PressAcademia
www.pressacademia.org/journals/rjbm
RJBM
ISSN 2148-6689
Research Journal of Business and Management YEAR 2024 VOLUME 11 ISSUE 2

UTILIZING TEXT MINING TO EXTRACT KNOWLEDGE AND CLASSIFY INFORMATION IN INDUSTRIAL ENGINEERING

DOI: 10.17261/Pressacademia.2024.1948
RJBM- V.11-ISS.2-2024(2)-p.83-93

Fereidoon Bidollahkhany[1], Mehdi Basiri[2]
[1]Karabuk University, Faculty of Business Administration, Safranbolu Campus, Karabuk, Turkiye.
fbidollahkhany@karabuk.edu.tr , ORCID: 0009-0004-3064-9289
[2]Tehran Azad University, Department of Business Administration, Tehran, Iran.
basiri60@gmail.com , ORCID: 0000-0002-2519-4679

Date Received: April 25, 2024 **Date Accepted:** December 14, 2024

To cite this document
Bidollahkhany F., Basiri, M., (2024). Utilizing text mining to extract knowledge and classify information in industrial engineering. Research Journal of Business and Management (RJBM), 11(1), 83-93.
Permanent link to this document: http://doi.org/10.17261/Pressacademia.2024.1948

ABSTRACT

Purpose- In the modern information age, organizations face an overwhelming amount of textual data from various sources. The true potential of this data is realized when transformed into actionable knowledge. This study aims to create a text-based model for information classification and knowledge extraction in industrial engineering.
Methodology- The research follows a descriptive-survey approach, employing text-mining techniques to analyze a comprehensive dataset of scientific research articles from the Science Direct database between 2015 and 2020. Data preprocessing was performed using Excel, while analysis was conducted with MATLAB software. The proposed model employs the nearest neighbor and support vector machine algorithms for robust text classification and knowledge extraction.
Findings- The study demonstrates the model's effectiveness in systematically extracting valuable knowledge from diverse textual sources. It shows that this approach can facilitate information extraction without compromising data integrity, thereby contributing to knowledge management practices in industrial engineering.
Conclusion- The text-based model developed in this study provides a reliable method for extracting knowledge from extensive textual datasets. The approach can be applied to other fields beyond industrial engineering, indicating its broader relevance and utility in the contemporary information age.

Keywords: Text mining, knowledge extraction, industrial engineering, data preprocessing, support vector machines.
JEL Codes: C55, L23, O32.

1. INTRODUCTION

Knowledge management offers new approaches in response to the challenges faced by organizations to increase the effectiveness and efficiency of their processes. At the same time, it plays an essential role in the maintenance and continuity of organizational innovation. The paradigm that dominates the activity space of organizations considers knowledge management as the axis of organizational performance in achieving more sustainable competitive advantages. Meanwhile, knowledge creation is not only the first key step in most knowledge management measures but also creates extensive requirements in the next steps of the knowledge management process. This issue has made the issue of knowledge creation an important area in knowledge management processes (Geisler and Wickramasinghe,2009)

On the other hand, to the advancement of technologies and the widespread use of information and communication technology in organizations, as well as their increasing integration of electronic systems, the amount of electronic information has increased. Quick and accurate access to important and necessary resources has become a main concern in using these sources of information. The main reason for paying attention to the knowledge of text mining is the large amount of available data that must be extracted from this information. The science of text mining is a part of the superior science that has created a revolution in the age of technology. It can be said that text mining can be used in any space where data is produced. Textual information processing by machine learning method, NLP natural language processing, and intelligent information retrieval were considered from the beginning regarding structured information, and then a branch of science appeared called data mining. In any case, today we are faced with a large amount of information, and due to the abundance and diversity of this information, checking and analyzing the information is a difficult and time-consuming task.

In this way, useful information can be extracted and used in the combination of features to produce a subset of efficient features that provide the best and most useful features without interfering with the information. This article aims to present a model based on text mining for information classification to extract knowledge from various scientific texts using the SVM and K-NN algorithms.

To achieve this, the paper is structured as follows: The next section reviews the literature on data mining and text mining, highlighting their intersection and relevance to organizational innovation. Following this, the methodology section describes the proposed model and the algorithms used, specifically SVM and K-NN, detailing their implementation for text classification. The discussion section interprets the results, explaining how text-mining techniques and particle optimization methods for data mining were utilized. Exploring and experimenting with various optimization techniques can further enhance classification performance, reduce computational costs, and lead to more robust and effective information classification and knowledge extraction systems.

