BILINGUAL PUBLISHING GROUP

Forum for Linguistic Studies

https://journals.bilpubgroup.com/index.php/fls

ARTICLE

Automated University Complaint Management System by Leveraging Machine Learning and Natural Language Processing for Enhanced Efficiency and Accuracy

Hanad Mohamud Mohamed * **, *Abdishakur Hassan Mohamed*** **, *Sabrin Omar Hashi*** **, *Hassan Nur Hassan*** **,** ***Khadar Abdirahman Barre*** **, *Bashir Abdinur Ahmed***

Faculty of Computer & IT, Jamhuriya University of Science and Technology, Mogadishu, Somalia

ABSTRACT

The goal of this research is to create an automated system that works with machine learning (ML) and natural language processing (NLP) to automate university complaint management. Students get dissatisfied when traditional complaint handling techniques, such physical suggestion boxes, are ineffective and prone to delays. Designing and implementing a system that automates the submission, classification, and analysis of student complaints especially those made in Somali is the aim of this project. The suggested approach greatly lessens the manual workload of university administrators by classifying issues into Academic, Finance, and Equipment using a machine learning model trained on complaint data. The system has an administrative dashboard for tracking and handling complaints, as well as an easy-to-use interface for filing complaints. The primary results show that the system improves the accuracy and efficiency of resolving complaints, which results in quicker resolution times and pleased students. Proactive decision-making is made possible by the system's integration of data analytics, which also offers insightful information on persistent problems. According to the project's findings, automated complaint handling can greatly enhance the entire university experience by creating a more accommodating and student-focused atmosphere.

Keywords: Automated Complaint Management; Machine Learning; Natural Language Processing; University Administration; Student Satisfaction

*CORRESPONDING AUTHOR:

Hanad Mohamud Mohamed, Faculty of Computer & IT, Jamhuriya University of Science and Technology, Mogadishu, Somalia; Email: hanad@just.edu.so

ARTICLE INFO

Received: 6 December 2024 | Revised: 12 January 2025 | Accepted: 13 January 2025 | Published Online: 5 February 2025
DOI: https://doi.org/10.30564/fls.v7i2.8162

CITATION

Mohamed, H.M., Mohamed, A.H., Hashi, S.O., et al., 2025. Automated University Complaint Management System by Leveraging Machine Learning and Natural Language Processing for Enhanced Efficiency and Accuracy. Forum for Linguistic Studies. 7(2): 373–354.
DOI: https://doi.org/10.30564/fls.v7i2.8162

1. Introduction

The effectiveness of student complaint management helps ensure a comfortable atmosphere to the university users, motivates learners, and upholds the reputation of the organization. Traditionally, universities have employed manual systems such as suggesting boxes and in person complaint submission to cater to student issues. These methods have proven effective in the past; however, their importance is diminishing as they are at the moment unable to sufficiently attend to the growing variety and volume of student complaints.

These manual methods frequently lack the efficiency and scalability required for successful complaint classification and resolution. Consequently, delays in response times, ineffective complaint categorisation, and administrative bottlenecks frequently occur, resulting in student displeasure and a diminished trust in the institution's capacity to resolve their issues. The significance of cultivating effective systems for institutional administration is reinforced by research on distributed leadership and digital innovation, which underscores the need for strong frameworks to navigate operational difficulties in educational settings [1].

In addition, the challenges associated with complaint handling are not unique to the university environment but are also observable in other educational institutions. It has been noted that students' tendency to overreport can undermine the effectiveness of the feedback tool, which calls for automated and unbiased systems [2]. These challenges are solved by the implementation of ML and NLP in the systems for management of the students' complaints, as these tools allow for automation of the classification process and increase the accuracy and speed of unresolved complaints [3].

In order to address these challenges, the adoption of modern technology such as machine learning (ML) and natural language processing (NLP) has emerged as a practical solution. Streamlining this process through automation and efficient categorization of complaints will enable universities to address these issues quickly, reducing administrative strains in the process. Research in parallel areas, such as bibliometric studies of professional learning and mentorship, suggest that systematic frameworks can greatly enhance institutional processes and outcomes.

This study presents a new idea of developing automated university complaint management system through machine learning and natural language processing techniques. The proposed system addresses the complexities that arise in addressing complaints in Somali. The key objectives of this research are the optimization of efficiency in and accuracy of the processes of managing complaints, the elimination of deficiencies of conventional systems and the demonstration of the feasibility of deployment such systems in low-resource language contexts.

2. Literature Review

According to [2] was focused on an individual-level classification of aggregated complaints submitted to a Portuguese governmental institution. The researchers explored several machine learning algorithms, including k-NN, SVM, Naive Bayes, and deep learning approaches (e.g., BERT). They addressed problems related to small and unbalanced data sets by incorporating classes and generating new instances by translating into various languages. Over the research determined SVM and BERT-based models were outperformed by alternative methods, specifically in managing classes with a limiting amount of instances.

Different analysis was performed on the application of sentiment analysis for educating students in online postgraduate programs using NLP technology. While analyzing social media content of students enrolled in a course, the Google Cloud Natural Language API was used, helping to define feelings and opinions about the course content. The work demonstrated that sentiments as an indicator can be determined at a higher level of understanding regarding a student when combined with online classroom services. While the conventional sentiment analysis in education rests solely on the end of course assessments, this one is focused on the evolution of student emotions over time [4].

