International Journal of Artificial Intelligence and Mathematical Sciences
June 2025 Vo4. Issue No. 01
e-ISSN: 2663-7073
DOI: https://doi.org/10.58921/ijaims.v4i1

# Machine Learning-based Smart Students Complaint Resolution System

Rimsha Javed[1*], Shaukat Wasi[1], Muhammad Hussain Mughal[2], Zulfiqar Ali Bhutto[3]
[1]Department of Computer Science, Mohammad Ali Jinnah University, Karachi, Pakistan
[2]Department of Computer Science, Sukkur IBA University, Sukkur, Pakistan
[3]Department of Information Technology, University of Sindh, Jamshoro, Pakistan
rimsha.javed@jinnah.edu, shaukat.wasi@jinnah.edu, muhammad.hussain@iba-suk.edu.pk,
zulfiqar.bhutto@usindh.edu.pk

Received: 5-03-2025; Accepted: 15-06-2025; Published: 1-7-2025

**Abstract**
For enhancing the quality of any educational institution, student feedback and complaints play an important role. Effectively handling complaints is crucial for ensuring student happiness and expediting the resolution of issues. Traditional manual complaint handling processes, however, are frequently cumbersome and ineffective, resulting in delays that irritate employees and students alike. Our research presents an academic facilitation system that automatically categorizes student complaints by department and particular aspect to address these issues. Our study focuses only on complaints pertaining to the academic setting, in contrast to many other approaches that use supervised and unsupervised learning techniques for common complaint management. We are manually an-notated the data across four departments and forty-two expectations after conducting a survey to gather actual student input because there was no previous dataset of academic complaints accessible. Afterwards, machine learning and deep learning models for classification and aspect identification were trained using this annotated dataset. Bidirectional Encoder Representations from Transformers (BERT), Convolutional Neural Networks (CNN), Support Vector Machines (SVM), Random Forest (RF), Decision Trees (DT), and One Vs All classifiers were among the several algorithms we tested with. Standard performance measures were used to evaluate the models. Among these, the Random Forest model performed well, with 65% accuracy across all our test data, and the One Vs All classifier reached 94% accuracy in both department classification and aspect identification. The study did have several drawbacks, though. The medium-length complaint letters that made up our dataset did not specifically address latent meanings. In the future work, we intend to increase the size of the dataset, incorporate longer complaint texts, and investigate cutting-edge methods for identifying hidden or implicit meanings in student complaints.
.

**Keywords:** Department Classification; Aspect Identification; Random Forest; Natural Language Processing; Student Complaints; OneVsALL Approach; Complaint Resolution System.

## 1. INTRODUCTION

A student's method of communicating their dissatisfaction to the appropriate party, whether it be a specific department, service, or event, is known as complaining in the context of a student complaint system. On the plus side, it also functions as perceptive criticism, offering details on problems related to academic procedures or services.

Email: rimsha.javed@jinnah.edu

The phrase "complaining" suggests a breach of expectations and suggests that the student expressing their complaint is irritated or unhappy with the group or educational institution.

There is a huge volume of complaint texts because many students use the Internet to express their displeasure with different facets of their education. To promote a positive learning environment, educational institutions must effectively address student concerns. Conversely, manual complaint processing techniques are typically time-consuming, error-prone, and delay the resolution of issues. For example, the Mohammad Ali Jinnah University student facilitation site has traditionally handled student complaints manually. This could cause problems if the complaints are sent to the wrong department. This manual approach is frustrating to both staff and students because it requires a lot of resources and takes a long time to respond to inquiries.

These systems use current technology to automate the classification, routing, and resolution of complaints, minimizing response times, increasing student happiness, and optimizing resource use. Department categorization and identification are closely related jobs because, when a student files a complaint, it is critical to identify both the department and the specific aspect associated with the problem. In this research, we proposed and built an academic facilitation system that automatically categorizes student complaints by department and aspect, designed for educational institutions. Department classification involves directing a complaint to the proper department, which in this study is divided into four categories: examination, IT support, registration, and accounts. Aspect identification is an important part of this procedure since it allows for the categorization and prioritization of complaints. Table III illustrates how each department includes different aspects.

Previous research has focused on complaint classification in a variety of fields; however, multi-department classification combined with aspect identification remains an open question. Our contribution is the creation of a student complaint dataset with aspect annotations, which has not been addressed in previous work. For this study, we gathered student complaints directly from students through a survey since there are no publicly accessible datasets of student complaints. Natural language processing methods were used to preprocess the obtained dataset, which involved cleaning and modifying the unprocessed text to prepare it for analysis. After preprocessing, we manually annotated the features, highlighting particular aspects or subjects within each instance, and categorized each complaint according to the appropriate department.

Bidirectional Encoder Representations from Transformers (BERT), Convolutional Neural Networks (CNN), Support Vector Machine (SVM), Random Forest (RF), Decision Tree (DT), and One Vs All classifiers were the six machine learning and deep learning models we used for department classification and aspect identification. Metrics like accuracy, precision, recall, and F1-score were used to assess performance. The results show that the One Vs All strategy worked best for department classification with an accuracy of 94%, and the Random Forest model performed best for aspect identification with an accuracy of 65%.

The remainder of this paper is organized as follows: the Literature Review section presents related work, gap analysis, and methodological differences. The Methodology section describes the dataset and the proposed approach for department classification and identification. This is followed by experimental results and their discussion, and finally, we present the conclusion along with identified limitations and future directions.

Department classification and identification are closely related tasks because, when a student states a complaint, it's crucial to figure out which department and aspect is associated with the issue. In this research work, we proposed and developed an academic facilitation system that automatically classifies academic complaints of the students to the specific department and categorizes their aspects, specifically designed for educational institutions. Department classification involves assigning a particular department to address a complaint. It has been divided into four classes: Examination, IT support, Registration, and Accounts. Aspect identification is the key component of the departments; it helps us categorize and prioritize complaints.