2. RESEARCH LITERATURE

2.1. Data Mining

In today's era of rapid information and communication expansion, an immense volume of data including numbers, satellite images, photos, and text is continuously proliferating. Access to information is no longer merely a necessity but is recognized as a potent force, emerging as a crucial factor in production dynamics. On the other hand, a big difference in data production processes, such as the paper-based analog method and computer-based digital method, has made the use of data mining more important. Many solutions have been presented to collect, store, organize, and manage information and achieve more valuable results (Romero, Ventura, Pechenizkiy, and Ryan S.J.d. Baker,2010). Data mining is one of the recent developments in data management technologies. Data mining is a set of techniques that allows a person to extract valuable hidden information from the mass of data. It can be said that in scientific and medical studies, the nature of data mining is not considered a new technology because extracting information and knowledge from abundant data has a long history. However, what has led to the interest in this approach is the convergence and sharing of several disciplines, which has created excellent conditions for data mining in the world of science, business, and economics for the user. It can be said that the field of statistics is very important because without statistics data mining has no concept (Gupta, Vishal, Gurpreet S Lehal,2009).

Data mining is a new field of science that has been realized by conducting research in the fields of statistics, machine learning, and computer science (especially database management). One of the most important fields is classical statistics, which includes concepts such as regression analysis, standard distribution, standard deviation, variance, cluster analysis, and confidence intervals. Therefore, it can be said that classical statistical analysis plays an important role in data mining techniques. The next field that is directly related to data mining is machine learning algorithms. Machine learning algorithms can be called the evolution of artificial intelligence because machine learning is a community of statistical analysis and innovative methods of artificial intelligence, and by allowing computer programs to use statistics to understand its concepts and methods Artificial intelligence initiatives work to achieve the goal.

2.2. Text Mining

In today's world, a significant portion of our knowledge exists in the form of audio, video, and text. To unravel and derive insights from this wealth of information, individuals must comprehend the essence of the data and process it effectively to uncover interconnections between concepts. In the realm of this discipline, terms such as "text analysis," "exploring textual data," and "knowledge discovery in text" are commonly used.

Text mining operates similarly to data mining, albeit with a notable distinction: while data mining tools are tailored for structured database data, text mining excels in handling semi-structured or unstructured data, such as emails, full-text documents, and HTML files. Consequently, text mining emerges as a superior solution for organizations grappling with diverse data formats. However, it's noteworthy that the bulk of research and resources have traditionally been devoted to data mining techniques focused on structured data (Hotho, Andreas, Nurnberger, Andreas, Paab, Gerhard, 2005).

2.3. The Difference Between Data Mining and Text Mining

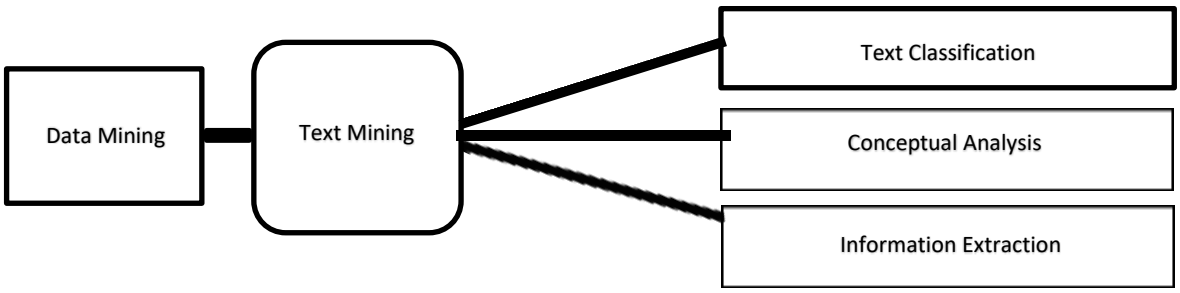
Data mining and text mining both seek to discover knowledge from data. Now the question arises, what is the relationship between data mining and text mining?

Databases can be an example of structured data; But text mining deals with texts that are mostly unstructured, such as web pages, notes, invoices, and emails, or semi-structured texts, such as HTML and XML pages. Therefore, in text mining, to extract information, the texts must first be structured by methods.

As a result, the text itself is data and text mining is one of the sub-branches of data mining. The use of data mining on text gave rise to another branch of artificial intelligence science called text mining. In text mining, because it works on text, natural

language processing methods should be used. In text mining, pre-processing tasks are performed using natural processing methods, and processing tasks are performed by data mining. (Romero, Cristobal, Sebastian Ventura, Senior, 2010)

Figure 1: Data Mining and Text Mining (Relationship)



2.4. Text Mining Process

Text mining is done in the following three steps:

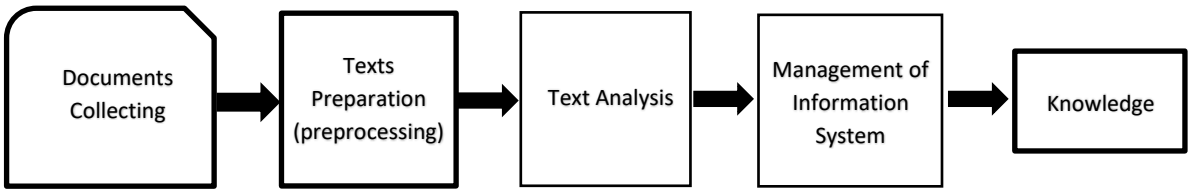
Preparing the text- The database on which the text mining process is supposed to take place is selected with the help of suitable software and performs the cleaning and preliminary processing of the text.

