

# Grasshopper KUWAHARA and Gradient Boosting Tree for Optimal Features Classifications

Rabab Hamed M. Aly<sup>1,\*</sup>, Aziza I. Hussein<sup>2</sup> and Kamel H. Rahouma<sup>3</sup>

<sup>1</sup>The Higher Institute for Management and Information Technology, Minya, 61768, Egypt

<sup>2</sup>Department of Electrical and Computer Engineering, Effat University, Jeddah, KSA

<sup>3</sup>Electrical Engineering Department, Faculty of Engineering, Minia University, Minia, 6111, Egypt

\*Corresponding Author: Rabab Hamed M. Aly. Email: rabab.cse2010@gmail.com

Received: 07 December 2021; Accepted: 27 February 2022

**Abstract:** This paper aims to design an optimizer followed by a Kawahara filter for optimal classification and prediction of employees' performance. The algorithm starts by processing data by a modified K-means technique as a hierarchical clustering method to quickly obtain the best features of employees to reach their best performance. The work of this paper consists of two parts. The first part is based on collecting data of employees to calculate and illustrate the performance of each employee. The second part is based on the classification and prediction techniques of the employee performance. This model is designed to help companies in their decisions about the employees' performance. The classification and prediction algorithms use the Gradient Boosting Tree classifier to classify and predict the features. Results of the paper give the percentage of employees which are expected to leave the company after predicting their performance for the coming years. Results also show that the Grasshopper Optimization, followed by "KF" with the Gradient Boosting Tree as classifier and predictor, is characterized by a high accuracy. The proposed algorithm is compared with other known techniques where our results are found to be superior.

**Keywords:** Metaheuristic algorithm; KUWAHARA filter; Grasshopper optimization algorithm; and Gradient boosting tree

## 1 Introduction

Nowadays, many companies are solving problems about their employees' performance by using artificial intelligence for prediction to get practical decisions in the companies. Many companies depend on the prediction of employees' performance which helps the companies to make quick and reasonable decisions. In addition, this drives the company to be successful. Organizations are paying attention to how to reduce the usage of paper in their decisions. It costs numerous resources for them. The first step to reduce this problem is identifying which employee will resign by using prediction techniques [1].



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Optimization techniques play an important role in prediction. The optimization process helps to get the prediction values more accurately and faster than any other methods. Optimization refers to the process of finding optimal solutions to a specific problem. Optimization techniques are applied in prediction methods using Machine Learning (ML) and Deep Learning (DL) [2]. Prediction with optimization is considered a technique of analyzing data.

On the other hand, there are many kinds of datasets that are high dimensional and contain irrelevant features. These datasets have useless information and affect the performance of prediction methods. Many authors introduced a set of methods to solve these problems. Feature selection is one of the methods which solve the problems of high-dimensional datasets [3].

Note that, the accuracy of classification and prediction does not depend on the large selection features. The classification is divided into two groups: a) Binary classification. b) Multi classifications [4]. The classification is very more practical with optimization method. In this paper, we will use optimization for classification based on the feature selection. The main category is using the binary classification based on Grasshopper Optimization as a classifier in the prediction model [5,6]. The work of this paper is divided into some parts. The first part is collecting datasets of the company employees. The second part is clustering, and visualizing data based on hierarchical clustering with principal components analysis. The optimizer is built to select optimal data features. The optimizer type is called "Grasshopper Optimize".

A KUWAHARA Filter (GOKF) follows the optimizer. This new design of optimizer helps to select the optimal features based on KUWAHARA Filter (KF). KF is a non-linear smoothing filter used in image processing for adaptive noise reduction. The fact that any edges are preserved when smoothing makes it especially useful for feature extraction and segmentation. KF is based on placing asymmetric square neighborhood around each pixel of images or data of datasets and dividing it into four square sub-regions. The value of the central pixel or data is replaced by the average of general data over the most homogeneous sub-region. The sub-region refers to the lowest standard deviation values. This filter helps the optimizer to rapidly select the best solution and get the best performance.

Both prediction and classification are based on Gradient Boosting Tree. The results of the proposed technique will be compared with other results based on Gradient Boosting Classifier Tree (GBT) obtained by using Quadratic Discriminant Analysis Function (QDF).

The rest of the paper is organized as follows: Section (2): briefly introduces the literature review. Section (3): shows methodology. Section (4): discusses the empirical results of design. Eventually, conclusions are drawn in Section (5).

## 2 Literature Review

Several theories have been proposed for optimization techniques. Some techniques focus on how to use them in classifications and feature extractions, while others concentrate on predictions. In this section, we will show a number of the previous research which focused on different studies of optimization in different fields. Various authors have been focusing on growing ML for business studies and predicting the performance of the target of work [7].