**TABLE I:** Dataset descriptive statistics.

| Departments | Aspects |
|---|---|
| Registration | Courses Add/Drop, Courses Withdrawal, Course Schedule, Re-admission, Course Registration, Semester Freeze, Grade Improvement, Course and Credit Transfer, Course Unavailability, Advising problems, student ID Clash of Courses, Unaware |
| Examination | Attendance issues, Transcript Request, Degree Verification, Invigilation issues, Result Delay, UFM Cases, Rescheduling, Grading, Exam Timetable, Time mismanagement, Short Attendance, Unrespectful, Not Responding, Sitting arrangement, Exam Accommodation, Exam issues, Unjustice in exam |
| IT Support | Printers, Software, Lab Equipment's, Internet Issues, Wi-Fi, and Online Portal issues |
| Accounts | Scholarship Issue, Fee defaulter, Fine charged, Fee issue, Financial Issue, Fee refund |

Each department has different aspects, as shown in Table III. The existing research has focused on complaint clas-sification in various domains. However, multiple departments' classification and aspect identification are still a gap that needs to be covered. Our contribution to this research work is to create a student complaint dataset and aspect identification that has not been explored in previous research. Student complaint datasets are not available on the Internet, so we decided to survey to collect the data from the students for our study. The collected dataset was pre-processed using natural language processing techniques—this involved cleaning and transforming the raw text data to ensure its suitability for analysis. After preprocessing, we assigned labels to each complaint based on the relevant department to which it pertained.
Additionally, we manually labeled the aspects of the complaints, identifying specific elements or topics within each complaint (Paramesh & Shreedhara, 2018), (Bencke, Cechinel, & Munoz, 2020), (Al-Hawari & Barham, 2021). For department classification and aspect identification, we used six different machine learning and deep learning models: Bidirectional Encoder Representations from Transformers (BERT), Convolutional Neural Networks (CNN), Support Vector Machine (SVM), and Random Forest (RF), Decision Tree (DT), One Vs All. To evaluate the performance of the models, we used some evaluation metrics, such as accuracy, precision, recall, and F1-Score. Performs better for department classification, Random Forest performs better for department classification, and Random Forest performs. We received the result that the One Vs All approach performs better for department classification, and Random Forest performs better for aspect identification with accuracies of 0.94 and 0.65, respectively. The paper is organized as follows: Literature review, which describes the previous related work, Gap analysis, and methodological differences. Section Methodology discusses the dataset description and methodology for department classification and aspect identification. Next, we describe the experimental results and their discussion, and finally, we present the conclusion along with limitations and future directions.

## 2. MATERIALS AND METHODS

Automated ticket classification and complaint categorization systems are crucial in managing the increasing volume of tickets and complaints across various domains. In this literature review, we analyze and summarize the existing research in this field to gain insight into the advancements made in automated systems. We searched Springer, IEEE Xplore, ACM, Google Scholar, and Science Direct databases. We limited the search to papers in the English language. Our inclusion criteria are focused on the keywords Machine Learning, Natural language processing, Ticket Classification, Service Desk, Automated Customer Service, complaints, Preprocessing, and related topics within the context of service departments, especially targeting the 'complaint classifications. We considered papers published within the last 5 years to ensure they reflect the latest advancements in the field. Well-known Conferences or Journals: Only papers published in conferences or journals categorized as "W" or" X" were included. This factor adds to the chosen literature's reliability and trustworthiness. On the other hand, the use of precise exclusion criteria enhanced the selection process. Publications that were judged irrelevant were eliminated to guarantee accessibility and understanding. As previously mentioned, a comprehensive search using the keywords and an article date filter (articles published between 2019 and 2024) yielded 500 or more articles in the databases.

