THE ROLE OF N-GRAMS IN ALBANIAN TEXT MINING AND CLASSIFICATION

MARKELA MUÇA[1], BORA LAMAJ (MYRTO)[2], FREDERIK DARA[3]

[1,3]Department of Applied Mathematics, Faculty of Natural Science, University of Tirana, Albania

[2]Department of Computer Science, Faculty of Information Technology "Aleksander Moisiu", University of Durres, Albania

e-mail: markela.muca@fshn.edu.al

Abstract

This research investigates the efficacy of the Naïve Bayes classification method combined with n-gram language models for text categorization in the Albanian language. Despite Naïve Bayes being esteemed for its simplicity, computational efficiency, and competitive efficacy in natural language processing tasks, its utilization in low-resource and morphologically intricate languages like Albanian has not been thoroughly examined. This research addresses the lack by implementing and evaluating Naïve Bayes models utilizing unigram, bigram, trigram hybrid n-gram tokenization approaches. The influence of different n-gram representations on classification efficacy is evaluated using common measures, such as accuracy, precision, recall, and F1-score. This study advances machine learning approaches for under-resourced languages and offers empirical evidence to bolster the development of computational linguistics resources for Albanian. The experimental results reveal that hybrid models combining unigrams and bigrams outperform single-order n-gram configurations, achieving the highest accuracy and F1-score. Conversely, trigram-based models exhibit performance degradation due to data sparsity, highlighting the trade-off between contextual richness and feature dimensionality in small datasets. Furthermore, the study demonstrates that traditional machine learning approaches remain robust in low-resource settings, offering competitive results without the computational overhead of deep learning models. Beyond classification accuracy, this research emphasizes practical implications for sentiment analysis applications, such as identifying recurring negative themes in Albanian-language reviews to support decision-making for businesses and content creators. The findings contribute to the advancement of NLP for under-

resourced languages and provide methodological guidelines for optimizing feature representation in similar linguistic contexts.

Key words: *Text Classification, Naive Bayes, n-Gram.*

Përmbledhje

Ky punim ka për qëllim analizimin e algoritmit Naïve Bayes të kombinuar me sekuencat n-gram për klasifikimin e tekstit në gjuhën shqipe. Megjithëse Naïve Bayes vlerësohet për thjeshtësinë, efikasitetin llogaritës dhe performancën konkurruese në klasifikimin e tekstit, përdorimi i tij në gjuhë me pak burime dhe me morfologji të ndërlikuar, si shqipja, nuk është shqyrtuar gjerësisht. Ky punim e adreson këtë mungesë duke zbatuar dhe vlerësuar modele Naïve Bayes që përdorin qasje hibride të n-gram (unigram, bigram, trigram). Ndikimi i n-gram në efektivitetin e klasifikimit të tekstit vlerësohet me anë të treguesve të performancës, si saktësia (accuracy), saktësia e parashikimit (precision), rikthimi (recall) dhe F1-score. Punimi avancon qasjet e të mësuarit të makinës për gjuhë me pak burime dhe ofron prova empirike për të mbështetur zhvillimin e burimeve të gjuhësisë për shqipen. Rezultatet eksperimentale tregojnë se modelet hibride që kombinojnë unigramet dhe bigramet tejkalojnë modelet me n-gram me rend të vetëm, duke arritur saktësinë dhe F1-score më të larta. Ndërkaq, modelet e bazuara në trigram shfaqin rënie të performancës për shkak të rrallësisë së të dhënave, duke nxjerrë në pah kompromis midis pasurisë kontekstuale dhe dimensionalitetit të veçorive në sete të vogla të dhënash. Për më tepër, studimi demonstron se qasjet tradicionale të të mësuarit të makinerisë mbeten të qëndrueshme edhe në mjedise me pak burime, duke ofruar rezultate konkurruese pa ngarkesën llogaritëse të modeleve të të nxënit të thellë (deep learning). Përtej saktësisë së klasifikimit, ky kërkim thekson implikimet praktike për aplikimet e analizës së ndjenjave, si identifikimi i termave negative të përsëritura në vlerësimet në gjuhën shqipe, me qëllimin për të mbështetur vendimmarrjen e bizneseve dhe departamenteve të shitjes. Gjetjet kontribuojnë në avancimin e NLP-së për gjuhën shqipe dhe ofrojnë udhëzime metodologjike për optimizimin e përfaqësimit të veçorive në kontekste gjuhësore të ngjashme.