Authors in [7], presented three main experiments to predict employee attrition. The first experiment focused on Support Vector Machine (SVM) and K-Nearest Neighbors (KNN), and the second experiment showed the usage of Adaptive Synthetic (ADASYN) to overcome the class imbalance. At the same time, the third experiment involved using manual under-sampling to balance the classes. The results were achieved using 12 features based on the random forest as a feature selection method.

Furthermore, certain authors described techniques of ML to classify the best employees in companies. Authors in [8], presented different methods of ML algorithms. ML algorithms are KNN (Neighbor K-Nearest), Naïve Bayes, Decision Tree, Random Forest in addition to two techniques were called stacking and bagging. The results showed that Random Forest was the best method of classification. In addition to that, the Random Forest, stacking, and bagging methods achieved withdrawals of 88%.

In [9], the author described the prediction techniques based on a Hybrid of K-means clustering and naïve Bayes classifier. The method achieved high accuracy in testing employee performance.

In [10], authors presented the prediction of employee attrition based on several ML models. The models were developed automatically and achieved high accurate results in prediction.

Numerous authors introduced the ML algorithms to describe the prediction of employee turnover. In [11], authors explored the application of Extreme Gradient Boosting (XGBoost) technique. That showed significantly higher accuracy for predicting employee turnover.

Moreover, authors in [12], introduced a study of how to design a system of an automatic job satisfaction based on an optimized neural network. This study was consisted of various parts. The initial part was preprocessing which was applied to convert data into numeric data. The second part was data analysis which was introduced by using three factors. Each factor described the details with the analysis of each employee. The third part showed how to determine the correlation between the factors. The authors added the genetic algorithm to enhance the quality of factors and described neural network to predict the employee satisfaction level.

On the other hand, the DL based on optimization is considered one of the more practical prediction techniques. The optimizations have been described in different research and have shown the benefit of several optimization designs such as pipeline applications. In [13], the authors described DL with pipeline optimization for the Korean language framework. The paper showed that the entity extraction and the classification were based on the F1-score. The accuracy and F1-score were 98.2%, 98.4% for intent classification, 97.4% and 94.7% for entity extraction. The authors showed that it is the best accuracy through the experiment of this model.

ML and DL are playing a vital role in the early diagnosis as it is important in treating diseases. There are different methods to diagnose several cases of different diseases. In [14], authors demonstrated methods by making a survey of ML techniques for diagnosing several diseases.

Likewise, in [15] authors described discrete wavelet method to enhance the images of livers disease datasets based on Optimization of Support Vector Machines (OSVM) with Crow Search Algorithm (OSVCSA). OSVCSA is used for accurate diagnosis of livers diseases. The accuracy of classification 99.49%.

LSTM plays a significant role in predicting pandemic diseases. In [16], authors introduced studies of how to predict data of COVID-19. The prediction of data is based on LSTM method and GRU by using python. The paper showed that LSTM achieved higher accuracy than GRU in prediction of COVID-19 data.

In [17], the authors introduced new techniques of ML based on supervised learning and genetic optimization for occupational disease risk prediction. There were three ML methods which were introduced and compared. One of them was based on K-Means and another one was based on Support Vector Machines and K-Nearest Neighbours (KNN). The last approach was based on a genetic algorithm. The results described that the three techniques were clustering-based techniques that allowed a deeper knowledge, and they were helpful for further risk forecasting.

In [18], the authors described a new technique of segmentation for COVID-19 in chest X-rays. They introduced a multi-task pipeline with special streaming of classifications and that helped in growing advances of deep neural network models. That helped them to train separately specific types of infection manifestation. They evaluated the proposed models on widely adopted datasets, and they demonstrated an increase of approximately 2.5%. On the other hand, they achieved a 60% reduction in computational time.

Recently, certain authors involved DL in different complex medical research such as therapeutic antibodies. Authors, in [19], showed that the optimization with DL can be used in the prediction of antigen specificity from antibodies.

As it is known, ML is of a significant benefit in predicting future outcomes. In addition to that, there are numerous Occupational Accidents around the world. Some authors introduced ML to predict the Occupational Accidents such as in [20].

In [20], authors optimized ML to predict outcomes such as injury, near miss, and property damage using occupational accident data. They applied different methods of ML and optimizations such as genetic algorithm (GA) and particle swarm optimization (PSO) to achieve higher degree of accuracy and robustness. They also introduced case study to reveal its own potentiality and validity.

In addition, there are some filters had been used in different application and approved practical results and helps in classification and predictions techniques such as KF [21].

In [21], authors introduced KF as filter with K-means cluster to extract the optimal features from images of tumors to help in segmentation process. The design help to extract tumor and help in classification process to achieve result near to 95%. Based on the review of the literature presented above, in the following section, the new method which is based on optimization with filter for employee performance will be identified, and a new optimization technique will be introduced.