Text processing- This method uses natural language processing and data mining algorithms (such as decision trees, neural networks, and genetic algorithms) and data preparation and compression to detect hidden information.

Text analysis- In text analysis, the obtained data are evaluated to check the correctness of knowledge discovery. (Zong and his colleagues,2021)

The purpose of performing these steps is to extract data limitations and discover the strengths and weaknesses of the presented algorithm.

Figure 2: Text Mining Process



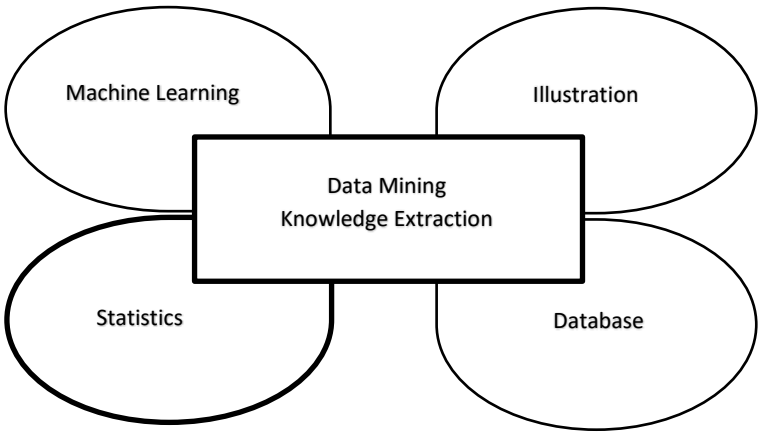
2.5. Discover Knowledge

Knowledge is extracted by categorizing information. Knowledge discovery is related to the entire knowledge extraction process; including how data is stored and accessible, how to analyze huge data sets using efficient and scalable algorithms, how to interpret and visualize results, and how to model and support interactions It is between man and machine.

Knowledge discovery is done in two ways:

- Data mining or extracting knowledge from databases.
- Text mining or extracting knowledge from the text (Hotho, 2005).

Figure 3: Effective Factors on Knowledge Extraction



2.6. Research Background

Ali Soltaninejad and Mohammad Ahmadinejad (2016) in research titled: Text mining of computer engineering articles based on the documents retrieved from the Web of Science database using the famous K-means algorithm to create an algorithm for extracting Persian words have given. The method of this research is a descriptive-analytical type, which was carried out by a survey method and considered the text mining approach. The research community was the computer engineering degrees indexed in the Web of Science database, which reported 6186 records in the period from 2004 to 2014. The collected data were analyzed using Heist Site and Excel version 2013 software as well as RapidMiner version 7.3 software. The results of the research showed that using the words selected as keywords in the clusters can help the user save time and retrieve relevant information.

Seyed Alireza Mir Mohammad Sadeghi and Mehdi Moghan (2014) in research titled Examining the Place of Data Mining in the Process of Knowledge Management and presenting a conceptual model for extracting knowledge presented a systematic process for the principled use of many data sources. This approach aimed to leverage a multitude of data sources effectively to enhance organizational decision-making processes.

Ahmed Agha Kardan and Mina Kihaninejad (2013) in research titled Providing a Model for extracting information from Text documents, based on Text mining in the field of e-learning, examined the types of text mining methods and how to combine and use them in the field of e-learning.

Zeng and his colleagues (2016) in an article investigated sentiment prediction by text mining using particle swarm optimization based on a feature selection search. Sentiment prediction by text documents and in medical text mining using feature selection based on particle swarm search as an important topic of machine learning to gain insights from unstructured texts has gained new popularity in the healthcare industry. Text mining has been one of the basic data analysis methods for predicting emotions for a long time. One of the popular pre-processing steps in text mining is converting the text string into a word vector, which is represented as a thin matrix with high dimensions. This sparse matrix is a computational challenge to induce an accurate sentiment prediction model. Feature selection is a popular dimensionality reduction method that provides a subset of all features of the sparse matrix, which is predicted to increase the accuracy and precision of the model. In this article, a new feature selection method called search-based swarm optimization (OSS-FS) is applied. This method is a group of search functions that selects an ideal subset of features for increasing classification accuracy.

