University of Tehran Press

**Iranian Journal of Management Studies (IJMS)**
Home Page: https://ijms.ut.ac.ir

Online ISSN: 2345-3745

# Computational Cost Reduction Strategies for Business Cases

**Genesis Sembiring Depari**

Faculty of Economics and Business, Universitas Pelita Harapan, Indonesia. Email: genesissembiring@gmail.com

## ARTICLE INFO

**Article type:**
Research Article

**Article History:**
Received 24 January 2022
Revised 17 October 2022
Accepted 05 November 2022
Published Online 18 June 2023

**Keywords:**
*Evolutionary Algorithm,*
*Filtering Methods,*
*Grid Search,*
*Parameter Optimization.*

## ABSTRACT

Feature selection and parameter optimization are vital techniques in the data mining process, significantly impacting the computational costs of machine learning. Computational cost is a critical consideration in business analytics, making feature selection and parameter optimization research crucial for reducing operational costs. This study investigates the performance of 10 dimensionality reduction methods and 2 parameter optimization techniques in various business applications. The evaluation focuses on predictive accuracy and run time. The analysis reveals distinctive tendencies among the filtering methods, highlighting time-consuming behaviors in different business scenarios for Weight by Rule (WRul) and Weight by Relief (Wrel). Additionally, the study proposes a cost-effective approach to parameter optimization by utilizing grid search and evolutionary algorithms, particularly when the optimal parameter range is unknown.

**Cite this article:** Depari , G. S. (2023). Computational Cost Reduction Strategies for Business Cases. *Iranian Journal of Management Studies* (*IJMS*), 16 (3), 757-768. DOI: http//doi.org/10.22059/ijms.2022.337917.674929

 **Publisher:** University of Tehran Press.
DOI: http//doi.org/10.22059/ijms.2022.337917.674929

## 1. Introduction

With the rapid advancement of information technologies in business activities, recording large datasets has become feasible, offering companies the opportunity to extract valuable insights and improve efficiency. However, the computational cost and run-time required to extract such insights from large datasets pose significant challenges. To reduce computational costs, it is necessary to address both the dimensionality of the model and the tuning parameters of the algorithm used. High-dimensional datasets often contain irrelevant and redundant attributes that can impede the extraction of meaningful insights. Feature selection plays a crucial role in creating a more robust model and enhancing its generalization ability (Pal and Mitra, 2004). Researchers are actively exploring feature selection and parameter optimization to improve model performance (Garg, 2020). Therefore, parameter optimization and dimensionality reduction are essential techniques for reducing computational costs and run-time, ultimately leading to more efficient machine learning applications.

When dealing with high-dimensional data, feature selection techniques may incur additional processing time, and certain techniques may be less suitable (Kou et al., 2020). The performance of a feature selection technique often depends on data characteristics such as data quality, number of dimensions, and number of observations. Therefore, it is crucial to examine multiple feature selection techniques in relation to the specific data available. Furthermore, over the past decades, three categories of feature selection (Guyon and Elisseeff, 2003) have been introduced: filter methods, wrapper methods, and embedded methods. Wrapper and embedded methods typically require more computational resources compared to filter methods, which is a concern in business operations.

The No Free Lunch theorem emphasizes that there are no shortcuts to reducing the computational cost of parameter optimization issues (Igel, 2014). Hyperparameter optimization typically involves searching a large parameter space, making it a costly process to tune (Feurer and Hutter, 2019). Researchers must possess preliminary knowledge, especially when using a grid search algorithm, to identify influential parameters that can improve the performance of a machine learning algorithm without extensively exploring a large parameter space (Yu and Zhu, 2020). Grid search and evolutionary parameter optimization techniques are two well-known techniques in machine learning. Grid search often achieves high learning accuracy but comes with a high computational cost. Additionally, when using grid search, it is necessary to have prior knowledge of the best parameter range to reduce computational costs.

In comparison to previous research, our study focuses primarily on reducing computational costs for business datasets by examining ten filtering methods: Weight Information Gain (WIG), Weight by Information Gain Ratio (WIGR), Weight by Rule (WRul), Weight by Deviation (WD), Weight by Correlation (WC), Weight by Chi-Squared Statistics (WCSS), Weight by Gini Index (WGI), Weight by Uncertainty (WU), Weight by Relief (WRel), and Weight Principal Component Analysis (WPCA) for dimensionality reduction. We also explore two potential parameter optimization techniques: grid search and evolutionary methods. The objective of this study is to identify the most efficient and cost-friendly filtering methods and parameter optimization techniques that reduce computational costs, making them accessible for small and medium-sized companies with limited available resources.

The remaining sections of this research are organized as follows: Section 2 discusses the feature selection methods used in this study. Section 3 delves into the concept behind the support vector machine algorithm. Section 4 covers the datasets and methodology employed. Section 5 presents the analysis and results. Finally, Section 6 concludes with recommendations.

