QUALITATIVE RESEARCH JOURNAL FOR SOCIAL STUDIES

ISSN Online: 3006-4686, ISSN Print: 3006-4678
Volume No: 02 Issue No: 02 (2025)

SENTIMENT ANALYSIS FOR MOVIE RECOMMENDATION SYSTEM

Ubaid Riaz[1*], Fawad Nasim[1] ,Arfan Jaffar[1]
[1] Faculty of Computer Science and Information Technology, The Superior University, Lahore, 54600, Pakistan
***Corresponding Author : Ubaid Riaz Email : ubaidriaz99@gmail.com**

Abstract
This paper discusses the application of sentiment analysis to recommend movies. The means of identifying and interpreting the sentiments of online users across numerous and diverse sources utilise sentiment analysis, which can also be described as the process of extracting the emotional tone of online text. The research aims to conduct sentiment analysis of film reviews, identify existing research questions and gaps, and implement the optimal possible strategy for completing the given work. By analyzing textual data, businesses can predict the developing market trends, be more effective with their corporate strategies and decision making since they will be able to determine whether the emotional tone is neutral, positive or negative. Specifically, this research looks at the precision, and effectiveness of Naive Bayes classification models, including Bernoulli Naive Bayes, Multinomial Naive Bayes and Complement Naive Bayes in both predicting positive and negative movie reviews.

Keywords: Naive Bayes; movies reviews; and sentiment analysis

1. Introduction

This work studies the application of sentiment analysis[1],[2] in terms of the movie recommendation systems. Sentiment analysis was first originated on printed paper documents in the 1950s but is now commonly deployed to find subjective details within an assortment of online content formats, as blog posts, reviews, tweets, and text. Through such information, organisations are able to discover new chances as well as effectively deliver their messages to the intended groups of people. Sentiment analysis finds its roots in text subjectivity analysis in the 1990s and early 20th century text analysis of opinion.

Recommendation systems predict or filter the items according to the behaviour of users. Youtube, Spotify, and Netflix are the prominent ones of the 20th century. These systems are employed by numerous firms with the aim of increasing the income and assisting the consumers. As an example, Facebook recommends friends, LinkedIn recommends jobs, Spotify and Pandora recommends music, Netflix recommends movies and Amazon recommends products.

2. Research Objectives:

The three overall objectives of the study are given below:

1. To carry out sentiment analysis for movie review.
2. To identity the gaps and challenges of the previous research.
3. To identity and implement the best methodology for sentiment analysis of movie reviews.

Research Questions:

This research will answer the following questions

1. Which emotions should be examined and who is involved?
2. How can we tell from the opinions whether a consumer is feeling positively or negatively?
3. How accurate are we to capture those emotions?

The primary tool to be used in this research is Python because it simplifies all the data collection processes, preprocessing, processing, feature extraction, and machine learning tasks[3-5].

3. Literature Review

There is continuous study in the wide subject of computer science, especially in the area of sentiment analysis methods. This section examines earlier research and writers' viewpoints that are pertinent to the project. Numerous studies have investigated different methods for sentiment analysis and movie recommendation systems:

Using K-means and collaborative filtering, a study sought to develop a movie recommendation system. It reported better accuracy and shorter execution times than traditional techniques. A monolithic hybrid recommender system that included a fuzzy expert system and the content-based SVD algorithm was proposed by another study, attaining a precision of more than 80%. High accuracy and precision were demonstrated by a hybrid recommender system that combined Collaborative Filtering on a Content-Based Approach with a neural network (Self-organising Map), improving the effectiveness of the conventional Collaborative Filtering technique [6].

The best concepts from CF and CBF were coupled with sentiment analysis of tweets from microblogging sites to create a hybrid movie recommendation system that produced more accurate suggestions than previous models [7].

Using the TMDB dataset with SVM, one study built a sentiment analysis system that achieved 85% accuracy in predicting whether user reviews will be positive or negative [8]. Using real-time tweets and RNN for categorisation, a multilingual opinion mining movie recommendation system reported 91.67% accuracy and 92.6% precision [9].

When four different recommendation models were compared for movie selection—Backpropagation (BPNN), Singular Value Decomposition (SVD), Deep Neural Network (DNN), and DNN with Trust—it was found that the DNN with trust model outperformed the others, achieving an exceptional accuracy rate of 83 percent [10].