Song and his colleagues (2015) investigated the use of learning-based network resources for automatic text classification. In this article, a new learning class using a resource allocation-based learning network (SLRAN) is presented for text classification. Considering learning progress, SLRAN is divided into a preliminary learning stage and a refined learning stage. In the first stage, a hierarchical K-Means method is used to reduce the sensitivity of the input data and to create the initial algorithm from the hidden layers. After that, a new criterion for adjusting the centers of hidden layers dynamically is proposed. In the last phase, a least squares method is used to increase the convergence rate of the network and improve the classification ability. Such an approach based on learning and its aggregate structure reduces the computational complexity of the network and increases its learning ability. To implement SLRAN for text classification, a semantic similarity approach that reduces the scale of neural network input and hidden meanings between text features is used. Data sets of 20 expertise groups have been tested in experiments and extensive experimental results show that the dynamic learning process of SLRAN improves its classification function.

3. DATA AND RESEARCH METHODOLOGY

This research utilized a blend of SVM and K-NN algorithms, with MATLAB software for data analysis. The study involved 51 English-language scientific articles within the realm of industrial engineering. The selection of English articles was influenced by software constraints. Each criterion was weighted using a scoring system (1=not at all, 3=weak, 5=average, 7=good, 9=very good), and repetition within criteria was addressed by removing duplicate articles.

To validate the model, collected articles underwent scrutiny which details such as authors' names, affiliations, references, journal titles, and submission/acceptance dates were excluded to ensure impartial assessment. University professors served as experts in evaluating the articles based on predefined criteria.

4. FINDINGS AND DISCUSSION

It assigns natural language texts to a previously introduced category based on the content of the text. The work procedure is such that in the first step, data is pre-processed using natural language processing techniques, and the obtained results are interpreted with the help of a machine learning algorithm. The machine learning algorithm consists of two phases, training, and testing, in the training phase, previously defined categories are used for machine learning, and it specifies the meaning of each category. In the test phase, unknown documents are given to the system. The system automatically classifies those data based on the greatest similarity. In the modeling of the machine learning algorithm, choices must be made about the type of training data, the objective function, its display, and an algorithm for learning. Selection of features and samples is one of the important tasks for learning algorithms.

Linguistic Preprocessing- To implement the weighting algorithm, several linguistic pre-processing is needed, which is a matrix of document vectors. The failures of the document that lead to its constituent sentences are: Finding the root of words and removing redundant words.

Remove Redundant Words- Words that connect words in texts and have many uses, such as "if", "and", "or", although these words are highly used, have little semantic value, and for this reason, they are removed in the pre-processing phase. The process of removing these words is such that a list of words is prepared in advance and after seeing these words in the text, they are removed from the document.

Finding The Roots of Words- Exploring the roots of words involves condensing words to their foundational roots. For instance, "computing," "compute," and "computer" all trace back to the root "compute." This process simplifies understanding and categorization. However, it's noteworthy that information retrieval systems typically don't employ root-finding techniques. One prominent root finder algorithm in the English language is the "Martin Porter" algorithm.

Feature Extraction- The operation that takes place on the data to discover its specific and unique features is called feature extraction. In this research, important words are extracted according to textual data.

Create The Final Document-Dictionary Matrix- After discovering the features, we use a matrix of 0 and 1 to train the recognition system, whose columns represent the extracted features, and each row represents a pattern.

Data Reduction- In this research, feature selection and combination algorithms have been used with the help of the IG criterion to weigh the data and remove inappropriate data.

$$IG = P(t_k, C_i) \log \frac{P(t_k, C_i)}{P(t_k) \cdot P(C_i)} + P(\overline{t_k}, C_i) \log \frac{P(\overline{t_k}, C_i)}{P(\overline{t_k}) \cdot P(C_i)}$$

The Information Gain formula helps identify how much information a term provides about the class, which is essential for feature selection in text mining and classification tasks. Higher IG values indicate terms that are more informative about the class, helping to reduce data by focusing on significant features.

4.2. Classification Assessment

To evaluate the obtained results, we use the criterion of accuracy and the amount of data reduction.

4.2.1. Experimental Results

All the tests have been run on a machine with a 4.0GHz CPU and 16.00 GB RAM, the operating system used is Windows 10. The proposed method has been implemented in MATLAB R2013a. In the next sections, the data set and the results of the implemented tests are explained.