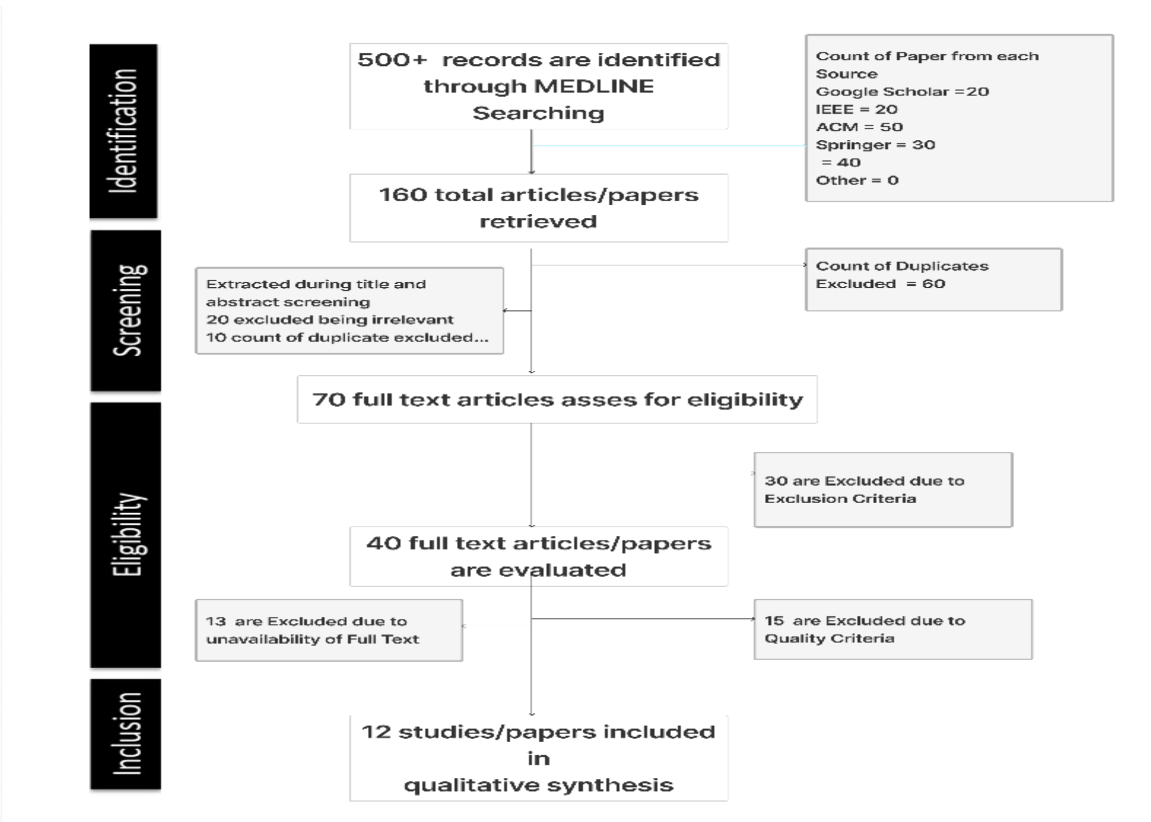

**Fig. 1:** Systematic literature review of academic facilitation system

Figure 1 depicts the distribution of these articles among the four selected databases. Only distinct copies were retained after eliminating standard papers from the chosen database. In the Identification phase, a total of 160 publications were scrutinized. This involved reviewing the titles, abstracts, and selected keywords to determine their clear relevance to the chosen topic. Any articles that appeared irrelevant or were duplicated across selected databases were eliminated. As a result, 70 publications remained after this screening stage. These 70 publications were then assessed. During the eligibility check stage to determine if they fulfilled the inclusion, exclusion, and quality criteria, some are excluded due to the unavailability of full text. After the eligibility phase, 15 publications met our qualitative synthesis's inclusion-exclusion and quality assessment criteria. Many researchers performed the classification of complaints through supervised and unsupervised machine learning techniques (Paramesh & Shreedhara, 2018; Al-Hawari & Barham, 2021; Bastani, Namavari, & Shaffer, 2019; Bencke, Cechinel, & Munoz, 2020; HaCohen-Kerner, Miller, & Yigal, 2019).

Several studies have focused on creating efficient IT service desk systems using machine learning techniques (Paramesh & Shreedhara, 2018). These systems use historical ticket data and a variety of classification techniques to deal with unstructured and noisy data sets. Building on this, machine learning-based help desk solutions have been proposed to associate tickets with the appropriate services and improve decrease resolution time and user happiness (Al-Hawari & Barham, 2021). Additionally, efforts have been undertaken to automatically classify complaint letters into relevant service provider categories (HaCohen-Kerner, Miller, & Yigal, 2019), allowing for effective classification and faster complaint management processes. Furthermore, real-time ticket assignment and resolution technologies have been created to enhance customer assistance and help desk operations (Feng, Senapati, & Liu, 2022). These solutions speed up issue resolution times by recommending pertinent ticket resolvers using deep learning techniques. Using user-provided natural language descriptors, supervised machine learning algorithms have been applied to predict ticket categories in competent helpdesk automated ticketing systems (Feng, Senapati, & Liu, 2022). The SVM classifier is one of these technologies that successfully raises customer happiness and ticket classification accuracy.

The dispatch and ticket assignment processes are crucial to the service delivery industry and offer a great deal of automation and optimization potential. A complete end-to-end automated helpdesk email ticket assignment system that is offered as a service is presented in this article. The main objective is to determine the nature of the problems in incoming email tickets and promptly forward them to the relevant team or resolver group for resolution. The proposed method utilizes a rule engine that can be configured and an ensemble classifier. The challenge of creating an accurate classifier is addressed, but just as much focus is placed on creating a system that is robust and flexible enough to accommodate shifting business requirements. The study examines the primary design issues related to email ticket assignment automation and offers details on the approaches taken. The system's design choices prioritize high accuracy, comprehensive coverage, scalability, business continuity, and the most efficient use of computational resources.

Using three main service providers, the system handles over 90,000 emails per month with an accuracy of over 90% and coverage of at least 90% of email tickets. This invention translates into accuracy comparable to that of a human and yields a significant yearly net savings of over 50,000 man-hours. To date, the deployed system in production has successfully handled over 700,000 tickets, proving its value in the service delivery environment (Mandal, Malhotra, Agarwal, Ray, &

Sridhara, 2019). Performance results can be compared by using supervised machine learning techniques on the Azerbaijani text corpus, which reveals how well-known methods function in the Azerbaijani language. This study advances the study of linguistic subtleties in Azerbaijani through the lens of natural language processing by offering information on the potential and limitations of applying machine learning approaches to this linguistically diverse context (Suleymanov & Rustamov, 2018).

The authors looked at how traditional ticketing systems classified social media posts about smart cities automatically. These systems efficiently classify communications into several smart city elements using natural language processing (NLP) and machine learning, enabling more efficient information management. Furthermore, a variety of topics have been extracted from customer complaints using topic modeling approaches like Latent Dirichlet Allocation (LDA), which have yielded significant insights into the challenges and concerns of customers (Bastani, Namavari, & Shaffer, 2019). The authors developed layer-by-layer semantic matching algorithms and text augmentation techniques to solve the issue of unbalanced customer complaints (Feng, Senapati, & Liu, 2022). Convolutional neural networks (CNNs) were used to automatically classify and handle citizen transit requests (Banga & Peddireddy, 2023). These techniques outperformed machine learning algorithms. Furthermore, sentiment analysis and complaint identification have been implemented using multitasking learning techniques (Fuchs, Drieschner, & Wittges, 2022). The system uses bidirectional long short-term memory (BiLSTM) networks to record contextual information in customer complaints in order to enhance performance, customer sentiment, and customer knowledge. Finally, there are plans to use cloud-based decision support and complaint management systems to improve sustainable farming practices (Kgomo, 2024). These technologies provide prompt assistance and facilitate decision-making by applying knowledge discovery and analytical techniques to agricultural data and historical farmer complaints.

In the field, the development of automated incident management solutions is the primary topic of discussion; customer sentiment prediction and request escalation rank second and third, respectively. Interestingly, their research showed that Random Forest and Support Vector Machine were the most effective classification algorithms in this field (Fuchs, Drieschner, & Wittges, 2022). Text preparation is an important step in text mining that must be handled carefully. In this initial phase, every document must go through a comprehensive data curation process that includes important operations like case folding, to-kenizing, filtering, and stemming. The results of this preprocessing effort have a significant impact on the accuracy of document classification. The stemming process needs to be improved due to the widespread issue of over- and under-stemming in Bahasa Indonesia papers. This work proposes integrating Sastrawi libraries to improve the results of earlier studies that have not yet attained optimal preprocessing results, especially in the domains of filtering and stemming. The investigation's findings show that the Sastrawi library effectively handles the problems of over-stemming and under-stemming and processes information more quickly than a Tala stemmer (Rosid, Fitrani, Astutik, Mulloh, & Gozali, 2020).

The recommended approach (Banga & Peddireddy, 2023) increases the efficiency of handling customer complaints by utilizing automation and artificial intelligence techniques. It uses a natural language processing model that was trained on a set of client complaints in order to classify and comprehend them. The system responds quickly and accurately based on classification. In order to spot patterns and recurring issues, it also monitors complaints. It also makes improvements to its goods and services to improve the consumer experience. Customer satisfaction, response time, and classification accuracy are among the metrics used to compare the system's efficacy with traditional complaint-handling methods. Automated financial complaint processing will take the place of live agents (Naik, Prashanth, Chandru, Jaganath, & Balan, 2023) to boost output and cut costs. The system will be developed as a web application using Jupyter Notebook and Python, utilizing NLP, AI, ML, and DL. A comparison of ML, DL, and Ensemble approaches will help with model selection based on accuracy and time efficiency. Data pretreatment techniques like Word2Vec, TF-IDF, lemmatization, and augmentation improve the complaint classification system.