Fjalë kyçe: *Klasifikimi i tekstit, Naive Bayes, n-Gram.*

Introduction

Text categorization is a fundamental and extensively studied topic in Natural Language Processing (NLP), serving as a cornerstone for numerous

applications in downstream areas. It entails the automated allocation of predetermined categories or semantic labels to textual material, facilitating the effective interpretation and organization of unstructured information by systems. Common applications include spam identification in electronic communications, topic categorization in digital news aggregation, sentiment analysis on social media platforms, and information retrieval in extensive text libraries. The efficacy of such systems is based upon the effectiveness of feature representation methods and the algorithms' capacity to identify the fundamental linguistic patterns of a certain language. Naïve Bayes is a strong and widely used baseline among classic supervised learning methods, owing to its theoretical simplicity, minimal computing expense, and surprisingly competitive efficacy relative to more complex classifiers. Despite its assumption of probabilistic independence, it frequently demonstrates exceptional performance when utilized with high-dimensional text data, rendering it an appropriate benchmark for assessing feature extraction techniques.

A crucial factor in the effectiveness of text categorization is the representation of textual features, which dictates the encoding of semantic and syntactic information for computational modeling. N-gram language models have been extensively utilized to capture local word dependencies and contextual relationships that unigram isolated single-word representations do not preserve. N-grams, by modeling sequences of n contiguous words, provide a more refined comprehension of language context, facilitating improved classification and sentiment analysis. The bigram "not good" expresses a negative attitude that cannot be precisely deduced from the individual tokens "not" and "good." (Shah & Rohilla, 2018). Their implementation and evaluation in low-resource languages including Albanian remain limited. The lack of annotated corpora, specialized lexicons, and standardized preprocessing techniques presents distinct problems for efficient feature extraction and model generalization.

Consequently, it is crucial to examine if the advantages of N-gram representations in high-resource environments are applicable to languages with constrained computational resources and smaller datasets. This work provides a systematic evaluation of Naïve Bayes classifiers using unigram, bigram, trigram, and hybrid n-gram representations for Albanian text categorization. An empirical analysis of the trade-offs between n-gram size, contextual richness, and data sparsity in a low-resource language environment. A contribution to low-resource NLP research, expanding the understanding of

feature engineering strategies and demonstrating how traditional machine learning techniques can be effectively adapted to under-resourced languages.

2. Literature review

Text classification has been extensively studied in Natural Language Processing (NLP), with N-gram-based models serving as a foundational approach for capturing local contextual dependencies. (Jurafsky & Martin, 2025) provide a comprehensive theoretical framework for N-gram language models, emphasizing the Markov assumption and techniques such as smoothing and interpolation to address zero-probability issues. They also introduce perplexity as a key intrinsic metric for evaluating language models, linking it to cross-entropy and entropy concepts, which remain critical for assessing model performance across languages. Building on this foundation, (Shannaq, 2025) investigates the impact of N-gram length on text classification for English and Arabic, two languages with distinct morphosyntactic structures. The study demonstrates that English achieves optimal performance with bigrams ($F1 \approx 0.47$), while Arabic requires longer N-grams (6-grams) to capture its morphological complexity, achieving an F1 score of approximately 0.84.

These findings underscore the importance of language-sensitive optimization in multilingual text classification tasks. Beyond N-gram optimization, recent research has explored advanced feature selection and deep learning approaches. (Al Katat et al., 2024) conducted a systematic review of Arabic sentiment analysis techniques, highlighting that deep learning models with multi-level embeddings outperform traditional methods, particularly when combined with N-gram features to capture contextual nuances. Similarly, (Al-Shalif et al., 2024) reviewed metaheuristic-based feature selection methods, concluding that these approaches significantly enhance classification accuracy and reduce computational complexity compared to conventional techniques.

