engineering proceedings

MDPI

Proceeding Paper

An Improved Multi-Dimensional Data Reduction Using Information Gain and Feature Hashing Techniques †

Usman Mahmud [1], Abubakar Ado [2,*], Hadiza Ali Umar [3] and Abdulkadir Abubakar Bichi [4]

[1] Department of Computer Science, Al-Qalam University, Katsina 820101, Nigeria; usmanmahmud@auk.edu.ng
[2] Department of Computer Science, Yusuf Maitama Sule University, Kano 700252, Nigeria
[3] Department of Computer Science, Bayero University, Kano 700241, Nigeria; haumar.cs@buk.edu.ng
[4] Department of Software Engineering, Yusuf Maitama Sule University, Kano 700252, Nigeria; engraabubakar@gmail.com
* Correspondence: aarogo@yumsuk.edu.ng
† Presented at the 5th International Electronic Conference on Applied Sciences, 4–6 December 2024; https://sciforum.net/event/ASEC2024.

Abstract

Sentiment analysis is a sub-field within Natural Language Processing (NLP), concentrating on the extraction and interpretation of user sentiments or opinions from textual data. Despite significant advancements in the analysis of online content, a continuing challenge persists in the handling of sentiment datasets that are high-dimensional and frequently include substantial amounts of irrelevant or redundant features. Existing methods to address this issue typically rely on dimensionality reduction techniques; however, their effectiveness in removing irrelevant features and managing noisy or redundant data has been inconsistent. This research seeks to overcome these challenges by introducing an innovative methodology that integrates ensemble feature selection techniques based on information gain with feature hashing. Our proposed approach aims to enhance the conventional feature selection process by synergistically combining these two strategies to more effectively tackle the issues of irrelevant features, noisy classes, and redundant data. The novel integration of information gain with feature hashing facilitates a more precise and strategic feature selection process, resulting in improved efficiency and effectiveness in sentiment analysis tasks. Through comprehensive experimentation and evaluation, we demonstrate that our proposed method significantly outperforms baseline approaches and existing techniques across a wide range of scenarios. The results indicate that our method offers substantial advancements in managing high-dimensional sentiment data, thereby contributing to more accurate and reliable sentiment analysis outcomes.

Keywords: sentiment analysis; feature selection; feature hashing; information gain

check for updates

Academic Editor: Carlos Cruz

Published: 14 July 2025

Citation: Mahmud, U.; Ado, A.; Umar, H.A.; Bichi, A.A. An Improved Multi-Dimensional Data Reduction Using Information Gain and Feature Hashing Techniques. *Eng. Proc.* **2025**, *87*, 92. https://doi.org/10.3390/engproc2025087092

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Movie review is an important way of gauging the acceptability and performance of a movie by providing numerical ratings of the movie. However, movie review is associated with problems of massive data generation, textual jargon, and tokens that contain no sentiment values. A key element in sentiment classification is feature selection. Feature selection involves choosing and identification of the most essential attributes within noisy data, thereby eliminating irrelevant characteristics and reducing the data's dimensionality [1]. In [2], the authors proposed an approach for analyzing sentiment in movie reviews; they enhanced the classifier's performance by incorporating Chi-Square feature selection,

resulting in an increased accuracy of 85.35%. Furthermore, when AdaBoost was applied, the accuracy improved even further to 87.74%. Some recent studies have focused on feature selection methods for textual data, specifically utilizing information gain as a technique for selecting relevant features [3–5]. The common weakness of all the stated approaches is that they have not considered data redundancy. Ref. [6] proposed a method of eliminating data redundancy by hybridizing information gain with an ontology-based method; a limitation of the approach is that information gain does not consider redundancy between features, while the ontology-based approach involves much more human intervention for the redundancy removal. Ref. [7] proposed a new approach by introducing feature hashing. This approach showed a remarkable performance, especially when dealing with online applications. Ref. [8] proposed a lower and upper bond to overcome the problems associated with the existing feature hashing. The results show that their approach enabled the reduction bias by almost 50%. Ref. [9] proposed a new feature hashing approach by utilizing term weight to overcome the problem of collision affecting the existing feature hashing methods. The experimental results show that the approach achieved significant performance improvement. This result reduces bad "collision" by almost 30%. The existing approaches are still inefficient. Hence, this research proposes a new approach using an ensemble of information gain (IG) and feature hashing (FH) to overcome the problem of redundancy associated with existing methods. Table 1 gives a pictorial view of the related literature in the text classification, while Table 2 illustrates related literature utilizing feature hashing techniques for dimensional reduction in the textual data domain.