The automated classification of social media posts about smart cities in traditional ticketing systems was examined by the authors (Bencke, Cechinel, & Munoz, 2020). These systems use machine learning and natural language processing to efficiently classify communications into different smart city features, allowing for more effective information management. Additionally, topic modeling techniques like Latent Dirichlet Allocation (LDA) have been used to extract a variety of topics from customer complaints, which have produced important insights into the difficulties and worries of customers (Bastani, Namavari, & Shaffer, 2019).

Due to the increasing number of customer complaints in all industries, businesses are looking for automated solutions to speed up complaint handling and improve customer service effectiveness. Because traditional manual methods are frequently inaccurate and time-consuming, machine learning (ML) approaches are being used for text classification. Numerous studies have shown that machine learning algorithms can effectively classify customer complaints into relevant themes or product categories, reducing the workload of service personnel and facilitating faster response times (Kgomo, 2024).

These findings had an impact on several machine learning methods, cloud-based strategies, deep learning, and natural language processing, resulting in more accurate and effective systems. Studies show that there is room for improvement in several areas, such as ticket administration, complaint classification, and decision support, which could lead to better customer service procedures (Paramesh & Shreedhara, 2018; Al-Hawari & Barham, 2021; Bastani, Namavari, & Shaffer, 2019; Bencke, Cechinel, & Munoz, 2020; HaCohen-Kerner, Miller, & Yigal, 2019). Our investigation focuses on categorizing complaints and the traits associated with them across departments.

## GAP ANALYSIS

Numerous significant gaps in the classification of complaints and the identification of their various aspects across departments are revealed by a thorough analysis of the literature. While previous studies (Paramesh & Shreedhara, 2018) and (Munir, Ismail, Ghalwash, & El-Bakry, 2022) focused on the complaints category, they only examined one department and did not look into how complaint features were categorized across departments. This represents a gap, as aspect identification is essential for understanding the nature of complaints and addressing them across various departments.

Some studies have proposed machine learning-based help desk systems (Al-Hawari & Barham, 2021), but the aspect-specific classification of complaints for multiple departments has not been explored. Another identified gap is the limited consideration of balanced complaint texts, specifically for numerous departments. Imbalanced data means most of the data belongs to one department, which has many examples as compared to others. There is a lack of research on domain-specific complaint classification and aspect identification for multiple departments. At the same time, studies have focused on specific domains such as transportation (Banga & Peddireddy, 2023) and building quality problems (Mandal, Malhotra, Agarwal, Ray, & Sridhara, 2019). The application of complaint classification and aspect identification to multiple departments remains largely unexplored in the literature.

The best-performing model was BERT, a deep learning model that Google developed. BERT achieved an accuracy of 92.79% on the NLP tasks. The next best-performing model was CNN, which achieved an accuracy of 89.38%. The other models performed significantly worse than BERT and CNN. The lowest-performing model was LDA, which achieved an accuracy of only 32.50%. The results of this study suggest that deep learning models are currently the best-performing models for NLP tasks. However, it is essential to note that the study was conducted on a limited dataset, and the results may not be generalizable to other tasks or datasets.

## 3. PROPOSED METHODOLOGY

To address the gaps identified in the gap analysis section, we proposed a comprehensive methodology for complaint classification and aspect identification in Figure 2.

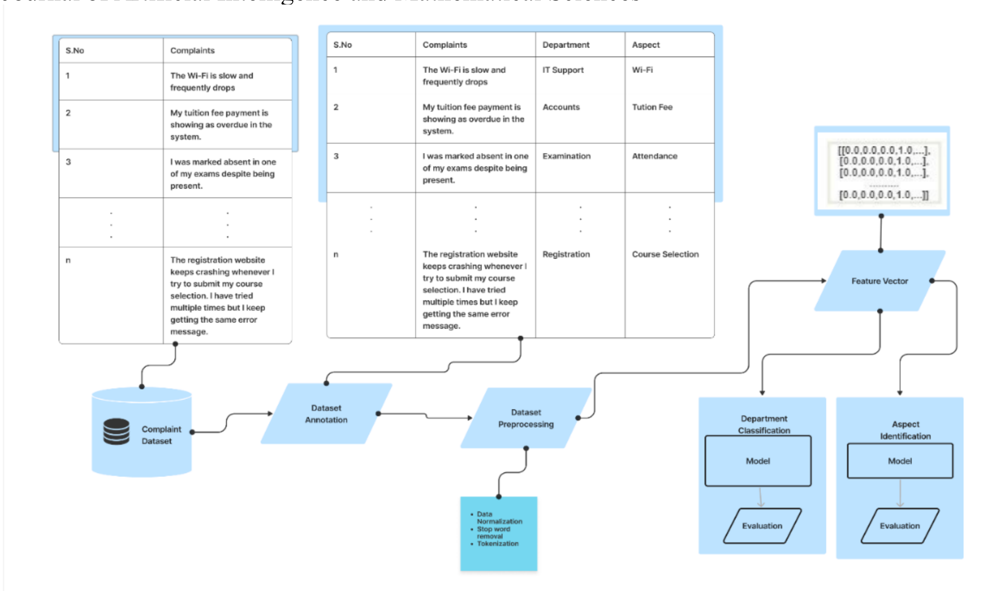


**Fig. 2:** Process flow diagram.

### *A. DATASET DESCRIPTION*

The dataset used for our research consists of complaint text, department, and aspects, as shown in Table II.

