1. Sentiment Analysis of Movie Review Comments

The study evaluates the effectiveness of different machine learning algorithms and feature extraction methods for sentiment analysis of movie review comments. The authors conducted experiments using supervised (Naive Bayes, Decision Trees, Maximum-Entropy) and unsupervised (K-Means Clustering) learning algorithms with various feature extractors, such as plain bag-of-words and bag-of-words using frequencies from the movie reviews corpus. Results showed that Naive Bayes and Maximum-Entropy algorithms performed best for classifying sentiments in comments, with around 45% accuracy in distinguishing between positive, negative, and neutral sentiments. For binary sentiment classification, Naive Bayes and Decision Trees performed equally well, achieving around 67% accuracy.

Further experiments focused on the use of the existing movie review corpus for training classifiers. The results revealed that the performance of classifiers trained on a subset of 1000 positive and 1000 negative sentences from the movie review corpus was lower than expected, achieving around 58% accuracy. This was attributed to the difference in the number of training and test sentences and the presence of many words in the training corpus that did not appear in the test corpora. Additionally, the study highlighted the challenges associated with using small, manually annotated corpora and the potential for errors in annotations.

The conclusion emphasizes that while Naive Bayes combined with a bag-of-words feature extractor using negated words was the most effective method, the study's findings should be interpreted cautiously due to several threats to validity, such as small corpus size, potential annotation errors, and topic-specific language. The authors suggest that further research with larger and more diverse corpora is needed to confirm their findings and explore the applicability of these techniques to other types of comments and communities.

1. Sentiment Analysis of Movie Reviews Using Machine Learning Techniques

The research paper titled "Sentiment Analysis of Movie Reviews Using Machine Learning Techniques" explores various methods for classifying the sentiment of movie reviews. The study begins by examining different machine learning algorithms, including Naive Bayes, K-Nearest Neighbors (KNN), and Random Forest. The dataset used for the analysis consists of 2000 movie reviews from IMDb, equally divided into positive and negative sentiments. The reviews were preprocessed and converted into a format suitable for the WEKA tool, which was then used for performing the sentiment analysis. The primary focus was to identify the polarity of the reviews, i.e., whether they are positive or negative, using the aforementioned algorithms.

The results of the experiments showed that the Naive Bayes classifier performed the best, achieving an accuracy of 81.45%. The Random Forest classifier followed with an accuracy of 78.65%, while the K-Nearest Neighbors algorithm lagged behind with 55.30% accuracy. These results highlight the effectiveness of the Naive Bayes algorithm in handling the sentiment analysis task for movie reviews. Additionally, the paper emphasizes the need for further experimentation with different algorithms or hybrid methods to potentially increase the accuracy of sentiment classification.

The paper concludes that accurate sentiment analysis can significantly benefit various domains by enabling the development of intelligent systems that provide comprehensive reviews of movies, products, and services. These systems can help users make informed decisions without the need to read through individual reviews. The research also acknowledges that testing a broader range of algorithms and incorporating more diverse datasets could enhance the robustness and accuracy of sentiment analysis models.

1. Sentiment Analysis on Nepali Movie Reviews using Machine Learning

The paper "Sentiment Analysis on Nepali Movie Reviews using Machine Learning" by Ashok Kumar Pant and Abhimanu Yadav presents a study on using machine learning techniques to classify the sentiment of Nepali movie reviews as positive or negative. The researchers created a dataset of 500 movie reviews, with an equal split between positive and negative sentiments. They employed various natural language processing (NLP) techniques for preprocessing the text, such as removing noise, handling negations, and performing part-of-speech tagging. The core of their approach was using a Naive Bayes classifier to perform the sentiment classification, achieving precision, recall, and F-score metrics of approximately 79%.

The methodology involved several key steps: data acquisition from online sources, preprocessing to clean the text, feature extraction to identify sentiment-bearing words, and finally classification using the Naive Bayes algorithm. The preprocessing phase included steps like removing special symbols, expanding abbreviations, and stemming. Feature extraction focused on identifying terms with strong sentiment orientation, using measures like term frequency-inverse document frequency (TF-IDF) and counting polar words. The Naive Bayes classifier then used these features to distinguish between positive and negative reviews.

The results of the study showed that the Naive Bayes classifier could effectively classify the sentiment of Nepali movie reviews with relatively high accuracy. The authors concluded that their approach could be beneficial for various applications, such as marketing and product selection, where understanding public sentiment is crucial. They also suggested that future work could explore using larger datasets and other machine learning techniques like support vector machines (SVM) and neural networks to potentially improve classification performance.