## 2. Feature Selection Methods

Various methods have been discussed in the literature for reducing model dimensionality. These methods can be classified into three categories (Guyon and Elisseeff, 2003): filter methods, wrapper methods, and embedded methods. However, wrapper and embedded methods are often associated with high computational costs. Therefore, in this study, we examined 10 filtering methods to identify the most efficient filtering methods that offer affordable computational costs.

### 2.1 Weight by Information Gain

To be able to calculate the information gain, the average entropy of a given attribute must be computed first. This calculation can be performed by applying the following formula.

$$\mathrm{E}(\mathrm{T},\mathrm{X}) = \sum_{x} P(x)E(x)$$

P(x) is the probability of class-x. The higher the entropy value, the more content or information in this Attribute. Before we have the Average Entropy with respect to the Attribute x, E(x) needs to be computed first by using the Shannon entropy formula:

$$E(x) = -\sum_{i=1}^{n} P(x_i) log_2 P(x_i)$$

After calculating the Average Entropy with respect to Attribute x, then we have to calculate the entropy of the target attribute denoted by H(x) formulated as below.

$$H(x) = -Therefore \sum_{l=1}^{n} P(x_l) log_2 P(x_l)$$

Therefore, information gain can be calculated by deducting the target entropy attribute and average entropy formulated below.

$$\text{Information Gain} = H(x) - \mathrm{E}(\mathrm{T},\mathrm{X})$$

### 2.2 Weight by Information Gain Ratio

Because Information Gain suffers from large values used, a normalization technique was then applied to the result of Information gain. This technique is called Information Gain Ratio. This technique is applied by normalizing the information gained by using the formula below.

$$\mathrm{SplitInfoM}(\mathrm{D}) = -\sum_{J=1}^{V} \left(\frac{|D_j|}{|D|}\right) Log_2 \left(\frac{|D_j|}{|D|}\right)$$

The formula above depicts the generated information by splitting feature M within the training data D with respect to the v-label output. Therefore, we can define the information gain ratio formula as follows

$$\text{Information Gain} Ratio_M = \frac{Gain(M)}{SplitInfo(M)}$$

### 2.3 Weight by Rule

Weight by rule calculates the weight of an attribute by utilizing a single rule that takes into account the attribute label and calculates the error for each attribute. The attribute with the highest weight is considered the most relevant attribute in relation to the attribute label. This toolbox is available in Rapidminer2020 9.7.002.

### 2.4 Weight by Deviation

The weight by deviation toolbox computes the weight of an attribute based on the normalized standard deviation of each attribute, considering the attribute label. Various techniques are available for normalizing the standard deviation, such as normalization by the maximum, average, and minimum values of the features. The standard deviation can be formulated as follows.

$$S = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (x_i - \mu)^2}$$

Where $S$ is the standard deviation, $x_i$ is the value of sample items, and μ is the mean of the observed sample.

### 2.5 Weight by Correlation

The weight by correlation toolbox in Rapidminer2020 9.7.002 computes the weight of an attribute based on the correlation of each Attribute which respects the attribute label. The correlation technique can be described as follows.

$$r_{xy} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}}$$

where $r_{xy}$ is the correlation of Attributes x and y, $x_i$ is the value of Attribute x in a sample, $\bar{x}$ is the mean of Attribute x, $y_i$ is the value of attribute y in a sample, and $\bar{y}$ is the mean of attribute y in a sample.

**2.6 Weight by Chi-squared Statistics**

The weight of the Chi-squared Statistics toolbox in Rapidminer2020 9.7.002 computes the weight of attributes based on the Chi-squared statistics of each Attribute which respects the attribute label. The chi-squared statistic is a famous technique in the field of statistics that is used to gauge if the dispersion of determined frequencies varies from the theoretically expected frequencies. The correlation technique can be formulated as follows.

$$x^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$$

where squared is $x^2$ is Chi-squared Statistics results, $O_i$ is the number of measurements of type *i*, n is observations in total, $E_i$ is the type *i* expected.

**2.7 Weight by Gini Index**

The Gini index was first used to assess the impurity of features for categorization purposes (Zhu et al., 2014). Suppose T is an example of n classes, then Gini (T) is defined as follows.

$$Gini(T) = 1 - \sum_{i=1}^{N} p_i^2$$

where $p_j$ is the probability of *i* in *T*. Therefore, if dataset *B is* into two subsets, $T_1$ and $T_2$, then the Gini index could be defined as follow:

$$\text{giniB}(\text{T}) = \frac{T_1}{T} gini(\text{T1}) + \frac{T_2}{T} gini(\text{T2})$$

Therefore, the reduction in impurity is defined as

$$\Delta Gini\ (T) = gini\ (T) - gini_B(T)$$

The feature which achieves the lowest $\text{gini}_{\text{split}}$ (T) or generates the highest reduction in impurity is then selected to split the node.