### 3 Methodology

The work of this paper consists of several stages show Fig. 1 as follows:

1. Data preparation.
2. Building Optimization and prediction model.

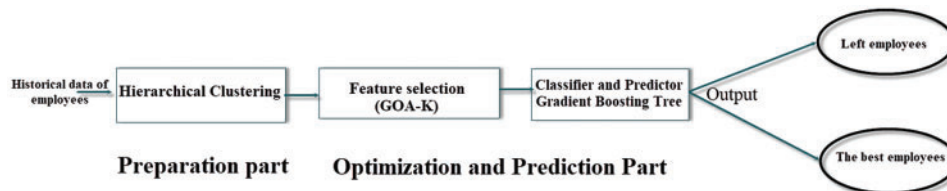


Figure 1: The system block diagram

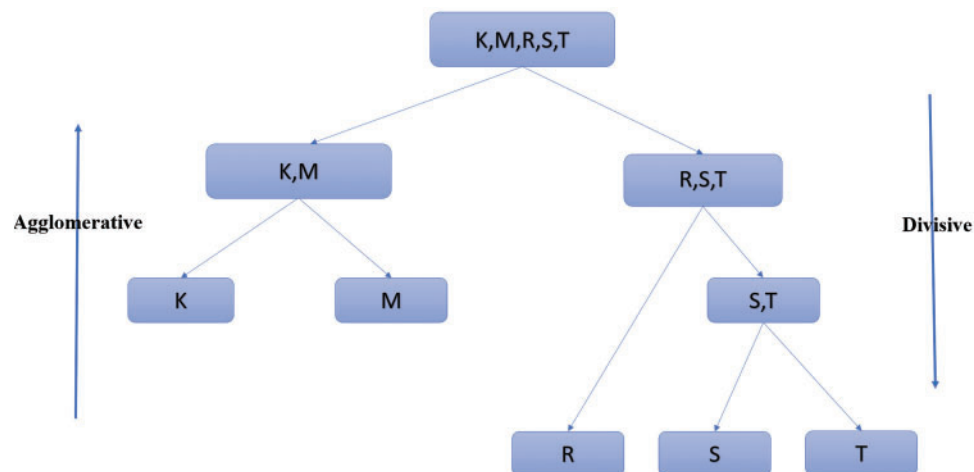
#### 3.1 Data Preparation

The first part of data preparation is based on the clustering analysis. As known, the filtration of data is the most frequent data manipulation operation. The filtration of this part after using the data is based on the library of python called “pandas”. The filtration and analysis with pandas are based on summarizing characteristics of data such as patterns, trends, outliers, and hypothesis testing using descriptive statistics and visualization [22,23].

The clustering analysis of data is based on “hierarchical clustering”. This method is used to seek and build a hierarchy of clusters. Hierarchical clustering is considered an update of the performance of K-means clusters. K-means clusters are based on four stages:

- First, decide the number of clusters ( $k$ )
- Second, select  $k$  as a random point from the data as centroids
- Third, how to assign all the points to the nearest cluster centroid
- The last stage is to calculate the centroid of newly formed clusters, and then repeat the last two steps.

The problem in K-Means clusters is the necessity of predefining the number of clusters because there are certain challenges with K-means which try to make clusters of the same size. The hierarchical clustering presented to improve this problem, so it is more practical, especially in the biggest data. There are two methods into hierarchical clustering as shown below in Fig. 2:



**Figure 2:** General example of agglomerative and divisive hierarchical clustering methods

1. Agglomerative hierarchical clustering
2. Divisive hierarchical clustering.

In this paper, the most similar points or clusters in hierarchical clustering were processed by a series of fusions of the  $n$  objects into groups which called agglomerative.

The mathematical formula of an agglomerative method is as the following: -

-  $P_n, P_{n-1}, \dots, P_1$  are observations of clusters for an agglomerative hierarchical clustering where  $P_n$  contains  $n$  single object clusters and  $P_1$  consists of a single collection involving all  $n$  cases.

At each stage, the most two similar are combined. Note that, for the introductory stage each cluster has an individual object, and the mounts are joined and there are different aspects of defining distance (or similarity) between clusters [23].

- Single linkage agglomerative method: it represents the distance between the closest pair of objects, where only pairs consisting of one object from each group are taken into consideration. The distance  $D(r,s)$  is determined as (1).

- Complete linkage agglomerative method: it shows the distance between the furthest pair of objects, one from each group. The distance  $D(r,s)$  is measured as (2).

$$D(r, s) = \text{Min}\{d(k, m)\} \quad (1)$$

$$D(r, s) = \text{Max}\{d(k, m)\} \quad (2)$$

where  $r, s$  are distance between the two clusters  $(k, m)$ .