Table 1. Illustrating some related literature utilizing information gain and others filter algorithms as feature selection on textual data.

Author	Year	Aim	Finding
[10]	2016	Implemented a feature ranking method for the context of aspect-level sentiment analysis using information gain.	Results showed that restricting the number of features during selection does not considerably affect accuracy.
[11]	2017	Presented a feature selection method that integrates information gain with word embeddings.	Experiments on Chinese text classification demonstrated enhanced results.
[12]	2018	Utilized Naïve Bayes and genetic algorithms for feature selection.	Naïve Bayes accuracy increased from 68.50% to 87.50% after applying genetic algorithm for feature selection.
[1]	2019	Proposed a method combining information gain with an ontology-based approach.	Information Gain effectively removed noise from the dataset, while the ontology-based method required more human intervention to eliminate redundancy.
[4]	2020	Applied Support Vector Machine with Information Gain for movie review classification.	Information gain enhanced the performance of the Support Vector Machine, but did not address feature redundancy.
[5]	2021	Presented a hybrid approach that integrate IG and T-test for selecting most informative features	The hybrid approach effectively selects the set of important features by complementing the backside of each method
[2]	2021	Proposed an approach using Chi-Square for feature selection in movie review analysis.	Chi-Square feature selection increased accuracy to 85.35%, and further improved to 87.74% with the application of AdaBoost.

Table 2. Related literature utilizing feature hashing techniques for dimensional reduction in the textual data domain.

Author	Year	Aim	Finding
[7]	2009	Introduced a novel technique called feature hashing.	The approach demonstrated exceptional performance, particularly in online applications.
[8]	2009	Suggested solutions to resolve challenges in current feature hashing techniques by introducing lower and upper bounds.	Results indicated a reduction in bias by almost 50%.
[13]	2012	Created an effective method for dimensionality reduction in text classification tasks by integrating hashing and abstraction techniques.	Effectively handled high-dimensional text data without compromising classification performance, demonstrating its usefulness for large-scale text data applications, including news categorization, web content analysis, and scientific document classification.
[14]	2018	Unveiled Cuckoo Feature Hashing (CCFH), a groundbreaking technique aimed at tackling issues related to feature hashing for large-scale sparse datasets.	Cuckoo Feature Hashing addressed feature collisions, decreased parameter dimensions, and preserved model sparsity, resulting in enhanced performance. Experimental results validated CCFH's ability to reduce parameters while maintaining performance
[15]	2020	Conducted a study titled "HASHING" to tackle dimensionality issues in large-scale image categorization by employing a taxonomy hierarchical structure.	Offered both theoretical and empirical evidence for the method's efficacy in lowering computational and storage demands while preserving classification accuracy.
[9]	2021	Proposed a new feature hashing method utilizing term weight to mitigate collision issues in existing feature hashing methods.	Experiments showed significant performance improvements, reducing collision issues by nearly 30%.
[16]	2023	Aimed to optimize sentiment analysis on the Stanford movie dataset using Feature Hashing for dimensional reduction.	The proposed method consistently achieved superior accuracy and precision across three classifier algorithms, though it negatively impacted recall.

2. Materials and Methods

This section presents the proposed method in conjunction with experimental settings.

2.1. Proposed Framework of the Proposed Method

Figure 1 shows the framework of the proposed approach. The method consists of two main stages: the feature reduction and model training stages. Firstly, the input document (dataset) goes directly to pre-processing, which performs stop word removal, tokenization, and stemming, among other strategies. Next, the framework invokes IG to compute the important features and apply a threshold of 70% to select the most informative features. This aim is to remove the relevance between features and noise from the original feature set. The feature score computation is described in Section 2.1.1. Next, FH is deployed and hashes the resulting reduced features obtained from IG into reduced feature space. This aims to reduce non-informative features and remove redundancy under the indices of the n^{14} index. The feature hashing is described in Section 2.1.1. Lastly, the optimal feature set obtained is passed to the machine learning algorithm for class label prediction.