This research examined sentiment analysis in digital education using machine learning and natural language processing techniques. This research focused on understanding student attitudes in the context of digital education. The project aimed to categorize student input and identify recurring patterns in their learning journeys by employing diverse NLP techniques and machine learning algorithms. The results emphasized the effectiveness of utilizing machine learning for real-time sentiment analysis, providing educators with valuable insights to improve course content and delivery [5].

Investigated the use of unsupervised methods for sentiment classification in restaurant reviews. The study categorized reviews into food, service, price, ambiance, and others, using co-occurrence and spreading activation methods as well. This research accurately demonstrates the usefulness of unsupervised learning in environments with little labelled data, for example, in case of large and heterogeneous data [6].

Explored the utilization of the Google Natural Language Processing API for conducting sentiment analysis with regards to students' social engagement on a postgraduate course. Their study analyzed 300 posts on social media to determine the patterns of students' sentiments when taking particular modules within an online course. The conclusions showed considerable variation of emotions within the time frame indicating useful information regarding student's feelings and their involvement in the course. The capabilities of sentiment analysis software to assist people with autisms in interpreting thoughts embedded in written text. Through design science as well as thematic analysis, they gathered information from autistic individuals in order to determine functional and design requirements. Their results find out a glaring need for intuitive and flexible designs coupled with sophisticated analysis in order to minimize chances of distorting one's emotions [7].

According to [8] created an online complaint management system for the English Study Program at Victory University. Manual and paperwork procedures were completely replaced by the efficient digital system that was created to manage effectively the students' complaints. This study demonstrates the real benefits that the complaint management systems offer towards the integration of more sophisticated systems which are such as advanced analytical ones.

Ref. [9] presented a thorough analysis of approaches for performing sentiment analysis and emotion classification in written texts. It was shown that various machine learning techniques have the capability to mine a wide range of emotions and sentiments from large databases. Much focus was also placed onto issues such as the need of large cover labelled databases and the complexity of the automation of the interpretation of emotions for the improvement of complaint management systems. This study [8] investigated the relationship between students' health issues and their perceived stress levels at the institution. The analysis highlighted the need to incorporate health issues into the context of overall care for students. The results explain greatly the strained need for complaint management system in times of the health and stress issues towards students, thus are ordered to have features such as them to understand and solve the exact issues.

This research [10] analyzed the complaint behaviors of international students and determined shared phenomena underlying the process of their complaints and their resolution. The study noted the importance of the presence of a culturally sensitive and accessible complaint management system that is capable of serving the needs of international students. This suggests that there is a need for the development of systems that are not only effective but are also more inclusive and contextually sensitive to culture. This article [11] evaluated the uses of sentiment analysis in a digital form of education that is coupled with machine learning and natural language processing in categorizing student input and in finding patterns in their learning experience. The research demonstrated the ability of these techniques to enhance understanding of students' feelings, attitudes and engagement with these systems which is useful for the development of advanced complaint systems.

Ref. [12] investigated how sentiment analysis may be used as an aid for emotional perspective taking in written language, in particular for people with autism. Their research addressed the application of sentiment analysis techniques as supportive devices for people with disabilities in understanding written text, thus helping them to interact better in communication. This study is pertinent to our research as it underscores the significance of customized NLP applications that address certain user requirements, such as students with communication difficulties.

According to [13] Exploration progressed in techniques of emotion detection and sentiment analysis. The authors explored the tiers of sentiment analysis (document, sentence, aspect) and finessed the problem of distinguishing specific emotions. Their study demonstrates the valuable role in opinion mining and human-computer interaction of combining emotions detection with natural language processing which is increasingly important. This [14] Studied stress perception in relation to health problems in university students in the UK and Egypt. The results showed that psychiatric symptoms were the primary stress factor and therefore interventions need to be very specific to tackle both the physical and the psychological aspects in different educational contexts.

This Examined the categorization of public administration complaints through machine learning techniques. They analyzed methodologies for addressing imbalanced datasets, including data augmentation via translation and the amalgamation of smaller classes. Findings demonstrated that SVM and BERT-based models provided enhanced performance, especially for under-represented categories [15].

Inspected higher education students' perceptions of emergency online learning during COVID-19 in South Africa, Wales, and Hungary. They recognized substantial variations in learning settings, engagement, and digital preparedness attributable to economic and cultural inequities, emphasizing the necessity of adaptive educational techniques [1].

This study [16] Performed a comparative comparison of sentiment analysis techniques using deep learning. They assessed models such as CNN, RNN, and hybrid architectures, demonstrating their efficacy in managing sentiment polarity and domain-specific datasets. The research underscored the capability of deep learning to enhance precision in intricate NLP jobs. Also, this study [7] examined letters of complaint authored by Jordanian university students, emphasizing rhetorical structures and persuasive techniques. The research employed critical discourse analysis, classifying tactics into ethos, pathos, and logos. Research indicated that emotional appeals (pathos) predominated, shaped by the socio-cultural values inside Jordanian society. This study enhances comprehension of complaint communication within academic environments and its cultural foundations. Proposed an innovative CNN-LSTM architecture that integrates TF-IDF weighted GloVe word embeddings for the analysis of product reviews. The research shown enhanced predictive precision relative to conventional techniques, highlighting the significance of sophisticated embedding and neural network models in sentiment analysis [12].

3. Methodology

This study deals with the shortcomings of existing systems for the management of complaints in universities which of particular concern are related to the processing of complaints in the Somali language. Currently machine learning and natural language processing techniques often fail to perform well largely because of lack of language resources since traditional methods of dealing with a large volume and variety of complaints are also not sufficient.

