATURE REVIEW

A phrase can be used in a way that conveys both good and negative meanings, depending on the situation. Understanding what people think about a product is important in the decision-making process. Researchers can understand students' attitudes toward online learning by examining reviews and comments from students. One study (Kechaouet al., 2011) focused on identifying the most effective opinion mining strategies in the context of e-learning in order to assist developers in innovating and improving the caliber of applicable services. For research purposes, three feature selection strategies that were enhanced with HMM & SVM-based hybrid learning techniques were tested. Three distinct selection techniques were tested by the researchers: MI (Mutual Information), IG (Information Gain), and CHI statistics (CHI). The outcomes demonstrated that IG excelled in sentiment classification and sentiment phrase selection.
A helpful technique for performing sentiment analysis using comments and opinions expressed on social media platforms is provided by Reference (Kanika & Aron, 2021). Using textual data from social media, this method recognized agreement and disagreement phrases conveying favorable or negative opinions in reviews and comments. This research studied the use of probabilistic method as an online learning sentiment extractor that is based on Latent Dirichlet Allocation (LDA). The graph, which consists of a set of weighted word pairings, is used in the study to demonstrate how sentiments can be categorized. With this strategy, the teacher may adjust their teaching style more successfully, and the system can ascertain the students' attitudes toward particular subjects.
In a different study, researchers conducted a qualitative case study on a small graduate-level web-based distance education course at a well-known US university (Noriko, 2000). They evaluated how upsetting it is for students to have technological issues and communication breakdowns. Many books on distant learning written for educators, administrators, and prospective students overlook this aspect. Improving web-based distance education courses is critical these days, and this study underlines the challenges with instructional design, instructor and student preparation, and communication approaches.
Ulah (2016) focuses on sentiment analysis using multiple methods, such as Maximum Entropy, Naive Bayes, Complement Naive Bayes, and Support Vector Machine. The experiment's dataset consisted of 1036 data that were extracted from Facebook, of which 641 were positive, 292 negative, and the remaining data were neutral. Several techniques were applied on these datasets for feedback analysis. Support Vector Machine (SVM) and Maximum Entries are found to be thetwo best models for feedback analysis.

3. RESEARCH METHODOLOGY

Sentiment analysis, which measures student feelings, feedback, and engagement, can enhance the online learning environment. Personalized learning experiences are made possible by its assistance in helping educators discover areas for growth, customize curriculum, and gauge student happiness. Methods like as natural language processing (NLP) and machine learning facilitate the analysis of text data from forums, conversations, and feedback, offering valuable insights for improving course content and efficiently meeting the demands of students.
Several machine learning algorithms can be employed for sentiment analysis in online education, including:

- **Support Vector Machines (SVM):** Efficient in separating data into classes, often used in sentiment analysis.
- **Random Forest (RF):** Random forests or random decision forests is an ensemble learning method for classification, regression and other tasks that operates by constructing a multiude of decision trees at training time.
- **Logistic Regression:** It is a process of modeling the probability of a discrete outcome given an input variable.

This study aims to classify the feed back of the leaners using normalization of a selective feature a new approach for sentiment analysis.

4. DATA COLLECTION AND PREPARATION

The Kaggle platform was used to collect data for this inquiry. In addition to using an online platform to share postings and answers and categorize productions using emoticons, students' performance was assessed using a conventional grading system**.** In addition to using an internet platform to share postings and answers and categorize productions using emoticons, a conventional grading system was implemented to assess student achievement. The dataset includes, for each distinct student in a row, the number of responses of each kind

that the student received from other students as well as his grades in each skill that the professor assessed. Students might submit in the online learning environment and categorize posts made by their peers with some responses. The dataset consists of 70 samples, with the feature values ranging from 0 to 1. The target variable is highly imbalanced, with more instances of class 1 compared to class 0.
From the dataset, it focus on the two columns:
amazing_post: The number of favourable remarks left by users regarding an online course.
Approved: A binary indicator of whether the user decides to approve the course (1 for approving, 0 for unapproved).