4.2.2. Data Collection

In this research, 51 scientific articles in English in the field of industrial engineering, which contain scientific-research articles, were used. One of the reasons for choosing English articles was software limitations. To give weight to each criterion, options were assigned to those options that each option has a specific score (1=not at all, 3=weak, 5=average, 7=good, 9=very good) and to reach the desired answer to the questions Repetition was considered in the criteria, and in case of occurrence, the selected article will be removed from the research process. The selected articles were model validation. To validate the model, the collected articles were reviewed by researchers, and information such as authors' names, authors' information, references, journal names, and date of submission and acceptance of the article were removed to better review the articles. After preparing the articles, their evaluation criteria were evaluated by different researchers.

4.2.3. Evaluation Method

In this research, only words are not enough, and key phrases are also considered. After the keywords and phrases were identified for each criterion, it was found that there were similar keywords and phrases for some criteria, as a result, some special attributes were recognized as common keywords and phrases between several criteria. Finally, after removing similar traits, 121 traits were selected for this research. These 121 keywords and phrases are the criteria that the software should use to perform the learning and testing process.

As stated earlier, one of the goals of this research is to present a model that can be used to evaluate a large volume of articles based on a series of predetermined criteria, and the articles that meet these criteria should be distinguished from the articles in which these criteria are not met, therefore, in this research, classification methods have been used to achieve this purpose. Because in this research, data with an imbalance in the class was used, the accuracy measure cannot have a completely correct evaluation of the created models. For this reason, measures such as sensitivity, accuracy, or ROC curves should be used.

Among all the methods available for classification, three methods of Bayesian, Auto MLP, and k nearest neighbors were used (the reason for using the above three models is that after the text processing stage, the records that are used to perform the stage Data-mining is in hand, they have many special features, for this reason, one should use models that have a high ability to use this type of records and can obtain acceptable answers) and to validate the model The cross-validation method and ROC diagram have been used, and finally, bagging, “ADABOOST “and voting techniques have been used to improve the classification performance.

The results obtained from the application of Bayesian, Auto MLP, K nearest neighbor, and improved methods of these methods are shown in tables 1, 2, 3, and 4 (disturbance tables).

Table 1: Disturbance Matrix (Method of BAYESIAN)

	The second category (Prediction)	The first category (Real)	Class accuracy (percentage)
The second category (prediction)	6	2	75
The first category (real)	4	6	60
Level of sensitivi (Percentage)	70	75	-

According to Table number 1, the model placed 6 articles in the second category (articles that did not meet the criteria) which are really in the second category data (professors and doctoral students put these articles in this category), and also 4 articles were placed in the first category (articles in which the considered criteria were observed) and these articles were placed in the second category by experts. On the other hand, this model has placed 2 articles in the second category, and these two articles have been placed in the first category by the experts and have placed 6 articles in the first category, and the experts' opinions about These are 6 similar articles. According to the above explanation and the percentage of the values in the table, the level of sensitivity and accuracy of the classes is also clear in the last row and column of Table 2 (other tables have a similar interpretation).

4.2.4. Test Review

After performing the pre-processing steps and selecting the features and samples in the text classification system that was explained in the previous parts, it is time to implement and compare the classification algorithms. The tests performed using two classification methods (SVM and K-NN) can be seen in related tables. These results indicate the improvement of system performance after using the proposed method, especially using SVM as a classifier.

In Tables (2 and 5), the performance of basic learners has been implemented on 4 different datasets and the accuracy of classification and the amount of data reduction have been shown. As can be seen, in most cases the results show improvement in classification performance. This increase in efficiency is about 3%, which is more in SVM than in other classifications. As can be seen from the results in the table, it can be said that the proposed method has a better efficiency for data classification compared to using feature and sample selection separately. In addition to the increase in accuracy, we see a reduction in data that has a significant effect on speeding up the learning stage. For example, in the DVD dataset, the classification accuracy using SVM has increased from 0.78 to 0.811, while the number of features has decreased from 4657 to 161. Of course, as it is clear from the results in the tables and figures, this improvement can be seen in other classifications as well, which shows the efficiency of using the proposed method.

Table 2: The Accuracy of Different Classifications For "DVD".

Technique	Without using data reduction		FS		FU-	
	ACCURACY	#Fe	accuracy	#Fe	accuracy	#Fe
SVM	0.767	4657	0.780	500	0.811	161
K-NN	0.640	4657	0.677	500	0.71	164

Chart 1: The Accuracy of Different Classifications for The Data of "DVD"