**2.8 Weight by Uncertainty**

Symmetrical uncertainty with respect to the class was measured in order to weigh the Attribute. This process is called weight by uncertainty. The calculation process is formulated as follows.

$$Attribute\,Relevant = \frac{2\big(P(Class) - P(Class|Attribute)\big)}{P(Class) + P(Attribute)}$$

**2.9 Weight by Relief**

The relief algorithm was initially developed by Kira and Rendell (1992) to address binary classification problems with numerical attributes as inputs. When evaluating attribute quality, the Relief algorithm stands out as one of the most effective and straightforward algorithms. The main idea behind this algorithm is to predict attribute quality by considering the differences in attribute values between the closest identical instances (*near-hits*) and the closest distinct classes (*near-misses*) that are in proximity to each other.

**2.10 Weight by PCA**

PCA can be calculated by computing the eigenvectors of the covariance matrix of the input variables. PCA can convert high dimensional data into lower dimensional data, which are independently orthogonal and uncorrelated (Cao et al., 2003). In doing so, first, we calculate the eigenvalues of the attribute matrix by computing the covariance matrix p x p, where p is the number of attributes illustrated below.

$$\begin{bmatrix} Cov(k,k) & Cov(k,l) & Cov(k,m) \\ Cov(l,k) & Cov(l,l) & Cov(l,m) \\ Cov(m,k) & Cov(m,l) & Cov(m,m) \end{bmatrix}$$

where *k*, *l*, and *m* are the attributes. Attributes that are highly correlated depict a redundant attribute that needs to be eliminated from the model. Afterward, finds the eigenvalues by solving the following formula.

$$\text{Eigenvalue} = \det(\Sigma - \lambda I) = 0$$

where $\Sigma$ is a covariance matrix, λ is *lamda*, I is an identity matrix. Furthermore, sort the eigenvalues by abandoning the smallest values. Then the selected eigenvalues are employed to convert the data into eigenspace (eigenvectors) by applying the formula below.

$$i^{th} \text{ eigenvector} = \Sigma e_i = \lambda_i e_i$$

## 3. Machine Learning Algorithm

The Support Vector Machine (SVM) was initially introduced by Vladimir N Vapnik and Alexey Chervonenkis in 1963 and later further developed by Boser et al. (1992) to incorporate the kernel trick, which maximizes the margin hyperplane. SVM is a supervised learning algorithm commonly used for classification and regression problems. The concept of SVM is illustrated in Figure 1.

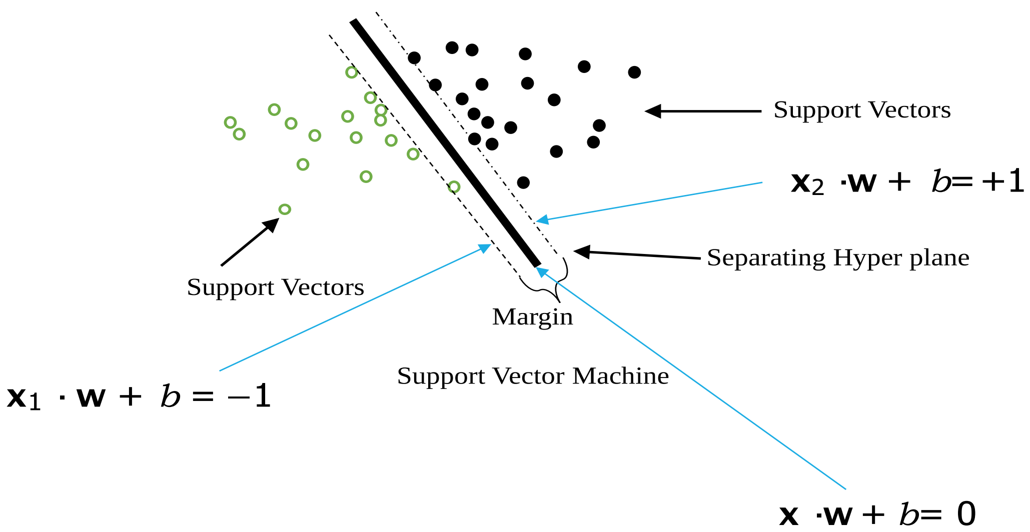


**Figure 1.** Support Vector Machine

The distance between two separating hyperplanes can be computed following the formula below.

$$d(w) = x_2 - x_1 \frac{w}{w} = \frac{2}{w}$$

Maximizing $d(w)$ is identical with minimizing $\|w\|$. Therefore, the constrained optimization problem in the Support Vector Machine can be formulated in the *Lagrangian* function as follow.

$$\text{Min}\, L(\mathbf{w}, b, \{\alpha_i\}) = \frac{1}{2} w^2 - \sum_{i=1}^{n} \alpha_i \left[ y_i (x_i \cdot w + b) - 1 \right]$$

Subject to $\alpha_i \geq 0,\ y_i(\mathbf{x}_i \cdot \mathbf{w} + b) - 1 \geq 0,$

$\alpha_i[y_i(\mathbf{x}_i \cdot \mathbf{w} + b) - 1] = 0,\ \forall i = 1, \ldots, N$

