Aspect-based Sentiment Analysis of IMDb Movie Reviews Using Machine Learning Techniques

K. A. V. K. Karunanayake

2024

Aspect-based Sentiment Analysis of IMDb Movie Reviews Using Machine Learning Techniques

A Thesis Submitted for the Degree of Master of Computer Science

K. A. V. K. Karunanayake

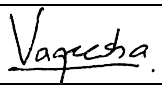
University of Colombo School of Computing

2024

DECLARATION

Name of the student: K.A.V.K Karunanayake
Registration number: 2020/MCS/041
Name of the Degree Programme: Degree of Master of Computer Science
Project/Thesis title: Aspect-based Sentiment Analysis of IMDb Movie Reviews Using Machine Learning Techniques

1. The project/thesis is my original work and has not been submitted previously for a degree at this or any other University/Institute. To the best of my knowledge, it does not contain any material published or written by another person, except as acknowledged in the text.
2. I understand what plagiarism is, the various types of plagiarism, how to avoid it, what my resources are, who can help me if I am unsure about a research or plagiarism issue, as well as what the consequences are at University of Colombo School of Computing (UCSC) for plagiarism.
3. I understand that ignorance is not an excuse for plagiarism and that I am responsible for clarifying, asking questions and utilizing all available resources in order to educate myself and prevent myself from plagiarizing.
4. I am also aware of the dangers of using online plagiarism checkers and sites that offer essays for sale. I understand that if I use these resources, I am solely responsible for the consequences of my actions.
5. I assure that any work I submit with my name on it will reflect my own ideas and effort. I will properly cite all material that is not my own.
6. I understand that there is no acceptable excuse for committing plagiarism and that doing so is a violation of the Student Code of Conduct.

Signature of the Student	**Date (DD/MM/YYYY)**
	11.09.2024

Certified by Supervisor(s)\

This is to certify that this project/thesis is based on the work of the above-mentioned student under my/our supervision. The thesis has been prepared according to the format stipulated and is of an acceptable standard.

Supervisor 1	**Signature**	**Date**
Dr. B.H.R Pushpananda		13 - 09 - 2024

I would like to dedicate this thesis to My Beloved Parents for always being with me with their immense love and support throughout this project.

Thank you for all the sacrifices you made; you made me walk through this amazing endeavour and milestone. I genuinely consider you both to be a blessing to my life.

ACKNOWLEDGEMENTS

First and foremost, I would like to sincerely thank Dr. B.H.R Pushpananda, Senior Lecturer at the University of Colombo School of Computing, who served as my research supervisor. He gave me the go-ahead to carry out the study and gave me vital advice throughout this research. His extensive knowledge and sensible approach to thinking have given me important support, direction, and counsel on how to carry out the research effectively.

I genuinely appreciate all the scholars who helped me with their discoveries in my literature review and pointed me in the right direction for research techniques. I can never say how much I appreciate my family—especially my parents—for giving me the support I required during this journey...

Finally, I would like to thank all the people who supported me to complete the research work, directly or indirectly.

ABSTRACT

Among the ever-growing field of Internet movie reviews, one cannot stress the significance of Internet movie reviews. As digital channels grow, reviews have become an increasingly potent insight into the opinions of viewers, cultivating the narrative surrounding films and impacting the choices made by consumers, studios, and directors alike. The objective of this study is to analyze and assess the sentiments connected to particular aspects or features of movies by using an Aspect-Based Sentiment Analysis (ABSA) technique using IMDb movie reviews. Using machine learning techniques, the trained model classifies movie aspects such as kid-friendliness, character development, directing, acting, story, and music, then categorizes the sentiment connected to each movie aspect. This research executes aspect-based sentiment analysis using sophisticated machine learning models such as Multinomial Naïve Bayes (MNB), Logistic Regression (LR), Support Vector Machines (SVM), KNNeighbors (KNN), Random Forest (RF), Gradient Boosting Machines (GBM) and Multilayer Perceptron (MLP) on a dataset that includes a wide range of user evaluations. The outcomes of this study offer insightful information about the advantages and disadvantages of films from the viewpoint of the audience, which helps filmmakers improve their work and empowers audiences to make wise choices. Additionally, the study looks into how different movie aspects could affect overall user fulfilment, providing insight into those aspects that have a significant impact on the opinions of the audience.

This study contributes to the field of sentiment analysis while also offering filmmakers and movie enthusiasts a valuable tool to help them better comprehend the complex dynamics found in IMDb movie reviews and develop a greater appreciation for the richness of cinematic experiences.

Key Words: Machine Learning, Aspect-based Sentimental Analysis, Sentiment Polarity Classification, Aspect Classification

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER

1 INTRODUCTION

In today's world, movies play a vital role in the entertainment industry. If a person feels stressed and anxious, watching films will help them to cope with their stress. Other than that, movies help to get people motivated to change their lives in a positive direction, allow people to learn new things, especially foreign languages, and help them to deal with difficult situations. A good documentary movie can improve knowledge of important historical events and important information about the outside world.

1.1 Motivation

In the digital age, it is critical to comprehend audience sentiment, which is why researching aspect-based sentiment analysis of IMDb reviews is essential. Given the explosive growth of user-generated material, consumers must interpret complex sentiments regarding particular aspects of film productions to identify what is best for them. The need to piece together the complex web of emotions revealed in user reviews is what motivates this study, which seeks to provide a thorough grasp of how viewers evaluate and respond to different aspects of films and television series. The driving force is closing the knowledge gap between unprocessed text data and valuable insights, enabling consumers to find enjoyable film productions.

1.2 Statement of the problem

With the development of the World Wide Web, the distribution of movies has increased globally. Due to the large number of movie releases yearly, people needed a way to distinguish good films from bad ones. IMDb is one of the popular websites that was introduced solely to address that issue. With the help of the IMDb website, people were able to look into movie reviews and find enjoyable movies to watch. As a result, it is not simply word of mouth that attracts people to the cinemas. These consumer reviews on sites like IMDb play a critical role in the success of films or product sales (Agarwal and Mittal, 2016). But with the increasing number of internet users in today's world, there are large number of reviews for a single film, and one movie review has more than two paragraphs which will take a considerable amount to read.

The concern mentioned above raises the question, ‘How can a simple internet user watch an entertaining movie without spending much time finding (reading a large number of reviews) it?’. Also, some of the most noticeable facts are that people would like to find movies based on the movie’s characters and whether the movie is child-friendly. How attractive the characters in the film are, how well they are developed, and whether the film is child-friendly are some of the major factors people consider for an interesting movie to watch. Sentimental analysis of movie reviews aids in opinion summary by extracting and assessing the reviewers’ feelings (Ikonomakis, Kotsiantis and Tampakas, 2005). Although the reviews include important and valuable material, a new user cannot read all of them and discern whether they are favourable or unfavourable.

In summary, there is a need to identify a way to find an enjoyable film without wasting time. More specifically, the following research questions need to be addressed:

1. What are the most important aspects that people mention in movie reviews, and how do they affect the overall sentiment of the movie review?
2. How accurate is aspect-based sentiment analysis in predicting the sentiment of different aspects of a movie?
3. How do different aspect-based sentiment analysis techniques compare in terms of their accuracy and efficiency in analyzing movie reviews?

1.3 Research Aims and Objectives

1.3.1 Aim

The aim of this project is to create a system that uses an analysis of IMDb movie reviews to assist users in finding enjoyable films. This definition of review analysis is the process of determining a review’s sentiment based on several components of the film. By analyzing the attitudes of the IMDb reviews, the current study seeks to thoroughly evaluate the literature on the subject and develop an approachable system that can generate valuable results. The results of this study will help scholars interested in the field of sentimental analysis as well as regular internet users to advance and enhance aspect-based sentimental analysis of IMDb website evaluations.

1.3.2 Objectives

This research aims to develop a system that helps users find entertaining movies by analyzing movie reviews from the IMDb website. The review analysis is defined herein as the process of identifying the sentiment of a movie review according to different movie aspects. The current study aims to produce a comprehensive review of the literature concerning the sentimental analysis of IMDb reviews and outline a user-friendly system that can provide convenient output by analyzing the sentiments of the IMDb reviews. Specifically, the study has the following sub-objectives:

1. To identify important aspects that help the people to determine the quality of the movie.
2. To train several machine-learning models that can process people's opinions and, produce a sentiment output according to identified movie aspects and determine the best-performing machine-learning model.
3. Incorporate the best-performing machine-learning model and develop a system that can produce a reliable result that can help people to find an enjoyable movie.

The outcome of this research will be beneficial to the everyday internet user as well as the researchers who are interested in the sentimental analysis field to improve and move ahead with the aspect-based sentimental analysis of the IMDb website reviews.

1.4 Scope

The following scope describes the effort put into developing a system that is capable of producing a reliable result by applying aspect-based sentimental analysis on the IMDb website movie reviews. A large number of people in the world look into the movie characters of a film and whether the movie is child-friendly to decide whether the film is enjoyable. Whether the characters in the movie are likeable, the development of the movie characters, and whether the film is child-friendly throughout the movie are some of the factors that stand out the most when people are looking for a film. Other aspects, such as directing, acting, story, and music are also considered in this research.

Moreover, people's opinions about films are categorized as positive, negative, and neutral based on the sentiment. Several machine learning models are created for aspect and sentiment classification to ease the burden of finding an enjoyable film. For the training purposes of the machine learning model, a dataset containing 70,000 movie reviews from the IMDb website will be used. Once all the machine learning models get trained, the best-performing machine-learning model is selected to develop a system that can produce a reliable result that helps people find an enjoyable movie.

This study divides its broad scope into the following tasks in order to simplify it,

1. Research project coordination
2. Data Preparation and Research
3. System Development and Evaluation
4. Research Documentation

1.5 Structure of the Thesis

1.5.1 Literature review

Our goals in this chapter are to locate, assess, and compile important research in the fields of machine learning approaches, aspect-based sentiment analysis, and IMDb reviews sentimental analysis. By showcasing what has been done, what is emerging, what is widely accepted, and what the state of knowledge is currently on aspect-based sentimental analysis, it will shed light on how review analysis using machine learning techniques has developed within the sentimental analysis field. In addition to these, a research gap, such as an area that has not received enough attention or investigation, will be identified with the help of a literature review.