In addition, (Kumar & Thirumaran, 2024) proposed leveraging higher-order N-grams (4-grams and 5-grams) for English word prediction, addressing sparsity challenges through smoothing techniques. Their work illustrates that expanding context windows can improve predictive performance in languages with simpler syntactic structures, though it introduces trade-offs between accuracy and efficiency. Collectively, these studies converge on the notion that optimal N-gram length, robust feature selection, and integration with advanced models such as transformers are critical for improving text

classification performance, especially in low-resource languages. Future research should explore hybrid approaches that combine traditional features (TF-IDF, N-grams) with transformer-based embeddings to enhance generalization and practical applicability in multilingual and cross-domain settings.

In light of this body of literature, our work seeks to extend the analysis to Albanian text classification and to determine the interplay between N-gram size, data sparsity and model performance in a low-resource setting providing empirical insights into the suitability of various n-gram configurations for Albanian text categorization.

3. Methodology

Machine learning algorithms and n-grams are used in text data categorization into different classes or categories based on their features. Naïve Bayes is used to predict the class labels of new and unseen data once we have trained them. In this section, we present the methodology steps to analyze the impact of n-grams on text classification.

3.1 Dataset Description

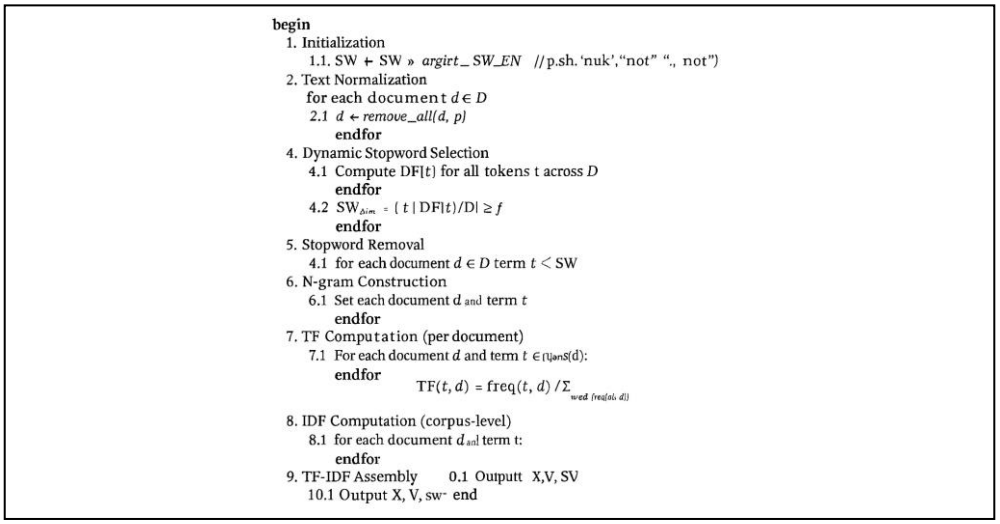
We analyzed a dataset comprising 800 Albanian-language sentences manually labeled into two categories: Positive and Negative. The data incudes user reviews on a movie in Albanian language. Each data instance consists of a single sentence or a short paragraph. For this specific dataset of Albanian movie reviews, the model addresses the problem of uncovering actionable insights from unstructured text, helping stakeholders understand what drives negative sentiment and where improvements are needed.

3.2 Preprocessing

In this session is presented the algorithm that formalizes the text preprocessing pipeline which used to convert raw Albanian reviews into machine‑readable features. Regarding preprocessing in **Figure 1**., it standardizes the text (lowercasing, punctuation removal), applies whitespace tokenization, and constructs a hybrid stopwords list that merges curated Albanian function words with corpus‑specific high‑frequency terms, while deliberately preserving negation tokens (e.g., “nuk”) due to their impact on sentiment. Finally, it generates unigram/bigram/trigram features and computes TF‑IDF

weights, producing a sparse feature matrix and vocabulary ready for downstream modeling.

Figure 1. Preprocessing & TF-IDF N-gram Feature Extraction



3.3 Classifier

For all experiments, we used the Naive Bayes classifier due to its effectiveness in high-dimensional sparse data, such as text. Naive Bayes is a classification algorithm that relies on Bayes' theorem and assumes that the events are independent of each other in our case the reviews are independent. The classification's competitive performance is astonishing, given that the premise of conditional independence, on which it relies, is seldom accurate in real-world scenarios.