3.1. Proposed Method

The suggested approach for automating the university complaint management system encompasses several essential stages: data collection, preprocessing, model training, complaint categorisation, and system installation. Every stage is essential to guarantee the precision, efficacy, and dependability of the system (**Figure 1**).

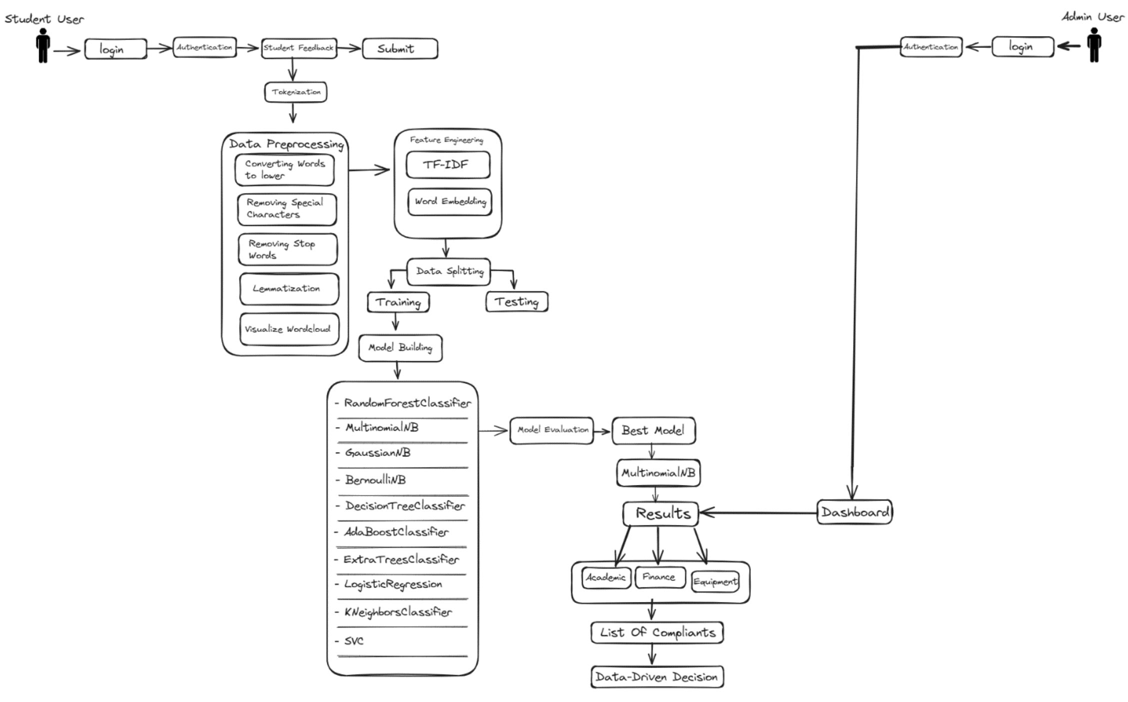


Figure 1. Proposed system.

The proposed solution is the development of a fully automated complaint management system based on machine learning and natural language processing capable of categorizing and prioritizing students' issues especially those pertaining to Somali complaints. This approach streamlines the management of normally cumbersome categorization and provides administrators of universities with useful trends in regard to the expectation and dealing with complaints.

3.2. Data Collection

The initial phase of the suggested methodology involves the collection of complaint data from students. Complaints are made through Google Forms, which offers an easy and accessible platform for students to express their problems. Data is gathered in real-time and retained in a MongoDB database, enabling scalability and adaptability in managing substantial amounts of unstructured data. This data constitutes the training set for the machine learning model.

3.3. Data Preprocessing

Data preprocessing is a critical step to prepare the raw complaint data for model training. The preprocessing pipeline includes the following steps:

- Text Cleaning: Elimination of unneeded characters, punctuation, and special symbols that do not enhance the sense of the complaint.
- Tokenization: The text is divided into discrete tokens (words) to enable analysis.
- Stopwords Removal: Frequently occurring words that lack substantial meaning (e.g., "kan", "kaas", "iyo") are eliminated
- Lemmatization: Words are transformed into their base or root form (e.g., "nin" becomes "niman") to maintain consistency in data representation.

Data Labeling and Annotation

Due to the significance of precise complaint categorization, the dataset was annotated by three domain experts, each specializing on one of the primary categories: Finance, Academic, and Equipment. The annotating procedure was as follows:

- Finance Expert Annotation: The university's Finance Officer, tasked with overseeing all financial activities, examined and commented on complaints pertaining to financial issues. These encompassed matters such as tuition prices, scholarships, and financial assistance. The Finance Officer meticulously categorised all complaints in this domain, demonstrating a high level of proficiency in comprehending and classifying financial grievances.
- Academic Expert Annotation: The Academic Officer, responsible for supervising academic programs and student achievement, was assigned to annotate complaints concerning academic difficulties. This encompassed apprehensions regarding course material, instructional quality, examination outcomes, and other academic-related issues.
- Equipment Expert Annotation: The Equipment Officer, tasked with maintaining and accessibility of university buildings and equipment, documented complaints concerning equipment and infrastructure. This included difficulties related to classroom space, laboratory equipment and other physical resources. Of all the officers, the Equipment Officer was trusted such a high level of skill which assured that all complaints about the equipment were captured and properly classed and labelled.