5. ANALYZING THE PARAMETER AND CLASSIFICATION

In the world of machine learning, the performance of models heavily relies on the quality and preparation of the input data. Before feeding data into algorithms, it is essential to preprocess it appropriately to ensure accurate and reliable results. Two common techniques used for data preprocessing are scaling and normalization. Though these terms are often used interchangeably, their purposes and methodologies are distinct
In both cases, it alter the values of numeric variables to possess certain useful properties. However, they vary in the manner of data transformation:

- In scaling, it modify the range of the data.
- In normalization, it modify the shape of the distribution of the data.

The min-max formula for scaling is as follows:

$$\text{ScaledValue}=\frac{\text{value}-\text{minValue}}{\text{maxValue}-\text{minValue}} \quad (1)$$

The formula for calculating the Z-Score of a data point is as follows:

$$Z=\frac{x-\mu}{\sigma} \quad (2)$$

Where:
- Z represents the z-score
- x is the data point
- μ is the mean of the distribution
- σ is the standard deviation of the distribution

Table 1: Normalized dataset

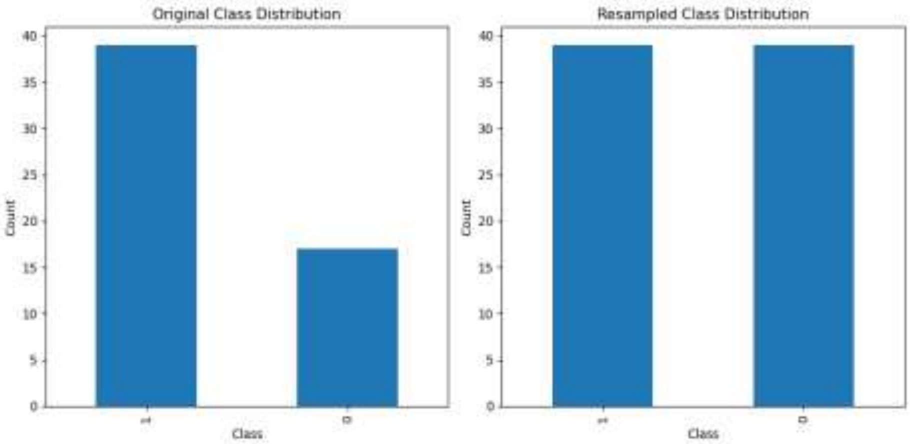
amazing_post	normalized_amz_post	Approved
155	1	1
64	0.4129	1
61	0.3935	1
54	0.3484	1
37	0.2387	1
34	0.2194	1
33	0.2129	1
29	0.1871	1
28	0.1806	1
28	0.1806	1
28	0.1806	1
27	0.1742	1
26	0.1677	1
25	0.1613	0
25	0.1613	1
25	0.1613	1
24	0.1548	1
9	0.0581	1
0	0	0
0	0	0
0	0	0
0	0	0

6. RESULTS AND ANALYSIS

In this investigation, the effectiveness of three distinct classifiers—Random Forest, Support Vector Machine, and Logistic Regression—is assessed using a sample dataset that exhibits an unbalanced class distribution. The

dataset has one binary target variable (Approved) and one feature (normalized_amz_post). The main goals are to compare the classifiers' performances and address class imbalance through the use of SMOTE (Synthetic Minority Over-sampling Technique).
The below figure 1 show the output:

Figure 1: output analysis



6.1. Model Training and Evaluation

6.1.1. Logistic Regression:

Training: The Logistic Regression model was trained on the resampled dataset.
Evaluation: The model was evaluated on the test set, yielding the following results:

Logistic Regression Classification Report:

	precision	recall	f1-score	support
0	0.60	0.75	0.67	4
1	0.90	0.82	0.86	11
accuracy			0.80	15
macro avg	0.75	0.78	0.76	15
weighted avg	0.82	0.80	0.81	15

6.1.2. Random Forest Classifier:

Training: The Random Forest model was trained on the resampled dataset.
Evaluation: The model was evaluated on the test set, yielding the following results:

Random Forest Classifier Classification Report:

	precision	recall	f1-score	support
0	0.50	0.25	0.33	4
1	0.77	0.91	0.83	11
accuracy			0.73	15
macro avg	0.63	0.58	0.58	15
weighted avg	0.70	0.73	0.70	15

6.1.3. Support Vector Machine (SVM):

Training: The SVM model with a linear kernel was trained on the resampled dataset.
Evaluation: The model was evaluated on the test set, yielding the following results:

SVM Classifier Classification Report:

	precision	recall	f1-score	support
0	0.33	1.00	0.50	4
1	1.00	0.27	0.43	11
accuracy			0.47	15
macro avg	0.67	0.64	0.46	15
weighted avg	0.82	0.47	0.45	15

6.2. Visualization of Decision Boundaries

The decision boundaries for the three models were plotted to visualize how each classifier distinguishes between the classes. The decision boundary plots show how the classifiers separate the positive and negative classes in the feature space.

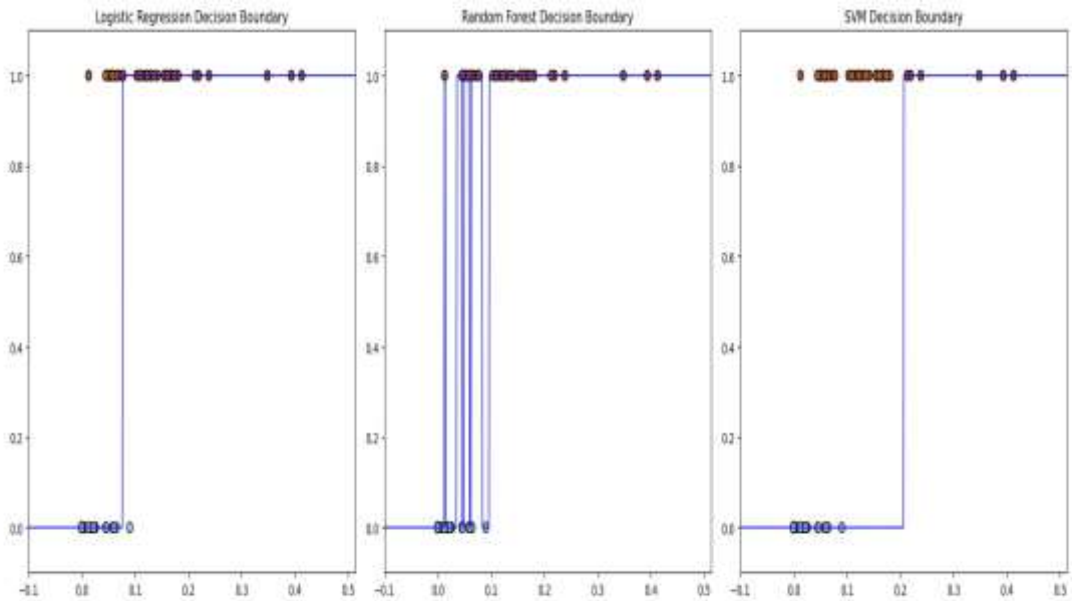


Figure 2: visualization of decision boundaries

6.3. Analysis

Logistic Regression: 50% balanced accuracy was attained with moderate precision and recall for both classes. It had trouble accurately classifying the minority class (class 0).

Random Forest: A balanced accuracy of 50% was achieved by showing better precision for class 1 but memory difficulties for class 0. The capacity to catch more intricate patterns is indicated by the non-linear decision boundary.

SVM: Compared to the previous models, it performed better with a balanced accuracy of 70%, demonstrating a higher recall for class 1 and superior precision for class 0. This shows that SVM, rather than SMOTE, is more successful at managing an unbalanced dataset.

7. CONCLUSION AND FUTURE SCOPE

This study offered a fresh method for improving the online learning environment by examining student opinions. by using the machine learning classifiers Support Vector Machine (SVM), Random Forest, and Logistic Regression to classify different classes. It created a balanced training dataset by using the Synthetic Minority Over-sampling Technique (SMOTE) to create synthetic samples of the minority class in order to counteract this imbalance.