The model estimates the likelihood of an instance belonging to a specific class based on its features. The conditional probability of target class, (P(Y|X)) is computed by the formula (3)

$$P(Y/X) = \frac{P(X/Y)*P(Y)}{P(X)} \quad (3)$$

where: P(Y/X) is the probability of class Y given input X, P(X/Y) is conditional probability X given class Y, P(X) is the prior probability of input X and P(Y) is the prior probability of class Y (Wan & Gao, 2016).

Naïve Bayes algorithm is used for classification problems and has proven particularly effective in text classification, where datasets are typically high-dimensional (Breiman & Cutler, 2004). This algorithm is applied in spam filtering, sentiment recognition, and rating classification. The main advantages of Naïve Bayes are:

- Efficiency: It is computationally fast, both in training and prediction.
- Scalability: It handles large feature spaces (e.g., thousands of n-grams) with ease.
- Simplicity: The model is easy to implement and interpret

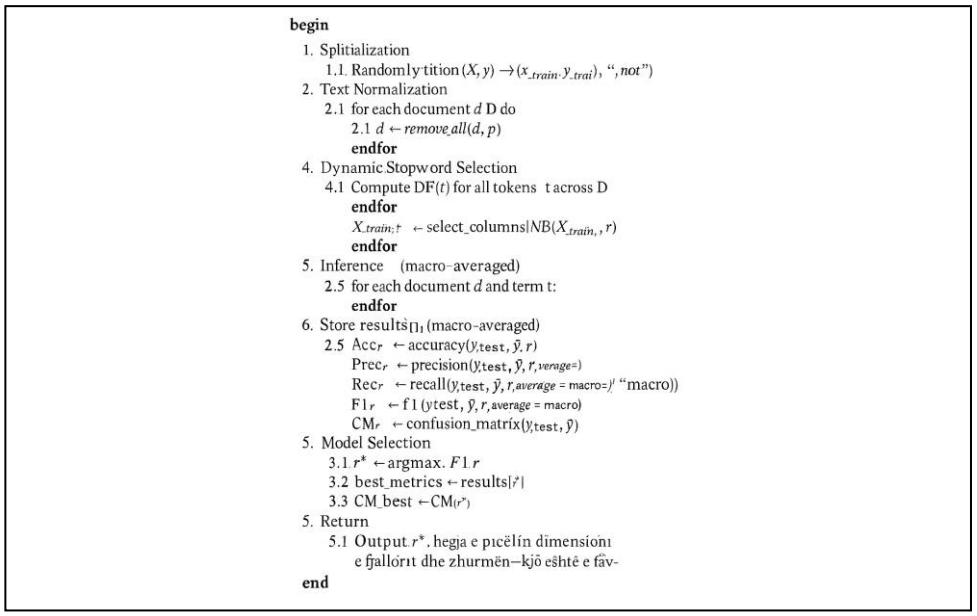


Figure 2. Training & Evaluation (Naïve Bayes with N‑grams)

In **Figure2**. the algorithm outlines the supervised learning procedure used to train and assess the sentiment classifier. After a stratified random split (80% train / 20% test), the model is trained across multiple N‑gram configurations to quantify the trade‑off between context and sparsity. For each configuration, it computes Accuracy, Precision, Recall, and macro‑averaged F1, and records

the confusion matrix. The best configuration is selected by maximizing F1, ensuring robust performance on balanced Albanian sentiment data.

Evaluation measures

To evaluate the performance of a classification model impacted by n-grams is essential to ensure its accuracy and effectiveness. Accuracy is important but is only one of the measures. Additional evaluation criteria are considered to enhance the overall comprehension of our model's performance. This paper examines these indicators and demonstrate how they might assist you in making informed decisions to enhance the predictive capability of your model (Shah & Rohilla, 2018).

Accuracy, recall, F1-Score, and precision are metrics that can be utilized to evaluate methods of classification. Based on those measures will be determined the impact of n-gram for text classification making the right decisions to improve the model's predictive power (Kadriu et al., 2019).