The experts were the exclusive sharding users and were given the relevant part of the dataset pertaining to their area of expertise. The experts, on their part, were given the responsibility of reviewing the complaints and assigning each one of them to the correct class which was appropriate within their area of jurisdiction. This approach improved the dataset considerably and ensured that the categories were well defined and relevant to the academic setting.

3.4. Feature Extraction

After preprocessing the text, features are extracted to ready the data for machine learning. The Term Frequency-Inverse Document Frequency (TF-IDF) method is utilized to transform textual input into numerical features suitable for model processing. TF-IDF measures the significance of a term within a document in relation to a set of documents (the corpus). This study selected TF-IDF for its simplicity and interpretability; however, future research may explore advanced feature extraction techniques, such as word embeddings (e.g., Word2Vec, GloVe, or BERT), to more effectively capture semantic nuances, especially for under-represented or ambiguous complaint categories.

3.5. Model Training

The essence of the suggested methodology is the machine learning model designed to classify complaints into established categories: Academic, Finance, and Equipment. The algorithm used in the model is a Support Vector Machine (SVM) using a linear kernel, selected for its performance in text categorization tasks. The model is trained on a labelled dataset, wherein each complaint is classified according to previous data. The SVM method aims to identify the hyperplane that optimally divides the complaint categories within the feature space.

3.6. Dataset Sources

The dataset used for training our machine learning model was derived from genuine complaints lodged by students at the university over the preceding three years. Although this yielded a comprehensive and varied dataset, enabling the model to address multiple complaint scenarios, we recognize a possible ethical issue about the utilization of this student data. It is imperative to confirm that ethical permission was secured and that data privacy and consent were adequately managed. Anonymization methods were employed on all complaints to eliminate any personally identifiable information. Moreover, clearance from the institutional review board (IRB) was secured before data collection, assuring compliance with ethical norms in the utilization of this data.

3.7. Initial Data Analysis

Prior to commencing model training, we examined the distribution of complaints across various categories. Approximately 36.81% of the grievances pertained to academic issues, 25.55% were associated with financial matters, and the remaining 37.64% concerned equipment.

3.8. Complaint Categorization

Upon completion of training, the model is deployed to automatically classify incoming complaints. Upon submission of a new complaint, it is subjected to preprocessing and feature extraction before to being analyzed by the SVM model, which forecasts the category. The system thereafter directs the complaint to the appropriate university department (Academic, Finance, or Equipment) for settlement.

3.9. System Implementation

The system is executed using a web-based and mobile application, leveraging a backend constructed with Node.js and Express, and a frontend created with contemporary web technologies, including React.js and Flutter for mobile application development. The complaint data and the trained model are preserved in a MongoDB database. A Flask API facilitates the connection between the machine learning model and the web application, providing smooth integration and real-time categorization of complaints.

This is the administrative dashboard (**Figure 2**), offering a summary of user-submitted complaints. The dashboard presents essential indicators, including the quantity of complaints classified by type: Academic Complaints, Finance Complaints, and Equipment Complaints. Visual representations, including bar charts and pie charts, provide a rapid and intuitive comprehension of complaint distribution.

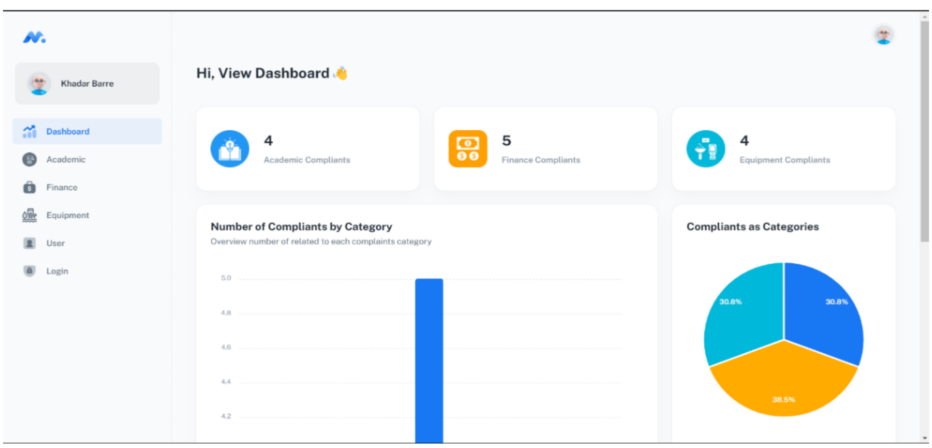


Figure 2. Dashboard page for Admins.

Figure 3 depicts specialized pages for managing complaints in particular areas. The Finance Complaints Page, restricted to finance administrators, offers a consolidated inventory of user-submitted financial grievances, facilitating efficient assessment and swift resolution. The Academic Complaints Page is intended for academic administrators to oversee complaints concerning teaching quality, attendance policies, and other educational issues.

This form (**Figure 4**) is intended for students who have received authorisation to access the system. Upon logging in, students can use this portal to express their complaints (**Figure 5**). The classification feature automatically categorises complaints according to their content, hence optimising the complaint management process for administrators. This tool employs natural language processing (NLP) and machine learning techniques to precisely categorise complaints into pertinent classifications, including Academic, Finance, or Equipment.