Furthermore, in dealing with new datasets, hard and soft margin techniques can be possibly selected. Once the hard margin technique is applied, a new hyperplane will be produced consequently the separating hyperplane becomes narrower, therefore vulnerable to falling into an overfitting problem. In tackling this problem, a soft margin approach was performed in this study. The soft margin approach allows some data to violate the constraints denoted as *slack variables* ($\{\varepsilon_i\}$). Therefore, the *lagrangian* optimization problem changes as follows.

$$\operatorname{Min} L(\mathbf{w},b,\alpha)=\frac{1}{2}w^2+C\sum_i \varepsilon_i-\sum_{i=1}^{n}\alpha_i\left[y_i\left(x_i\cdot w+b\right)-1+\varepsilon_i\right]-\sum_{i=1}^{N}\beta_i\varepsilon_i$$

*C* is the penalty parameter imposed to penalize each violation towards the constraints.

## 4. Data and Methodology

Five types of business-related datasets were collected from various sources, as described in Table 1. All datasets were transformed into numeric features, except for the label, which remained in binary form. To ensure efficient run-time and memory usage, stratified random sampling was employed to extract 10% of the data from the population. In order to ensure consistent results across all tested models, local random seeds were set throughout the analysis in this study.

**Table 1.** Data Characteristics

| Name | Attribute | Observations | Purpose | Source |
|---|---|---|---|---|
| Bank telemarketing dataset | 17 | 45211 | New subscribers' prediction | UCI Machine Learning Datasets |
| Bitcoin dataset | 27 | 729 | Predict bitcoin daily movement | Various sources |
| Absenteeism Dataset | 21 | 740 | Work absenteeism Prediction | UCI Machine Learning Datasets |
| Social media dataset | 11 | 500 | Post-performance Prediction | Indonesian Educational Service Company |
| Online news dataset | 61 | 39797 | Social networks popularity prediction | UCI Machine Learning Datasets |

The first dataset, known as the bank telemarketing dataset, was collected by the Portuguese Bank. Its purpose is to predict whether a customer will subscribe to a banking deposit or not. The second dataset, called the bitcoin dataset, was collected from various sources and aims to predict the daily movement of bitcoin. The third dataset, referred to as the absenteeism dataset, is used to predict employees' work absenteeism at the workplace. The fourth dataset, known as the social media dataset, was obtained from the company's social media page and is utilized to predict the performance of social media content. Lastly, the fifth dataset was acquired from the UCI Machine Learning Datasets and focuses on predicting the popularity of online news.

Rapidminer2020 9.7.002 was utilized to address the research objectives. The analysis followed the CRISP-DM procedure, developed by a consortium of data mining users, practitioners, and suppliers (Shearer, 2000). Firstly, an understanding of the dataset's domain and business nature was gained, aligning with the research objectives. Relevant attributes and methodologies were then selected accordingly. Secondly, the data was cleaned and normalized using the Z-transformation technique.

To evaluate the performance of each filtering method, a LibSVM library was employed for a classification task. Parameter optimization was carried out for two key parameters: Gamma and C. Gamma controls the curvature of the decision boundary and influences the level of curvature. A higher gamma value indicates a higher level of curvature. On the other hand, C is used to regulate the trade-off between errors in the training and testing data. For the banking dataset, the RBF kernel function was chosen due to its non-linear characteristics, which align well with the dataset's properties. The RBF kernel function can be described as follows.

$$K\left(X_1,X_2\right)=exp\left(-\frac{\left\|X_1-X_2\right\|^2}{2\sigma^2}\right)$$

where $\mathrm{K}(X_1, X_2)$ is the kernel function, $X_1 - X_2$ is the Euclidean distance between two distances of $X_1$ *and* $X_2$, and $\sigma$ is the variance. However, for the bitcoin and absenteeism datasets, the linear kernel function was chosen since these datasets form linearly separable data. Furthermore, to optimize the parameters, grid search and evolutionary algorithm were compared and assessed. Moreover, to be able to avoid the problem of over-fitting, a 10-fold cross-validation technique was performed on the trained support vector machine model. Additionally, Weight Information Gain (WIG), Weight by Information Gain Ratio (WIGR), Weight by rule (WRul), Weight by Deviation (WD), Weight by Correlation (WC), Weight by Chi-Squared Statistics (WCSS), Weight by Gini Index (WGI), Weight by Uncertainty (WU), Weight by Relief (WRel), and Weight Principal Component Analysis (WPCA) were also examined in reducing the model dimensionality. The complete sequence is depicted in figure 2.