Using the lexical technique and fuzzy emotions, researchers created a movie recommendation system based on review emotions. The results showed that rating-based RS has a 30% accuracy rate, whereas emotion-based RS has a 70% accuracy rate [11].

It was discovered that a recommendation system based on hybrid deep neural networks performed better than other techniques including KNN, SVD, CF, and CBF because it could learn extremely complex characteristics and achieve remarkable prediction accuracy [12].

Various strategies and their accuracy across diverse datasets were emphasised in a review of sentiment analysis technologies [13]. For example, on the product review training dataset, SVM and co-reference resolution achieved 73.6% accuracy, whereas on the Twitter data, Naive Bayes and SVM demonstrated 86.4% and 73.5% accuracy, respectively. When it came to sentiment analysis of Arabic reviews, Naive Bayes achieved the highest accuracy of 97%.

4. Challenges and Gaps

Despite the advances, sentiment analysis continues to be faced with a variety of challenges:

• **Sarcasm:** Sarcasm is more difficult because, it relies on tone of voice, context, previous understanding and other variables, which is beyond the capabilities of robot. Sarcasm (an acid remark which is supposed to be compliment) is very difficult to detect yet a very intriguing issue in NLP study because it has an unconscious meaning often negating the actual meaning in a bid to offend or mock.

• **Informal style of writing** One of the major obstacles to any NLP initiative is informal style of writing present in online reviews and messages that consists of emoticons, acronyms and hard to read abbreviations. Most local and shifting acronyms are a problem, whereas global ones can be controlled.

• **Spelling and grammar:** As often encountered in casual writing, spelling and grammatical mistakes are many times hard to correct. Natural language processing and sentiment analysis tasks can be more accurate in case of identification and elimination of these errors.

ISSN Online: 3006-4686, ISSN Print: 3006-4678
Volume No: 02 Issue No: 02 (2025)

• **Computational Cost:** The more training data and complex it is, the more the model is computationally expensive; large corpus of information might require the help of an expensive graphics card.AI Models [14],[15] like Support Vector Machines (SVM) and Naive Bayes (NB) have been found out not to be computationally expensive as neural networks and attention models.

• **Availability of data:** Data may be an occasional problem due to cybersecurity [16-18] and the current boom in NLP and sentiment sciences technology. Even in the case of Twitter providing data available to sentiment analysis, high-quality training data is very hard to come by, to the supervised learning algorithm. The generation of the ABSA training data should be manual since it is very challenging to get it online. As an example, a model trained on a hotel reviews dataset does not permit predicting the attitude in a stock or mutual fund dataset and vice versa. The training data in one field cannot be applied and used in another one.

Language Adjustments: Language adapt to movement of people; though the basic language does not change much, there are changes in the language like literacy level, dominant language, how language is pronounced among other elements. Take an example of the English language that is widely used among many individuals around the globe and has been divided into different dialects (American, British, Indian, and so on). Most words in different countries have a number of definitions (such as thong in Australia and the UK). Like the case, regional impositions on the same word (in this case, colour and colour) would attract multiple spellings despite bearing the same meaning which may render it redundant (wastes computational cost and accuracy of the model). The language barrier is considered as the thorniest NLP issue because the number of languages with access to NLP techniques are limited to five and ten in comparison to English.

Degree intensifiers and adverbs : An affect can also be quantified by the incorporation of adverbs like slightly, barely, and moderately (e.g. - The food is wonderful fantastic or The food is awfully wonderful or good). Transformed into simple language, this is what it means: ("The food is hardly all right"). Intensifiers such as the words very and too are also used to increase the positive or negative features of a token (i.e., it is not good, but too good). The presence of intensifiers and degree adverbs complicate the aggregation of the sentiment values and contrast two sentences of a certain feeling instead of distinguishing between two sentences with opposite polarities.

Methodological Issues:

Disadvantages of the SVMs are that they cannot process large datasets, training is long, as the number of features increases so does their complexity, and have poor performance on high noise levels.

• **Neural Network Issues**: Neural networks need much data, consume more time to build, and are costly in computations.

• **Machine Learning Issues**: Most machine learning algorithms need much data to achieve their optimal functioning. Other issues are low-rating datasets, unlabelled training data and overfitting and underfitting considerations.