1.5.2 Methodology

We will address the following queries in this chapter: how did I do this research? This will cover the theoretical approach, important techniques, and frameworks utilized in the field of machine learning, as well as the methodologies used to gather and analyze the data collected related to this research requirements. This would guide the selection of techniques as well as the interpretation strategy for aspect-based sentimental

analysis of the gathered data. All things considered, the methodology section provides a thorough manual for carrying out aspect-based sentiment analysis methodically while also bringing transparency and repeatability to the study process.

1.5.3 Evaluation and results

The project's findings are presented and discussed in this chapter. The performance of the applied models is examined in detail in the results section, which also covers the metrics for accuracy, precision, recall, and F1-score. It demonstrates how well the approach captures and categorizes sentiment related to various review aspects. This section also delves into significant patterns and trends that sentiment research has revealed, providing a more comprehensive insight into audience preferences and perceptions. By providing a solid foundation for comprehending audience attitudes, this systematic assessment of aspect-based sentiment analysis of IMDb reviews hopes to support well-informed decision-making processes that depend on user input and preferences.

1.5.4 Conclusion and future works

This research paper's conclusion and future works section summarize key findings, implications, and possible directions for future investigation into aspect-based sentiment analysis applied to IMDb reviews. The study's main findings are first synthesized in the conclusion, together with the knowledge gained about audience attitudes towards particular aspects of motion picture productions. It highlights how important these insights are for guiding decision-making in all sectors of the economy that depend on customer input. Furthermore, potential directions for further study and development in this area are delineated in the section on future studies. It points out things like improving sentiment analysis models for more accuracy, examining how sentiment changes over time, and modifying the process for use on other review-based platforms. These recommendations offer as a road map for scholars and professionals who want to expand on the groundwork our study established.

CHAPTER

2 LITERATURE REVIEW

This research paper's literature review segment provides an in-depth analysis and synthesis of previous academic studies, placing the study in the larger context of sentiment analysis, aspect-based sentiment analysis, and user-generated content analysis. It begins by giving a general review of sentiment analysis approaches, including both conventional methods and more contemporary developments based on machine learning techniques. This lays the groundwork for comprehending how sentiment analysis techniques have evolved and how they are used in various fields.

The literature review then delves into the field of aspect-based sentiment analysis, examining research and theoretical models aimed at analyzing opinions about particular aspects or characteristics in reviews. It clarifies a number of methods, including aspect extraction strategies, sentiment classification models, and difficulties that arise in this specific type of sentiment analysis. Additionally, the section explores the particular field of user-generated content analysis in the context of entertainment portals such as IMDb. It summarizes previous studies on sentiment analysis of reviews for films and television shows, emphasizing significant discoveries, approaches, and research constraints found in these studies.

2.1 Related Works

This section will examine the evolution of aspect-based sentimental analysis within the Machine Learning field of study, highlighting previous research discoveries.

2.1.1 Sentimental Analysis

Over the last few decades, Sentiment Analysis has gained popularity as a field of study in the Machine Learning (ML) realm. Multiple research on Sentiment Analysis has been carried out targeting several domains such as online user reviews (movies, TV series). There is a significant amount of multiple research conducted in the area of study, sentimental analysis of IMDb Reviews. (Kumar and Benitta, 2022) used an approach to sentiment analysis through the use of deep learning techniques, a combination of long short memory (LSTM) and recurrent neural network (RNN) used to classify the sentiments with a high degree of accuracy in a short period to predict

the classification of reviews from the IMDb dataset. Another example is the research study that was conducted with the title of Application of Machine Learning for Sentiment Analysis of Movies Using IMDB Rating (Rathor and Prakash, 2022) by Sandeep Rathor and Yuvraj Prakash to identify a framework for applying machine learning and data mining methods to the analysis of customer sentiment. This study presents a method for doing sentiment analysis on online user reviews using supervised machine learning classifiers to help users choose based on the popularity and interest of the reviews.

Moreover, the research conducted on Sentiment Analysis on IMDb Movie Reviews, such as (Shaukat *et al.*, 2020) and (Tarimer, Coban and Kocaman, 2019), all focus on the overall sentimental analysis of online reviews. A more sophisticated method of evaluating IMDb reviews is becoming more and more in demand as people become aware of the shortcomings of traditional sentiment analysis. This acknowledgement has spurred interest in examining certain review aspects, opening the door to a more thorough and in-depth comprehension of audience opinions. This led to the introduction of aspect-based sentiment analysis (ABSA), which enables a more detailed examination of particular components inside IMDb evaluations.

2.1.2 Aspect-based Sentimental Analysis

A significant paradigm in the field of Machine Learning (ML) is Aspect-based Sentiment Analysis (ABSA), which provides a detailed and nuanced assessment of how users feel about different aspects of a review. Because Overall Sentiment Analysis usually delivers an overall favourable or unfavourable categorization without exploring the underlying components leading to these feelings, it is frequently unable to capture the nuances of thoughts stated in reviews. One of Aspect-based Sentiment Analysis's main benefits over Sentiment Analysis is its capacity to offer a thorough analysis of certain features, illuminating both the positive and negative feelings connected to each component. For example, a film may be praised for having a gripping storyline yet criticized for having an inferior performance. For filmmakers and other industry professionals looking to improve their production and creative processes, this level of information is crucial.

The researchers such as (Kumar and Garg, 2020) have introduced a way to use deep learning Convolutional Neural Networks (CNN) for the classification of sentiment analysis by aspects base ontology. Four other researchers (Tripathy et al., 2019) came together to introduce an aspect-based approach for document-level sentiment classification of movie reviews, which helps the viewer to identify the specific movie to watch. (Hu and Liu, 2004) work highlights the significance of extracting views from user-generated information, and ABSA is a logical step forward in this regard by offering a methodical framework for sentiment extraction related to different aspects.

2.1.3 Feature Extraction

A crucial element of aspect-based sentiment analysis (ABSA) is aspect extraction, which finds and examines particular elements or characteristics of a review to enable a more detailed comprehension of the sentiments expressed in user reviews. Through the process of extracting features that people bring up in their reviews, sentiments about certain features like storyline, cinematography, and soundtrack can be thoroughly examined. According to (Hammi, Hammami and Belguith, 2022), feature extraction can be implemented using three feature extraction methods depicted in Figure 1.

1. Named Entity Recognition (NER)
2. Topic Modelling
3. Hand-crafted Lexicon

Named Entity Recognition (NER) seeks to identify and classify substantial information chunks in text, including people, places, organizations, and so on. The second feature extraction method, Topic Modelling, uses word co-occurrence analysis to reveal hidden topics inside text. Since it is unsupervised, labelled data is not required. It is capable of identifying some features of sentiment analysis; nevertheless, polysemy and domain dependency pose obstacles. Lastly, aspect extraction uses Hand-crafted Lexicons to construct lists of terms associated with particular features (e.g., lighting for location). Even though they are precise and domain-specific, they take a lot of work to develop and maintain if features expand significantly.

The study conducted for Sentiment Analysis on IMDb Movie Reviews (Kumar, Harish and Darshan, 2019) has identified a method for using hybrid features, which are derived by concatenating Lexicon features (Positive-Negative Word Count, Connotation) with Machine Learning features (TF, TF-IDF) to more reliably determine the overall polarity of user reviews. (Ur Rehman *et al.*, 2019) conducted an exploration to build a hybrid CNN-LSTM model for improving the accuracy of movie review sentiment analysis. This hybrid model uses a profound architecture of convolutional layers to extract the local features of a text. The researchers have used the Convolutional Neural Network (CNN) layer to extract the high-level features, and the Long Short-Term Memory (LSTM) layer detects long-term dependencies between words.

2.1.4 Sentiment Classification at the Feature Level

At the aspect level of sentiment classification, perspectives and feelings represented in a text are examined with an emphasis on specific features or characteristics of the subject. A more sophisticated understanding of sentiment in many domains, such as online user reviews, is made possible by this fine-grained approach. According to (Tripathy *et al.*, 2019), Sentiment Analysis can be performed at several levels. The following types of sentiment classification processes are possible:

1. Binary
    - The method is referred to as binary since just two classes are taken into consideration when the system's output is thought of as either positive or negative kinds. Considerations of approved and rejected classes are possible for positive and negative classes, respectively.
2. Multi-class
    - Multi-class refers to a system's output that has more than two classes in it. This kind of classification is generally applied when reviews are divided into many categories, such as positive, negative, and neutral.

For sentiment categorization at the aspect level, researchers use a variety of methods, such as machine learning models like Support Vector Machines (SVM)(Ramadhan and Ramadhan, 2022), Logistic Regression (Banerjee, Mazumder and Datta, 2022) and Convolutional Neural Networks (RNN) (Başarslan and Kayaalp, 2023). These

models are able to precisely categorize sentiments at a granular level because they are trained on annotated datasets that describe attitudes associated with specific features. Let's examine some extra past research on sentiment classification at the feature level and dive into studies that look at sentiment analysis at a more detailed level. These investigations use advanced machine learning methods to examine reviews at a fine point, providing insightful information for focused enhancements in a variety of applications like recommendation systems.

The study conducted by (Onalaja, Romero and Yun, 2021) examines a comparison of classification models used to detect aspect-based separated text sentiment and predict binary sentiments of movie reviews with genre and aspect-specific driving factors. This research used five machine and deep learning algorithms for extensive classification analysis. The algorithms that were used are Logistic Regression (LR), Naïve Bayes (NB), Support Vector Machine (SVM), and Recurrent Neural Network Long-Short-Term Memory (RNN LSTM). The authors of this research have covered a more in-depth understanding of sentiment analysis, including various aspects of movie reviews such as casts, music, location, technology, and quality. These aspects will help when reviewers are interested in the quality or cast of a movie. In this research, reviews are split and reduced by the dominant aspect, specific for each review. Then, the text information is converted into vectors as the input for the machine-learning models mentioned above to produce the outcome. The researchers created a dataset of 3,000 movie reviews by scraping the IMDb website. The dataset was created with an equal number of positive and negative reviews. In this research, reviews are split and reduced by the dominant aspect, specific for each review. Then, the text information is converted into vectors as the input for the machine-learning mentioned above models to produce the outcome. The authors were able to observe 67% as the highest accuracy of this research.