- Average linkage agglomerative method: it reflects the mean of distances between all pairs of objects. Each pair includes one object from each group. The distance  $D(r,s)$  is computed as (3).

$$D(r, s) = T_{rs}/(N_r * N_s) \quad (3)$$

where the sum of all pairwise spaces between cluster  $r$  and cluster  $s$  “ $T_{rs}$ ”. In addition to that, the size of clusters are  $(N_r, N_s)$ .

In this paper, this average linkage clustering method was applied hierarchical clustering, and after that the features of principal component analysis were added to reduce the dimensionality and increase interpretability based on the mathematical formula of it which is introduced in [24,25].

### 3.2 Building Optimization and Prediction Model

In this paper, this part will focus on the design of the optimizer and prediction or classifier model. The prediction model will be built based on different parts. One of these parts is visualizing data to see the performance of employees before the prediction. In this part, the visualization is divided into two categories. The first category is based on the number of employees with several projects through a set of years. The second category is creating Label Encoder Object (LEO) with splitting the data of datasets. The last part is building a model of optimization and prediction. The optimizer is based on Grasshopper Optimization (GO) followed by KF to select optimal features which help in the classifier stage.

The optimization part is based on GOKF. The first part is extracting the features of values and select optimal features based on the constructions of GOKF.

The GO is decreasing the dimensionality of data or to select the optimal feature vectors with using KF. The last part is the classifier using GBT which was applied as a predictor for the performance of employees [26].

The datasets in this paper are collected from two datasets of the online employee databases [23,27]. The datasets were collected from the HR department to study the performance of employees and that helps in their decisions about employees after four years of working and that will be shown in the section of results in details. After the collection of data, the clustering method was achieved. Then the features were extracted which were optimized by using GOKF to extract and decrease the dimensionality of data or to select the optimal feature vectors. The cause of using KF with GO is the enhancement of it is more suitable in feature extraction technique [21].

The GO depends on three components (gravity  $Gr$ , social relationship  $Si$ , and horizontal wind movement  $Wi$ ) which affect the flying route of grasshoppers.

The search process is based on the following equation: -

$$S_i = \sum_{j=1, j \neq i}^M s(P_{i,j}) \widehat{P}_{i,j} \quad (4)$$

where  $s$  is the strength of social forces and  $P_{ij}$  is the distance between  $i$ th and  $j$ th grasshopper that is estimated as  $P_{ij} = |x_j - x_i|$ . The unit vector of  $i, j$  is indicated by  $\widehat{P}_{ij}$  as shown in (5).

$$\widehat{P}_{ij} = \frac{x_i - x_j}{P_{ij}} \tag{5}$$

We replaced this part by using KF equations as follows: -

As known, the KF filter will be applied by using the concept of his equation by dividing the regions into four regions. The regions have based on arithmetical mean  $m_i(x, y)$  and standard deviation  $\sigma_i(x, y)$  and the output of the KF filter  $P(x, y)$  for any point  $(x, y)$  as shown in Eq. (6) [28–29].

$$P(x, y) = \left\{ \begin{array}{l} m_1(x, y) \text{ if } \sigma_1(x, y) = \min_i \sigma_i(x, y) \\ m_2(x, y) \text{ if } \sigma_2(x, y) \text{ min}_i \sigma_i(x, y) \\ m_3(x, y) \text{ if } \sigma_3(x, y) \text{ min}_i \sigma_i(x, y) \\ m_4(x, y) \text{ if } \sigma_4(x, y) \text{ min}_i \sigma_i(x, y) \end{array} \right\} \tag{6}$$

The social relation of direction of swarm is  $s$  [26–28]. The equation of  $s$  can be described as follows:

$$S(r) = e^{-\frac{r}{L}} - e^{-r} \tag{7}$$

where  $b$  is the attractive force,  $r$  is the distance between grasshoppers and  $L$  is attractive length. Fig. 3 shows the primitive corrective patterns of GO. On the other hand, the mathematical expression of grasshopper interaction can be presented by (8).

$$X_i^k = c \left( \sum_{j=1, j \neq i}^M c \frac{up_k - Ip_k}{s} s(x_j^d - x_i^d) \frac{x_i - x_j}{P_{ij}} \right) + \widehat{T}_k \tag{8}$$

Notably,  $up_k, Ip_k$  are lower and upper bound and  $c$  is a coefficient which is used for reducing the comfort, repulsion, and attraction regions and it is considered the target to get the best solutions. In addition to that,  $k$  is a dimension which indicates  $\widehat{T}_k$ .