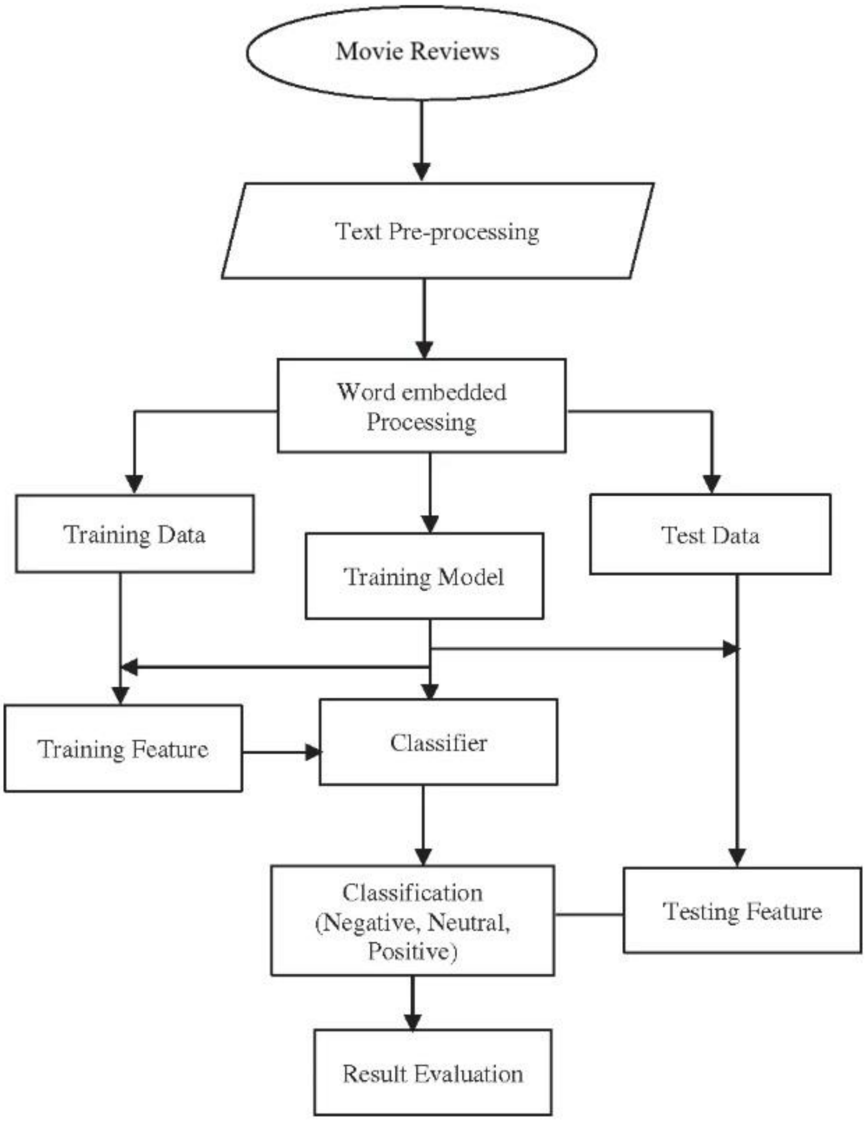
• **Naive Bayes Problems:** A term known as zero-frequency issue arises when Naive Bayes Algorithm is applied to data that contains a categorical variable, having a category that did not occur in the training dataset. It operates on a set of assumptions that every characteristic is independent and this is hardly ever true in real life. The fact that Naive Bayes is vulnerable to the quality of data provided means that only biased and uninformed results might be achieved in case the data is noisy, imbalanced, incomplete, or involves irrelevant qualities**.**

5. Methodology

Research methods, more precisely a contextual framework of research, a circumstantial system based on opinions, beliefs and values on which one reaches a decision, is called methodology. The research design and the system used in conducting this research is elaborated in this section, where more focus is given to the tools and the process undertaken.