Accuracy refers to the percentage of documents that are correctly classified by the system. A good accuracy level is typically considered to be above 70% (Hegde & Padma, 2017).

Precision is a metric that measures the accuracy of a system in retrieving relevant information. It is calculated as the percentage of true positive (TP) documents to the sum of true positive and false negative (TP + FN) documents. Precision value needs to be at minimum 70-80% for a model to be useful. The metric only measures the model, and not the underlying data (Hegde & Padma, 2017).

Recall is the measure of the system's ability to accurately retrieve relevant documents, expressed as a percentage. A good value for recall needs to be at minimum 70-80%.

F1-Score is a comprehensive metric that combines precision and recall providing a worldwide assessment of the performance of an information retrieval system (Hegde & Padma, 2017). A performance with a score ranging from 0.8 to 0.9 is considered to be of high quality, whereas a score between 0.5 and 0.8 is considered to be average. A model is said to have poor performance if its F1 score drops below 0.5 (Hegde & Padma, 2017).

4. Analysis &discussion

The experiments were conducted using a custom-labeled dataset of 800 Albanian-language reviews related to a movie, equally distributed among the two sentiment classes, 400 positive reviews and 400 negative reviews. To ensure balanced evaluation, the dataset was split into training and testing subsets, with 80% (640 reviews) used for training and 20% (160 reviews) reserved for testing. The split was performed randomly, without applying any thematic or chronological criteria, to minimize bias and ensure that both subsets represent the overall distribution of sentiments. This random partitioning approach is widely adopted in supervised learning tasks to guarantee generalization and avoid overfitting. For feature extraction, we applied TF-IDF weighting with different n-gram configurations: unigram, bigram, trigram and hybrid-gram. Classification was performed using the Naïve Bayes algorithm, which is well-suited for handling high-dimensional, sparse text data. Evaluation was done using the following metrics: Accuracy, Precision, Recall, and F1-score.

4.1 Experiments

The proposed approach integrates general text preprocessing with N-gram feature extraction in the Naïve Bayes framework. The preprocessing pipeline consisted of case folding, symbol removal, duplicate removal, tokenization, stopwords removal, and stemming. The performance evaluation also involved an examination of the classification reports generated by the machine learning (ML) model. The classification report gives detailed information about the values of each measure to study the performance which encompasses recall, precision, and F1-Score for each class (negative and positive), as well as accuracy and other combined metrics taking in consideration 1-gram, 2-gram, 3-gram and hybrid-gram.

The model demonstrated balanced performance across both classes, with precision and recall values near 0.87 and an overall accuracy was 0.87, indicating robust classification capabilities for Albanian language reviews.

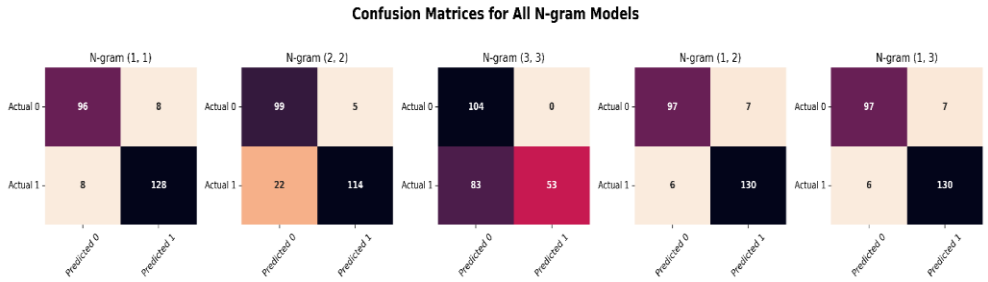


Figure 3. Confusion matrix *of Naïve Bayes model a.1-gram b.2-gram, c. 3-gram,d.1-2-gram, e.1-3-gram*

Confusion matrices for all N-gram configurations (**Figure 3.**) illustrate the influence of feature selection on sentiment classification. The unigram model achieved balanced performance with minimal misclassification, while the bigram model reduced false positives but increased false negatives, indicating a bias toward negative predictions. The trigram model exhibited poor results due to high sparsity, misclassifying most positive instances. Hybrid models combining unigrams and bigrams delivered the best accuracy, confirming that short-range contextual features enhance performance, whereas adding trigrams provides no significant improvement.