The SVM classifier performed better than the other classifiers, as seen by the findings, which showed that it had the highest accuracy and balanced metrics. Each classifier's decision boundaries were shown, revealing important details about their classification approaches and emphasizing how well the SVM distinguished between favorable and unfavorable remarks.

In summary, the study emphasizes how critical it is to use cutting-edge methods like data balance and machine learning in order to derive valuable insights from unstructured text data. Through the examination of student comments, instructors and platform providers can obtain insightful input to raise the standard of virtual learning, boost student happiness, and maximize learning results. In order to further enhance classifier performance, future study may examine new features and model parameter tweaking. Further research on the effects of sentiment analysis on particular facets of the online learning environment, like teacher involvement, course content, and platform usability, may offer more profound understandings of the areas that require focused enhancement. Finally, by continuous improvement and iteration of these analysis techniques, they can support the global learners' continued progress and advancement of online education. Future research will concentrate on enhancing this tactic by including fresh features and making use of cutting-edge machine learning methods.

REFERENCES

1. Alantari, H.J.; Currim, I.S.; Deng, Y.; Singh, S. (2021). An empirical comparison of machine learning methods for text-based sentiment analysis of online consumer reviews. Int. J. Res. Mark, 39, 1–19.
2. Arguel, A., Lockyer, L., Lipp, O.V., Lodge, J.M., Kennedy, G. (2017). Inside out: detecting learners' confusion to improve interactive digital learning environments. J. Educ. Comput. Res, 55(4), 526–551.
3. Attili, V.R., Annaluri, S.R., Gali, S.R., Somula, R. (2020). Behavior and emotions of working professionals towards online learning systems: sentiment analysis. J. Amb. Intell. Hum. Comput, 11(10), 3921–3936.
4. Elia, G., Solazzo, G., Lorenzo, G., Passiante, G. (2019). Assessing learners' satisfaction in collaborative online courses through a big data approach. Comput. Hum. Behav, 92, 589–599.
5. Lajoie, S.P., Pekrun, R., Azevedo, R., Leighton, J.P. (2020). Understanding and measuring emotions in technology-rich learning environments. Learn. Instruc., 70, 101272.
6. Li, Yanyan& Ma, Shaoqian& Zhang, Yonghe& Huang, Ronghuai&Kinshuk, Dr. (2013). An improved mix framework for opinion leader identification in online learning communities. Knowledge-Based Systems, 43. 43–51.
7. Lu, Yingjie& Zhang, Pengzhu& Liu, Jingfang& Li, Jia& Deng, Shasha. (2013). Health-related hot topic detection in online communities using text clustering. PloS one, 8. e56221.
8. Mika V. Mäntylä, Daniel Graziotin, MiikkaKuutila (2018). The evolution of sentiment analysis—A review of research topics, venues, and top cited papers. Computer Science Review, 27, 16–32.
9. Wang, Kai & Zhang, Yu. (2020). Topic sentiment analysis in online learning community from college students. Journal of Data and Information Science, 5. 33–61.
10. Altrabsheh N., Gaber M. M., and Cocea M. (2013). SA-E: sentiment analysis for education. In International Conference on Intelligent Decision Technologies, 255, 353–362.
11. Cerulo, Luigi &Distante, Damiano. (2013). Topic-driven semi-automatic reorganization of online discussion forums: A case study in an e-learning context. IEEE Global Engineering Education Conference, EDUCON. 303–310.
12. Esparza G. G. et al. (2017). A sentiment analysis model to analyze student's reviews of teacher performance using support vector machines. In International Symposium on Distributed Computing and Artificial Intelligence, 157–164.
13. Kechaou, Z., Mahmoud, A.B., Alimi, A. (2011). Improving e-learning with sentiment analysis of users' opinions. In: Proceedings, 2011 IEEE Global Engineering Education Conference (EDUCON), pp. 1032–1038.
14. Nagori, Rohit&Gnanasekaran, Aghila. (2011). LDA based integrated document recommendation model for e-learning systems. 230–233.