**TABLE II:** Dataset descriptive statistics.

| Fields | Qty |
|---|---|
| NO. OF COMPLAINTS | 1952 |
| NO. OF DEPARTMENTS (CATEGORIES) | 4 |
| NO. OF ASPECTS (CATEGORIES) | 42 |

a. **Dataset Collection:** Student complaint datasets were unavailable on the Internet, so we decided to survey to collect the data from the students for our research. We have collected a total of 1952 complaint datasets. To collect the concerns and feedback of the students, we used a survey form specially created to capture their complaints, as shown in Figure 3. The survey form allowed students to describe their complaints across different departments, including examinations, IT support, accounts, and registration. The model's responses are based on learned patterns from diverse internet data, and while it can generate coherent and contextually relevant texts (Herbold, Hautli-Janisz, Heuer, Reinke, & Grabowski, 2023).

b. **Dataset Annotations:** The annotation process involved the detailed labeling of the dataset, such as the departments and their aspects being categorized and identified. Annotations of the dataset are essential as they provide the structured information necessary for training the models. As seen in Figure 4, data annotation was our next objective to accomplish after dataset gathering. The dataset, aspect information, and each complaint's associated department (examination, IT support, accounts, or registration) were all carefully annotated. In order to record every element mentioned in each complaint, we found and assigned pertinent labels. After annotation, we proceeded to annotation verification. So, I contacted my colleagues to verify the annotation of the dataset. They corrected, checked, and made me make some little changes. So, I followed their instructions and revised the annotation process, and after verification, I added them to my dataset, as shown in Figure 5.

c. **Dataset Preprocessing:** Data preprocessing is essential to handle all unwanted and noisy data and ensure that the data is suitable for data mining. The collected data was analyzed to understand its characteristics. In Figure 6, the complaint text was normalized by converting it all to lowercase and breaking it into n-gram tokens, allowing us to do feature extraction by dividing it into separate words. After converting text into tokens, we remove punctuation and extra white space from the complaint. Next, stop word removal was used to remove the most frequently occurring terms like conjunctions, articles, and prepositions that 'don't have much meaning. It helped us to reduce noise and focus on the more meaningful words in the complaints. Our dataset was then divided into training and test datasets.

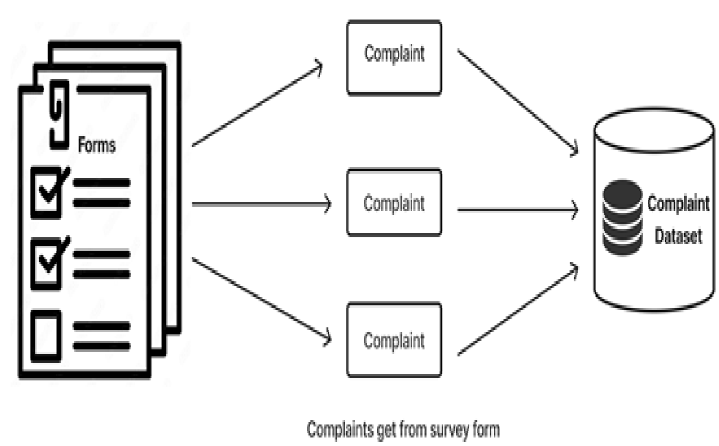


**Fig. 3:** Dataset collection.

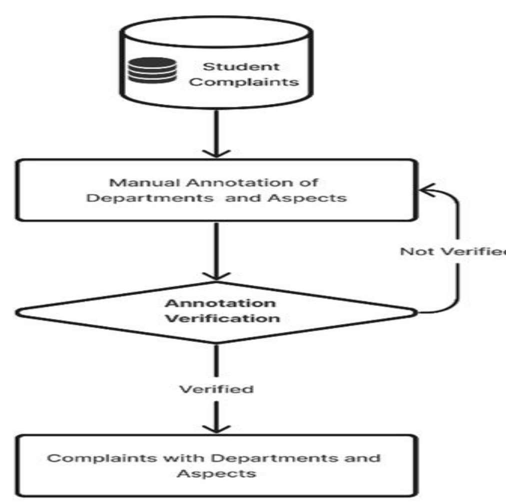


**Fig. 4:** Dataset annotation.

**TABLE III:** Complaint Types

| S.No | Complaint | Departments | Aspects |
|---|---|---|---|
| 1 | I applied for a course change two weeks ago and I haven't received any updates on the status of my request. This is causing me a lot of stress and anxiety. | [“Registration”] | [“Courses Add/Drop”] |
| 2 | My answer sheet was not checked properly and I received a low grade despite writing correct answers. I want to check my answer sheet but the teacher was not showing me my answer sheet. | [“Examination”] | [“Grading”, “Not Responding”] |
| 3 | The registration department charged me extra fees for a course that I never enrolled in. I've been attempting to receive a refund, but they haven't responded. | [“Registration”, “Account”] | [“Fee_issue”] |
| 4 | The campus Wi-Fi network has been persistently problematic, resulting in frequent disconnections and sluggish speeds. This has made it difficult to attend online classes, access internet materials, and submit assignments on time. I request that the university’s IT department take immediate action to improve the campus Wi-Fi infrastructure and provide a more stable and consistent internet connection for students. | [“IT”] | [“Wi-Fi”] |
| 5 | I was unable to complete my exam due to officials warning me about unpaid dues during the exam. | [“Examination”, “Account”] | [“Exam issues”, “Fee_issue”] |

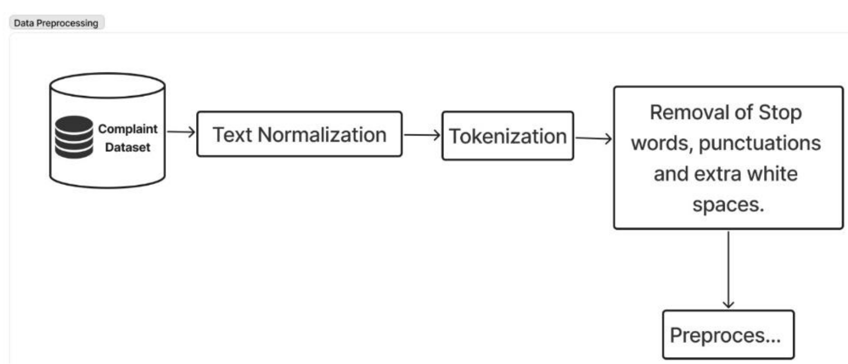


**Fig. 5:** Dataset preprocessing.

d. **MODEL DEVELOPMENT:** Our research is divided into two sub-tasks: classifying a department and aspects. We split our dataset into train and test datasets with a 70% and 30% ratio. The training dataset was then fed into the multiclass classification models tested using the dataset, such as BERT, CNN, Support Vector Machines (SVM), Random Forest, and Decision Tree for both subtasks. Another Approach used for department classification is OneVsAll for multiclass classification. Our 'model's performance is evaluated using a variety of performance metrics.