Table 1. N-Gram evaluation measures

Modell (N-gram)	**Precision**	**Recall**	**F1-Score**	**Accuracy**
(1, 1)	0.93	0.93	0.93	0.933
(2, 2)	0.89	0.9	0.89	0.887
(3, 3)	0.78	0.69	0.64	0.654
(1, 2)	0.95	0.94	0.94	0.946
(1, 3)	0.95	0.94	0.94	0.946

At the **Table 1.** are shown the evaluation metrics for 1-gram, 2-gram, 3-gram and hybrid-grams.

4.2 Results

Naïve Bayes performance varies significantly with N-gram representation. Unigrams provide a strong baseline (Accuracy 0.93, F1 0.93), while bigrams

slightly reduce accuracy (0.88) due to sparsity. Trigrams perform poorly (0.65), confirming that higher-order N-grams are unsuitable for small datasets. The best results come from hybrid unigram-bigram models (Accuracy 0.94, F1 0.94), offering the optimal balance between context and generalization. Adding trigrams does not improve performance. Lexical analysis of negative reviews reveals recurring issues: boredom and lack of engagement, weak acting and characterization, technical flaws (e.g., poor effects), and incoherent plots. Frequent phrases like "waste of time" highlight dissatisfaction with the film's value. These insights help filmmakers prioritize improvements in storytelling, acting, and production quality. Using TF-IDF and bigram frequency analysis, this study transforms subjective feedback into actionable patterns, providing a data-driven foundation for creative decisions in the entertainment industry.

Figure 4. *Bigrams Cloud graphic of negative reviews*

In the Figure 4 is shown the bigram cloud graphic of negative reviews that reveals several recurring themes that indicate where the movie struggles most: The most frequent bigrams are *"ia vlen"* (worth it) and *"nuk ia"* (not worth it), suggesting that many viewers question whether the movie is worth watching. This indicates a general dissatisfaction with the film's perceived value.

4.3 Key Insights

Based on the dataset, the n-gram analysis highlights several shortcomings in the movie that contribute to negative reviews:

Boredom and Lack of Engagement: Phrases like *"film mërzitshëm"* (boring film) and *"mërzitshëm nuk"* appear prominently, pointing to a common complaint that the movie fails to keep the audience engaged. This is a critical issue for entertainment content. *Weak Acting and Characterization*: Bigrams

such as *"aktrim dobët"* (poor acting), *"aktrim qesharak"* (ridiculous acting), and *"personazhi kryesor"* (main character) suggest dissatisfaction with the acting quality and possibly the depth or believability of characters. *Technical and Artistic Shortcomings*: Mentions of *"efekte speciale"* (special effects) and *"realizim dobët"* (poor execution) indicate that viewers found the technical aspects, such as visual effects or production quality, lacking.

Plot and Narrative Issues: Frequent phrases like *"pa kuptim"* (meaningless) and *"pa përmbajtje"* (without substance) highlight criticism of the storyline, suggesting that the plot may be incoherent or superficial. *Time and Effort Concerns*: Expressions such as *"humbje kohe"* (waste of time) and *"mos humbisni"* (don't waste) reinforce the perception that watching the movie was not a good investment of time. In summary, this analysis shows that the Naïve Bayes model, coupled with N-gram TF-IDF features, not only accurately classifies sentiment but also identifies recurring themes in negative reviews.

Like many text datasets, it presents several challenges for automated sentiment analysis which by using the combination of TF–IDF weighting, N-gram feature representation, and Naïve Bayes classification solves:

1. *High Dimensionality and Sparsity*: Each review contains many unique words and expressions, resulting in a high-dimensional, sparse feature space that complicates traditional machine learning approaches.

2. *Contextual Ambiguity*: Single words (unigrams) often fail to capture the sentimental context. For example, the bigram "nuk pëlqej" ("do not like") expresses negativity that cannot be inferred from its individual words.