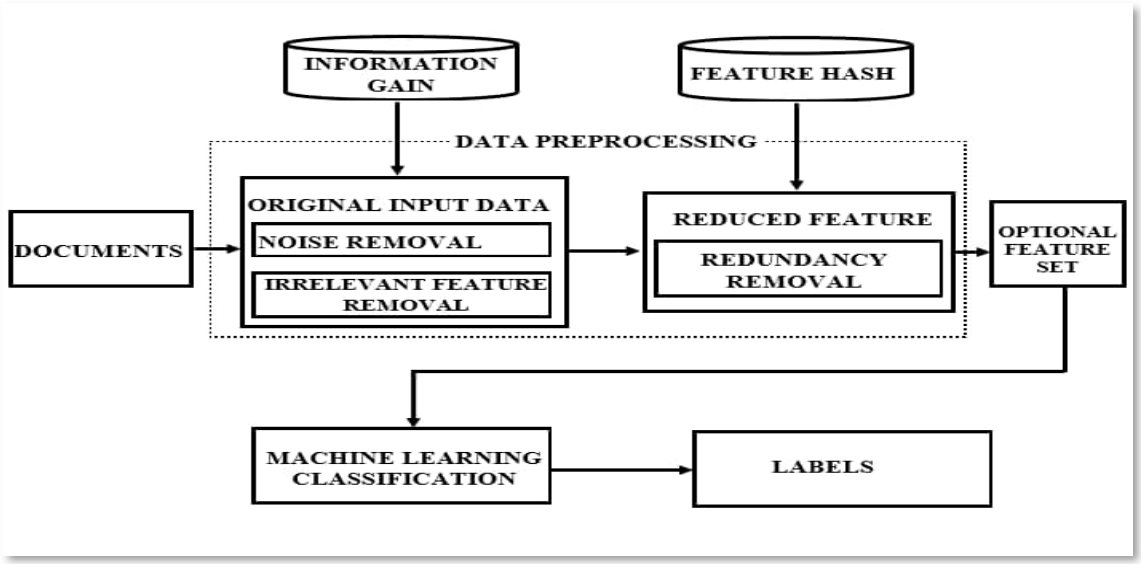

Figure 1. The framework of the proposed method.

2.1.1. Information Gain on Movie Review

Ref. [10] defines information gain as a metric that assesses the relevance of features within sentiment analysis, while ref. [11] specifically evaluates how the significant attribute B concerns class D. Mathematically, information gain can be represented as follows:

$$I(D, B) = H(D)\ H(D - B) \quad (1)$$

In this formula, H(D) denotes the entropy of the class, which is calculated as [12]:

$$H(D) = -\sum c \in C p(D) \log P(D) \quad (2)$$

Here, H(D − B) refers to the conditional entropy of the class given the attribute B [13]. The entropy $H(D - B)$ can be expressed as:

$$H(D - B) = -\sum d \in DP(D - B) \log P(D - B) H(D - B) = -\sum d \in Dp(D - B) \log P(D - B) \quad (3)$$

Given that the Stanford movie review dataset is balanced, there is an equal probability of 50% for a review to be either positive or negative, resulting in H(D) = 1. Thus, the knowledge gain can be reformulated as:

$$I(D, B) = 1 - H(D - B) \quad (4)$$

The minimum value of I(D,B) occurs when H(D − B) = 1, indicating that attribute B has no relationship with class D. Conversely, the objective is to identify the attribute B that is predominantly present in one class, whether positive or negative. This implies that the optimal features are those that are distinctive to a single class. The highest information gain, I(D,B), is obtained when the probabilities P(B) and P(B − D1) are equal, resulting in both $P(D_1 - B)$ and H(D1 − B) being 0.5. When $P(B) = P(B - D_1) P(B) = P(B - D_1)$, it leads to $P(B - D_2)$ giving $P(D_2 - B) = 0$ and $H(D_1 - B)=0$. Therefore, the value of I(D,B) ranges starts between 0 and 1/2.

The calculation of information gain is based upon the concept of entropy, which quantifies the impurity or randomness of a dataset relative to its target classifications [14]. The formula for information gain can be written as

Information Gain = Entropy (Before Split) MINUS Weighted Average of Entropies (After Split). Where:

Entropy (Before Split) refers to the entropy of the target variable prior to the division of the dataset based on a particular feature.