The equation of parameter  $c$  can be described as follows: -

$$C = c_{max} - l \frac{c_{max} - c_{min}}{N} \tag{9}$$

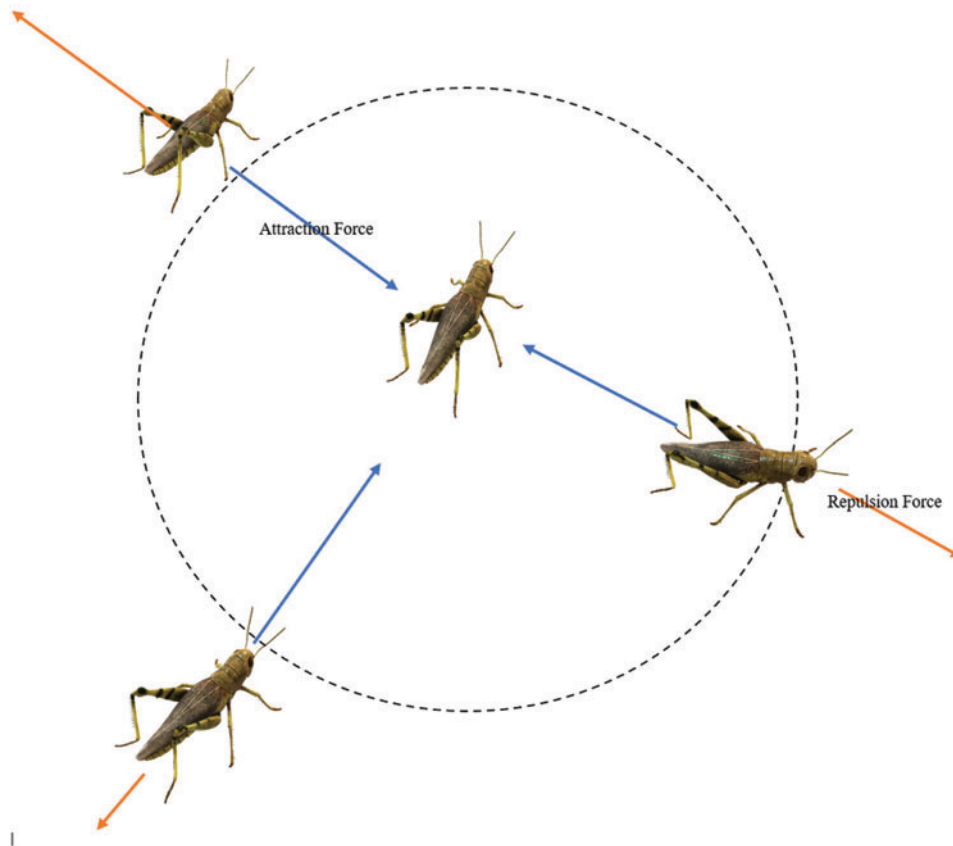
where  $N$  is the maximum iterations.

Then, the GBT will applied to classify the features which extracted from optimal solution of optimization technique. As known, GBT involves subsampling the training dataset and training individual learners on random samples created by subsampling. The GBT design in some steps: -

- The first step in the GBT was to initialize the model with some constant value. The building of the model is used to predict the observations in the training features. For simplicity we take an average of the target column and assume that to be the predicted value.
- The difference in classification is in the calculation of average of the target column by using the log of values to get the constant value after initializing the model with some constant values based on Eq. (10)

$$L = - \sum_{i=1}^n y_i \log(p) + (1 - p) \log(1 - p) \tag{10}$$

where  $L$  is loss function,  $p$  is probability of prediction and  $y_i$  is the observed value. The python programming library “SKLEARN” is applied to achieve the results of the GBT and GO with K filter algorithms [30].



**Figure 3:** The primitive corrective patterns of Grasshopper optimization

### 3.3 The General Pseudo-Code of Design of GOKF

#### Algorithm:

- 1: Generate the initial population of Grasshopper  $P_i (i = 1, 2, \dots, n)$  based on the KF with a few steps:
  - Build sub-windows for data input as the same work for data from images.
  - Calculation of averages and variances on sub-windows.
  - Choice of the index with minimum variance.
  - Build the filtered features by using nested loop
  - Extract  $P(x, y)$  for data input.
- 2: initialize  $C_{max}$ ,  $C_{min}$  and maximum number of iteration  $N$
- 3: Evaluate the fitness  $f(P_i)$  for each Grasshopper  $P_i$  based on  $P(x, y)$  data points.
- 4:  $T$  is the best solution
- 5: While ( $L > N$ ) do
- 6: Update  $C_1$ ,  $C_2$  using Eq. (8).
- 7: For  $i = 1$  to  $M$  (all  $M$  grasshopper in the population using Eq. (7))

(Continued)