System Architecture:

First of all, raw data was collected from Kaggle which contains 50 thousand IMBD movies reviews and that data needed to be cleaned and pre processed for the sake of better results. The data was imported in python using the pandas library and then it was cleaned by using pandas as well as numpy library in python. Exploratory data analysis was done by using matplotlib, seaborn and plotly and data was analysed descriptively. The outliers, null values and duplicates were removed. This cleaned data was used for applying machine learning algorithms but the data was splitted into training and test data and then machine learning algorithm was applied in python. Machine learning algorithm gave us the results which were evaluated further. The architecture diagram of the whole process is given below:



QUALITATIVE RESEARCH JOURNAL FOR SOCIAL STUDIES

ISSN Online: 3006-4686, ISSN Print: 3006-4678
Volume No: 02 Issue No: 02 (2025)

Figure 1: System Architecture

Naive Bayes Sentiment Analysis of Film Reviews: Here we need to apply the Naive-Bayes classification based models to predict the number of good and negative reviews on the basis of feelings.
The following figure explains the framework of sentiment analysis process:

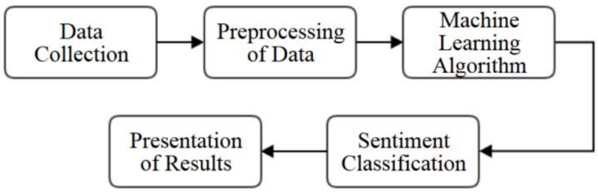


Figure 2: Sentiment Analysis Framework

The whole process is going to be explained as follows:
1. Preprocessing of text data: Before applying the method the text data should be preprocessed. This involves such processes as lemmatisation, stemming, tokenisation, stop-word removal. The data set we have has too many errors, and thus we have to clean it up, such as turning it to lowercase, removing links of URLs, special characters, punctuation, numbers, emojis, and abbreviations. Occasionally, stop words that can eliminate relatively less useful words are used (they are a collection of commonly recognized words in any language (e.g., a, the, is, are, etc.).
Data preprocessing involves the following steps:

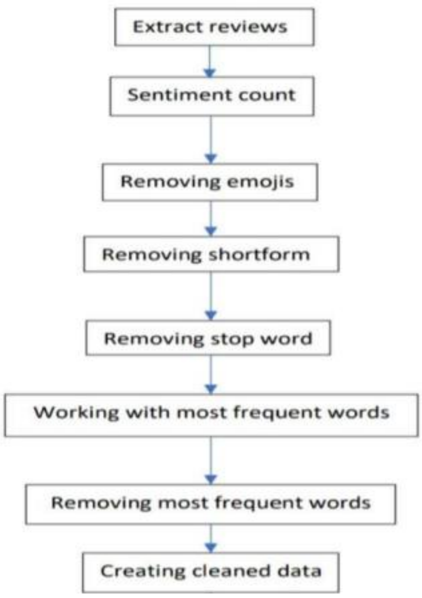


Figure 3: Data Preprocessing

2. Feature extraction: Before the text data can be fed in the algorithm, it needs to be converted to the format that allows it: a feature vector. The most common feature extraction method is using a bag-of-words where each document is represented by a vector of word frequencies.
3. Data splitting: The training and testing sets of data should be divided. The training set is used to train the model and the testing set can be used to evaluate the model. The IMDB dataset has 50K movie reviews, 25,000 positive reviews, and 25,000 negative reviews, giving it a balanced distribution as a training and a testing set (each has 25,000 movie reviews that are

QUALITATIVE RESEARCH JOURNAL FOR SOCIAL STUDIES

ISSN Online: 3006-4686, ISSN Print: 3006-4678
Volume No: 02 Issue No: 02 (2025)

marked highly polar). The train-test split is done in which the cleaned data is assigned a binary value (negative 0; positive 1). The data will also be randomised in order to avoid the biasness. We can use training set to train a model and test its accuracy on the testing set with train test split feature of Scikit Learn.

4. Model training: The NB model is trained by estimating the likelihood of all features considering each of the classes, using the training set. This involves determination of the probability of each characteristic with each of the classes and prior probability of each of the classes.

5. Evaluation of model performance: The performance of the model is evaluated using measures such as the accuracy, precision, recall and F1-score of the testing set.

6. Predicting using the model: Once the model is trained, the model may be used to new text data giving predictions. The text data first undergo prepossessing to get it in the format of a feature vector, and then input into the trained model to obtain the predicted class label. NB is a simple and yet powerful algorithm having shown its good performance in various NLP applications such as topic classification, spam, and sentiment analysis. However, it has its weaknesses such as relying on feature independence assumption that may not be true all the time. It is therefore paramount to perform a careful examination of the effectiveness of the model before implementing it in a practical environment.

5. Results and Discussion

This section contains the results of the application of different Naive Bayes models with the accompanying explanation.