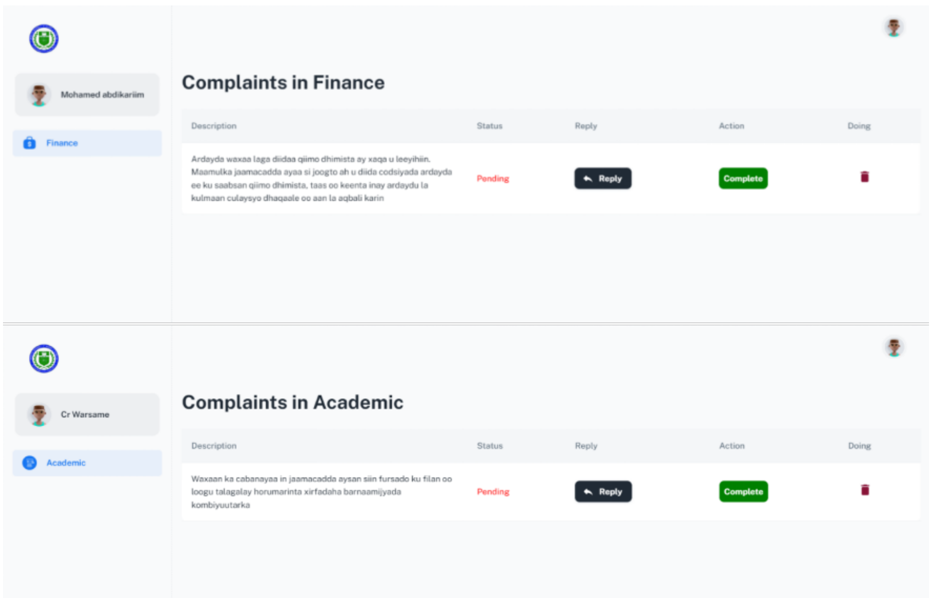


Figure 3. Complaints of Finance and Academic.

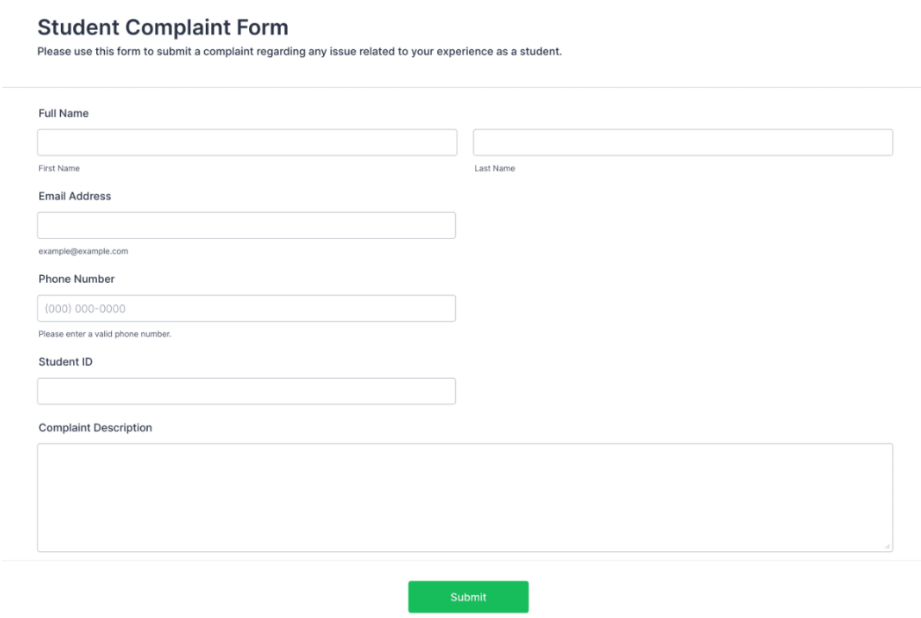
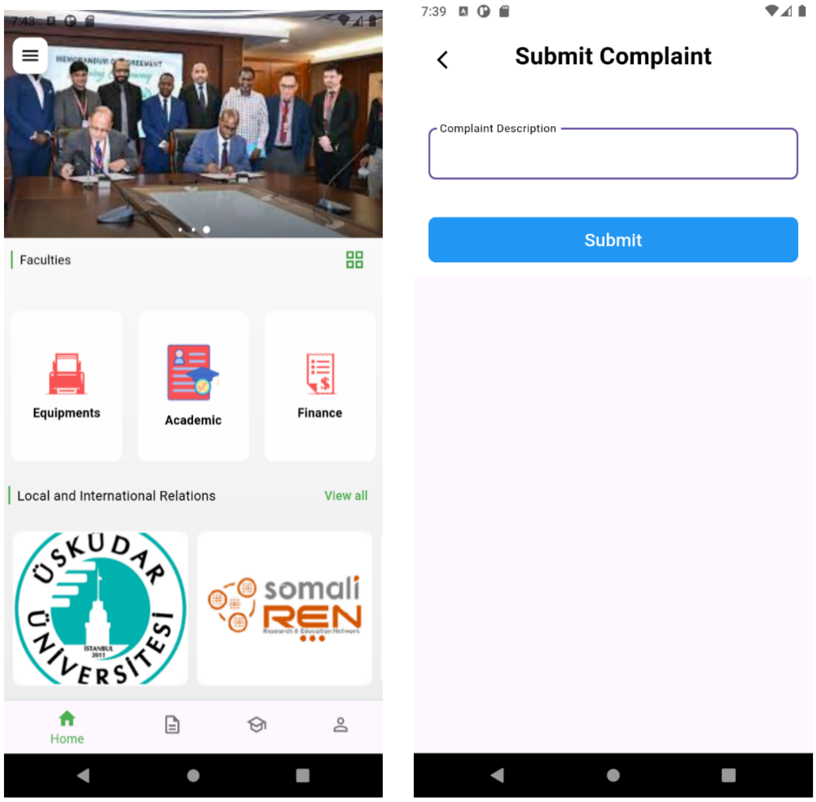


Figure 4. Complaint submission form for students.