**BERT:** BERT stands for " Bidirectional Encoder Representations from Transformers". It is a free and open-source Natural Language Processing (NLP) framework. BERT was inspired by a model called Transformer, which looks at how words are connected in a sentence. Unlike Transformer, BERT only needs the encoder part to do its job. 'BERT's primary goal is to add meaning to unclear words in data by contextualizing the information around them. The architecture of the BERT-based model is used for both department classification and aspect identification. The input complaint text is tokenized and passed through the BERT encoder to extract contextual embeddings (hidden states), which are then fed into a classifier layer to predict the corresponding categories. We used the Adam optimizer, and CrossEntropyLoss is defined to train the BERT model. We used BertForSequenceClassification, which is employed to classify the sequences of complaints into various departments and aspects, as shown in Figure 6.

**CNN:** CNN, which stands for Convolutional Neural Network, is a type of neural network developed to learn the representation of words. CNN model can adjust its parameters when it is used in text classification. the architecture of the CNN-based system, consisting of a preprocessing stage with word segmentation and stop word removal, followed by a convolutional neural network pipeline that includes word embedding, convolutional layers, and an MLP classifier to generate department and aspect labels. In this, we first preprocess the complaint text and do word segmentation, which converts the complaints into phrases with limited semantics for classification without the adverbs, mood words, and function words. then this text is fed to our CNN model in vectors and word embedding; after this, we mapped semantic features to the individual labels through a multi-layered perceptron classifier. This was implemented on both the subtasks: department and aspect classification.

**Support Vector Machine (SVM):** The Support Vector Machine (SVM) classifier is a discriminant-based method that prioritizes instances near the border or discriminator, ignoring others. The 'classifier's complexity depends solely on the number of support vectors, not the dataset size, making it efficient for large datasets. Kernel-based methods in SVM involve convex optimization, which aims to find the best possible solution for complex problems. We used linear SVM for both subtasks, suitable for our problems where a straight line can effectively separate the classes or hyperplane in the feature space. The decision function for a linear Support Vector Machine (SVM).

**Forest Classifier:** Random Forest Classifier is an ensemble tree-based learning algorithm that functions by constructing a collection of decision trees called a forest. Each decision tree is built on a subset of the training set, employing attribute selection indicators such as information gain, gain ratio, or Gini index. Structure of the Random Forest model, where multiple decision trees are trained on different subsets of the dataset. Each tree outputs a result, and the final prediction is determined through majority voting or averaging of individual tree outputs. In classifying departments and aspects, the algorithm gathers votes from individual decision trees to determine the final class for the test set. The research discussed involves the application of the Random Forest Classifier with both default and hyperparameters to explore its performance in various settings.

**Decision Tree:** A decision tree is a supervised learning technique for classification and regression that partitions the feature space recursively using binary decisions. It chooses characteristics and thresholds for each node to optimize criteria like information gain or Gini impurity. The resulting tree gives a clear and understandable decision-making approach.

**One Vs All:** Approach for Department Classification: The One-vs-All approach is a traditional machine learning technique that involves training numerous binary classifiers, each of which distinguishes between a specific class and the others. A one-versus-all classification system in which pre-processed complaint vectors are fed into distinct models for each department. Each model individually forecasts the relevance of its respective department, and the final findings are aggregated to produce the final categorization output. We used Random Forest Decision Trees and Logistic Regression for each department class, assigning 0 and 1 labels. We chose the Random Forest model since it performs well in all categories. As seen in Figure 10, we then combined the results of the four departments' individual class labels into a single label.

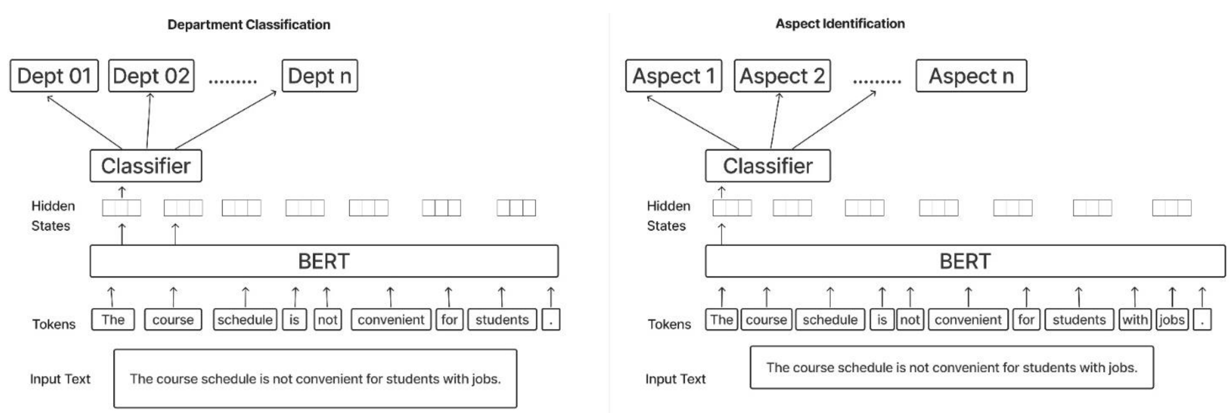


**Fig. 6:** BERT classifier

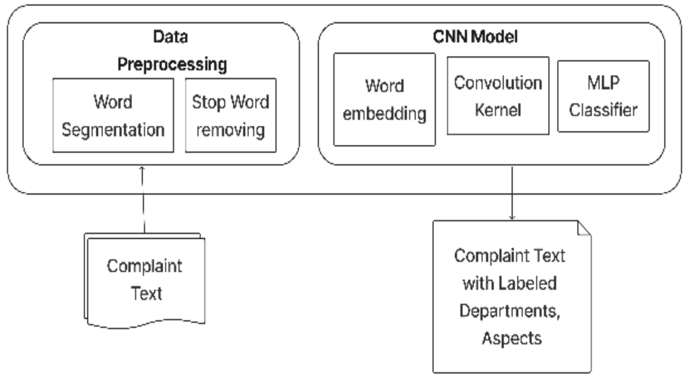


**Fig. 7:** CNN Model

## 4. RESULTS

Metrics like accuracy, precision, recall, and F1-score were used to assess the outcomes of both subtasks. True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) values were used to compare these metrics.