A research study was conducted for the classification of movie reviews using the term frequency-inverse document frequency and optimized machine learning algorithms (Naeem *et al.,* 2022). This study presents a method for doing sentiment analysis on movie reviews using supervised machine learning classifiers to help users choose movies based on the popularity and interest of the reviews. The researchers of this paper used four machine learning methods for sentiment analysis, including Decision

Tree (DT), Random Forest (RF), Gradient Boosting (GBC), and Support Vector Machine (SVM), which are trained on a preprocessed dataset. Furthermore, the usefulness of four feature extraction techniques, including BoW, TF-IDF, GloVe, and Word2Vec, in extracting relevant and effective features from reviews is evaluated in this research. The study demonstrates that the sentiment prediction models used in this research were more accurate when more significant driving variables were assigned to particular aspects and movie genres. When trained and assessed, SVM obtains the most excellent accuracy of any classifier, with an accuracy of 89%.

The study conducted for Sentimental Analysis based on IMDB aspects from movie reviews using SVM (Support Vector Machine) (Ramadhan and Ramadhan, 2022) has provided a way to get the sentiment opinion of movie reviews using IMDb website star rating and SVM method. With the help of the IMDb website rating, the sentiment was set as positive for a rating greater than five, and negative was set for a rating lower than five. This study uses a movie review dataset from the squid game film. This dataset uses a random sampling technique as the data retrieval technique. First, the dataset is preprocessed, undergoing the Tokenization, removing non-alphabetics and stemming as the preprocessing steps. Then, the dataset is used for the classification of a support vector machine model with the selected kernel, which is linear. The dataset is divided into 70% training data and 30% testing data to get the results of accuracy, precision, and recall of the trained model. The trained classification model was able to produce 79% accuracy, 75% precision, and 87% recall after training.

A research study was conducted with the title of Importance Evaluation of Movie Aspects: Aspect Based Sentiment Analysis (Wang, Shen and Hu, 2020) by researchers named Yanqing Wang, Gufeng Shen, and Liangyu Hu to identify the most important aspect of a movie. The experiment was conducted with a dataset containing movie reviews of more than 19000 for 1000 movies. The dataset was preprocessed with the following steps: remove punctuation and numbers, remove stop words, and stemming as the preprocessing steps. Once the data is preprocessed, the Sentiment Intensity Analyzer in VADER is used for classification. VADER makes use of human-validated, valence-based, gold-standard lexicons. Preprocessed data is the input of the VADER classification algorithm, and the ratings based on compound sentiment scores

are the output. Researchers considered the linear relationship between aspect ratings and overall scores to assess the importance of the following aspects of the movie: actors, directors, plot, and music. As a result, correlation coefficients between each aspect's score and the overall evaluations are determined. The authors were able to observe 81% as the highest accuracy and recall as 80% of this research.

An approach to aspect-based sentiment analysis through the use of deep learning was introduced for aspect prediction and sentiment prediction of a given review (Bo and Min, 2021). This study used two training datasets, each including about 550 reviews of laptops and restaurants, annotated with the relevant characteristics and polarity. Only subtasks pertinent to the restaurant domain are included in the following sections, as researchers have concentrated on the out-of-domain and in-domain ABSA in this experiment. An aspect model and a sentiment model comprise the two components of the proposed system's breakdown. A two-layer neural network has been used by researchers to create the aspect model. Convolutional neural networks (CNNs), which only require sentiment at a sentence level, have been chosen for the sentiment model. An input sentence is fed into the aspect model, which produces an output probabilistic distribution across the aspects. Sentiment analysis uses the sentiment model to extract the sentiment from a given sentence. The authors were able to observe 52.6%, 50.1%, and 51.3% for Precision, Recall, and F1 Measure of the trained model.

Research called 'Comparative Study on Sentiment Analysis on IMDB Dataset' (Banerjee, Mazumder and Datta, 2022) was conducted by three researchers to get a comprehensive qualitative understanding of different facets of the movie. The sentiment analysis experiment was performed using a dataset that contained 50000 data records. Lower case conversion, HTML tag removal, emoji removal, expanding the contraction, punctuation removal, number removal, URL removal, stop words removal, and lemmatization were executed as the preprocessing steps of the dataset. The dataset was split into two halves by the authors, with 80% of the data being utilized for training and 20% for testing. The researchers of this paper used five machine learning methods for sentiment analysis, including KNNeighbors (KNN), Random Forest (RF), Decision Tree (DT), Logistic Regression (LR), and Support Vector Machine (SVM), which are trained on the preprocessed dataset. When

compared to other classifiers, the LR classifier performs better on the performance evaluation across the board for all performance metrics. LR outperforms SVM, KNN, DT, and RF in terms of accuracy (74%), sensitivity (83%), and specificity (76%) in that order.

Finally, a thorough examination of the several research projects devoted to Sentiment Classification at the Feature Level reveals a diverse range of approaches. Prior research work focuses mostly on sentiment analysis with small datasets through aspect-based methods. By using a larger dataset, our work deviates from the others. Critical elements like character development and movie kid-friendliness have also been overlooked in previous research. It is a significant departure in methodology as our approach combines aspect and sentiment analysis into a single model, unlike previous research that uses separate models for each task. We are motivated to investigate a more complete and nuanced analysis paradigm based on these gaps in the literature.

2.2 Presentation of Scientific Material

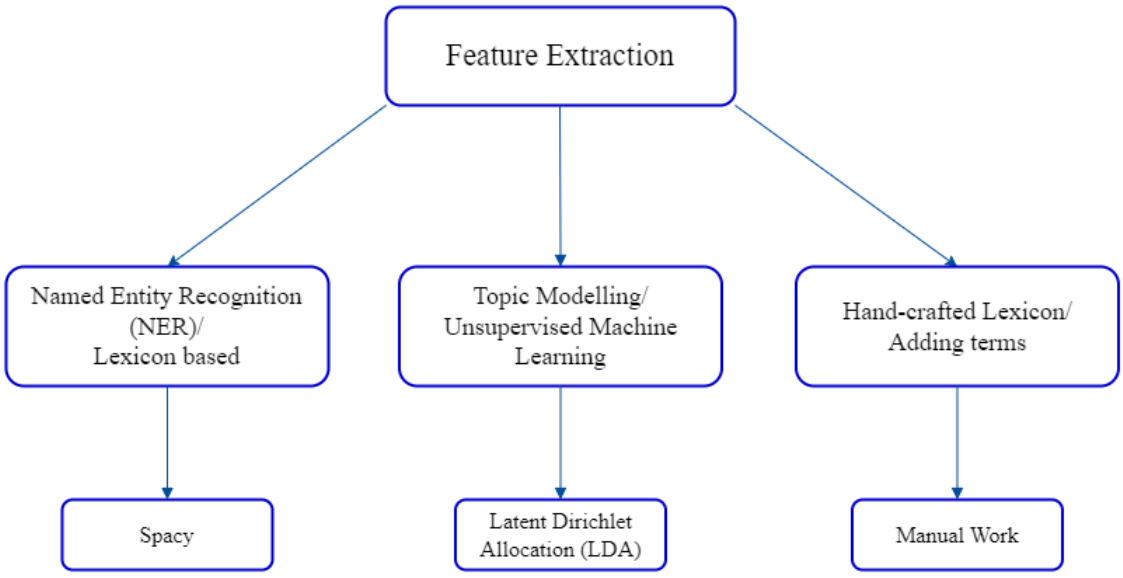


Figure 2-1 - Three Different Methods of Feature Extraction

Table 2-1 - Summary of Multiple Aspect-based Sentiment Analysis Research

Author	*Research Paper Title*	*Dataset*	*Result (%)*
S. Onalaja, E. Romero 2021	Aspect-based Sentiment Analysis of Movie Reviews	IMDb Dataset with 3000 data records	Precision: Recall: F1 score: Accuracy: 67.19
A. Tripathy, Ch. C. Rao, P. Jha, G. Satyanarayana 2019	Aspect-based approach for document-level sentiment classification of movie reviews	IMDb Dataset with less than 23000 data records	Precision: 87.5 Recall: 92 F1 score: Accuracy: 89.88
N. G. Ramadhan, T. I. Ramadhan 2022	Analysis Sentiment based on IMDB aspects from movie reviews using SVM	IMDb Dataset with less than 3000 data records	Precision: 75 Recall: 87 F1 score: Accuracy: 79
Y.Wang, G. Shen, L. Hu 2020	Importance Evaluation of Movie Aspects: Aspect-based Sentiment Analysis	IMDb Dataset containing around 19000 data records	Precision: Recall: 80 F1 score: Accuracy: 81
B. Wang, M. Liu 2021	Deep Learning for Aspect-Based Sentiment Analysis	SemEval-2015 Task 12 dataset containing around 550 data records	Precision: 52.6 Recall: 50.1 F1 score:51.3 Accuracy:
H. M. K. Kumar, B. S. Harish, H. K. Darshan 2019	Sentiment Analysis on IMDb Movie Reviews Using Hybrid Feature Extraction Method	IMDb Dataset with 5000 data records	Precision: Recall: F1 score:83.7 Accuracy: 83.9
D. Banerjee, S. Mazumder, S. Datta 2022	Comparative Study on Sentiment Analysis on IMDB Dataset	IMDb Dataset with 50000 data records	Precision: Recall: 83 F1 score: Accuracy: 74

CHAPTER

3 METHODOLOGY

This chapter outlines the proposed solution for the research problem that has been addressed in this research. The proposed solution includes data collection, preprocessing, aspect extraction, and sentiment classification. Each action taken is thoroughly explained later in this chapter.

3.1 Research Methodology

This study explores a thorough methodology that includes data gathering, data preprocessing, aspect term extraction, and sentiment and aspect category categorization. In order to improve the accuracy of the aspect-based sentimental analysis that follows, the first stage is careful data preprocessing. After that, aspect extraction is carried out using natural language processing (NLP) methods in order to determine which aspect is more prevalent, which is often denoted by a noun or noun phrase. Subsequently, this study integrates sophisticated methods for classifying sentiment and aspect categories by augmenting the sentiment analysis structure. After aspect words are extracted together with their orientations, the system uses machine learning techniques to categorize sentiments into neutral, positive, or negative. This categorization helps to quantify and classify the general attitude that is conveyed toward every recognized aspect. At the same time, aspect category classification entails classifying every extracted aspect into predetermined categories such as kid-friendly, character, acting, directing, story, music, and others, which facilitates an organized arrangement of the many aspects that are being examined.