(**a**) Mobile Home Screen (**b**) Complaint Submission Screen

Figure 5. (**a**) The home screen functions as the initial page upon user login. The app offers a summary of its functionalities, including expedited submission of new complaints, examination of complaint types (Academic, Finance, and Equipment), and navigation to various sections of the app. (**b**) This interface enables students to file new grievances. Users may provide information of their grievance and submit it for evaluation.

4. Results and Discussion

4.1. Hardware and Software Setup

To implement and test our automated complaint management system, we used a fairly powerful setup. The hardware included an Intel Core i7 processor, 16 GB of RAM, a 1TB SSD for storage, and an NVIDIA GeForce RTX 2070 GPU (**Table 1**).

Table 1. Software requirements.

Tools	Version	Description
Pandas	1.5.3	Pandas-profiling is used to display sample, correlation, and duplicated data easily
Microsoft Excel	2019	Used data preparation tasks during data are crawled from Facebook pages and sorting the gather data. Also, used to manage the annotation task.
Anaconda Navigator	2.3.1	Allows us to launch development applications and easily manage condominium packages, environments, and channels without the need to use command-line commands.
Python	3.8.9	Powerful programming language to develop a Machine learning application. It is also easy to process natural language.
Nltk	3.8.1	Natural Language Toolkit is a suite of libraries and programs for symbolic and statistical for English written in the Python programming language. This study uses it for data reading, manipulation, writing, and handling the data frame.
NumPy	1.24.2	Python library for topic modeling document indexing and similarity retrieval with large corpora. This Study uses it to handle text-to-number conversions Functionality and training and test data model is a comprehensive library for creating static, animated, and interactive visualizations in Python.
Matplotlib	3.3.2	Matplotlib makes easy things easy and hard things possible This study uses it for data and results visualization.

4.2. Important Findings

There are several significant findings which contribute to the improvement of automated systems for managing student complaints. Below is a detailed explanation of the key insights derived from the research:

1. **Data Diversity Helps**: A key finding from the study was the significance of data variety. The researchers used a dataset of actual student complaints, enhancing the robustness and adaptability of their algorithm. The model was subjected to a diverse array of complaint types by integrating grievances related to administrative issues, teacher performance, infrastructural concerns, and financial matters. This diversity enabled the system to comprehend the intricacies and specificities of various complaint kinds, resulting in enhanced adaptability when faced with new or previously unencountered complaint subjects. The incorporation of real-world data not only improved the model's precision but also guaranteed the system's efficacy across diverse real-world situations where the form of complaints may differ.
2. **Effective Feature Extraction**: The TF-IDF (Term Frequency-Inverse Document Frequency) technique for feature extraction significantly influenced the system's performance. TF-IDF enabled the model to discern the most salient words and phrases in the complaints by evaluating their frequency inside individual complaints and throughout the full dataset. This method enabled the system to comprehend the essential aspects of each complaint and categorise them appropriately. The application of TF-IDF allowed the model to identify significant traits while avoiding the influence of irrelevant or excessively common phrases. The system successfully extracted significant patterns and enhanced its classification accuracy, underscoring the critical role of effective feature engineering in natural language processing (NLP) applications.
3. **Category Overlap**: Notwithstanding the encouraging outcomes, the study revealed that the model occasionally encountered difficulties with category overlap, specifically between finance-related complaints and those pertaining to equipment. Complaints pertaining to financial matters and those concerning university equipment exhibited similar wording and underlying difficulties, resulting in the model occasionally misclassifying complaints into incorrect categories. This overlap indicated the necessity for additional refining in the model's categorization logic. It proposed that improved discrimination of categories may be attained by augmenting the training data, optimizing feature extraction methods, or employing more advanced algorithms adept at managing ambiguous or overlapping complaint types. Confronting this obstacle will enhance the precision and accuracy of the automated system, facilitating more precise categorization of complaints.
4. **Real-Time Performance**: The system's real-time performance was one of the most promising discoveries. The model efficiently categorized student concerns, illustrating its potential for practical application in academic environments. The capacity to address complaints in real time is crucial for sustaining the efficacy of complaint management systems, particularly in major colleges with a substantial influx of complaints. The capacity to swiftly categories complaints facilitate expedited response times from university personnel and guarantees timely attention to students' issues. The study demonstrated that an automated complaint management system may function both accurately and efficiently in practical applications, rendering it a significant asset for colleges aiming to optimize their complaint resolution procedures.

Evaluation Metrics for System Performance

Evaluation metrics are essential for assessing the performance of the automated complaint management system. These metrics typically include accuracy, precision, recall, and F1 score, which measure how well the system classifies complaints into the appropriate categories.

1. **Precision:** Precision measures the proportion of true positive predictions among all positive predictions made by the model. It is defined as:

   $$Precision = \frac{True\ Positives(TP)}{True\ Positives(TP)+False\ Positives(FP)}$$

   A high precision indicates that the system accurately classifies complaints into their respective categories with few false positive predictions.
2. **Recall:** Recall (also known as sensitivity or true positive rate) measures the proportion of true positive predictions out of all actual positive instances. It is defined as:

   $$Recall = \frac{True\ Positives(TP)}{True\ Positives(TP)+False\ Negatives(FN)}$$

A high recall indicates that the system correctly identifies a large number of complaints in each category, ensuring that minimal complaints are missed.