3. Low-Resource Language: Albanian has limited annotated corpora and pre-existing NLP tools, making feature extraction and model generalization more difficult.

4. Noise in Text Data: Reviews may contain punctuation, symbols, repeated words, or domain-specific terms, which can reduce classification accuracy if not properly preprocessed.

Conclusions

The experiments show that N-gram representation strongly influences Naïve Bayes performance in text classification. Unigrams provide a solid baseline due to their frequency and compatibility with Naïve Bayes' independence assumption. Adding bigrams in a hybrid (1–2) model delivers the best overall results by capturing short context without causing sparsity. In contrast,

trigrams alone significantly reduce accuracy and F1-score, as higher-order N-grams introduce excessive sparsity and overfitting in small datasets. The hybrid (1–3) approach performs similarly to (1–2) but offers no meaningful improvement, confirming that trigram features add little value. Overall, combining unigrams and bigrams achieves the optimal balance between statistical robustness and contextual insight.

Analysis of negative reviews highlights recurring issues: boredom and lack of engagement, weak acting and poor characterization, technical flaws such as unimpressive special effects, and narrative problems like incoherent or meaningless plots. Frequent phrases such as "waste of time" emphasize dissatisfaction with the film's value. These findings provide actionable guidance for filmmakers improving plot coherence, character depth, acting quality, and technical execution can enhance audience satisfaction. This study demonstrates how NLP and bigram analysis transform subjective feedback into measurable patterns, offering a data-driven foundation for creative decisions in the entertainment industry.

Future work

Future research could explore several directions to further enhance the performance of Naïve Bayes classifiers with n-gram representations. First, experiments with character-level n-grams may provide robustness against spelling variations, morphological changes, and noisy text data. Second, applying feature selection techniques (e.g., Chi-square, Mutual Information, or Information Gain) could help reduce sparsity by retaining only the most discriminative n-gram features, particularly when using bigrams or trigrams.

Additionally, future studies could examine the impact of TF-IDF weighting compared to raw frequency counts within the Naïve Bayes framework, as this may improve the influence of rare but informative n-grams. Another promising direction is the integration of smoothing techniques and prior adjustments to mitigate zero-probability issues in sparse feature spaces. Finally, hybrid approaches that combine n-gram features with word embeddings or pre-trained language model representations (e.g., BERT-derived features) could be investigated to capture both statistical frequency patterns and semantic relationships, potentially improving generalization across diverse text domains.

Acknowledgements

I would like to thank Erion Çano for creating and releasing the AlbMoRe dataset, presented in the paper AlbMoRe: A Corpus of Movie Reviews for Sentiment Analysis in Albanian. The availability of 800 sentiment-annotated Albanian movie reviews has been essential for the development and evaluation of this work. I also acknowledge the contribution of the arXiv platform and its open-access infrastructure, which enables transparent and reproducible research. The DOI (10.48550/arXiv.2306.08526) ensures long-term accessibility and citation stability, which I greatly appreciate (Çano, 2023).

References

Abbas, M., Memon, K., Jamali, A., Memon, S., & Ahmed, A. (2019). Multinomial naive Bayes classification model for sentiment analysis. International Journal of Computer Science and Network Security, 19(3), 62–67.

Al Katat, S., Zaki, C., Hazimeh, H., Bitar, I., Angarita, R., & Trojman, L. (2024). Natural language processing for Arabic sentiment analysis: A systematic literature review. IEEE Transactions on Big Data. https://doi.org/10.1109/TBDATA.2024.3366083

Al-Shalif, S. A., Senan, N., Saeed, F., Ghaban, W., Ibrahim, N., Aamir, M., & Sharif, W. (2024). A systematic literature review on meta-heuristic based feature selection techniques for text classification. PeerJ Computer Science. https://doi.org/10.7717/peerj-cs.2084

Andrew, L., Raymond, E., Peter, T., Dan, H., Andrew, Y., & Christopher, P. (2011). Learning word vectors for sentiment analysis. Stanford University.