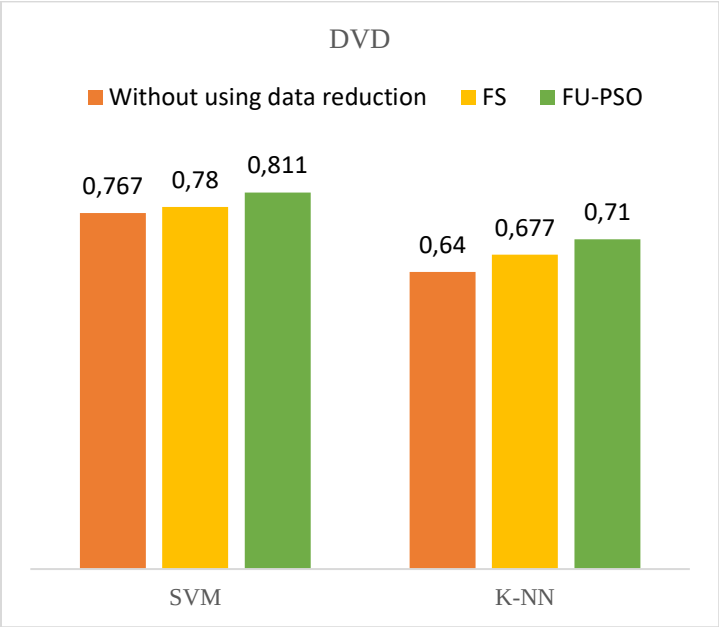


Table 3: The Accuracy of Different Classifications for the Data" BOOK"

Technique	**Without using data reduction**		**FS**		**FU-PSO**	
	Accuracy	#Fe	Accuracy	#Fe	Accuracy	#Fe
SVM	0.762	8457	0.780	500	0.821	200
K-NN	0.647	4657	0.677	3000	0.712	955

Chart 2: The Accuracy of Different Classifications for the Data of Book

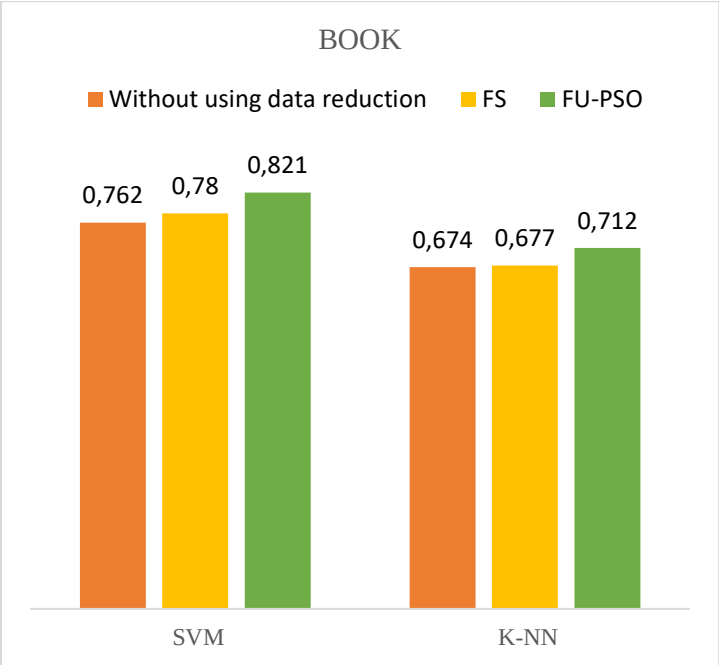


Table 4: The Accuracy of Different Classifications for the Data" Electronic"

Technique	Without using data reduction		FS		FU-PSO	
	Accuracy	#Fe	Accuracy	#Fe	Accuracy	#Fe
SVM	0.722	4216	0.785	1500	0.803	500
K-NN	0.725	4216	0.723	1000	0.753	330

Chart 3: The Accuracy of Different Classifications for the Data of Electronic

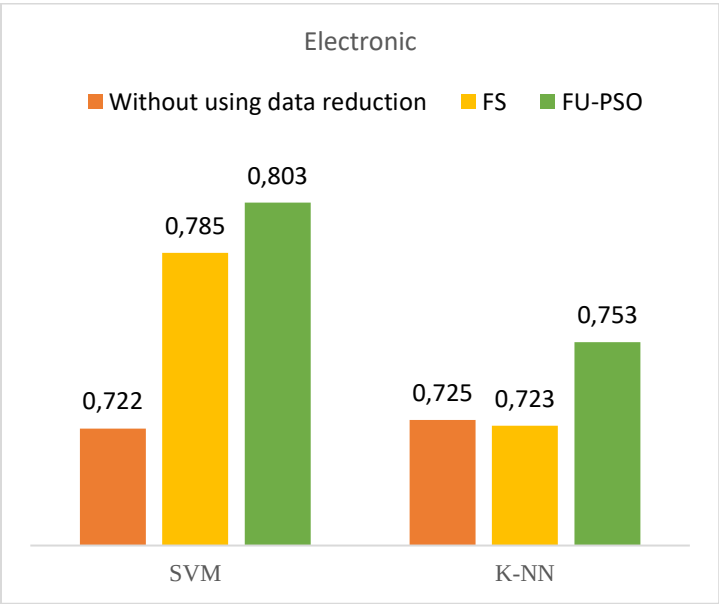


Table 5: The Accuracy of Different Classifications for the Data of KITCHEN

Technique	Without using data reduction		FS		FU-PSO	
	Accuracy	#Fe	Accuracy	#Fe	Accuracy	#Fe
SVM	0.820	3738	0.822	500	0.843	200
K-NN	0.720	3738	0.722	2000	0.747	750