Naive Bayes Modelling Method We shall review some of the Naive Bayes models here: Compare the accuracy of Bernoulli, Multinomial and Complement NB models.

a) Strength of the NB Model:

The specificity of the Complement NB model is 86.33 percent. The average number of predictions done correctly by the model is 86 predictions out of a hundred.

Confusion matrix: A table of two rows and columns that indicate the number of true negatives, false positives, false negatives, and true positives is called a confusion matrix.

- True Positive: The model generated 4349 negative reviews as the model predicted, compared with the real value of 4349 as well.
- The number of predicted favourable reviews by the algorithm was 4284 as compared to 4284. This is really negative.
- False Positive: The number of a predicted negative review by the model was 628, whereas the correct value was 628 positive reviews.
- False Negative: The model will decide 739 good reviews yet the true figures are 739 bad reviews.

Classification report:

The precision of a negative review is 0.85 because it is the ratio of correctly predicted negative review observations and total predicted negative review observations.

The accuracy of positive reviews is 0.87, that is the proportion of accurate predictions of positive review observations to the number of all the anticipated positive review observations.

Recall (Negative review): In the cases where observations have been made in the actual class, there is a ratio of 0.87 of the properly forecasted negative review observations.

Recall (Positive review): 0.85 is the fraction of correct positive review observations to complete set of true class observations.

ISSN Online: 3006-4686, ISSN Print: 3006-4678
Volume No: 02 Issue No: 02 (2025)

The F1 Score refers to the weighted mean of the precision and the recall. In the scenarios of both negative and positive evaluation, F1 scores are the same, which equals 0.86.

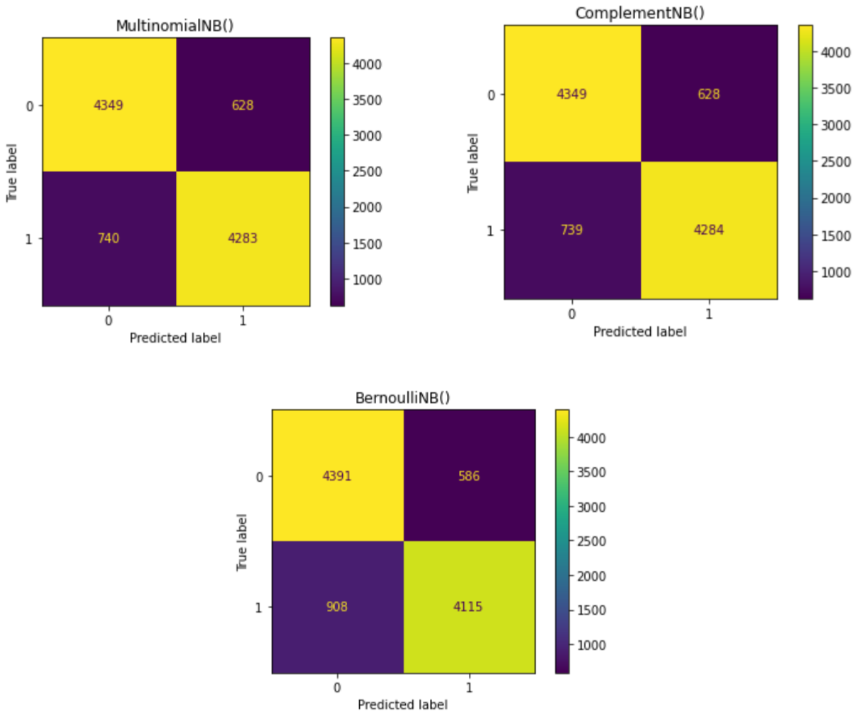
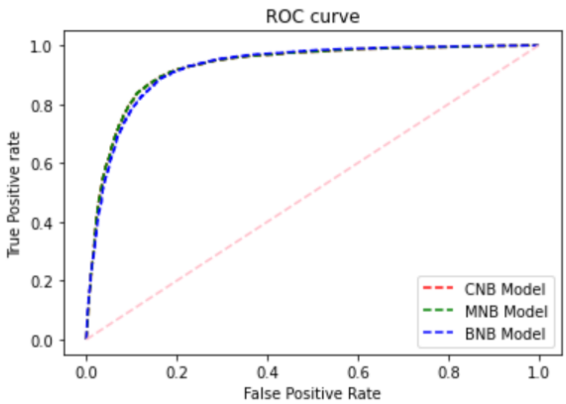


Figure 4: Classification Report

- **Multi nominal NB model accuracy is 86.32%**.
- **Compliment NB model accuracy is 86.33%**.
- **Bernoulli NB model accuracy = 85.06%**.