Techniques based on machine learning are used in both sentiment classification and aspect category classification in order to determine the aspect category and underlying sentiment polarity. This dual-classification method allows for a more structured and comprehensible study by methodically organizing views around particular features in addition to improving the sentiment analysis output. The objective of this research is to improve the granularity and depth of the sentiment analysis process by introducing several new aspects that are popular in the IMDb movie reviews. The utilization of machine learning models guarantees flexibility with respect to various datasets and promotes a more robust understanding of the complex interrelationships between sentiments and aspects in the text under analysis. The research will

build upon a proposed solution encapsulated in the flowchart below, specifically centred around the development and implementation of several aspect-based sentiment analysis models.

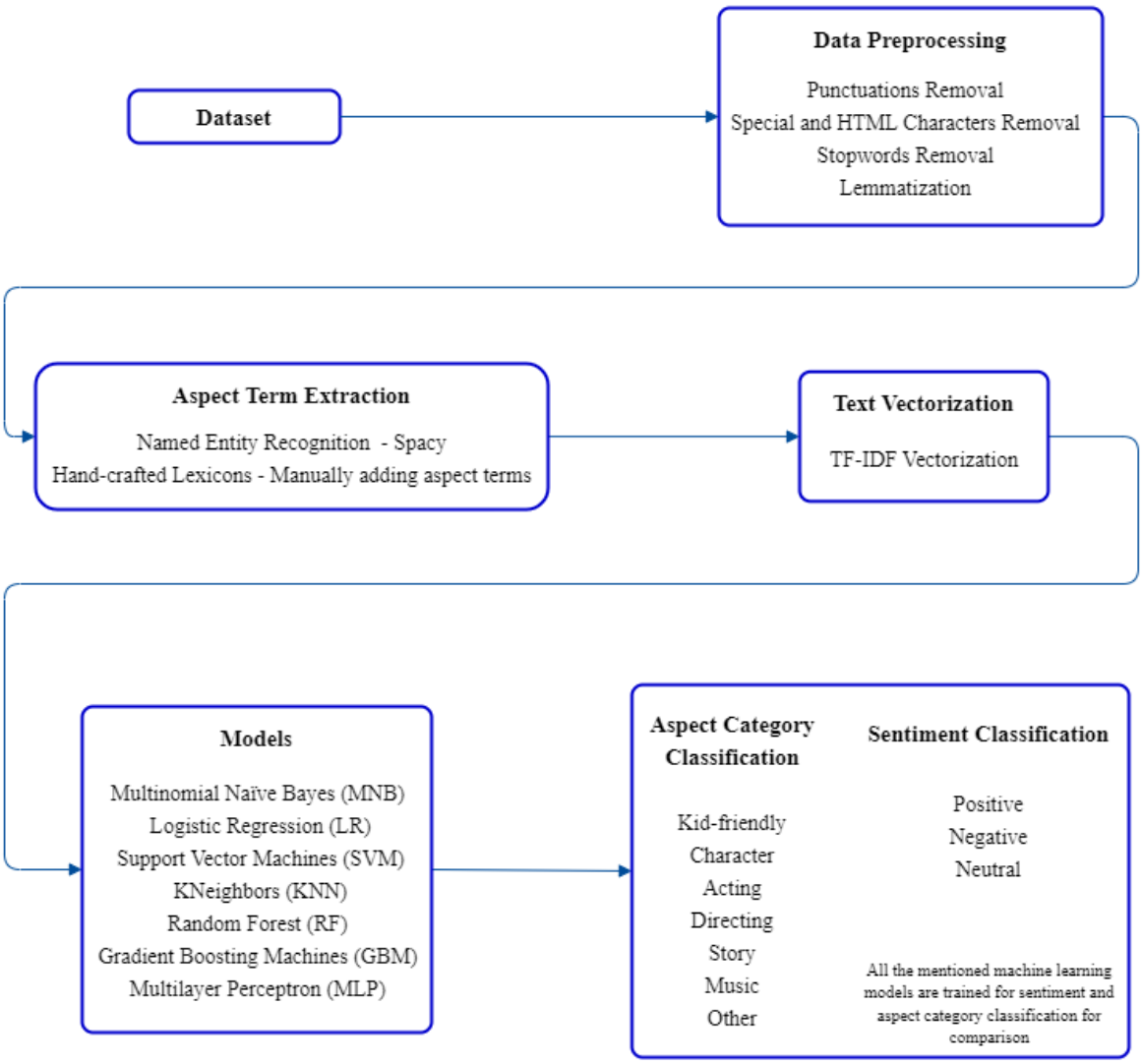


Figure 3-1 - Proposed Solution for the Research Problem

3.2 Data Collection

The proposed research uses the ACL Internet Movie Database (IMDb) dataset (Maas *et al.*, 2011) as the data collection source. This dataset was created for learning word vectors. The dataset consists of 50,000 textual reviews of movies from the IMDb website. This IMDb dataset stands as a pivotal resource in the realm of sentiment analysis. This dataset is extremely significant since it contains a large collection of user-submitted movie reviews from the IMDb website, which represent a wide variety of opinions. Realistic user-generated material is one of the IMDb dataset's most notable features. The genuine variety seen in the viewpoints of internet users is reflected in the vast range of writing styles and sentiment expressions found in these reviews. The realism of the dataset makes it more applicable to real-world situations, where sentiment analysis models are frequently used to examine user-generated content on websites, discussion boards, and social media.

3.3 Data Preprocessing

An essential first step in getting raw data ready for analysis is data preparation. Cleaning and transformation are required in order to improve the quality and utility of the dataset. By lowering noise and inconsistencies, data preparation improves data quality and usability and establishes the groundwork for efficient analysis and modelling, which raises the precision and dependability of ensuing data-driven activities.

3.3.1 Lowercase Conversion and Punctuation Removal

Lowercasing ensures consistency and simplifies the data by changing all of the text's letters to lowercase. Punctuation removal ensures that the non-semantic symbols are removed from the text, improving text clarity and expediting future natural language processing actions.

3.3.2 Special and HTML Characters Removal

Eliminating HTML and special characters from web-based sources guarantees cleaner text data for analysis and modelling while avoiding any disruptions to subsequent processes.

3.3.3 Stopwords Removal

The technique of removing stopwords refines text analysis by removing common but uninformative words, enabling a more concentrated investigation of significant material. The nltk.corpus package will be used to eliminate stopwords.

3.3.4 Lemmatization

Lemmatization ensures consistency by reducing words to their most basic form. This improves text coherence and allows for more meaningful analysis in tasks involving natural language processing by reducing word variants to their root forms.

3.4 Aspect Term Extraction

A crucial sentiment analysis component is called Aspect Term Extraction, which is finding certain phrases or entities in a text that correspond to the characteristics or elements under discussion. Using the Spacy package to leverage Named Entity Recognition (NER) is one method of Aspect Term Extraction. A natural language processing library called Spacy makes it easier to accurately identify items like people, places, and various types of named entities. Spacy's pre-trained models make Aspect Term Extraction more effective by enabling the automatic identification of context-relevant elements. Hand-crafted Lexicons are a valuable technique for Aspect Term Extraction when used in conjunction with NER via Spacy. With this manual method, collections of words associated with certain interest areas or characteristics are curated. These lexicons act as domain-specific knowledge bases, which improve the model's recognition of aspect phrases that might not be picked up by models that have just been trained before. Adding keywords linked to certain aspects is the primary step involved in creating a Hand-Crafted Lexicon.

The proposed research solution uses a dual-strength approach to aspect term extraction by using Spacy for NER and Hand-Crafted Lexicons. Spacy offers efficiency and generalization, while Hand-Crafted Lexicons provide domain-specific expertise by identifying subtleties that statistical models could overlook. The overall effectiveness of sentiment analysis models in comprehending and interpreting user opinions across a variety of domains is improved by this hybrid technique, which enables a more thorough and accurate extraction of aspect words.

3.5 Text Vectorization

An essential stage in natural language processing is text vectorization using Term Frequency-Inverse Document Frequency (TF-IDF), which allows textual data to be represented in a numerical manner appropriate for machine learning methods. TF-IDF prioritizes phrases that are important inside a text and distinct throughout the dataset by allocating weights to words based on their frequency in a particular document in relation to their occurrence throughout the whole corpus. The Term Frequency (TF) component calculates a term's frequency in a document and offers context-specific information. Conversely, the Inverse Document Frequency (IDF) component gives phrases that are less common across documents a higher weight by evaluating a term's uniqueness over the whole dataset. By combining TF and IDF, a vector representation is produced that accounts for the broader relevance of terms in a document while capturing their significance.

3.6 Sentiment and Aspect Category Classification

A vital task in natural language processing is Sentiment and Aspect Category Classification, which tries to identify the sentiment that is communicated as well as the particular aspect category that is stated in a text. Advanced machine learning classifiers, such as Multinomial Naïve Bayes (MNB), Logistic Regression (LR), Support Vector Machines (SVM), KNNeighbors (KNN), Random Forest (RF), Gradient Boosting Machines (GBM) and Multilayer Perceptron (MLP) are used to accomplish this task. Because these classifiers are trained on labelled datasets, they can recognize aspect categories and predict sentiment polarity with accuracy. Making use of these classifiers' advantages guarantees the development of a robust model that can extract subtle aspects and sentiment information from a variety of textual data sources.

3.6.1 Multinomial Naïve Bayes (MNB)

For text classification problems, Multinomial Naïve Bayes (MNB) is a well-liked probabilistic machine learning technique. It is a member of the Naïve Bayes classifier family, which is predicated on feature independence and the Bayes theorem. MNB is widely used in natural language processing applications like sentiment analysis and document categorization. It was created explicitly for data with discrete characteristics. In reality, MNB frequently exhibits astonishingly good performance,

particularly when handling sparse and high-dimensional datasets common to text classification issues. One of MNB's main advantages over more complicated models is that it is computationally inexpensive due to its efficiency and simplicity. It scales effectively as the number of features rises and is especially well-suited for jobs involving massive datasets.

3.6.2 Logistic Regression (LR)

One of the most popular and fundamental machine learning algorithms is logistic regression (LR), especially for its simplicity and effectiveness in modelling the probability of an event occurring. As the name suggests, LR is used for classification as opposed to regression. Using the logistic (sigmoid) function to restrict the output between 0 and 1, it simulates the possibility that an instance belongs to a specific class. LR works especially effectively in situations when there is a roughly linear relationship between the binary output and the aspects. Interpretability, which is the ability to determine how each attribute affects the probability of the desired result, is one of LR's main advantages. Due to its computational efficiency and ability to handle large datasets, LR is the preferred method for situations involving feature spaces that range from moderate to high in size.