3. **F1-score:** The F1-score is the harmonic mean of precision and recall, providing a balance between the two, especially when the dataset is imbalanced. It is defined as:

$$F1 - score = 2x\frac{Precision\ x\ Recall}{Precision + Recall}$$

The F1-score is useful when balancing the need for both precision and recall, ensuring that the system performs well in identifying and classifying complaints while minimizing errors.

4.3. Comparison to Existing Approaches

In comparing our Automated University Complaint Management System to other cutting-edge methodologies, some significant differences and benefits become apparent. The primary distinction resides in the methodologies operated for classifying complaints. Many existing complaint management facilities in academic institutions still employ rule based classification or even worse keyword matching as easy solutions. These approaches are good for simple complaints but they fail in view of the diversity and complexity of actual complaint data. On the other hand, our system focuses on modern systems such as machine learning (ML) and natural language processing (NLP) providing greater able to learn from past complaint data and provide real time and more precise classification. This results to better treatment of vague or type dependent complaints which was a limitation of older systems which were based on a fixed rule. Moreover, this paper goes beyond narrowing the scope of the complaint to classification because the use case, submission, analysis and consolidation of complaints has been automated reducing the manual work and room for error greatly.

4.3.1. Traditional Methods: Rule-Based Categorization & Keyword Matching

Traditional Methods which can be termed conventional make use of already available rules and/or a dictionary of keywords in order to classify the complaints. These systems expose key terms or patterns within the complaint text, and favourably leave it aligned with dominant categories. For instance, the incorporation of the word "tuition" may cause the complaint to be classified under "Finance" automatically. Systems based on rules work quite well in bounded environments or for classic-level complaints; however, their strength begins to dwindle where the complexity level crosses the boundaries, for instance, with complaints in different tones, variants of English, and eclectic contexts.

4.3.2. Limitations of Traditional Methods

- **Accuracy**: Rule-based systems may have a lower accuracy rate, especially when complaints are verbose and the wording used is rather vague. This makes it difficult for complaints that have multiple subject matters or require a lot of definitions.
- **Scalability**: Especially as the number of complaints goes upwards, there is a greater propensity for rule-based systems to become unwieldy and to lose effectiveness. Continuously updating the rules or keywords in accordance with the type of complaints being received is time-consuming and prone to error thus restricting the scalability of the system.
- **Context and Nuance**: Custom keyword solutions are limited in knowing the totality of situations in which a word may reside. For instance, the word "budget" might mean money in one sense. Such intricacies are often overlooked in traditional methods.

4.3.3. Our Approach: Machine Learning with MNB and TF-IDF

Conversely, our Automated Complaint Management System uses a more advanced methodology leveraging Multinomial Naive Bayes (MNB) in conjunction with TF-IDF for feature extraction. This machine learning approach provides substantial enhancements in accuracy, scalability, and contextual comprehension.

- **Multinomial Naive Bayes (MNB)**: The MNB model is a Bayesian classifier that is highly appropriate for text categorization issues. It calculates the likelihood of every complaint being categorized in a specific category by looking at the incidences of words in the text of the complaint, while also taking into account the correlation between relevant terms in different categories. This allows the model to categorize complaints that fall within the same category more effectively, especially in non-specific cases or where there is overlap of categories.
- **TF-IDF Feature Extraction**: Use of the TF-IDF empowers the model to prioritize the most important phrases in the complaints while down weighting common or generic

words especially those that do not assist in differentiating between categories. This enhances the model's ability to identify keywords and understand the intricacies present in complaints that are missed by the rule-based techniques.

4.3.4. Key Advantages of Our Approach

- **High Accuracy**: The MNB model achieved an impressive overall accuracy of 94% which demonstrates TCB's great performance in terms of classifying complaints into all the possible categories such as Academic, Finance, and Equipment. This is considerably better than the accuracy achieved in conventional systems, which is quite hard pressed to remain optimal under normal operational conditions when dealing with complex complaint messages.
- **Contextual Understanding**: In contrast to keyword matching where only words are looked at, the MNB model focuses on the context in which words are used. The word "fee" if used in one context can mean tuition while in another it can be the cost of rental equipment. The MNB model working with TF-IDF can assist in understanding these cases by looking at the context of the words being discussed.
- **Scalability**: Models of machine learning like MNB are very scalable meaning that all the systems can handle a large amount of complaints without a big change in their performance. As the dataset expands, the model can persistently adapt and assimilate new information, rendering it significantly more versatile in response to evolving conditions than rule-based systems that necessitate manual revisions.

5. Conclusions

This study created an automated complaint management system for universities utilizing a Multinomial Naive Bayes (MNB) model and natural language processing (NLP). The algorithm proficiently classified complaints into Academic, Finance, and Equipment categories, reaching an overall accuracy of 94%. This verifies that the integration of machine learning with natural language processing enhances the efficiency and precision of complaint management.

The results indicate that our methodology can substantially improve the management of complaints by universities, resulting in swifter and more precise responses. Nonetheless, there exist opportunities for enhancement, especially in minimizing the redundancy between the Finance and Equipment categories. Subsequent research may investigate more sophisticated models to enhance the system's precision. This research provides an effective approach for enhancing complaint handling in educational institutions, hence improving student satisfaction and administrative efficiency.