ROC Curve A graph of the performance of a classification model with all classification thresholds is referred to as a receiver operating characteristic curve, or a ROC curve. The two parameters that appear in this graph are the True Positive Rate and the False Positive Rate. The ROC curve of a random classifier would be a 50 percent line, or, in other words, it would perform just like guessing which ones of the examples are considered to be positive and which ones are considered to be negative. Having ROCs in range of 0.7-0.8 is considered satisfactory and 0.8-0.9 is considered excellent and more than 0.9 is exceptional. ROC curves of the models (CNB, MNB, and BNB) are good since they indicate that the models are strong in terms of predictive power**.**



QUALITATIVE RESEARCH JOURNAL FOR SOCIAL STUDIES

ISSN Online: 3006-4686, ISSN Print: 3006-4678
Volume No: 02 Issue No: 02 (2025)

Figure 5: ROC Curve

6. Conclusion

Various Naive Bayes classification models were used in this research in an attempt to analyze sentiment of movie reviews successfully. The paper concluded that some of the greatest barriers to sentiment analysis are sarcasm recognition, colloquialisms, and syntax errors, computational expenses, information access, and linguistic processes. It also focused on the methodological weaknesses of the popular algorithms in machine learning such as SVM, Neural Networks and Naive Bayes. IMDB movie review set was preprocessed carefully to train the models and evaluate them. The Complement Naive Bayes model and Multinomial Naive Bayes model recorded impressive accuracies of 86.33% and 86.32% respectively with the Bernoulli Naive Bayes model finishing second with 85.06%. The effectiveness of these Naive Bayes approaches in the sentiment classification on movie reviews is evidenced by the outstanding nature of the precision, recall and F1 scores along with the rich performance on the ROC curve. To enhance accurateness and user-friendliness of movie recommendation systems, this study provides an effective platform to incorporate sentiment analysis.

7. Future Work

The study that will follow the identification of the issues, and it can be developed into future research will focus on addressing the issues in sentiment analysis. Building upon advancements in deep learning, the visualization and classification of malware images[19] and other complex tasks [20]have shown promising results. .This is by designing more sophisticated models[21], which can handle sarcasm and informal writing style more accurately, possibly through the introduction of deep learning techniques designed specifically to understand contexts. There also will be attention, how to robustly handle grammatical defects and changing regional language adaptations as well. Research will also investigate how to reduce the cost of complex models in computing to guarantee a higher quality and domain-specific training datasets as well as the process of acquiring data. In addition to improving the underlying sentiment analysis, future work will be to search personalised recommendations based on changing emotional user state, fuse results of sentiment classification more adaptively into the real-time movie recommendation system and potentially combine multimodal data (e.g. audio and video cues in reviews) to obtain a more advanced view of user sentiment.

References:

[1]. Haroon, Muhammad, Zaheer Alam, Rukhsana Kousar, Jawad Ahmad, and Fawad Nasim. "Sentiment analysis of customer reviews on e-commerce platforms: A machine learning approach." *Bulletin of Business and Economics (BBE)* 13, no. 3 (2024): 230-238.

[2]. Noor, Hajira, Jawad Ahmad, Ammar Haider, Fawad Nasim, and Arfan Jaffar. "A Machine Learning Sentiment Analysis Approach on News Headlines to Evaluate the Performance of the Pakistani Government." *Journal of Computing & Biomedical Informatics* 7, no. 02 (2024).

[3]. Zainab, Hira, Muhammad Ismaeel Khan, Aftab Arif, and Ali Raza A. Khan. "Deep Learning in Precision Nutrition: Tailoring Diet Plans Based on Genetic and Microbiome Data." Global Journal of Computer Sciences and Artificial Intelligence 1, no. 1 (2025): 31-42.

[4]. Zainab, Hira, Ali Raza A. Khan, Muhammad Ismaeel Khan, and Aftab Arif. "Innovative AI Solutions for Mental Health: Bridging Detection and Therapy." Global Journal of Emerging AI and Computing 1, no. 1 (2025): 51-58.