3.6.3 Support Vector Machines (SVM)

One effective and adaptable machine learning approach for both regression and classification applications is Support Vector Machines (SVM). In order to function, SVM searches a high-dimensional space for the ideal hyperplane that optimally divides data points belonging to various classes. Support vectors determine this hyperplane, which guarantees strong generalization to unknown data. SVM, also known as the kernel technique, performs very well in situations when the data is not linearly separable by projecting it into a higher-dimensional space. SVM is a powerful tool for tasks like text categorization and picture classification because of its ability to handle high-dimensional data effectively. Additionally, the technique offers a range of flexible kernel function choices, including non-linear and linear ones, which enables it to capture intricate correlations in a variety of datasets.

3.6.4 KNNeighbors (KNN)

A versatile and user-friendly machine learning approach for classification and regression applications is KNNeighbors (KNN). Based on the majority class or average of the K-nearest data points in the feature space, KNN predicts things based on the proximity principle. Since it doesn't presume a particular underlying probability distribution for the data, this approach is non-parametric. KNN is widely used because of its simplicity and ease of implementation, particularly in situations where interpretability is critical. The performance of KNN is greatly influenced by the selection of the parameter K, which stands for the number of neighbors. A smoother decision boundary is produced by a bigger K, although over-smoothing might occur if the K is too large. A smaller K produces a more sensitive model. When data show local patterns and decision boundaries are irregular, KNN performs exceptionally well. Due to the issue of dimensionality, KNN may present difficulties when dealing with high-dimensional data, necessitating meticulous preparation or dimensionality reduction strategies.

3.6.5 Random Forest (RF)

Renowned for its resilience and adaptability in both classification and regression applications, Random Forest (RF) is a potent ensemble learning method. Built on the basis of decision trees, RF combines the predictions of many trees in order to reduce overfitting and improve accuracy. In order to promote variety across the component trees, the approach adds randomization by training each tree on a part of the dataset and using a random feature selection for each split. The capacity of Random Forest to handle high-dimensional datasets and yield insightful information about feature relevance is one of its most remarkable strengths. It promotes generalization to fresh data by efficiently mitigating the overfitting propensity of individual decision trees. Random Forest is widely used in many different fields because of its interpretability, resistance to outliers, and predictive effectiveness. Furthermore, Random Forest is appropriate for real-world datasets with insufficient information since it provides a built-in technique for addressing missing data. The technique has become a well-liked machine learning algorithm due to its adaptability to capture intricate links and interactions within the data.

3.6.6 Gradient Boosting Machines (GBM)

Gradient Boosting Machines (GBM) is an advanced ensemble learning method that performs exceptionally well in predictive modelling, showing great results in both regression and classification problems. GBM progressively assembles a collection of weak learners, usually decision trees, to create a powerful prediction model. With an emphasis on reducing the total residual error, each tree fixes the mistakes made by its predecessor. A strong and precise prediction model is produced by this repeated boosting procedure. The secret to GBM's success is its capacity to properly handle both numerical and categorical information and to grasp intricate connections within the data. It uses the optimization method of gradient descent, changing each tree's weight in order to minimize the loss function. GBM is very good at handling non-linear patterns and interactions in the data because of this method. Although it might be prone to overfitting if not properly adjusted, GBM provides regularization and tree complexity control settings. GBM is a well-liked machine learning method because of its capacity for handling big datasets and good predicted accuracy. GBM's interpretability of feature significance further increases its usefulness in a variety of tasks.

3.6.7 Multilayer Perceptron (MLP)

An essential part of deep learning, artificial neural networks are built on the Multilayer Perceptron (MLP). Multilayer pyramids (MLPs) are composed of several layers of linked nodes, or neurons, with an input layer, one or more hidden layers, and an output layer. Every node has an activation function, which is usually non-linear, and each link has a weight. This complexity allows the model to identify complicated patterns in the data. MLPs are incredibly flexible and can handle challenging jobs in both regression and classification. Because of its hierarchical nature, representation learning is made easier by the ability to extract hierarchical features from the incoming data. Backpropagation is used in MLP training, where mistakes are progressively reduced by modifying the weights using optimization techniques such as stochastic gradient descent.

CHAPTER
4 EVALUATION AND RESULTS

This chapter describes the dataset that was used to create the machine learning models, the experiments that were carried out throughout the final system's implementation, the outcomes of various approaches, and an assessment of those outcomes.

4.1 Dataset

This research uses the ACL Internet Movie Database (IMDb) dataset (Maas et al., 2011) as the data collection source. Movie reviews from the ACL dataset were broken down to sentence level, and aspect categories and sentiments were collected using crowdsourcing. The finalized dataset contained 10000 data records for each aspect category, totalling up to 70000 data records. This dataset is balanced accordingly since dataset balancing makes training the machine-learning model easier since it helps to prevent the trained model from becoming biased towards one specific class. To put it simply, the machine-learning model will no longer favour the majority class because it contains more data. That will help the machine-learning model produce more accurate results when trained. When annotating the dataset, aspect categories were represented by kid-friendly, character, acting, directing, story, music, and other, while the sentiment polarity was represented by neutral, positive, or negative. Then, in the following percentages, this annotated dataset is split into a training set and a testing set. Data for testing = 0.20, data for training = 0.80

4.2 Evaluation Metrics

Model assessment metrics are essential for assessing the effectiveness of prediction models in the field of machine learning. These metrics are essential tools that provide a quantitative understanding of the model's performance and ability to generalize to new data. It is necessary to use metrics that go beyond a simple binary evaluation of accuracy since models are meant to generate predictions and classifications. The careful choice of these metrics depends on the particular objectives and complexities of the machine learning activity that is being performed, guaranteeing a thorough comprehension of the model's performance in practical scenarios. The evaluation measures such as Accuracy, Recall, Precision, and F1-Score are often used to identify the machine learning model's performance.

4.2.1 Accuracy

When evaluating a machine learning model, accuracy is a fundamental indicator that shows the percentage of adequately predicted instances in the dataset. Although it offers a simple indicator of general accuracy, situations with unequal class distributions may restrict its usefulness. When one class far dominates the other, forecasting the majority class alone may provide a high accuracy while ignoring the model's ability to distinguish minority class occurrences. Accuracy can be calculated using the formula given below.

$$\text{Accuracy} = \frac{\text{True Positive} + \text{True Negative}}{\text{True Positive} + \text{True Negative} + \text{False Positive} + \text{False Negative}}$$

Figure 4-1 - Machine Learning Model Accuracy Formula

4.2.2 Recall

Recall is a crucial assessment parameter in machine learning, often referred to as Sensitivity or True Positive Rate. It measures a model's precision in identifying every occurrence of a given class in a dataset. Recall gauges explicitly the proportion of accurate positive predictions to the total number of actual positive cases, highlighting the model's capacity to catch all pertinent occurrences. Recall values that are high mean that the model performs well in identifying examples of the positive class, reducing the possibility of false negatives. This statistic is especially useful in situations where the model must maintain a sensitive and thorough approach to finding pertinent patterns in the data since missing positive examples might have serious repercussions. Recall can be calculated using the formula given below.

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}}$$

Figure 4-2 - Machine Learning Model Recall Formula

4.2.3 Precision

One of the most important metrics for assessing a model's ability to make accurate positive predictions is precision. It illustrates the accuracy of the model in properly identifying positive cases and is calculated as the ratio of true positive predictions to the total number of instances predicted as positive. Precision evaluates the model's ability to avoid misclassifying negative occurrences as positive, which is important in cases where minimizing false positives is critical. Precision helps to provide a more nuanced view of the model's performance by offering insights into the dependability of positive predictions. This is helpful in situations where the repercussions of false positives are substantial. Precision is also known as the Positive Predictive Value (PPV), which can be calculated as:

$$\textbf{Precision} = \frac{\text{True Positive}}{\text{True Positive + False Positive}}$$

Figure 4-3 - Machine Learning Model Precision Formula

4.2.4 F1-Score

The F1-Score is a harmonized statistic that evaluates a model's overall performance by combining Precision and Recall into a single measurement. The F1-Score, which is determined as the harmonic mean of memory and accuracy, finds a balance between the trade-offs that come with recall and precision. This statistic is especially helpful in situations when it's crucial to strike a balance between false positives and false negatives. Because it takes into account both dimensions, the F1-Score offers a thorough assessment that is ideal for uses where recall and precision are equally important. The F1 score, also known as the F-measure, can be calculated using the formula given below.

$$\textbf{F-measure} = \frac{2 * (\text{Precision} * \text{Recall})}{\text{Precision + Recall}}$$

Figure 4-4 - Machine Learning Model F-measure Formula

4.3 Evaluation of Machine Learning Models

An essential first step in evaluating a machine learning model's performance and generalization ability is to evaluate the training outcomes. For assessing the efficacy of the model entails using the previously indicated metrics: accuracy, precision, recall, and F1-score. The assessment phase provides insight into the model's capacity for accurate prediction on newly discovered data as well as how well it has learned from the training set. Advanced machine learning classifiers, namely Multinomial Naïve Bayes (MNB), Logistic Regression (LR), Support Vector Machines with Non-Linear Kernel (SVM Non-Linear), Support Vector Machines with Linear Kernel (SVM Linear), KNNeighbors (KNN), Random Forest (RF), Gradient Boosting Machines (GBM) and Multilayer Perceptron (MLP) are used to train eight machine learning models which are evaluated below.

4.3.1 Multinomial Naïve Bayes (MNB) Evaluation

The following code was employed to retrieve the evaluation parameters of the model.

```
sentiment_classifier = Pipeline(steps=[
                                ('pre_processing',preprocessor),
                                 ('nb_multi',MultiOutputClassifier(MultinomialNB()))
                                 ])

sentiment_classifier.fit(x_train,y_train)

print("MultinomialNB Evaluation ")
print("Accuracy " + str(sentiment_classifier.score(x_test,y_test)))

y_pred = sentiment_classifier.predict(x_test)

labels = ['aspect_category','polarity']
f1_scores = {}
precision_scores = {}
recall_scores = {}

for idx, label in enumerate(labels):
    f1_scores[label] = f1_score(y_test[label], y_pred[:, idx], average="macro")
    precision_scores[label] = precision_score(y_test[label], y_pred[:, idx], average="macro")
    recall_scores[label] = recall_score(y_test[label], y_pred[:, idx], average="macro")

average_f1 = np.mean(list(f1_scores.values()))
print(f"Average F1-score: {average_f1:.4f}")

average_precision = np.mean(list(precision_scores.values()))
print(f"Average Precision: {average_precision:.4f}")

average_recall = np.mean(list(recall_scores.values()))
print(f"Average Recall: {average_recall:.4f}")
```

Figure 4-5 - Code segment for calculating MNB model evaluation parameters

The Multinomial Naïve Bayes Model evaluation metric results are presented below, offering a concise yet informative overview of the model's performance across diverse metrics.