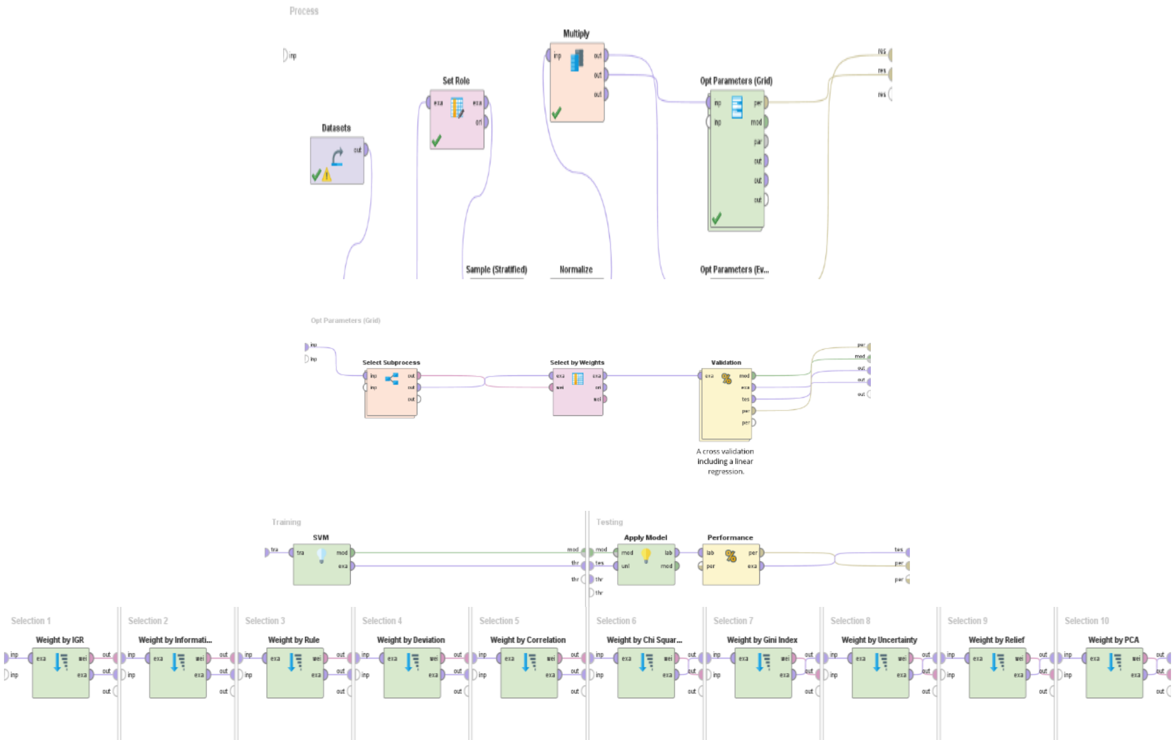

**Figure 2.** Data Analysis Processes

## 5. Analysis and Results

The main focus of this experiment is on dimensionality reduction and hyperparameter optimization, as they have a significant impact on the computational costs of machine learning applications. To achieve efficient classification processes and results, ten filtering methods (Weight by Information Gain - WIG, Weight by Information Gain Ratio - WIGR, weight by rule - WRul, Weight by Deviation - WD, Weight by Correlation - WC, Weight by Chi-Squared Statistics - WCSS, Weight by Gini Index - WGI, Weight by Uncertainty - WU, Weight by Relief - WRel, and Weight Principal Component Analysis - WPCA) and two parameter optimization techniques (Grid Search - GS and Evolutionary Algorithm - EA) are combined and tested.

The first experiment focuses on a bank telemarketing dataset, predicting whether a consumer will subscribe to a banking deposit or not. It was observed that WRel, both in GS and EA, is a time-consuming algorithm. In terms of accuracy, WRel in EA achieves the highest accuracy compared to other filtering methods. However, when grid search is used, WRel takes twice as much time as EA but with slightly different accuracy. In this experiment, WU achieves the highest accuracy and consumes the least time compared to other filtering methods when grid search is used. Therefore, it can be concluded that grid search can achieve higher accuracy but comes with higher computational costs.

The second dataset focuses on predicting the daily price movement of Bitcoin. In this dataset, it was observed that WRul is the slowest algorithm when used with both grid search and evolutionary algorithm. In terms of accuracy, WIG and WGI perform the best in grid search optimization, achieving an accuracy of 85.33%. WIG is also recognized as the most efficient filtering method in grid search parameter optimization, with the least processing time. On the other hand, when the evolutionary algorithm is used, WCSS is identified as the most accurate filtering method, with an accuracy of 84.64%. Additionally, the types of filtering methods do not significantly affect the accuracy achieved by both GS and EA, as they perform similarly across various filtering method types. The results are depicted in Figure 4.