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \tag{1}$$

$$Precision = \frac{TP}{TP+FP} \tag{2}$$

$$Recall = \frac{TP}{TP+FN} \tag{3}$$

$$F1-Score = \frac{2*(Recall * Precision)}{Recall + Precision} \tag{4}$$

Numerous labels are used in both identification and departmental classification. The results of our extensive dataset analysis, which concentrated on department classification and identification, are displayed in Tables III and IV..

The results of the department classification experiments are shown in Table IV. The OneVsAll algorithm received the highest score. The F1-Score for OneVsAll was 0.96, with a 94% accuracy rate. In comparison, the CNN algorithm lags behind with an accuracy of 0.92, demonstrating constant and commendable precision, recall, and F1-Score values of 0.91. The SVM, Random Forest, and Decision Tree algorithms demonstrate competitiveness.

| Algorithm | Accuracy | Precision | Recall | F1-Score |
|---|---|---|---|---|

| | | | | |
|---|---|---|---|---|
| **BERT** | 0.84 | 0.78 | 0.84 | **0.81** |
| **CNN** | 0.92 | 0.91 | 0.92 | **0.91** |
| **SVM** | 0.89 | 0.89 | 0.89 | **0.89** |
| **Random Forest** | 0.91 | 0.91 | 0.91 | **0.91** |
| **Decision Tree** | 0.90 | 0.90 | 0.90 | **0.90** |
| **OneVsAll** | **0.94** | **0.96** | **0.96** | **0.96** |

Accuracy levels are 0.89, 0.91, and 0.90, respectively. Each model strikes a compromise between precision, recall, and F1 score. On the other hand, despite reaching an accuracy of 0.84 and an F1 Score of 0.81, the BERT algorithm performs badly on department categorization.

**TABLE IV:** Department classification results.

Table V presents the findings of the aspect identification experiments. The Random Forest algorithm received the highest score. Random Forest had an F1-score of 0.61 and an accurate rate of 65%. In comparison, the SVM algorithm falls behind with an accuracy of 0.64, while the CNN and Decision Tree algorithms demonstrate competitive accuracy levels of 0.63 and 0.60, respectively. Each model strikes a balance between precision, recall, and F1 scores. On the other hand, while achieving an accuracy of 0.45, with an F1 Score of 0.35, the BERT algorithm performs poorly on aspect identification.

**TABLE V:** Aspect identification results.

| **Algorithm** | **Accuracy** | **Precision** | **Recall** | **F1-Score** |
|---|---|---|---|---|
| **BERT** | 0.45 | 0.32 | 0.45 | **0.35** |
| **CNN** | 0.63 | 0.60 | 0.63 | **0.59** |
| **SVM** | 0.64 | 0.65 | 0.64 | **0.63** |

| | | | | |
|---|---|---|---|---|
| **Random Forest** | 0.65 | 0.60 | 0.65 | **0.61** |
| **Decision Tree** | **0.60** | **0.60** | **0.59** | **0.59** |

## 5. DISCUSSION

Based on the findings, the OneVsAll and Random Forest models outperformed every other algorithm for department classification and aspect identification. They achieved the best scores in all the evaluation metrics due to their binary classification strategy, which addressed imbalanced class distributions and combined predictions from multiple decision trees, respectively.

In the case of department classification, CNN and Random Forest were close. However, CNN outperformed Random Forest in its ability to accurately detect more departments due to its ability to understand the context from both sides of a statement. In contrast, Random Forest combined predictions from multiple decision trees. SVM and Decision Tree performed comparably, as shown in Figure 11. In the case of aspect identification, the SVM algorithm closely trails behind with an accuracy of 0.64 because of its high-dimensional space to effectively separate classes. CNN and Decision Tree performed comparably. In both cases, BERT performed poorly due to the multiclass label, as shown in Figure 12. It can also be seen that department classification produces better results than aspect identification on the student complaint dataset.

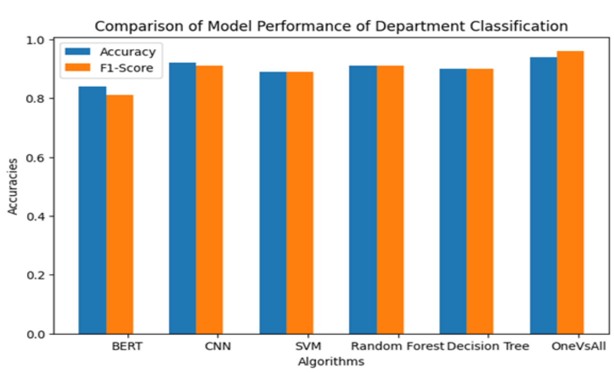


**Fig. 11:** Comparison of department classification model performance.

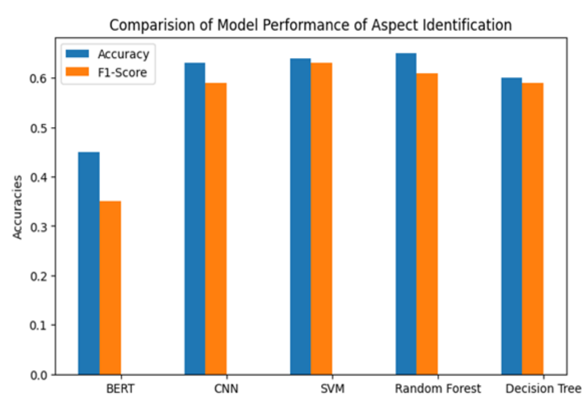


**Fig. 12:** Comparison of aspect identification model performance.

## 6. CONCLUSION

Student complaints play an essential role in any institution. It is crucial to manage them effectively to address delays and ensure the efficient classification of complaints across departments and their respective aspects. We proposed an academic facilitation system that automatically classifies the complaint departments and their respective elements.

We used six models for Department classification and five models for Aspect identification and evaluated the classifier performances on the complaint dataset. After comparing six models for department classification, it is evident that the One Vs All model outperformed the others due to its ability to merge the results of the different departments into one result, SVM, CNN, BERT, Random Forest, and Decision Tree lack. After comparing five models for aspect identification, it is evident that the Random Forest model outperformed the others due to its ability to combine predictions from multiple decision trees, which SVM, CNN, BERT, and Decision Tree lack. The student complaint dataset is unavailable in this research, and aspect identification of the student complaints has not been explored in the last study, making a new contribution in this domain. The academic facilitation system has many benefits, such as reducing human error, increasing throughput, and reducing staff workload.