Author Contributions

Conceptualization, A.H.M. and S.O.H.; methodology, A.H.M. and S.O.H.; data collection, A.H.M. and S.O.H.; writing—original draft preparation, A.H.M. and S.O.H.; data analysis, H.N.H. and K.A.B.; software development, H.N.H. and K.A.B.; literature review, B.A.A.; validation, B.A.A.; writing—review and editing, H.M.M.; supervision, H.M.M.; corresponding author, H.M.M. All authors have read and agreed to the published version of the manuscript.

Funding

This research received no external funding.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Data Availability Statement

The data that support the findings of this study are openly available in GitHub Repository at https://github.com/bashkahee-web/Full-University-Complaint-NLP.

Acknowledgments

We would like to express our sincere gratitude to our University, Dean, and supervisor for their unwavering support and guidance throughout this research. We also deeply appreciate their invaluable insights and financial assistance provided.

Conflicts of Interest

The authors declare no conflict of interest.

References

[1] Hossen, M.S., Chowdhury, M.N.A., Sristy, A.M., et al., 2022. Sentiment analysis using machine learning and NLP for digital education. In Proceedings of the 6th International Conference on Computational Methodologies and Communications (ICCMC 2022); Erode, India; 29–31 March 2022. pp. 902–908. DOI: https://doi.org/10.1109/ICCMC53470.2022.9754065

[2] Hassan, M.A., Hassan, Z.A., Ali, A.S., et al., 2024. Predicting customer churn in telecommunication companies in Somalia using machine learning. In Proceedings of the IEEE 2024 1st International Conference on Emerging Technologies for Dependable Internet of Things (ICETI); Sana'a, Yemen; 25–26 November 2024. pp. 1–9. DOI: https://doi.org/10.1109/ICETI63946.2024.10777125

[3] Pham, T.D., Vo, D., Li, F., et al., 2020. Natural language processing for analysis of student online sentiment in a postgraduate program. Pacific Journal of Technology Enhanced Learning. 2(2), 15–30. DOI: https://doi.org/10.24135/pjtel.v2i2.4

[4] Tyagi, P., 2020. Sentiment analysis of restaurant reviews. Journal of Natural Language Processing. 7(7).

[5] Manuhutu, M.A., Uktolseja, L.J., 2018. Design and implementation of online students' complaint (case study of English Study Program at Victory University, Sorong). International Journal of Computer Science and Engineering. 6(1), 228–232.

[6] Nandwani, P., 2021. A review on sentiment analysis and emotion detection from text. Social Network Analysis and Mining. 11(81). DOI: https://doi.org/10.1007/s13278-021-00776-6

[7] Bergman, S., Humble, N., 2022. Sentiment analysis for emotional navigation in written communication: What support do autistic people need? In: European Conference on the Impact of Artificial Intelligence and Robotics. Volume 4, No. 1, pp. 6–11.

[8] El Ansari, W., Oskrochi, R., Haghgoo, G., 2014. Are students' symptoms and health complaints associated with perceived stress at university? Perspectives from the United Kingdom and Egypt. International Journal of Environmental Research and Public Health. 11(10), 9981–10002. DOI: https://doi.org/10.3390/ijerph111009981

[9] Hart, D., Coates, N., 2011. International student complaint behaviour: Understanding how East-Asian business and management students respond to dissatisfaction during their university experience. International Journal of Management Education. 9(4), 57–66. DOI: https://doi.org/10.3794/ijme.94.380

[10] Caldeira, L.N., Ribeiro, R., 2022. Classification of public administration complaints. OpenAccess Series in Informatics. 104(9), 9:1–9:12. DOI: https://doi.org/10.4230/OASIcs.SLATE.2022.9

[11] Cranfield, D.J., Tick, A., Venter, I.M., et al., 2021. Higher education students' perceptions of online learning during COVID-19—A comparative study. Education Sciences. 11(8), 403. DOI: https://doi.org/10.3390/educsci11080403

[12] Dang, N.C., Moreno-García, M.N., De la Prieta, F., 2020. Sentiment analysis based on deep learning: A comparative study. Electronics. 9(3), 483. DOI: https://doi.org/10.3390/electronics9030483

[13] Al-Momani, K.R., 2014. Strategies of persuasion in letters of complaint in academic context: The case of Jordanian university students' complaints. Discourse Studies. 16(2), 1–24. DOI: https://doi.org/10.1177/1461445614546257

[14] Onan, A., 2020. Sentiment analysis on product reviews based on weighted word embeddings and deep neural networks. Concurrency and Computation: Practice and Experience. 33(23), e5909. DOI: https://doi.org/10.1002/cpe.5909

[15] Yadav, A., Vishwakarma, D.K., 2020. Sentiment analysis using deep learning architectures: A review. Artificial Intelligence Review. 53, 4335–4385. DOI: https://doi.org/10.1007/s10462-019-09794-5

[16] Jimale, M.A., Ali, M.A., Ahmed, B.A., et al., 2024. Comprehensive ATM security framework combining GSM alert messaging and GPS location tracking. Smart applications and sustainability in the AIoT era. In: SAS-AIoT 2024. Sustainable civil infrastructures. Springer: Cham, Switzerland. Volume 86, pp. 896–907. DOI: https://doi.org/10.1007/978-3-031-78276-3_86