---

```

do
    • Normalize distance between the grasshoppers based on Eqs. (3), (5).
    • Update the position of the current grasshopper based on Eq. (7).
    • Bring the current grasshopper back (outside boundaries).
End for.
8: Update T if there is the best solution
9: L = L + 1
10: End While
Return the best solution (The best solution is the features selection to classifier =  $y_i$  for GBT Eq. (10))

```

---

#### 4 Results and Discussion

This paper is applying prediction of employee performance under several stages: -

- Collecting the data of employees: data were collected from sample historical data for departments in the organization throughout the last five years of this sample. After that, the data is visualized by using python library to show the performance of employees based on the data collection as shown in Figs. 4, 5 and 6. Each figure shows the details of the structure of the collecting data. The Fig. 4 shows that the total number of employees which left the company through the last five years and Fig. 5 shows the total number of years spends in company. Contrast with the pervious figures, Fig. 6 shows the number of employees with the total number of projects which achieved the targets in point of time through the last five years. It will be noticed that the total numbers of employees are 6000 which includes both the current and left employees.
- The next step is extracting the features from the dataset based on the clustering operation. The clustering analysis of data is based on “hierarchical clustering”.
- The classification of data introduced in some steps; First, the dismissal of employees depends on a critical factor with total number of projects through the last five years. If an employee worked from 4-6 projects through the last years, he/she is less expected to leave the company. Second, the time of work through the company is important factor to take decisions about the performance of employees. The decisions are based on the total number of hours which an employee spent in company. Notably, there is a huge drop between 3 and 4 years experienced employees. On the other hand, the percentages of employees left are 25% of the total of employees. Most of the employees are receiving salary either medium or low. The tester of Information Technology (IT) department is having the maximum number of employees followed by customer support and developer.
- Building the prediction model: this part is based on GBT:
- First is extracting features from python after that, the data were saved in CSV file.
- Second part is building GOKF to extract optimal features for classification.
- Third part gives optimal features based on GOKF and applies the prediction function based on the GBT using python function. The accuracy of the classification got 96.7% based on Eqs. (11)–(13). It is considered higher and authentic accuracy. The classification report is shown in Tab. 1.

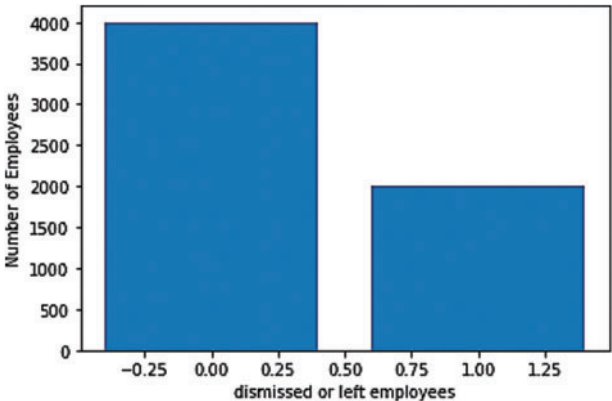


Figure 4: The total of dismissed or left employees

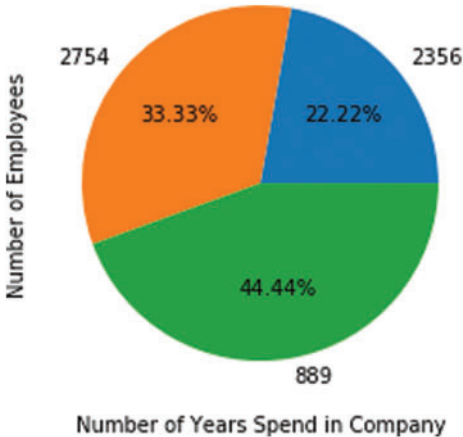


Figure 5: The total number of years spends in company

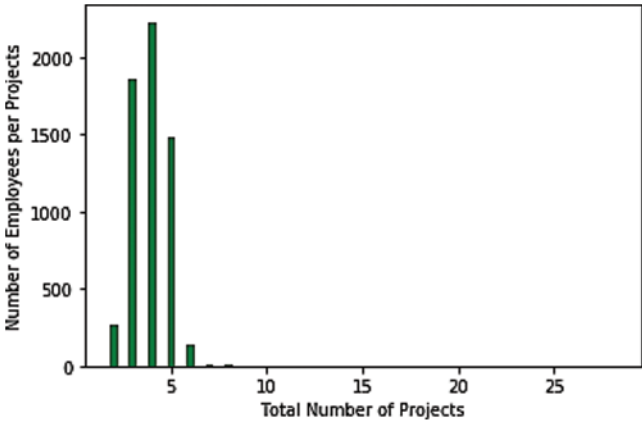


Figure 6: The Total number of employees per projects

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN} \quad (11)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (12)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (13)$$

where TP is True Positive, TN is True Negative; FP is False Positive and FN False Negative.

