Research Poster Outline:

I. Introduction

Explanation of feature selection in machine learning and statistics

Discussion of the importance of feature selection in preventing overfitting and increasing model accuracy

Introduction of the proposed algorithm: Mini Batch K-Means Normalized Mutual Information Feature Inclusion (KNFI) and Mini Batch K-Means Normalized Mutual Information Feature Elimination (KNFE)

II. Methodology

Description of data collection methods

Explanation of how the proposed algorithm was trained and tested

Comparison of the proposed algorithm to existing algorithms in terms of accuracy and computational efficiency

III. Results

Presentation of the results of the study, including any relevant graphs or statistics

Discussion of the implications of the findings

IV. Conclusion

Summary of the proposed algorithm and its effectiveness in feature selection

Discussion of future work and potential applications of the algorithm

V. References

List of references used in the poster, including the cited paper by Cai, Luo, Wang, and Yang (2018) on feature selection in machine learning.

Abstract:
This research proposes a new feature selection method for text classification and binary classification tasks, utilizing Mini Batch K-Means Normalized Mutual Information Feature Inclusion (KNFI) and Mini Batch K-Means Normalized Mutual Information Feature Elimination (KNFE) algorithm. This algorithm combines K-Means clustering and Normalized Mutual Information (NMI) to effectively select relevant features from high-dimensional text datasets. Empirical evaluation on various text classification tasks shows improved accuracy and computational efficiency compared to existing feature selection methods. The proposed algorithm can be applied to various text classification and binary classification problems.

Feature selection in machine learning and statistics is the process of identifying a subset of relevant features from an original feature set to improve model accuracy and prevent overfitting. The proposed Mini Batch K-Means Normalized Mutual Information Feature Inclusion (KNFI) and Mini Batch K-Means Normalized Mutual Information Feature Elimination (KNFE) algorithm utilizes a combination of K-Means clustering and Normalized Mutual Information (NMI) to perform feature selection in a computationally efficient manner. The algorithm is expected to improve model performance and generalization by reducing dimensionality and removing redundant or irrelevant features.

The proposed Mini Batch K-Means Normalized Mutual Information Feature Inclusion (KNFI) and Mini Batch K-Means Normalized Mutual Information Feature Elimination (KNFE) algorithm utilizes a combination of K-Means clustering, an unsupervised learning technique, which groups similar instances based on feature similarity, and Normalized Mutual Information (NMI), a feature relevance criterion that measures the mutual dependence between features and the target variable, to perform feature selection by identifying and eliminating redundant or irrelevant features in a computationally efficient manner, hence, resulting in improved model performance and generalization.

The methodology of this research includes collecting a large and diverse dataset of text samples that represent the target classification task, pre-processing the data and then using the proposed Mini Batch K-Means Normalized Mutual Information Feature Inclusion (KNFI) and Mini Batch K-Means Normalized Mutual Information Feature Elimination (KNFE) algorithm to perform feature selection, following by training and testing of various classification models such as SVM, Random Forest, and Naive Bayes with the selected features and comparing the performance, in terms of accuracy and computational efficiency, with the state-of-the-art feature selection methods such as Chi-Squared, Information Gain, and Mutual Information.

The results of the study, presented through relevant graphs and statistics, show that the proposed Mini Batch K-Means Normalized Mutual Information Feature Inclusion (KNFI) and Mini Batch K-Means Normalized Mutual Information Feature Elimination (KNFE) algorithm

outperforms existing feature selection methods in terms of accuracy as well as computational efficiency as it is evident from the Rand index, and when the data is pre-processed with min-max and standardized scaling. The improved performance is also evident in the reduced number of features selected by the proposed algorithm as compared to the existing algorithms. The findings of this study have significant implications in text classification and binary classification tasks, as the proposed algorithm can effectively reduce the dimensionality and improve the performance of various classification models.

The proposed Mini Batch K-Means Normalized Mutual Information Feature Inclusion (KNFI) and Mini Batch K-Means Normalized Mutual Information Feature Elimination (KNFE) algorithm provides a computationally efficient and effective solution for feature selection in text classification and binary classification tasks, by combining the clustering technique of K-Means and the feature relevance criterion of Normalized Mutual Information (NMI) to identify and eliminate redundant or irrelevant features, resulting in improved model performance and generalization. The results of this study show that the proposed algorithm outperforms existing feature selection methods in terms of accuracy and computational efficiency. Future work includes investigating the application of the proposed algorithm in other domains such as speech and image recognition, and exploring the use of other clustering techniques and feature relevance criteria in combination with the proposed algorithm to further improve its performance