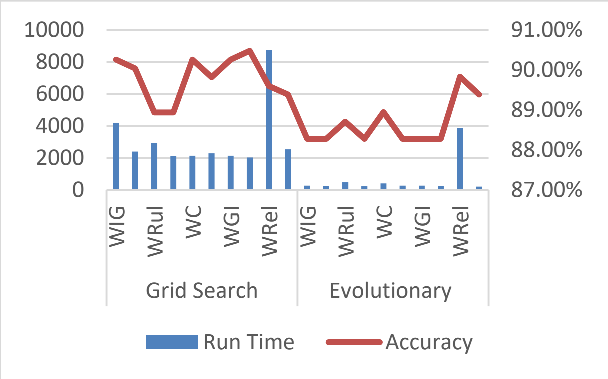


**Figure 3.** The performances of filtering methods and parameter optimization techniques of bank telemarketing dataset

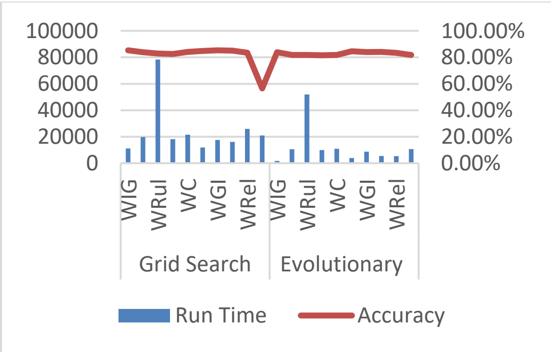


**Figure 4.** The performance of filtering methods and parameter optimization techniques of bitcoin dataset

When working with the absenteeism dataset, it was observed that WRel is the slowest filtering method when used with both grid search and evolutionary algorithm. On the other hand, WRul was found to be the most accurate filtering method for both grid search and evolutionary algorithm, achieving an impressive accuracy of 99.86%. Interestingly, when combined with the evolutionary algorithm, WRul can achieve equally efficient accuracy as the grid search algorithm. Therefore, WRul is proposed as an efficient computational approach for the absenteeism dataset.

Comparing the performance of grid search (GS) and evolutionary algorithm (EA) parameter optimization, EA demonstrates robust performance with significantly lower processing time while maintaining the same level of accuracy as GS. Consequently, when working with the absenteeism dataset, EA offers a friendly and efficient computational cost.

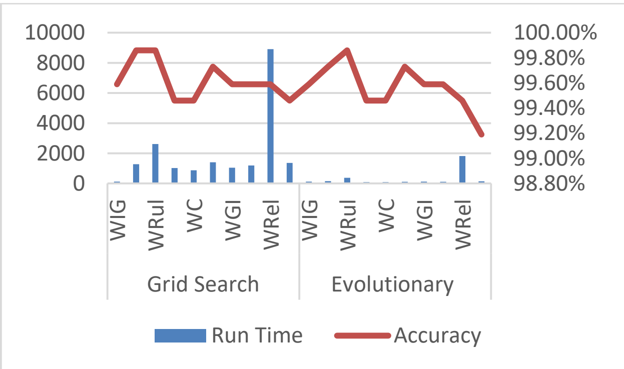


**Figure 5.** The performance of filtering methods and parameter optimization techniques of the absenteeism Dataset

When working with social media datasets, it was observed that the grid search algorithm outperforms the evolutionary algorithm. The grid search algorithm achieves a stable accuracy of around 86.4% across various filtering methods. However, the processing time varies among the filtering methods. WRel and WIGR were found to be the most time-consuming filtering methods.

In contrast, in the evolutionary algorithm, WIGR is the most accurate filtering method with an accuracy of 85.4%. However, this accuracy is lower than the performance achieved by the grid search algorithm. There may be a complementary relationship between grid search and evolutionary algorithm, especially when the optimal parameter range is unknown. The evolutionary algorithm can help determine the potential best range of parameters, which can then be used as input for grid search optimization. The results are depicted in figure 6.

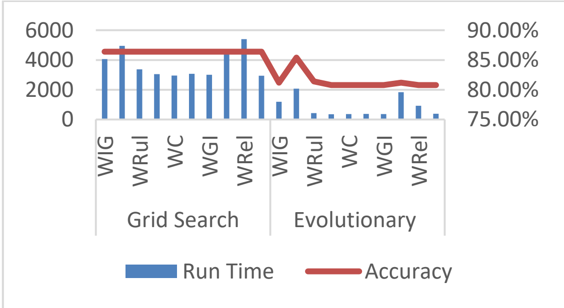


**Figure 6.** The performance of filtering methods and parameter optimization techniques of the social media dataset

When working with the online news dataset, it was observed that the grid search algorithm takes more time in the optimization process. WRel and WRul are two filtering methods that require more processing time compared to the other filtering methods. However, these two time-consuming filtering methods do not yield accurate classification results.

The highest accuracy is achieved by WCSS and WPCA, with a moderate level of processing time. In the evolutionary algorithm, WPCA emerges as a very promising filtering method, achieving an accuracy of 80.82% with relatively low processing time. Moreover, the evolutionary algorithm demonstrates its robustness as a parameter optimization algorithm, exhibiting two efficient characteristics: low processing time and high accuracy compared to the performance of grid search. The results are illustrated in figure 7.