**Table 1:** The report of GOKF system based on confusion matrix

	Precision	Recall	F1-score	Support
0	0.985	0.96	0.967	1168
1	0.93	0.98	0.95	632
Accuracy			0.967	1800
Macro-average	0.96	0.967	0.96	1800
Weighted-average	0.967	0.967	0.967	1800

The GO is based on hierarchical clustering, and CNN achieved a higher degree of accuracy in prediction as the other introduced methods in [25] which give more practical optimal solutions as shown in “Tab. 2”. In Tab. 2 the QDF refers to Quadratic Discriminant Analysis Function with K-means clusters [28] and GBT is Gradient Boosting Tree with K-means clusters [29].

In [29], authors introduced two unsupervised pattern recognition algorithms based on K-means clusters and called QDF and Gaussian Mixture Model (GMM). The accuracy was achieved 96% based on QDF and the same method was applied on the data of this paper and approved the same results but the method of GO-KF is faster and more practical for achieving the accuracy in more accurate time than the other method. Furthermore, in [30], authors applied the GBT for diabetes mellitus diagnosis system and achieved accuracy near to 97%. When the same method was applied for our datasets achieved the same performance when it was compared with the result of this paper. Tab. 2 shows the compare between pervious work methods and the method of this paper.

**Table 2:** Compare between the method of this paper with the other method from previous work

Method	Accuracy	Precision	Recall
QDF	96%	93%	98%
GBT	97%	95%	92%
GOKF-GBT (work of this paper)	96.7%	93%	98%

## 5 Conclusion

This paper introduced a technique for optimal classification and prediction of employees' performance. This technique is composed of a grasshopper optimizer followed by a Kawahara filter (KF). The employees' data are collected and then processed using a modified hierarchical K-means clustering method. The filter is used to obtain the best features of employees which match their best performance. This is done in two axes. Firstly, data of employees has been collected. From this data, the

performance of each employee is calculated and illustrated. Secondly, the classification techniques are applied to classify the employee performance and prediction techniques are carried out to predict this performance in the future. This is done by obtaining the employees features. The Gradient Boosting Tree classifier is utilized for the purposes of features classification and prediction. The model has been applied. The percentage of employees, who are expected to leave the company after predicting their performance for the coming years, is calculated. The results were found highly accurate. A discussion of the results and a comparison with the previous research methods are explained. The proposed algorithm is found to be superior.

**Acknowledgement:** The author would like to thank the editors and reviewers for their review and recommendations.

**Funding Statement:** The author received no specific funding for this study.

**Conflicts of Interest:** The authors declare that they have no conflicts of interest to report regarding the present study.

## References

- [1] M. Grządzielewska, "Using ML in Burnout Prediction: A survey," *Child and Adolescent Social Work Journal*, vol. 38, no. 2, pp. 175–180, 2021.
- [2] S. Serengil and A. Ozpinar, "Workforce optimization for bank operation centers: A machine learning approach," *Ijimai*, vol. 4, no. 6, pp. 81–87, 2017.
- [3] H. Chang, "Employee turnover: A novel prediction solution with effective feature selection," in *WSEAS Int. Conf. Proc. Mathematics and Computers in Science and Engineering*, vol. 6, no. 3, pp. 417–426, 2009.
- [4] J. Choi, I. Ko, J. Kim and S. Han, "MLFramework for multi-level classification of company revenue," *IEEE Access*, vol. 9, no. 358, pp. 96739–96750, 2021.
- [5] A. Saxena and R. Kumar, "Chaotic variants of grasshopper optimization algorithm and their application to protein structure prediction," *Springer Tracts in Nature-Inspired Computing. Springer*, vol. 5, no. 7, pp. 151–175, 2020.
- [6] H. Tao, N. Kadhim and K. Khedher, "River water level prediction in coastal catchment using hybridized relevance vector machine model with improved grasshopper optimization," *Journal of Hydrology*, vol. 598, no. 126488, pp. 1–13, 2021.
- [7] Th. Naidu, "Detection and identification of an employee attrition using MLAlgorithm," *Nveo-Natural Volatile and Essential Oils Journal*, vol. 8, no. 4, pp. 5563–5573, 2021.
- [8] P. Taqwa, I. Pratama and A. Yakobus Chandra, "Implementation of ML to determine the best employees using random forest method," *IJCONSIST Journal*, vol. 2, no. 2, pp. 53–59, 2021.
- [9] N. Sciences, "Hybrid of K-means clustering and naive bayes classifier for predicting performance of an employee," *Periodicals of Engineering and Natural Science*, vol. 9, no. 2, pp. 799–807, 2021.
- [10] A. Qutub, A. Al-mehmadi, M. Al-hssan, R. Aljohani and H. S. Alghamdi, "Prediction of employee attrition using ML and ensemble methods," *International Journal of ML and Computing*, vol. 11, no. 2, pp. 110–114, 2021.
- [11] R. Punnoose, "Prediction of employee turnover in organizations using MLAlgorithms," *Algorithms*, vol. 5, no. 9, pp. 22–26, 2016.
- [12] Y. Sun and j. Zhang, "Optimized neural network using beetle antennae search for predicting the unconfined compressive strength of jet grouting coalcretes," *International Journal for Numerical and Analytical Methods in Geomechanics*, vol. 43, no. 4, pp. 801–813, 2019.
- [13] M. Hwang and H. Cho, "KoRASA: Pipeline optimization for open-source korean natural language understanding framework based on deep learning," *Mobile Information Systems*, vol. 2021, no. 9987462, pp. 1–9, 2021.

