

ANALYSIS OF FEATURE SELECTION METHODS AND ALGORITHMS

Aktamov Shohrubbek Ulug'bek o'g'li

Doctoral Student, Tashkent University of Information Technologies named
after Muhammad al-Khwarizmi

aktamovshohrubbekk@gmail.com

Abstract:

This paper analyzes feature selection methods and algorithms for datasets. With the increasing volume of data, feature selection becomes crucial for improving algorithm efficiency, compressing data, and reducing the complexity of analysis. The paper discusses permissible feature selection approaches and automatic feature selection algorithms, including filter, wrapper, empirical, and machine learning methods. Factors to consider when selecting algorithms, such as dataset characteristics, algorithm purpose, efficiency, accuracy, and interpretability, are also discussed.

Keywords: Feature selection, algorithms, filter methods, wrapper methods, empirical methods, machine learning, efficiency, accuracy, interpretability, feature engineering

Introduction:

The problem of feature selection for datasets is becoming increasingly important with the growth of data volume. This problem leads to improved algorithm efficiency, data compression, reduced complexity of analysis, and the identification of additional dependencies between features. Feature selection is a crucial step for machine learning models, used to improve model accuracy and efficiency, optimize the inference process, and reduce various error-inducing factors. This paper will discuss common feature selection methods, their algorithms, and their importance in real-world applications.

Objective of the Article:

- To analyze feature selection methods and algorithms.

- To identify the advantages and disadvantages of each method and algorithm.
- To determine which methods and algorithms are optimal for different types of data.

Research Methods:

- Literature review
- Comparison of existing methods and algorithms
- Evaluation of the effectiveness of methods on various datasets

Feature Selection Methods

1. Permissible Approaches to Feature Selection:

- * Using all features: This is the simplest approach but can reduce algorithm efficiency.
- * Discarding all features: This approach increases algorithm efficiency but may lead to the loss of important features.
- * Manual feature selection: Experts select features based on their knowledge and experience. This is time-consuming and can be subjective.
- * Automated feature selection: Algorithms are used for feature selection.

This reduces subjectivity and saves time

2. Automated Feature Selection Algorithms:

Filter Methods: These methods evaluate feature importance and remove unimportant features.

- F-test: Evaluates the relationship between a feature and the target variable in a dataset.
- Chi-squared test: Evaluates the dependence between two features.
- ANOVA: Evaluates the effect of features on groups within a dataset.
- Mutual Information (MI): Measures the amount of mutual information between two features.
- Relief: Evaluates a feature's ability to distinguish between near and distant objects.
- Chi2: Evaluates the dependence between features.

Wrapper Methods: These methods add or remove features to optimize algorithm performance.

- Greedy Search: Selects the feature that improves algorithm performance at each step.

Conclusion

Feature selection is a complex problem that depends on the type of data, the purpose of the algorithm, and other factors. The methods and algorithms described in this article should be chosen appropriately depending on the dataset. The problem of feature selection is addressed through various methods. The chosen methods can improve efficiency and accuracy. Filter, wrapper, embedded, and hybrid methods all offer varying levels of importance and efficiency, with their effectiveness directly demonstrated through experimentation. Improving the efficiency and accuracy of feature selection continues to be a critical issue in healthcare, economics, financial indicators, marketing, and education.

In this context, mathematical algorithms are also a successful mechanism in the process of identifying inconsistencies. The key conclusion indicates that the classification of features and risks needs to be supplemented by a more efficient and accurate mechanism for linking factors.

REFERENCES

1. Vakhobova, N. M., Asadullaev, M. M., Raximbayeva, G. S., qizi Botirova, N. A., & Shohruhbek Ulug'bek ug'li, A. Lipoproteins and Their Receptors Association in Cerebrovascular Disease Pathogenesis: A Comprehensive Analysis. <https://doi.org/10.25163/angiotherapy.879786>
2. Актамов Ш., Олламберганов Ф. ТУРЛИ ТИПДАГИ МАЪЛУМОТЛАР ФАЗОСИДА ФИШЕР ТИПИДАГИ МЕЗОНГА АСОСЛАНГАН ИНФОРМАТИВ БЕЛГИЛАР МАЖМУАСИНИ ТАНЛАШ АЛГОРИТМИ //DIGITAL TRANSFORMATION AND ARTIFICIAL INTELLIGENCE. – 2024. – Т. 2. – №. 5. – С. 103-116.
3. Shohruhbek Ulug'bek o'g A. et al. THE ALGORITHM AND SOFTWARE TOOL A COMPLEX OF INFORMATIVE SIGNS IN THE CLASSIFICATION OF DISEASES OF THE CIRCULATORY SYSTEM //Journal of Modern Educational Achievements. – 2024. – Т. 5. – №. 5. – С. 192-198.
4. Shohruhbek Ulug'bek o'g A. et al. COMPARATIVE ANALYSIS OF AN ALGORITHM BASED ON RANDOM SELECTION IN THE CLASSIFICATION OF MEDICAL EMBLEMS //INNUC. – 2024. – Т. 3. – №. 3. – С. 106-120

5. Shohruhbek Ulug'bek o'g A. et al. CLASSIFICATION METHODS AND SOLUTIONS FOR CIRCULATORY DISEASES //Conferencea. – 2024. – T. 5. – №. 5. – C. 278-281.
6. Shohruhbek Ulug'bek o'g A. et al. ALGORITHM AND SOFTWARE TOOL FOR CHOOSING A COMPLEX OF INFORMATIVE SIGNS IN THE CLASSIFICATION OF CIRCULATORY DISEASES //Journal of Modern Educational Achievements. – 2023. – T. 7. – №. 7. – C. 225-229.