Table 4-1 - Multinomial Naïve Bayes Model Results

Evaluation Metric	*Score (%)*
Accuracy	59.75
Recall	75.89
Precision	77.83
F1 Score	75.50

Under several circumstances, the trained model has trouble classifying aspects correctly. For example, the algorithm classifies the review "I especially loved how the plot was so unpredictable, and the characters were like real-life superheroes." under the story aspect, ignoring the related character aspect category. This misunderstanding highlights how difficult it is to correctly identify and classify components, particularly when they blend together to form a single statement that expresses a range of sentiments. It highlights the difficult task of precisely identifying and categorizing subtleties, exposing the intricacy of interpreting complex expressions that capture a range of emotional tones in a single context. In the review, "However, I couldn't help but feel the story lacked a certain depth that could have made it truly exceptional." the trained model incorrectly labelled the story aspect with positive sentiment, neglecting the negative sentiment expressed for the story aspect earlier. The positive sentiment overshadows the subtle critique, making it challenging for a trained model to capture the mixed sentiment accurately. Despite that, the trained model was able to successfully classify the aspect category and the sentiment in a less complex review: "The acting was phenomenal, and the actors were truly outstanding in their performances." under the acting aspect and with a positive sentiment.

4.3.2 Logistic Regression (LR) Evaluation

The model's evaluation parameters were obtained by utilizing the subsequent code.

```
sentiment_classifier = Pipeline(steps=[
                                ('pre_processing',preprocessor),
                                 ('lr_multi',MultiOutputClassifier(LogisticRegression(max_iter=2000)))
                                 ])

sentiment_classifier.fit(x_train,y_train)

print("LogisticRegression Evaluation ")
print("Accuracy " + str(sentiment_classifier.score(x_test,y_test)))

y_pred = sentiment_classifier.predict(x_test)

f1_scores = {}
precision_scores = {}
recall_scores = {}

for idx, label in enumerate(labels):
    f1_scores[label] = f1_score(y_test[label], y_pred[:, idx], average="macro")
    precision_scores[label] = precision_score(y_test[label], y_pred[:, idx], average="macro")
    recall_scores[label] = recall_score(y_test[label], y_pred[:, idx], average="macro")

average_f1 = np.mean(list(f1_scores.values()))
print(f"Average F1-score: {average_f1:.4f}")

average_precision = np.mean(list(precision_scores.values()))
print(f"Average Precision: {average_precision:.4f}")

average_recall = np.mean(list(recall_scores.values()))
print(f"Average Recall: {average_recall:.4f}")
```

Figure 4-6 - Code segment for calculating LR model evaluation parameters

After training the Logistic Regression model, evaluation metrics such as accuracy, precision, recall, and F1-score were computed.

Table 4-2 - Logistic Regression Model Results

Evaluation Metric	*Score (%)*
Accuracy	68.32
Recall	82.98
Precision	82.44
F1 Score	81.09

The logistic regression model successfully identified the positive sentiment towards the acting and the story aspects in the review: "The acting performances are exceptional, breathing life into an exquisite story that unfolds with gripping intensity." However, it was difficult for the model to classify some elements correctly in some situations. Consider the following review: "While the vibrant colors and playful

animations create an inviting, kid-friendly atmosphere, the complex character relationships and nuanced performances introduce a layer of depth that transcends the typical children's movie experience." Here, the model classified the review under the kid-friendly aspect, disregarding the associated character aspect category. This misinterpretation emphasizes how difficult it may be to detect and classify elements correctly, particularly when they coexist in a single statement that expresses a range of sentiments. In the review, "The director's unconventional choices, while commendable, occasionally overshadow the brilliance of the cast, creating a delicate balance between outstanding performances and daring directing." the trained model correctly identified the acting and directing as the aspect categories, but the sentiment for directing aspect was incorrectly recognized as positive sentiment. The interplay between the positive portrayal of acting and the negative sentiment towards directing makes categorization complex for the trained model.

4.3.3 Support Vector Machines (SVM Non-Linear) Evaluation

The following code was employed to retrieve the evaluation parameters of the model.

```
sentiment_classifier = Pipeline(steps=[
                        ('pre_processing',preprocessor),
                         ('lr_multi',MultiOutputClassifier(SVC()))
                         ])

sentiment_classifier.fit(x_train,y_train)

print("SVC (non linear kernel) Evaluation ")
print("Accuracy " + str(sentiment_classifier.score(x_test,y_test)))

y_pred = sentiment_classifier.predict(x_test)

f1_scores = {}
precision_scores = {}
recall_scores = {}

for idx, label in enumerate(labels):
    f1_scores[label] = f1_score(y_test[label], y_pred[:, idx], average="macro")
    precision_scores[label] = precision_score(y_test[label], y_pred[:, idx], average="macro")
    recall_scores[label] = recall_score(y_test[label], y_pred[:, idx], average="macro")

average_f1 = np.mean(list(f1_scores.values()))
print(f"Average F1-score: {average_f1:.4f}")

average_precision = np.mean(list(precision_scores.values()))
print(f"Average Precision: {average_precision:.4f}")

average_recall = np.mean(list(recall_scores.values()))
print(f"Average Recall: {average_recall:.4f}")
```

Figure 4-7 - Code segment for calculating SVM Non-Linear model evaluation parameters

After training the SVM model with the kernel set as non-linear, evaluation metrics such as accuracy, precision, recall, and F1-score were computed. The evaluation metric results for the SVM model with the kernel set as non-linear are presented below, offering a brief yet informative overview of the model's performance across diverse metrics.

Table 4-3 - Support Vector Machines (SVM Non-Linear) Model Results

Evaluation Metric	*Score (%)*
Accuracy	67.45
Recall	81.17
Precision	81.90
F1 Score	81.24

The trained SVM Non-Linear model encountered difficulties accurately categorizing aspects in specific scenarios. Take the review: "The musical score weaves seamlessly through the great narrative, adding emotional resonance to key moments.". Here, the model classifies reviews under the story aspect, overlooking the associated music aspect category. The challenge of accurately detecting and classifying aspects becomes evident through this misinterpretation, especially when they converge within a solitary statement, encapsulating myriad sentiments. In review: "The unconventional musical choices, while adding a unique flavor, introduce a layer of complexity that makes it challenging to separate the impact of great actor performances from the music composition." The model was able to identify both the music and acting aspect categories. But it was only able to recognize the acting aspect sentiment correctly. Because acting and musical sentiments are interwoven, it is difficult for the SVC Non-Linear model to precisely understand and categorize sentiments for acting and music aspect categories. Even so, the trained model successfully classified the aspect category and the sentiment of the review: "The film undoubtedly took risks with its unconventional plot structure and avant-garde cinematography" under the story aspect and with a negative sentiment.

4.3.4 Support Vector Machines (SVM Linear) Evaluation

The model's evaluation parameters were obtained by utilizing the subsequent code.

```
sentiment_classifier = Pipeline(steps=[
                                ('pre_processing',preprocessor),
                                 ('svc_multi',MultiOutputClassifier(SVC(kernel="linear")))
                                 ])

sentiment_classifier.fit(x_train,y_train)

print("SVC (linear kernel) Evaluation ")
print("Accuracy " + str(sentiment_classifier.score(x_test,y_test)))

y_pred = sentiment_classifier.predict(x_test)

f1_scores = {}
precision_scores = {}
recall_scores = {}

for idx, label in enumerate(labels):
    f1_scores[label] = f1_score(y_test[label], y_pred[:, idx], average="macro")
    precision_scores[label] = precision_score(y_test[label], y_pred[:, idx], average="macro")
    recall_scores[label] = recall_score(y_test[label], y_pred[:, idx], average="macro")

average_f1 = np.mean(list(f1_scores.values()))
print(f"Average F1-score: {average_f1:.4f}")

average_precision = np.mean(list(precision_scores.values()))
print(f"Average Precision: {average_precision:.4f}")

average_recall = np.mean(list(recall_scores.values()))
print(f"Average Recall: {average_recall:.4f}")
```

Figure 4-8 - Code segment for calculating SVM Linear model evaluation parameters

The results of the evaluation metric for the Support Vector Machines Model with the kernel set as linear after training are as follows.

Table 4-4 - Support Vector Machines (SVM Linear) Model Results

Evaluation Metric	*Score (%)*
Accuracy	67.90
Recall	82.02
Precision	82.24
F1 Score	82.08

Under several circumstances, the trained model has trouble classifying aspects correctly. For example, the algorithm classifies the review "The story weaves an emotional tapestry, intricately threading through the musical score." under the story

aspect, ignoring the related music aspect category. This misinterpretation illustrates how challenging it is to accurately recognize and categorize aspects, especially when they combine to create a single statement that conveys various emotions. It draws attention to how difficult it may be to detect and classify nuances accurately, revealing how intricate it can be to decipher complicated expressions that convey a variety of emotional tones in the same context. In the review, “While some may appreciate the directorial brilliance, others might feel a sense of dissatisfaction with the open-ended conclusion, creating an intricate dance between visionary directing and an ambiguous story resolution.” the trained model incorrectly labelled the story aspect with positive sentiment, neglecting the negative sentiment expressed for the story aspect later. The SVM Linear model struggles to accurately interpret the dual nature of thought-provoking aspects and the sense of dissatisfaction towards the story aspect.