- [14] G. Kumar and R. Kalra, "A survey on MLTechniques in health care industry," *International Journal of Recent Research Aspects*, vol. 3, no. 2, pp. 128–132, 2016.
- [15] D. Devikanniga, A. Ramu and A. Haldorai, "Efficient diagnosis of liver disease using support vector machine optimized with crows search algorithm," *Endorsed Transactions on Energy Web*, vol. 7, no. 29, pp. 1–10, 2020.
- [16] W. Hastomo and N. Kalbuana, "Characteristic parameters of Epoch DL to predict Covid-19 data in Indonesia," *Journal of Physics: Conference Series*, vol. 1933, no. 1, pp. 1–6, 2021.
- [17] A. Noia and A. Rizzi, "Supervised MLTechniques and genetic optimization for occupational diseases risk prediction," *Soft Compute*, vol. 24, no. 6, pp. 4393–4406, 2020.
- [18] S. El-bana and M. Sharkas, "A multi-task pipeline with specialized streams for classification and segmentation of infection manifestations in COVID-19 scans," *PeerJ Computer Science*, vol. 6, no. 303, pp. 1–27, 2020.
- [19] D. Mason, S. Friedensohn, C. Weber, C. Jordi, B. Wagner *et al.*, "Optimization of therapeutic antibodies by predicting antigen specificity from antibody sequence via deep learning," *Nature Biomedical Engineering journal*, vol. 5, no. 6, pp. 600–612, 2021.
- [20] K. Govindan and T. Cheng, "Advances in stochastic programming and robust optimization for supply chain planning," *Computers and Operations Research*, vol. 100, no. 21, pp. 262–269, 2018.
- [21] R. Hamed, K. Rahouma and H. Fathy, "Brain tumors diagnosis and prediction based on applying the learning metaheuristic optimization techniques of particle swarm, ant colony and bee colony," *Procedia Computer Science*, vol. 163, no. 20, pp. 165–179, 2019.
- [22] B. Fernández and S. Luján-Mora, "Application of ML in predicting performance for computer engineering students: A case study," *Sustainability*, vol. 11, no. 10, pp. 1–18, 2019.
- [23] A. Ayoub, "Integrating grasshopper optimization algorithm with local search for solving data clustering problems," *International Journal of Computational Intelligence Systems*, vol. 14, no. 1, pp. 783–793, 2021.
- [24] D. da Silva, V. Klausner and A. Prestes, "Principal components analysis: An alternative way for removing natural growth trends," *Pure and Applied Geophysics*, vol. 178, no. 12, pp. 1–19, 2021.
- [25] S. Veerashetty and N. B. Patil, "Texture-based face recognition using grasshopper optimization algorithm and deep convolutional neural network," in *Int. Conf. on Communication, Computing and Electronics Systems: Proc. of ICCCES 2020*, Springer Nature, vol. 733, no. 43, pp. 43–59, 2021.
- [26] W. Tao and Y. Jiang, "Predicting shared-car use and examining nonlinear effects using gradient boosting regression trees," *International Journal of Sustainable Transportation*, vol. 15, no. 12, pp. 893–907, 2021.
- [27] Y. Meraihi and S. Mirjalili, "Grasshopper optimization algorithm: Theory, variants, and applications," *IEEE Access*, vol. 9, pp. 50001–50024, 2021.
- [28] E. Raptou and E. Manolas, "Consumption patterns and public attitudes toward organic foods: The role of climate change awareness," *Climate Change in the Mediterranean and Middle Eastern Region*, vol. 2, no. 19, pp. 365–393, 2022.
- [29] S. Dutta and S. Kumar, "A novel method for finding grasping handles in a clutter using RGBD Gaussian mixture models," *Robotica Journal*, vol. 2019, pp. 1–17, 2019.
- [30] R. Muthulakshmi, "Modeling the map reduce based optimal gradient boosted tree classification algorithm for diabetes mellitus diagnosis system," *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, no. 2, pp. 1717–1730, 2021.