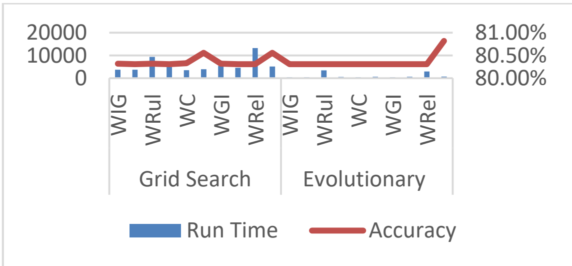


**Figure 7.** The performance of filtering methods and parameter optimization techniques of Online news dataset

Based on these experiments, it is evident that there are no filtering methods and parameter optimization techniques that consistently exhibit excellent performance across various datasets. The performance depends on factors such as data quality, characteristics, and adjustable parameters in the machine learning algorithm. However, certain tendencies can be observed for these filtering methods.

Both WRul and WRel consistently demonstrate a time-consuming behavior across the online news, absenteeism, bank telemarketing, and bitcoin datasets. This behavior indicates high computational costs, particularly in terms of processing time for the classification process. It is important to note that a time-consuming approach does not necessarily guarantee better accuracy. Therefore, these filtering methods should be avoided in machine learning models, especially when dealing with similar dataset characteristics.

In this experiment, it is evident that the grid search parameter optimization algorithm generally achieves higher accuracy compared to the evolutionary algorithm. However, it also requires more time for data processing. Researchers are advised to possess preliminary knowledge about specific machine learning algorithms before using them. This knowledge is used to set the possible best-range parameters to be input into the grid search algorithm.

On the other hand, the evolutionary-based optimization algorithm addresses the issue of inexperienced researchers by providing a large and automated search space for parameter optimization. It eliminates the need for prior knowledge of machine learning algorithms. However, the model accuracy is often slightly lower than that of the grid search algorithm, which is known for its drawbacks. This presents an opportunity for a mutually beneficial relationship between the grid search and evolutionary algorithms. The evolutionary algorithm can be utilized when researchers lack prior knowledge of certain parameters for optimizing specific machine learning models. In such cases, the evolutionary-based optimization can determine the best possible parameter range, which can then be used as input for the grid search algorithm.

## 6. Conclusion and Recommendation

Nowadays, achieving more efficient business operations is an important industrial objective. With the rapid changes brought about by globalization, there is a growing demand for the use of machine learning algorithms in business practices. These algorithms have the potential to enhance production and marketing activities, providing increased efficiency at affordable computational costs. In this context, computational cost refers to reasonable processing time and computational capability. By achieving reasonable computational costs and accurate performance from machine learning algorithms, small and medium-sized companies can improve their business operations, leading to enhanced effectiveness, productivity, and profitability.

To address the need for affordable computational costs in business implementation, this study analyzed five different business datasets: banking dataset (Moro et al., 2014), absenteeism dataset (Martiniano et al., 2012), bitcoin dataset, social media dataset, and online news dataset (Fernandes et al., 2015). In line with the objective, ten dimensionality reduction filtering methods were examined: Weight by Information Gain (WIG), Weight by Information Gain Ratio (WIGR), weight by rule (WRul), Weight by Deviation (WD), Weight by Correlation (WC), Weight by Chi-Squared Statistics (WCSS), Weight by Gini Index (WGI), Weight by Uncertainty (WU), Weight by Relief (WRel), and Weight Principal Component Analysis (WPCA). Additionally, two parameter optimization techniques, Grid Search (GS) and Evolutionary Algorithm (EA), were utilized. Support vector machines for classification, data normalization using Z-transformation, and 10-fold cross-validation were other techniques employed in this study.

When dealing with the banking datasets, Weight by Relief (WRel) emerged as the most time-consuming filtering method, regardless of whether it was used with grid search or the evolutionary algorithm. In terms of classification accuracy, WRel combined with the evolutionary algorithm exhibited the most robust performance, achieving 89.83% accuracy. On the other hand, in grid search, WU was identified as the most accurate filtering method with 90.48% accuracy. For the bitcoin dataset, WRul was identified as the slowest algorithm in both grid search and the evolutionary algorithm. However, it is worth noting that a time-consuming filtering method does not necessarily lead to higher accuracy. In grid search optimization, WIG and WGI achieved the highest accuracy, both with 85.33% accuracy.

In the analysis of the absenteeism dataset, WRul emerged as the most accurate filtering method in both grid search and the evolutionary algorithm, achieving 99.86% accuracy. However, WRel consumed more processing time when combined with both grid search and evolutionary parameter optimization.

Moving on to the social media dataset, the accuracy achieved by the grid search algorithm was consistently high across various filtering methods, with an average accuracy of 86.4%. This suggests that the different filtering methods used did not significantly impact the achieved accuracy. Similarly, the experiment conducted on the last dataset reinforced the previous research findings. WRel and WRul emerged as the two most time-consuming filtering methods, while WCSS and WPCA occupied the highest accuracy positions. This indicates that WRel and WRul are not favorable dimensionality reduction techniques due to their time-consuming nature.