[5]. Khan, Ali Raza A., Muhammad Ismaeel Khan, and Aftab Arif. "AI in Surgical Robotics: Advancing Precision and Minimizing Human Error." Global Journal of Computer Sciences and Artificial Intelligence 1, no. 1 (2025): 17-30.

ISSN Online: 3006-4686, ISSN Print: 3006-4678
Volume No: 02 Issue No: 02 (2025)

[6]. M. L. M. A. A. Yassine Afoudi, "Hybrid recommendation system combined content-based filtering and collaborative prediction using artificial neural network," Simulation Modelling Practice and Theory, vol. 113, 2021.

[7]. K. D. P. P. R. Sudhanshu Kumar, "Movie Recommendation System Using Sentiment Analysis From Microblogging Data," 2020.

[8]. S. V. K. K. S. Nimish Kapoor, "Movie Recommendation System Using NLP," in Fifth International Conference on Communication and Electronics Systems IEEE, 2020.

[9]. A. N. A. S. Tarana Singh, "Multilingual Opinion Mining Movie Recommendation System Using RNN," in First International Conference on Computing, Communications, and Cyber-Security, 2020.

[10]. S. N. M. A. K. J. Sasmita Subhadarsinee Choudhury, "Multimodal trust based recommender system with machine learning approaches for movie recommendation," International Journal of Information Technology, vol. 13, pp. 475-482, 2021.

[11]. S. C. A. K. Mala Saraswat, "Analyzing emotion based movie recommender system using fuzzy," International journal of information tecnology, 2020.

[12]. P. K. B. B. Kiran R, "DNNRec: A Novel Deep Learning based Hybrid Recommender System," Expert Systems With Applications, 2019.

[13]. P. A. A. A Review On Sentiment Analysis Methodologies, "Pooja Mehta, Dr.Sharnil Pandya," INTERNATIONAL JOURNAL OF SCIENTIFIC & TECHNOLOGY RESEARCH, vol. 9, 2020.

[14]. Zainab, Hira, Ali Raza A. Khan, Muhammad Ismaeel Khan, and Aftab Arif. "Ethical Considerations and Data Privacy Challenges in AI-Powered Healthcare Solutions for Cancer and Cardiovascular Diseases." Global Trends in Science and Technology 1, no. 1 (2025): 63-74.

[15]. Zainab, Hira, Muhammad Ismaeel Khan, Aftab Arif, and Ali Raza A. Khan. "Development of Hybrid AI Models for Real-Time Cancer Diagnostics Using Multi-Modality Imaging (CT, MRI, PET)." Global Journal of Machine Learning and Computing 1, no. 1 (2025): 66-75.

[16]. Khan, Muhammad Ismaeel, Aftab Arif, and Ali Raza A. Khan. "AI's Revolutionary Role in Cyber Defense and Social Engineering." International Journal of Multidisciplinary Sciences and Arts 3, no. 4 (2024): 57-66.

[17]. Arif, Aftab, Muhammad Ismaeel Khan, and Ali Raza A. Khan. "An overview of cyber threats generated by AI." International Journal of Multidisciplinary Sciences and Arts 3, no. 4 (2024): 67-76.

[18]. Khan, M. I., A. Arif, and A. R. A. Khan. "AI-Driven Threat Detection: A Brief Overview of AI Techniques in Cybersecurity." BIN: Bulletin of Informatics 2, no. 2 (2024): 248-61.

[19]. Tariq, Muhammad Arham, Muhammad Ismaeel Khan, Aftab Arif, Muhammad Aksam Iftikhar, and Ali Raza A. Khan. "Malware Images Visualization and Classification With Parameter Tunned Deep Learning Model." Metallurgical and Materials Engineering 31, no. 2 (2025): 68-73.https://doi.org/10.63278/1336.

[20]. Arif, A., A. Khan, and M. I. Khan. "Role of AI in Predicting and Mitigating Threats: A Comprehensive Review." JURIHUM: Jurnal Inovasi dan Humaniora 2, no. 3 (2024): 297-311.

[21]. Khan, Muhammad Ismaeel, Aftab Arif, and Ali Raza A. Khan. "The Most Recent Advances and Uses of AI in Cybersecurity." BULLET: Jurnal Multidisiplin Ilmu 3, no. 4 (2024): 566-578.