Asif, M., Ishtiaq, A., Ahmad, H., Aljuaid, H., & Shah, J. (2020). Sentiment analysis of extremism in social media from textual information. Telematics and Informatics, 48, 101345. https://doi.org/10.1016/j.tele.2020.101345

Balakrishnan, V., Selvanayagam, P. K., & Yin, L. P. (2020). Sentiment and emotion analyses for Malaysian mobile digital payment applications. In ACM International Conference Proceeding Series (pp. 67–71).

Breiman, L., & Cutler, A. (2004). Retools for predicting and understanding data. Interface Workshop, 1–62.

Çano, E. (2023). AlbMoRe: A corpus of movie reviews for sentiment analysis in Albanian. arXiv. https://doi.org/10.48550/arXiv.2306.08526

Hegde, Y., & Padma, S. K. (2017). Sentiment analysis using random forest ensemble for mobile product reviews in Kannada. In Proceedings of the International Conference on Advanced Computing and Communication (IACC) (pp. 777–782).

https://doi.org/10.1109/IACC.2017.0160

Jurafsky, D., & Martin, J. H. (2025). Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition with language models (3rd ed.). Stanford University.

Kadriu, A., Abazi, L., & Abazi, H. (2019). Albanian text classification: Bag of words model and word analogies. Business Systems Research Journal, 10(1), 74–87.

https://doi.org/10.2478/bsrj-2019-0006

Kumar, K. R., & Thirumaran, S. (2024). Enhancing automatic English word analysis and prediction using higher-order n-gram models. In Proceedings of the IEEE Conference on Science, Technology, Engineering and Mathematics (ICSTEM) (pp. 1–7).

Phienthrakul, T., Kijsirikul, B., Takamura, H., & Okumura, M. (2009). Sentiment classification with support vector machines and multiple kernel functions. In Neural Information Processing (pp. 583–592). Springer.

Rokach, L., Romano, R., & Maimon, O. (2008). Negation recognition in medical narrative reports. Information Retrieval, 11(6), 499–538.

https://doi.org/10.1007/s10791-008-9061-0

Setiawan, Y., Maulidevi, N. U., & Surendro, K. (2024). Optimization of n-gram feature extraction based on term occurrence for cyberbullying classification. Data Science Journal, 23(1), Article 031.

https://doi.org/10.5334/dsj-2024-031

Shah, N., & Rohilla, S. (2018). Description of the emot library (Version 2.2) [GitHub repository].

https://github.com/NeelShah18/emot

Shannaq, B. (2025). Optimizing n-gram lengths for cross-linguistic text classification: A comparative analysis of English and Arabic morphosyntactic structures. International Journal of Advanced and Applied Sciences, 12(4), 136–145.

https://doi.org/10.21833/ijaas.2025.04.015

Singh, M., Jakhar, A. K., & Pandey, S. (2021). Sentiment analysis on the impact of coronavirus in social life using the BERT model. Social Network Analysis and Mining, 11(1).

Sousa, M. G., Sakiyama, K., Rodrigues, L. S., Moraes, P. H., Fernandes, E., & Matsubara, E. T. (2019). BERT for stock market sentiment analysis. In Proceedings of the International Conference on Tools with Artificial Intelligence (ICTAI) (pp. 1597–1601).

Trandafili, E., Kote, N., & Biba, M. (2018). Performance evaluation of text categorization algorithms using an Albanian corpus. In Lecture Notes on Data Engineering and Communications Technologies (pp. 537–547). Springer.

https://doi.org/10.1007/978-3-319-75928-9_48

Wan, Y., & Gao, Q. (2016). An ensemble sentiment classification system of Twitter data for airline services analysis. In Proceedings of the IEEE International Conference on Data Mining Workshops (ICDMW) (pp. 1318–1325).

Yang, K., Yu, Z., Wen, X., Cao, W., Chen, C., Wong, H., & You, J. (2020). Hybrid classifier ensemble for imbalanced data. IEEE Transactions on Neural Networks and Learning Systems, 31, 1387–1400.

Yu, H., Lo, H., Hsieh, H., Lou, J., McKenzie, T., Chou, J., & Chung, P. (2010). Feature engineering and classifier ensemble for KDD Cup 2010. JMLR: Workshop and Conference Proceedings, 1–12.