In the analysis of parameter optimization performance, it has been established that grid search requires more processing time compared to the evolutionary algorithm. Additionally, prior experience in machine learning algorithms is necessary, particularly when determining specific parameter ranges for optimization. On the other hand, the evolutionary algorithm demonstrates promising performance in parameter optimization, with significantly lower processing time, albeit with slightly lower accuracy. Therefore, it is suggested to combine the use of grid search and the evolutionary algorithm. The evolutionary algorithm can be employed to identify the optimal parameter range, especially when the best range is unknown. This parameter range can then be utilized as input for the grid search algorithm to find the optimal combination of the filtering method and other optimized parameters. The results indicate a possible complementary relationship between grid search and the evolutionary algorithm. This complementary relationship can significantly improve accuracy and reduce processing time when appropriately selecting optimized parameters and influential dimensions.

However, when selecting filtering methods for dimensionality reduction, it is wise to test several techniques since no single technique fits all scenarios. The research results demonstrate that none of the filtering methods consistently outperforms the others. Performance depends on data characteristics and various factors such as the machine learning algorithm used, selected parameters, data quality, and feature selection methods, among others. Furthermore, the field of parameter optimization and dimensionality reduction in business datasets still holds promise for further study. Moreover, exploring the combination of a wrapper method for dimensionality reduction with various parameter optimization techniques is a hot research topic. The aim is to reduce computational costs and make them more affordable for small and medium-sized companies.

## Acknowledgments

The authors fully appreciate all parties for their valuable advice from the beginning to the end of the research process. Special gratitude goes to the Management Study Program of Universitas Pelita Harapan Medan Campus for financing these research activities. Moreover, the authors convey many thanks to the academic editors and reviewers of the **Iranian Journal of Management Studies** for creating a space for collaboration in the Business and Management Analytics field for the international research community.

## References

Boser, B.E., Guyon, I.M. and Vapnik, V.N., 1992, July. A training algorithm for optimal margin classifiers. In *Proceedings of the fifth annual workshop on Computational learning theory* (pp. 144-152). ACM.

Cao, L.J., Chua, K.S., Chong, W.K., Lee, H.P. and Gu, Q.M., 2003. A comparison of PCA, KPCA, and ICA for dimensionality reduction in support vector machine. *Neurocomputing*, *55*(1-2), pp.321-336.

Fernandes, K., Vinagre, P. and Cortez, P., 2015, September. A proactive, intelligent decision support system for predicting the popularity of online news. In *Portuguese Conference on Artificial Intelligence* (pp. 535-546). Springer, Cham.

Feurer, M. and Hutter, F., 2019. Hyperparameter optimization. In *Automated Machine Learning* (pp. 3-33). Springer, Cham.

Garg, A., 2020. Comparing Machine Learning Algorithms and Feature Selection Techniques to Predict Undesired Behavior in Business Processes and Study of Auto ML Frameworks.

Guyon, I. and Elisseeff, A., 2003. An introduction to variable and feature selection. *Journal of machine learning research*, *3*(Mar), pp.1157-1182.

Guyon, I. and Elisseeff, A., 2003. An introduction to variable and feature selection. *Journal of machine learning research*, *3*(Mar), pp.1157-1182.

Igel, C., 2014. No free lunch theorems: Limitations and perspectives of metaheuristics. In *Theory and principled methods for the design of metaheuristics* (pp. 1-23). Springer, Berlin, Heidelberg.

Kira, K. and Rendell, L.A., 1992, July. The feature selection problem: Traditional methods and a new algorithm. In *Aaai* (Vol. 2, pp. 129-134).

Kou, G., Yang, P., Peng, Y., Xiao, F., Chen, Y. and Alsaadi, F.E., 2020. Evaluation of feature selection methods for text classification with small datasets using multiple criteria decision-making methods. *Applied Soft Computing*, *86*, p.105836.

Martiniano, A., Ferreira, R.P., Sassi, R.J. and Affonso, C., 2012, June. Application of a neuro-fuzzy network in the prediction of absenteeism at work. In *7th Iberian Conference on Information Systems and Technologies (CISTI 2012)* (pp. 1-4). IEEE.

Moro, S., Cortez, P. and Rita, P., 2014. A data-driven approach to predicting the success of bank telemarketing. *Decision Support Systems*, *62*, pp.22-31.

Pal SK, Mitra P (2004) Pattern Recognit Algorithms Data Min, 1st edn. Chapman and Hall/CRC, London

Shearer, C., 2000. The CRISP-DM model: the new blueprint for data mining. *Journal of data warehousing*, *5*(4), pp.13-22.

Yu, T. and Zhu, H., 2020. Hyper-parameter optimization: A review of algorithms and applications. *arXiv preprint arXiv:2003.05689*.

Zhu, W., Feng, J. and Lin, Y., 2014, February. Using gini-index for feature selection in text categorization. In *2014 International Conference on Information, Business and Education Technology (ICIBET 2014)*. Atlantis Press