Weighted Average of Entropies (After Split) indicates the average entropy of the target variable following the data split, weighted according to the number of data points in each subgroup.

Mathematically, entropy is generally defined as follows:

$$\text{Entropy(S)} = -\sum_{1}^{k} (P_i * \log_2 P_i) \quad (5)$$

In this equation, S represents the collection of data points, and P_i denotes the proportion of distinct classes within the set S, with $\log_2$ indicating the base-2 logarithm [5].

2.1.2. Feature Hashing on Movie Review

The dimensionality reduction technique known as "feature hashing" was introduced by researchers such as [7,8,15,16]. It involves transforming high-dimensional input vectors y of size d into lower-dimensional feature vectors y^c of size c. Let X denote the set of all possible strings and c and ξ be two hash functions, such that c: X → {0, ⋯, d −1} and ξ: X → {±1}, respectively. Each token in a document is directly mapped, using d1, into a hash key, which represents the index of the token in the feature vector y^c, such that the hash key is a number between 0 and d − 1. Each index in y^c stores the value ("frequency counts") of the corresponding hash feature. The hash function ξ indicates whether to increment or decrement the hash dimension of the token, which renders the hash feature vector y^c as unbiased.

2.2. Dataset and Description

The Stanford dataset utilized in this study was obtained from Kaggle.com. It comprises fifty (50,000) movie review instances, evenly split into twenty-five (25,000) positive and negative reviews, respectively, for the classification and analysis tasks. To effectively assess the performance of machine learning algorithms for sentiment analysis, the dataset was divided into training and testing subsets. In this study, a split of 70% for training and 30% for testing was applied. Table 3 presents the properties of the dataset used.

Table 3. Dataset properties.

Dataset Name	Movie review dataset
Number of Instance	50,000
Number of Classes	2
Number of Instances per Class	25 instances per each class
Number of Attributes	45,584
Training	70%
Testing	30%

2.3. Classifiers

Naive Bayes (NB), Linear Support Vector Classifier (Linear SVC), and K-Nearest Neighbors (K-NN) are key classifiers used to evaluate model performance when information gain and feature hashing are applied to a dataset. Each of these techniques has distinct advantages when working with transformed data. Below is an explanation of their significance:

1. Linear Support Vector Classifier (Linear SVC): Linear SVC is a powerful linear classification method that aims to identify the hyperplane that optimally separates data points from different classes. When feature hashing and information gain are applied to the dataset, the dimensionality of the data can be significantly reduced. This reduction can enhance the efficiency of model training and inference while maintaining or even improving classification accuracy [17].

2. Naive Bayes: Naïve Bayes is a probabilistic classifier that is especially effective for binary and sparse datasets. By utilizing feature hashing, it can efficiently manage high-dimensional sparse data. The model is based on the assumption that features are conditionally independent given the class, which can be a valid assumption in some situations [18].

3. K-Nearest Neighbors (K-NN): k-NN is a straightforward yet effective classifier used for binary classification tasks. It is robust to high-dimensional data and can capture local patterns within the feature space. When the proposed approach is utilized, k-NN benefits from the reduced dimensionality while still performing well in classification tasks [19].

In summary, the classifiers Linear SVC, Naive Bayes, and K-NN are crucial for assessing model performance with feature hashing and information gain. They adapt well to the characteristics of transformed datasets, efficiently handle high-dimensional or sparse data, and deliver strong classification performance.

2.4. Feature Extraction

Word count vectors, often referred to as counting vectors, play a significant role in movie review sentiment analysis. These vectors are created by representing text data as numerical feature vectors, where each element corresponds to the count or weighted importance of a specific word in the text. They also reduce the dimensionality of text data by encoding it into numerical features based on word occurrences or importance. This simplifies the input data and makes it suitable for various machine learning algorithms. Word count vectors are commonly used as features in sentiment analysis models. They enable the training of classifiers, like Logistic Regression (LR), Random Forest (RF), and Convolutional Neural Networks (CNNs), to predict sentiment labels (e.g., positive, negative, neutral) for movie reviews [20].

2.5. Evaluation Metrics

AUC and accuracy are essential metrics for assessing the effectiveness of a classifier in sentiment analysis tasks.