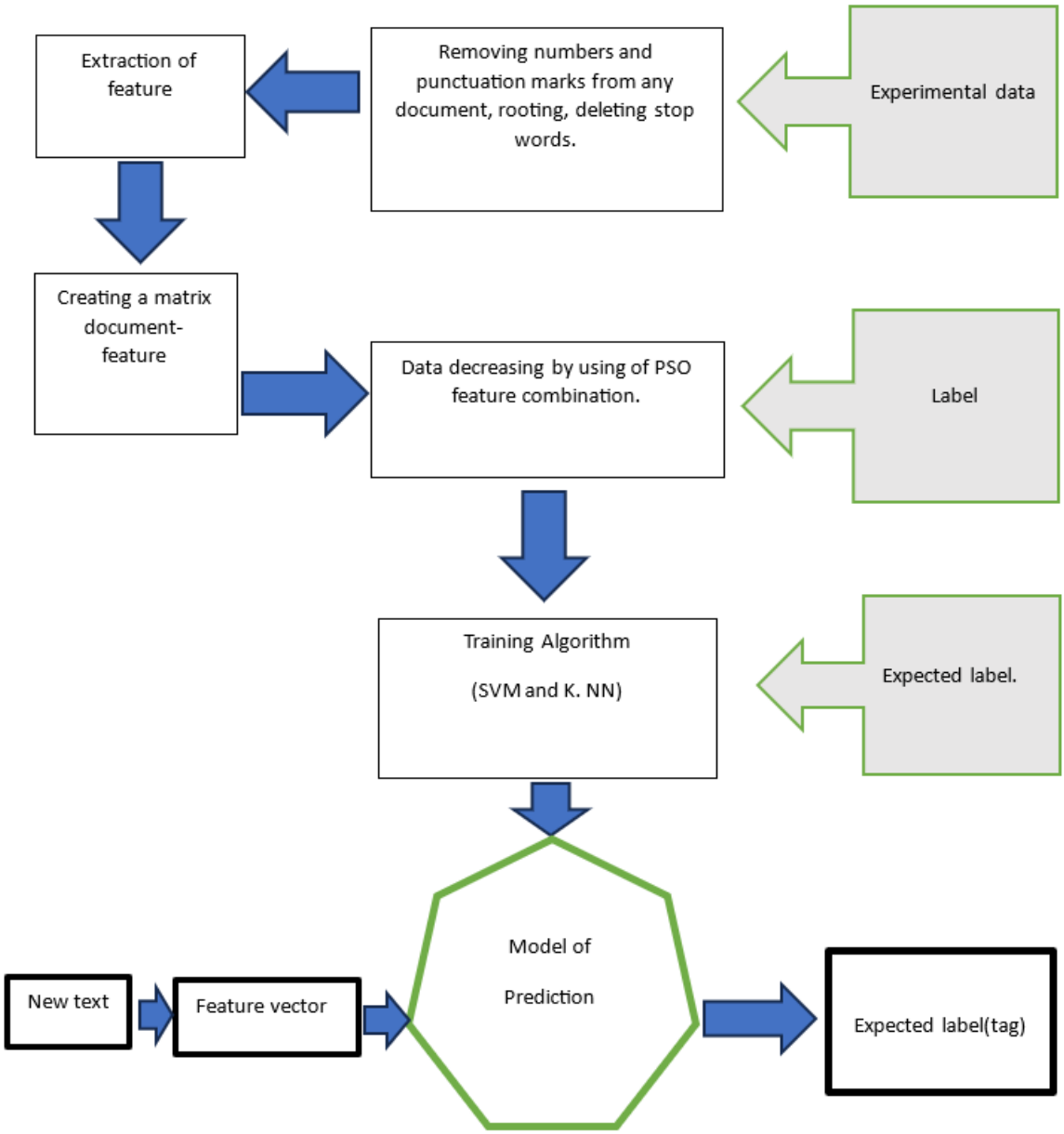
Chart 4: The Accuracy of Different Classifications for the Data of Kitchen

Classification of all datasets using the SVM algorithm has the highest accuracy. These results show that SVM has more classification power, which is consistent with previous research. The results of the experiments indicate the improvement of the classification performance by using the feature combination based on the particle optimization method in the field of data mining. It can also be said that the use of this method has a significant effect on increasing classification efficiency due to data reduction at the feature level. It can also be said that the effect of combining features with the K-NN algorithm is successful, which is due to the sensitivity of this algorithm to inappropriate data.