## 7. LIMITATIONS

There are certain limitations on our research. First off, medium-sized complaint texts make up the majority of the dataset we use. Longer complaint messages might cause confusion or mistakes in the models and analysis we developed. The model's attention span and ability to maintain contextual coherence may be put to the test in larger texts because they usually have multiple themes, overlapping elements, or more complex sentence structures. Therefore, there is a higher chance of misinterpreting the content, mislabeling portions of the text, or assigning it to the wrong department or category when we apply our current models to longer complaints. Second, it does not specifically address the problem of figuring out hidden intentions in written complaints. The nuanced meanings and implicit semantics of complaints were not examined by us, as we focused on the explicit content and surface-level aspects. For instance, a customer might sarcastically express their anger or hint to a problem without saying it out loud. Our models can miss crucial information or fall short of capturing the whole context of a complaint as they were not trained to recognize these nuanced or implicit meanings. Third, our models' ability to accurately represent various complaints may be impacted by the small size of our dataset. A limited dataset implies that the model has fewer examples to learn from, which could make it more difficult for it to understand the variety of ways people voice their grievances. In real life, complaints can range widely in tone, style, vocabulary, and context from succinct, direct messages to longer, more emotional statements. The model may find it challenging to account for this variability in a small dataset, which could result in skewed or inadequate predictions. However, increasing the size and diversity of the dataset could greatly improve the model's performance and stability. A larger dataset would enable the system to generalize more broadly across different types of complaints, improving its capacity to recognize patterns, handle exceptions, and generate more accurate forecasts. In this case, more complex models like BERT could be quite useful. Because BERT is built to comprehend richer language context and often performs better on larger datasets, it is ideal for handling more complex complaint messages.

## 8. FUTURE WORK

We explore hybrid approaches that integrate multiple models as we wrap up our investigation. To improve comprehension of complaint messages, it is necessary to address the size variance in these texts and look into methods for uncovering latent meanings. Future work will expand and rebalance the dataset beyond medium-length complaints to include long-form, multi-issue, multilingual/code-switched texts, with clearer annotation guidelines, double-coding and adjudication, and privacy-preserving release for replicability. On the modeling side, we will move to joint multi-task learning for department and aspect, explore long-sequence/hierarchical transformers, cast aspect identification as span or token-level tagging, and address class imbalance with focal loss, reweighting, threshold tuning, and calibrated probabilities; we will also leverage weak/semi-supervised learning, data augmentation, knowledge-aware label graphs, and model ensembles. Evaluation will report macro/micro and per-class scores, add hierarchy-aware metrics, and test robustness to length, noise, code-switching, and semester/domain shifts, alongside fairness analyses, uncertainty/OOD detection, and drift monitoring. Human factors will be assessed via user studies and A/B tests to quantify reductions in time-to-resolution and gains in satisfaction. For deployment and governance, we will implement human-in-the-loop triage with active learning feedback, MLOps pipelines for retraining and monitoring, explainability (e.g., SHAP/IG) for auditability, and policies for PII scrubbing, retention, and documentation (model cards/data statements). Lastly, in order to enhance generalization without exchanging raw data, we will investigate transfer and domain adaptation between departments and partner institutions, including privacy-preserving cooperation via federated learning.

## References

Al-Hawari, F., & Barham, H. (2021). A machine learning based help desk system for IT service management. Journal of King Saud University - Computer and Information Sciences, 33(6), 702–718.

Banga, D., & Peddireddy, K. (2023). Artificial intelligence for customer complaint management. International Journal of Computer Trends and Technology, 71(3), 1–6.

Bastani, K., Namavari, H., & Shaffer, J. (2019). Latent Dirichlet allocation (LDA) for topic modeling of the CFPB consumer complaints. Expert Systems with Applications, 127, 256–271.

Bencke, L., Cechinel, C., & Munoz, R. (2020). Automated classification of social network messages into smart cities dimensions. Future Generation Computer Systems, 109, 218–237.

Feng, L., Senapati, J., & Liu, B. (2022). TaDaa: Real time ticket assignment deep learning auto advisor for customer support, help desk, and issue ticketing systems. arXiv preprint arXiv:2207.11187.

Fuchs, S., Drieschner, C., & Wittges, H. (2022). Improving support ticket systems using machine learning: A literature review. In Proceedings of the 55th Hawaii International Conference on System Sciences (pp. 1893–1902).

HaCohen-Kerner, Y., Miller, Y., & Yigal, G. G. (2019). Automatic classification of complaint letters according to service provider categories. Information Processing & Management, 56(6), 102102.

Herbold, S., Hautli-Janisz, A., Heuer, U., Reinke, C., & Grabowski, J. (2023). A large-scale comparison of human-written versus ChatGPT-generated essays. Scientific Reports, 13, 18617.

Kgomo, T. (2024). Classification of customer complaints using machine learning algorithms. University of Cape Town, Faculty of Science, Department of Statistical Sciences.

Mandal, A., Malhotra, N., Agarwal, S., Ray, A., & Sridhara, G. (2019). Automated dispatch of helpdesk email tickets: Pushing the limits with AI. In Proceedings of the AAAI Conference on Artificial Intelligence, 33, 9381–9388.

Munir, K., Ismail, F., Ghalwash, A., & El-Bakry, H. (2022). AgroSupportAnalytics: A cloud-based complaints management and decision support system for sustainable farming in Egypt. Egyptian Informatics Journal, 23(1), 73–82.

Naik, P. K., Prashanth, T., Chandru, S., Jaganath, S., & Balan, S. (2023). Consumer complaints classification using machine learning & deep learning. International Research Journal on Advanced Science Hub, 5(05S), 116–122.

Paramesh, S., & Shreedhara, K. (2018). Automated IT service desk systems using machine learning techniques. In Data analytics and learning: Proceedings of DAL 2018 (pp. 331–346). Springer.

Rosid, M. A., Fitrani, A. S., Astutik, I. R. I., Mulloh, N. I., & Gozali, H. A. (2020). Improving text preprocessing for student complaint document classification using Sastrawi. IOP Conference Series: Materials Science and Engineering, 874, 012017.

Singh, A., Jha, S. K., Bansal, P., & Bharti, K. (2022). Multitask learning for complaint identification and sentiment analysis. Cognitive Computation, 14(1), 212–227.

Suleymanov, U., & Rustamov, S. (2018). Automated news categorization using machine learning methods. IOP Conference Series: Materials Science and Engineering, 459, 012006.