1. Accuracy: Accuracy quantifies the percentage of correctly predicted sentiments out of all predictions. In the context of sentiment analysis, accuracy reflects how well the model can accurately classify reviews as either positive or negative sentiments. Our approach's improved accuracy indicates its ability to provide more accurate sentiment predictions, which is vital for making decisions based on customer feedback.

2. AUC is commonly used as a performance metric for binary classification tasks like sentiment analysis. It can be calculated by:

$$\mathrm{AUC} = \frac{\mathrm{ROC} * \mathrm{TPR}}{\mathrm{FPR}} \tag{6}$$

where ROC stands for the Receiver Operating Characteristic (ROC) curve, which plots the True Positive Rate (TPR) against the False Positive Rate (FPR) at various thresholds. A higher AUC indicates a better ability of the classifier to distinguish between positive and negative reviews [21].

3. Result Discussion

In this section, we present the findings of our research, emphasizing the influence of the proposed methodology on sentiment analysis conducted on the Stanford movie dataset. The objective of our study is to minimize data redundancy, reduce noise, eliminate irrelevant features, enhance parameter tuning, and refine feature extraction processes, thereby achieving higher classification accuracy compared to single-stage methods. Tables 4 and 5 illustrate the results of the proposed method (IG-FH) and the single-stage methods obtained using three key machine learning models: Linear Support Vector Classifier (Linear SVC), K-Nearest Neighbors (K-NN), and Naïve Bayes. The best results are bolded. The best results are face bolded.

Table 4. Results comparison of the proposed ensemble method against single-stage methods based on accuracy.

Classifier	All Feature	FS_(IG)	FS_(FH)	Propose Method (IG + FH)
Linear SVC	86.12	86.67	87.13	**89.48**
KNN	63.12	64.11	66.08	**67.23**
Naïve Bayes	85.46	86.43	87.89	**89.47**

Table 5. Results comparison of the proposed ensemble method against single-stage methods based on AUC.

Classifier	All Feature	FS_(IG)	FS_(FH)	Propose Method (IG + FH)
Linear SVC	86.12	88.96	89.10	**89.56**
KNN	63.12	85.66	88.63	**93.08**
Naïve Bayes	85.41	86.66	88.68	**88.89**

From Table 4, it can be seen that the proposed method (IG-FH) recorded the best accuracy of 89.56%, 93.08%, and 88.89% with Linear SVC, KNN, and Naïve Bayes, respectively. The overall highest result is obtained when the proposed method is coupled with KNN. Furthermore, the single-stage approach, Feature Hashing (FH), supersedes IG by recording the second-highest accuracy, with Linear SVC reaching 89.10%, KNN at 88.63%, and Naïve Bayes at 88.68%. These results highlight the superior performance of the proposed method across all classifiers.

Table 5 presents the performance comparison based on AUC, indicating that the proposed ensemble method integrating information gain (IG) and feature hashing (FH) consistently outperforms all existing models. The combined approach yields the highest results with Linear SVC (89.48%), Naïve Bayes (89.47%), and K-Nearest Neighbors (KNN) (67.23%), respectively. These findings substantiate the effectiveness of the IG-FH method over single-stage methods (IG and FH) across all classifiers. Additionally, FH alone demonstrates superior performance compared to IG, achieving AUC scores of 87.13% (Linear SVC), 87.89% (Naïve Bayes), and 66.08% (KNN), respectively. Figure 2 provides a visual representation of the percentage improvements in both accuracy and AUC, further emphasizing the advantage of the proposed approach.