5. CONCLUSION AND RECOMMENDATIONS

The current research was conducted to provide a data mining-based method for information classification for knowledge extraction using the K-NN classification method and SVM neural network. The nearest neighbor algorithm has the task of finding the best and fastest answer from all the data, and the neural network is the supporting vector machine of pseudo-artificial intelligence. The example used in the SVM algorithm is presented with a linear structure and has a lot of similarities with the multilayer perceptron neural network or MLP. It can be said that these two models, despite their differences, have similar structures. How the MLP neural network works. It is such that it adjusts the error minimization parameters, but the SVM algorithm defines the risk caused by the lack of correct classification as an objective function, and the parameters will be adjusted and optimized according to the objective function.

Figure 4: Proposed a Text Mining-Based Conceptual Model



It this study, text-mining techniques were used, but advancements in data mining, natural language processing, and machine learning always is progressing. being updated with the latest methodologies and tools is important for improving the effectiveness of information classification and knowledge extraction.

The study mentions the use of particle optimization methods for data mining. Exploring and experimenting with various optimization techniques can further enhance classification performance and reduce computational costs.

Encourage collaboration between experts in industrial engineering, data science, and machine learning as Interdisciplinary teams can bring diverse perspectives and skills to improve the model and its applications. Additionally, collaboration and interdisciplinary approaches can lead to more robust and effective information classification and knowledge extraction systems.

The experimental results of this research were analyzed using two classification techniques (SVM and KNN) and data reduction (feature selection). From these results, it can be concluded that using the feature combination method increases the efficiency of classification and reduces the effect of this increase in reducing the efficiency of the classifier. The point that

should be considered in the following works is that the more the connection between the words selected to combine the features, the higher the accuracy of the system. Using a combination tactic with a hierarchical structure according to this structure can help with this goal.

REFERENCES

Agha Kardan, A., & Kihaninejad, M. (2013). Providing a model for extracting information from text documents, based on text mining in the field of e-learning.

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. NAACL-HLT 2019. https://arxiv.org/abs/1810.04805

Geisler, E., & Wickramasinghe, N. (2009). Principles of Knowledge Management: Theory, Practice, and Cases. New York, NY: Routledge. ISBN 9781315686448

Grootendorst, M. (2022). BERTopic: Neural Topic Modeling with Transformers and c-TF-IDF. Journal of Computational Social Science, 5(1), 345-360.

Gupta, V., & Lehal, G. S. (2009). A survey of text mining techniques and applications. Journal of Emerging Technologies in Web Intelligence, 1(1), 60-76.

Hotho, A., & Nürnberger, A. (2005). A brief survey of text mining. Journal for Language Technology and Computational Linguistics. https://doi.org/10.21248/jlcl.20.2005.68

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., ... & Stoyanov, V. (2019). Roberta: A Robustly Optimized BERT Pretraining Approach. arXiv preprint arXiv:1907.11692.

Müller, M., Hemati, W., & Reimers, N. (2021). Text Mining for Social Media Analysis: Trends and Tools. Information Systems Frontiers, 23(4), 1045-1060.

Romero, C., Ventura, S., & Ventura, S. (2010). Educational Data Mining: A Review of the State-of-the-Art Systems. CRC Press. ISBN 9780429130120

Sadeghi, S. A. M. M., & Moghan, M. (2014). Examining the place of data mining in the process of knowledge management and presenting a conceptual model for extracting knowledge. Presented a systematic process for the principled use of a large number of data sources.

Soltaninejad, A., & Ahmadinejad, M. (2016). Text mining of computer engineering articles based on the documents retrieved from the Web of Science database using the famous K-means algorithm to create an algorithm for extracting Persian words.

Sun, H., Yang, X., Zhang, Z., & Chen, Y. (2022). Advanced Sentiment Analysis with AI Models in Business Applications. Expert Systems with Applications, 200, 117164.

Song, Q., & Colleagues. (2015). A centrality-based label propagation algorithm for community detection in networks.

Tay, Y., Dehghani, M., Bahri, D., & Metzler, D. (2022). Transformers in Text Mining: An Overview. Communications of the ACM, 65(7), 72-80.

Wickramasinghe, N. (2008). Knowledge creation. In Knowledge Management Concepts, Methodologies, Tools and Application.

Zeng, & Colleagues. (2016). Sentiment prediction by text mining using particle swarm optimization based on a feature selection search.

Zong, C., Zhang, J., & Xia, R. (2021). Text Mining in Data Mining. Springer. ISBN 978-981-16-0100-21