4.3.5 KNNeighbors (KNN) Evaluation

The following code was employed to retrieve the evaluation parameters of the model.

```
sentiment_classifier = Pipeline(steps=[
                                    ('pre_processing',preprocessor),
                                     ('lr_multi',MultiOutputClassifier(KNeighborsClassifier()))
                                     ])

sentiment_classifier.fit(x_train,y_train)

print("KNeighborsClassifier Evaluation ")
print("Accuracy " + str(sentiment_classifier.score(x_test,y_test)))

y_pred = sentiment_classifier.predict(x_test)

f1_scores = {}
precision_scores = {}
recall_scores = {}

for idx, label in enumerate(labels):
    f1_scores[label] = f1_score(y_test[label], y_pred[:, idx], average="macro")
    precision_scores[label] = precision_score(y_test[label], y_pred[:, idx], average="macro")
    recall_scores[label] = recall_score(y_test[label], y_pred[:, idx], average="macro")

average_f1 = np.mean(list(f1_scores.values()))
print(f"Average F1-score: {average_f1:.4f}")

average_precision = np.mean(list(precision_scores.values()))
print(f"Average Precision: {average_precision:.4f}")

average_recall = np.mean(list(recall_scores.values()))
print(f"Average Recall: {average_recall:.4f}")
```

Figure 4-9 - Code segment for calculating KNN model evaluation parameters

After training the KNNeighbors model, evaluation metrics such as accuracy, precision, recall, and F1-score were computed. The evaluation metric results for the KNNeighbors model are presented below, offering a concise yet informative overview of the model's performance across diverse metrics.

Table 4-5 - KNNeighbors Model Results

Evaluation Metric	*Score (%)*
Accuracy	52.47
Recall	74.52
Precision	74.59
F1 Score	74.36

It was difficult for the KNNeighbors model to classify some elements correctly in some situations. Consider the following review: "The musical accompaniment, though ambitious, ventured into a melodic realm that seemed disconnected, creating a dissonance that posed a challenge in harmonizing the impactful directing with the discordant musical backdrop." Here, the model classified the review under the directing aspect, disregarding the associated music aspect category. This misinterpretation emphasizes how difficult it may be to detect and classify elements correctly, particularly when they coexist in a single statement that expresses a range of sentiments. In the review, "Within the intricate web of storytelling, there were cryptic plot mysteries that added a layer of complexity, leaving me torn between the admiration for the stellar acting and the perplexity induced by the enigmatic narrative twists." the trained model correctly identified the acting and story as the aspect categories, but the sentiment for acting aspect was incorrectly recognized as negative sentiment. The interplay between the positive portrayal of acting and the negative sentiment towards the story aspect makes categorization complex for the trained model. However, the KNNeighbors model successfully identified the positive sentiment towards the acting aspect in the review: "The acting was genuinely impressive, with the actors bringing a depth to their characters that made them feel remarkably authentic."

4.3.6 Random Forest (RF) Evaluation

The model's evaluation parameters were obtained by utilizing the subsequent code.

```
sentiment_classifier = Pipeline(steps=[
                                ('pre_processing',preprocessor),
                                 ('lr_multi',MultiOutputClassifier(RandomForestClassifier()))
                                 ])

sentiment_classifier.fit(x_train,y_train)

print("RandomForestClassifier Evaluation ")
print("Accuracy " + str(sentiment_classifier.score(x_test,y_test)))

y_pred = sentiment_classifier.predict(x_test)

f1_scores = {}
precision_scores = {}
recall_scores = {}

for idx, label in enumerate(labels):
    f1_scores[label] = f1_score(y_test[label], y_pred[:, idx], average="macro")
    precision_scores[label] = precision_score(y_test[label], y_pred[:, idx], average="macro")
    recall_scores[label] = recall_score(y_test[label], y_pred[:, idx], average="macro")

average_f1 = np.mean(list(f1_scores.values()))
print(f"Average F1-score: {average_f1:.4f}")

average_precision = np.mean(list(precision_scores.values()))
print(f"Average Precision: {average_precision:.4f}")

average_recall = np.mean(list(recall_scores.values()))
print(f"Average Recall: {average_recall:.4f}")
```

Figure 4-10 - Code segment for calculating RF model evaluation parameters

After training the Random Forest model, evaluation metrics such as accuracy, precision, recall, and F1-score were computed.

Table 4-6 - Random Forest Model Results

Evaluation Metric	*Score (%)*
Accuracy	63.37
Recall	78.50
Precision	79.77
F1 Score	78.42

The trained random forest model encountered difficulties accurately categorizing aspects in specific scenarios. Take the review: "The acting was not just entertaining but also resonated well with the younger audience, making it a perfect choice for a

family movie night.". Here, the model classifies reviews under the acting aspect, overlooking the associated kid-friendly aspect category. This mistake highlights the difficulty of precisely identifying and categorizing aspects, particularly when they come together in a single sentence that encapsulates a variety of emotions. In review: "The soundtrack, although ambitious in its own right, struck an unsettling chord, creating a dissonance that seemed at odds with the heart-pounding directing." The model was able to identify both the music and directing aspect categories. However, the model recognized both sentiments of the aspect categories incorrectly. Because words such as heart-pounding are vague, it is difficult for the SVC Non-Linear model to precisely understand and categorize sentiments for acting and music aspect categories. Even so, the trained model successfully classified the aspect category and the sentiment of the review: "Kudos to the director for keeping me on the edge of my seat with their suspenseful and gripping style." under the directing aspect and with a positive sentiment.

4.3.7 Gradient Boosting Machines (GBM) Evaluation

The following code was employed to retrieve the evaluation parameters of the model.

```
sentiment_classifier = Pipeline(steps=[
                                    ('pre_processing',preprocessor),
                                     ('lr_multi',MultiOutputClassifier(GradientBoostingClassifier()))
                                     ])

sentiment_classifier.fit(x_train,y_train)

print("GradientBoostingClassifier Evaluation ")
print("Accuracy " + str(sentiment_classifier.score(x_test,y_test)))

y_pred = sentiment_classifier.predict(x_test)

f1_scores = {}
precision_scores = {}
recall_scores = {}

for idx, label in enumerate(labels):
    f1_scores[label] = f1_score(y_test[label], y_pred[:, idx], average="macro")
    precision_scores[label] = precision_score(y_test[label], y_pred[:, idx], average="macro")
    recall_scores[label] = recall_score(y_test[label], y_pred[:, idx], average="macro")

average_f1 = np.mean(list(f1_scores.values()))
print(f"Average F1-score: {average_f1:.4f}")

average_precision = np.mean(list(precision_scores.values()))
print(f"Average Precision: {average_precision:.4f}")

average_recall = np.mean(list(recall_scores.values()))
print(f"Average Recall: {average_recall:.4f}")
```

Figure 4-11 - Code segment for calculating GBM model evaluation parameters

The results of the evaluation metric for the Gradient Boosting Machines Model after training are as follows.

Table 4-7 - Gradient Boosting Machines Model Results

Evaluation Metric	*Score (%)*
Accuracy	56.18
Recall	74.89
Precision	78.05
F1 Score	74.24

Under several circumstances, the gradient-boosting machine model has trouble classifying aspects correctly. For example, the algorithm classifies the review “So, the characters were this quirky bunch, and the cast did their thing amazingly.” under the acting aspect, ignoring the related character category. This misinterpretation highlights the challenge of correctly identifying and classifying components in a statement with diverse sentiments. In the review, “Yet, the acting, though good, lacked the depth needed for true exceptionality.” the trained model incorrectly labelled the acting aspect with positive sentiment, neglecting the negative sentiment expressed for the acting aspect earlier. The positive sentiment overshadows the subtle critique, making it challenging for a trained model to capture the mixed sentiment accurately. Despite that, the model successfully classified the aspect category and the sentiment in a less complex review: “The movie offered kid-friendly fun, with vibrant colors and playful moments.” under the kid-friendly aspect and with a positive sentiment.

4.3.8 Multilayer Perceptron (MLP) Evaluation

The model’s evaluation parameters were obtained by utilizing the subsequent code.

```
sentiment_classifier = Pipeline(steps=[
                                ('pre_processing',preprocessor),
                                 ('lr_multi',MultiOutputClassifier(MLPClassifier(max_iter=1000)))
                                 ])

sentiment_classifier.fit(x_train,y_train)

print("MLP (Neural Network) Evaluation ")
print("Accuracy " + str(sentiment_classifier.score(x_test,y_test)))

y_pred = sentiment_classifier.predict(x_test)
```

Figure 4-12 - First code segment for calculating MLP model evaluation parameters

```
f1_scores = {}
precision_scores = {}
recall_scores = {}

for idx, label in enumerate(labels):
    f1_scores[label] = f1_score(y_test[label], y_pred[:, idx], average="macro")
    precision_scores[label] = precision_score(y_test[label], y_pred[:, idx], average="macro")
    recall_scores[label] = recall_score(y_test[label], y_pred[:, idx], average="macro")

average_f1 = np.mean(list(f1_scores.values()))
print(f"Average F1-score: {average_f1:.4f}")

average_precision = np.mean(list(precision_scores.values()))
print(f"Average Precision: {average_precision:.4f}")

average_recall = np.mean(list(recall_scores.values()))
print(f"Average Recall: {average_recall:.4f}")
```

Figure 4-13 - Second code segment for calculating MLP model evaluation parameters

After training the Multilayer Perceptron model, evaluation metrics such as accuracy, precision, recall, and F1-score were computed.

Table 4-8 - Multilayer Perceptron Model Results

Evaluation Metric	*Score (%)*
Accuracy	64.47
Recall	80.52
Precision	80.52
F1 Score	80.52

The multilayer perceptron model successfully classified the aspect category and the sentiment in a less complex review: “The adventures in this movie were kid-friendly, with playful moments.” under the kid-friendly aspect and with a positive sentiment. However, the model has trouble classifying aspects correctly under several circumstances. For example, the algorithm classifies the review “The narrative had a unique flair, and the producer worked his magic, delivering an amazing experience.” under the story aspect, ignoring the related directing category. This misinterpretation highlights the challenge of correctly identifying and classifying components in a statement with diverse sentiments. In the review, “The music, though pleasant, lacked the depth needed for true exceptionality.” the trained model incorrectly labelled the acting music with positive sentiment, neglecting the negative sentiment expressed for the music aspect earlier. The positive sentiment overshadows a subtle critique, challenging accurate mixed sentiment captured by the trained model.