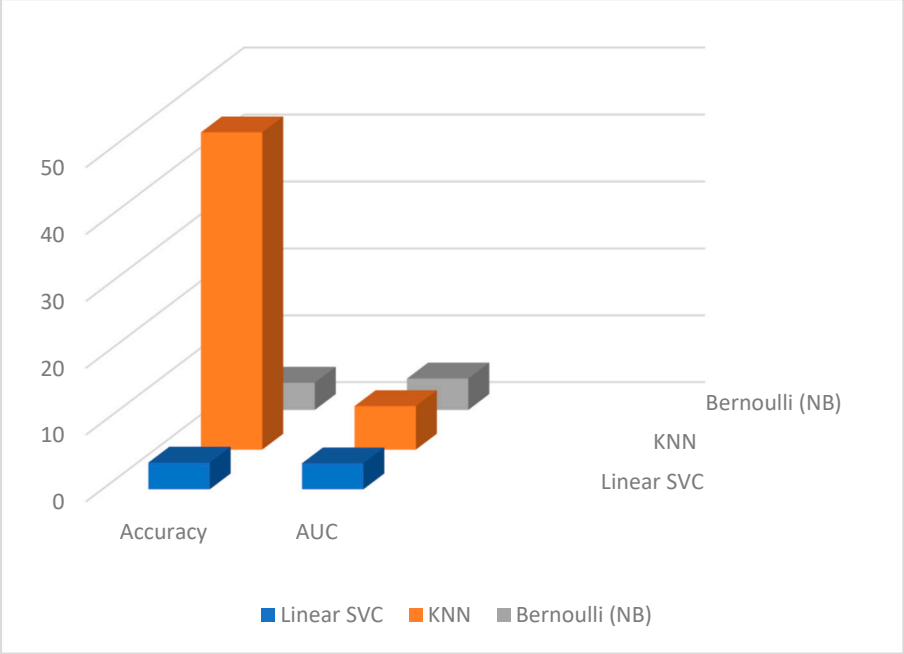

Figure 2. Percentage improvement of the proposed approaches based on accuracy and AUC.

4. Conclusions and Future Work

An approach was presented to minimize or reduce the dimensionality of high-dimensional, sparse feature vectors in movie review datasets. This technique leverages the combined strengths of information gain and feature hashing. Initially, information gain is used to filter out noisy and irrelevant features from high-dimensional datasets. Subsequently, feature hashing is applied to these datasets to eliminate data redundancy and reduce dimensionality by organizing hashed features with similar class distributions. Our experimental results validate the effectiveness of this method, which develops algorithms capable of transforming large-scale, sparse data into low-dimensional feature vectors in real-time. Potential applications of this method include faster classification of textual documents on social media platforms like Twitter and Snapchat. Additionally, future work should involve using other machine learning classifiers such as Random Forest and Maximum Entropy, as well as different dimensionality reduction techniques like Chi-Square and PCA, to evaluate and compare their performance levels.

Author Contributions: Methodology, U.M.; software, A.A.; validation, U.M., A.A. and H.A.U.; formal analysis, U.M.; writing—original draft preparation, U.M.; writing—review and editing, A.A.B.; supervision, A.A.; project administration, A.A.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The dataset is available at https://www.kaggle.com.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Ahmad, S.R.; Rodzi, M.Z.M.; Shapiei, N.S.; Yusop, N.M.M.; Ismail, S. A review of feature selection and sentiment analysis techniques in issues of propaganda. *Int. J. Adv. Comput. Sci. Appl.* **2019**, *10*, 240–245. [CrossRef]
2. Hamzah, M.B. Classification of Movie Review Sentiment Analysis Using Chi-Square and Multinomial Naïve Bayes with Adaptive Boosting. *J. Adv. Inf. Syst. Technol.* **2021**, *3*, 67–74. [CrossRef]
3. Pratiwi, A.I.; Adiwijaya, K. On the feature selection and classification based on information gain for document sentiment analysis. *Appl. Comput. Intell. Soft Comput.* **2018**, *2018*, 1407817. [CrossRef]
4. Maulana, R.; Rahayuningsih, P.A.; Irmayani, W.; Saputra, D.; Jayanti, W.E. Improved accuracy of sentiment analysis movie review using support vector machine based information gain. *J. Phys. Conf. Ser.* **2020**, *1641*, 12060. [CrossRef]
5. Abubakar, A.; Mustafa, M.D.; Noor, A.S.; Aliyu, A. A New Feature Filtering Approach by Integrating IG and T-test Evaluation Metrics for Text Classification. *Int. J. Adv. Comp. Sci. Appl.* **2021**, *12*, 500–511. [CrossRef]
6. Ahmad, I.S.; Bakar, A.A.; Yaakub, M.R. A review of feature selection in sentiment analysis using information gain and domain specific ontology. *Int. J. Adv. Comput. Res.* **2019**, *9*, 283–292. [CrossRef]
7. Shi, Q.; Petterson, J.; Dror, G.; Langford, J.; Smola, A.; Vishwanathan, S. Hash kernels for structured data. *J. Mach. Learn. Res.* **2009**, *10*, 2615–2637.
8. Weinberger, K.; Dasgupta, A.; Attenberg, J.; Langford, J.; Smola, A. Feature hashing for large scale multitask learning. In Proceedings of the 26th Annual International Conference on Machine Learning, Montreal, QC, Canada, 14–18 June 2009.
9. Ado, A. A new feature hashing approach based on term weight for dimensional reduction. In Proceedings of the 2021 International Congress of Advanced Technology and Engineering (ICOTEN), Taiz, Yemen, 4–5 July 2021.
10. Gray, R.M. *Entropy and Information Teory*; Springer Science and Business Media: Berlin/Heidelberg, Germany, 2011.
11. Zhu, L.; Wang, G.; Zou, X. Improved information gain feature selection method for Chinese text classification based on word embedding. In Proceedings of the 6th International Conference on Software and Computer Applications, ACM International Conference Proceeding Series, Bangkok, Thailand, 26–28 February 2017; pp. 72–76. [CrossRef]
12. Ernawati, S.; Yulia, E.R.; Frieyadie; Samudi. Implementation of the Naïve Bayes Algorithm with Feature Selection using Genetic Algorithm for Sentiment Review Analysis of Fashion Online Companies. In Proceedings of the 2018 6th International Conference on Cyber and IT Service Management, CITSM 2018, Citsm, Parapat, Indonesia, 7–9 August 2018; pp. 1–5. [CrossRef]
13. Caragea, C.; Silvescu, A.; Mitra, P. Combining hashing and abstraction in sparse high dimensional feature spaces. In Proceedings of the 26th AAAI National Conference on Artificial Intelligence, Toronto, ON, Canada, 22–26 July 2012; Volume 26, pp. 3–9. [CrossRef]
14. Gao, J.; Ooi, B.C.; Shen, Y.; Lee, W.-C. Cuckoo feature hashing: Dynamic weight sharing for sparse analytics. In Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence (IJCAI-18), Stockholm, Sweden, 13–19 July 2018; pp. 2135–2141. [CrossRef]
15. Ayesha, S.; Hanif, M.K.; Talib, R. Overview and comparative study of dimensionality reduction techniques for high dimensional data. *Inf. Fusion* **2020**, *59*, 44–58. [CrossRef]
16. Mahmoud, U.; Hassan, M.; Zanga, A.I.; Rogo, A.A.; Umar, U. Application of Information Gain Based Feature Selection for Sentiment Analysis Using Movie Review Dataset. *Niger. J. Comput. Engr. Tech. (NIJOCET)* **2023**, *2*, 124–138. Available online: https://nijocet.fud.edu.ng/wp-content/uploads/2023/09/NIJOCET-VOL2-ISSUE-1-010.pdf (accessed on 1 January 2025).
17. Forman, G.; Kirshenbaum, E. Extremely fast text feature extraction for classification and indexing. In Proceedings of the 17th ACM Conference on Information and Knowledge Management, New York, NY, USA, 26–30 October 2008; pp. 1221–1230.
18. Johnson, L.; Lee, K. Investigating the Statistical Assumptions of Naïve Bayes Classifiers. In Proceedings of the 2021 55th Annual Conference on Information Sciences and Systems (CISS), Baltimore, MD, USA, 24–26 March 2021.
19. Garcia, L.; Patel, S. k-Nearest neighbors in high-dimensional spaces: A comprehensive review. *Mach. Learn. Adv.* **2024**, *18*, 112–130.
20. Pang, B.; Lee, L.; Vaithyanathan, S. Thumbs up: Sentiment classification using machine learning techniques. In Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing-Volume 10, Stroudsburg, PA, USA, 6 July 2002; pp. 1606–1624.
21. Abubakar, A.; Abdulkadir, A.B.; Usman, H.; Mohammad, A.; Hahaya, G.S.; Romel, A.; Tayseer, A.; Theyazn, H.H.; Mahmoda, A.; Rami, S. An Improved Multi-Stage Framework for Large-scale Hierarchical Text Classification Problem Using a Modified Feature Hashing and Bi-filtering Strategy. *Int. J. Data Net. Sci.* **2024**, *8*, 2193–2204.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.