4.1 Final Evaluation

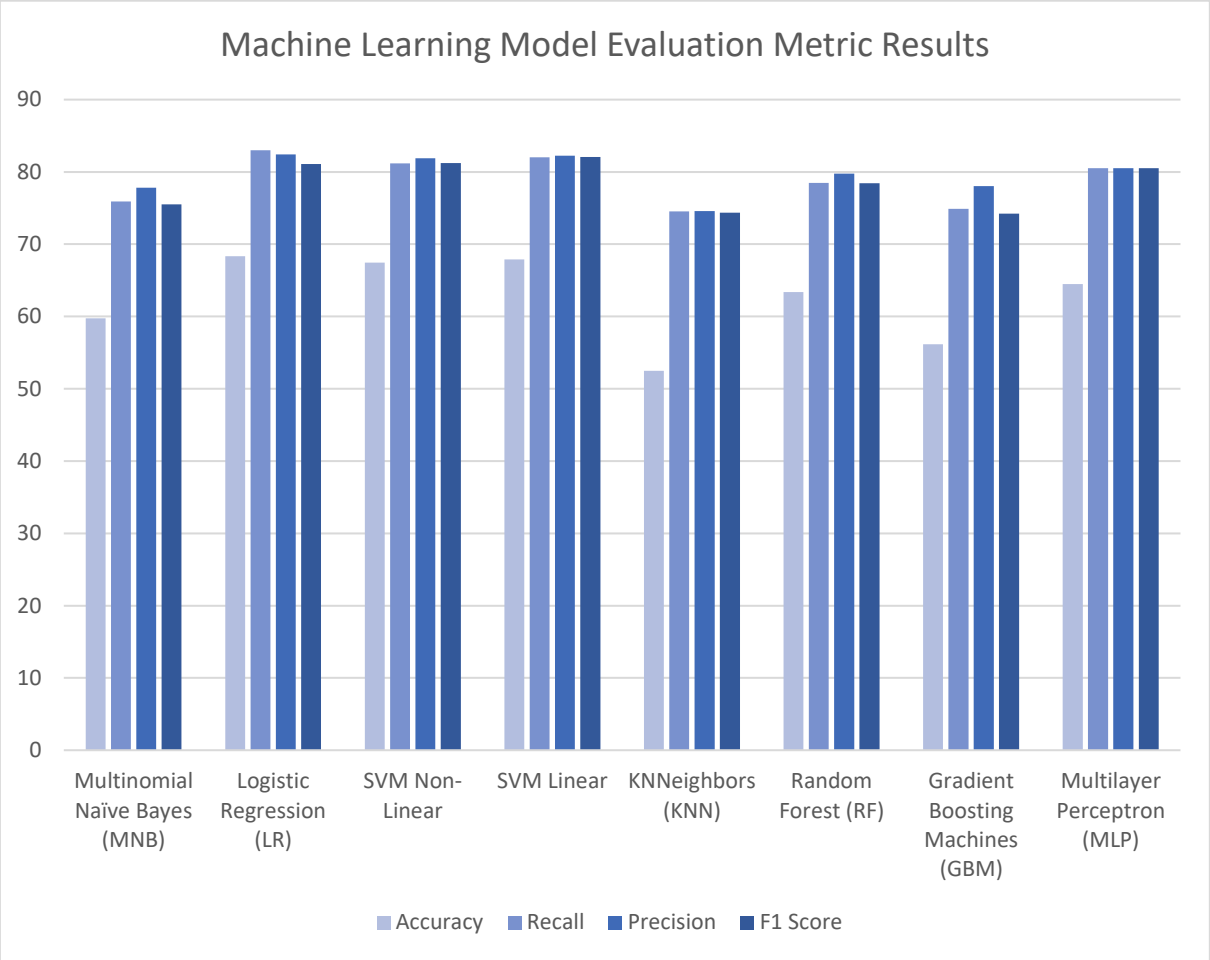


Figure 4-14 - Machine Learning Model Evaluation Metric Results

A comprehensive performance study using seven different classifiers has been used to thoroughly assess the effectiveness of the suggested sentiment analysis approach. Most remarkably, the Logistic Regression (LR) classifier was the top performer, surpassing its competitors on a number of performance metrics. When it came to accuracy, LR was superior to other classifiers by a whopping 68.32% when compared to Multinomial Naive Bayes (MNB), Support Vector Machine with Non-Linear kernel (SVM Non-Linear), Support Vector Machine with Linear kernel (SVM Linear), K-Nearest Neighbors (KNN), Random Forest (RF), Gradient Boosting Machine (GBM), and Multilayer Perceptron (MLP). This significant lead demonstrates LR's unmatched accuracy in sentiment classification and aspect category classification.

CHAPTER 5 CONCLUSION AND FUTURE WORK

In conclusion, aspect-based sentiment analysis has shown to be a useful and perceptive technique for gaining an understanding of the complex viewpoints of viewers when applied to IMDb movie reviews. We have investigated sentiments linked to particular elements or characteristics of films through this study, offering a more detailed comprehension of viewer perceptions. In addition to revealing the general attitude towards films, the research also clarified the specific aspects that influence viewpoints, whether negative, positive, or neutral. According to our research, audiences frequently have differing opinions about several aspects of a film, including kid-friendliness, character performance, directing, acting, story, and music. We have also been able to identify varied sentiments that may have gone unnoticed in more conventional sentiment analysis techniques with the help of aspect-based sentiment analysis. By concentrating on certain aspects, we can uncover hidden sentiments connected to certain parts of a film, offering a more thorough comprehension of viewer opinions. With this level of granularity, online users can easily find an enjoyable film they can watch without wasting time. Additionally, filmmakers, producers, and other industry participants may identify strengths and flaws with this degree of detail, which empowers them to make well-informed decisions that improve the entire cinematic experience.

In future research, utilizing innovative transformer models has enormous promise for aspect-based sentiment analysis of IMDb movie reviews. Additional investigation may entail optimizing transformer architectures such as BERT or GPT to augment the model's comprehension of cinematic subtleties and elevate sentiment prediction precision. Transformer-based models can be used to handle the intricacies of linguistic ambiguity and contextual differences in feelings associated with movies. Furthermore, as transformer architectures continue to evolve, chances to create more effective and context-aware models present themselves, opening the door for developments in multimodal, real-time, and customized aspect-based sentiment analysis in the ever-changing world of movie reviews.

6 REFERENCES

Agarwal, B. and Mittal, N. (2016) *Socio-Affective Computing Volume 2 : Prominent Feature Extraction for Sentiment Analysis*.

Banerjee, D., Mazumder, S. and Datta, S. (2022) 'Comparative Study on Sentiment Analysis on IMDB Dataset', *Ijarcce*, 11(3), pp. 280–286. Available at: https://doi.org/10.17148/ijarcce.2022.11348.

Ikonomakis, M., Kotsiantis, S. and Tampakas, V. (2005) 'Text classification using machine learning techniques', *WSEAS Transactions on Computers*, 4(8), pp. 966–974.

Kumar, R. and Benitta, A. (2022) 'IMDB Movie Reviews Sentiment Classification Using Deep Learning'.

Rathor, S. and Prakash, Y. (2022) 'Application of Machine Learning for Sentiment Analysis of Movies Using IMDB Rating', *Proceedings - 2022 IEEE 11th International Conference on Communication Systems and Network Technologies, CSNT 2022*, pp. 196–199. Available at: https://doi.org/10.1109/CSNT54456.2022.9787663.

Shaukat, Z. *et al.* (2020) 'Sentiment analysis on IMDB using lexicon and neural networks', *SN Applied Sciences*, 2(2), pp. 1–10. Available at: https://doi.org/10.1007/s42452-019-1926-x.

Tarimer, I., Coban, A. and Kocaman, A.E. (2019) 'Sentiment Analysis on IMDB Movie Comments and Twitter Data by Machine Learning and Vector Space Techniques', pp. 1–8.

Kumar, R. and Garg, S. (2020) *Aspect-Based Sentiment Analysis Using Deep Learning Convolutional Neural Network*, *Advances in Intelligent Systems and Computing*. Springer Singapore. Available at: https://doi.org/10.1007/978-981-13-7166-0_5.

Hu, M. and Liu, B. (2004) 'Mining and summarizing customer reviews', *KDD-2004 - Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 168–177. Available at: https://doi.org/10.1145/1014052.1014073.

Hammi, S., Hammami, S.M. and Belguith, L.H. (2022) 'Aspect Term Extraction Improvement Based on a Hybrid Method BT - Foundations of Intelligent Systems', in M. Ceci et al. (eds). Cham: Springer International Publishing, pp. 85–94.

Kumar, K., Harish, B.S. and Darshan, H.K. (2019) 'Sentiment Analysis on IMDb Movie Reviews Using Hybrid Feature Extraction Method', *International Journal of Interactive Multimedia and Artificial Intelligence*, 5(5), p. 109. Available at: https://doi.org/10.9781/ijimai.2018.12.005.

Ur Rehman, A. *et al.* (2019) 'A Hybrid CNN-LSTM Model for Improving Accuracy of Movie Reviews Sentiment Analysis', *Multimedia Tools and Applications*, 78, pp. 26597–26613. Available at: https://doi.org/10.1007/s11042-019-07788-7.

Tripathy, A. *et al.* (2019) 'Aspect based approach for document level sentiment classification of movie reviews', (May).

Ramadhan, N.G. and Ramadhan, T.I. (2022) 'Analysis Sentiment Based on IMDB Aspects from Movie Reviews using SVM', *Sinkron*, 7(1), pp. 39–45. Available at: https://doi.org/10.33395/sinkron.v7i1.11204.

Başarslan, M.S. and Kayaalp, F. (2023) 'MBi-GRUMCONV: A novel Multi Bi-GRU and Multi CNN-Based deep learning model for social media sentiment analysis', *Journal of Cloud Computing*, 12(1). Available at: https://doi.org/10.1186/s13677-022-00386-3.

Onalaja, S., Romero, E. and Yun, B. (2021) 'Number 3 Article 10 Part of the Applied Linguistics Commons, Business Intelligence Commons, Categorical Data Analysis Commons, and the Data Science Commons Recommended Citation Recommended Citation Onalaja', *SMU Data Science Review*, 5(3), p. 10.

Naeem, M.Z. *et al.* (2022) 'Classification of movie reviews using term frequency-inverse document frequency and optimized machine learning algorithms', *PeerJ Computer Science*, 8, pp. 1–28. Available at: https://doi.org/10.7717/PEERJ-CS.914.

Wang, Y., Shen, G. and Hu, L. (2020) 'Importance evaluation of movie aspects: Aspect-based sentiment analysis', *Proceedings - 2020 5th International Conference on Mechanical, Control and Computer Engineering, ICMCCE 2020*, pp. 2444–2448. Available at: https://doi.org/10.1109/ICMCCE51767.2020.00527.

Bo, W. and Min, L. (2021) 'Deep learning for aspect-based sentiment analysis'. Available at: https://doi.org/10.22075/IJNAA.2021.4769.

Maas, A.L. *et al.* (2011) 'Learning word vectors for sentiment analysis', *ACL-HLT 2011 - Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, 1, pp. 142–150.