Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.DOI

Prediction of the Employee's Work Preferences in Remote Work: A Metheuristic Approach for Parameter Tuning of the Binary Classification Algorithm

SALMA ALMOHTASIB[1], RANEEM QADDOURA[1], ISLAM ALOMARI[1], ABDELAZIM G. HUSSIEN[2,3,4,5] and Shuang Wang [6,7]

[1]School of Computing and Informatics, Al Hussein Technical University, Amman, 11831, Jordan (e-mail: Salma.Almohtaseb@htu.edu.jo, Raneem.Qaddoura@htu.edu.jo, Islam.Alomari@htu.edu.jo)
[2]Department of Computer and Information Science, Linköping University, Linköping, Sweden. (e-mail: abdelazim.hussien@liu.se)
[3]Faculty of Science, Fayoum University, Egypt.
[4]Applied Science Research Center, Applied Science Private University, Amman 11931, Jordan
[5]MEU Research Unit, Middle East University, Amman 11831, Jordan
[6]New Engineering Industry College, Putian University, Putian 351100, P.R. China. (e-mail: shuang.wang@fjsmu.edu.cn)
[7]School of Information Engineering, Sanming University, Sanming 365004, China

Corresponding author: Shuang Wang (e-mail: shuang.wang@fjsmu.edu.cn).

ABSTRACT For almost three years, the world has been introduced to a vicious virus that forced people to stay at home, introducing the world to remote work. People were forced to work remotely even though some of them do not prefer this option. This research paper aims to predict employees' work preferences for working from the office or remotely using optimization and parameter tuning to predict employees' work preferences based on specific factors. This study helps decision-makers in optimizing resource allocation, labor management, and scheduling which can all benefit from predictive modeling. It can also assist in finding opportunities for professional growth, personalizing rewards, and establishing customized work environments for employees. Businesses can enhance decision-making processes through the use of machine learning techniques, which will boost output, effectiveness, and worker satisfaction. Here, five algorithms were used throughout this study: Logistic Regression, Decision Tree, Random Forest, K-nearest neighbours, and Naive Bayes. Using the following evaluation metrics: Accuracy, Precision, Recall, and F1-score, the random forest showed the best performance compared to the other models. Metaheuristic Algorithms were used for parameter tuning to enhance the model. Using the following evaluation metrics: Accuracy, Precision, Recall, and F1-score, Random Forest outperformed the Decision Tree, but Random Forest outperformed the Decision Tree. Moreover, Random Forest's performance was much enhanced by the SSA method in all the measures. PSO improved performance as well, however, the Recall was marginally reduced. On the other hand, GWO performed less well than PSO and WOA, even if its recall was higher. This study not only supports organizational decision-making processes but also emphasizes how important it is to use predictive modeling to adjust to changing work environments, especially in light of the public's desire for office or remote work arrangements.

INDEX TERMS Classification; Prediction; Random Forest; Metaheuristic Algorithms; Remote Work; supervised learning

I. INTRODUCTION

Beginning in 2020, a pandemic that affected the whole planet forced people to adopt unconventional lifestyles, including remote work, to ensure everyone's safety. Remote working is doing your work and tasks outside your normal workspace. Employees, for example, do their jobs elsewhere besides their offices, but with the Covid-19 crisis, elsewhere was limited to a person's house. Some businesses and employees continued to work from home once the lockdown was lifted, while others chose blended work (working from home and the office). This idea caused some concerns in the workplace since it indicated that workers would not be as productive as they would be in an office setting. Many factors explain why an employee can become less productive, including the worker's gender, the number of family members who also live at home, the number of children, whether they have pets, and so on. Some studies have shown that certain factors that are present at home can affect their productivity [1]. In addition, the studies showed that such distraction could cause long working hours, leading to fewer production levels [2]. On the contrary, other research showed that a comfortable environment at home increases their productivity [3]. It was also mentioned that employees could use these hours for work instead of wasting time preparing to go to work or waiting for a bus.

Employees noticed both positive and negative effects of working remotely during the pandemic. Some workers felt lonely and isolated, which increased their stress levels and made them want to return to the office [4]. other workers said working from home was disruptive and made it more difficult to maintain a healthy work life. Contrarily, as previously noticed, employees reported favoring the work from home approach since they felt it reduced expenses and improved their health and well-being [5]. In a sense, employees' preferences for work after the pandemic are heavily influenced by their experiences while being forced to work from home during the lockdown [5].

The change in the working environment alerted the employers as they wanted insight into their employees' work preferences, studying this information, and deriving decisions for the company's growth. For such cases, machine learning models have usually been considered the go-to method to support human-decision making using algorithms that provide predictions or suggestions that are subsequently taken into account by a human decision-maker [6].

This paper discusses the factors that affected the teleworkers' work preferences and explores the reasons behind their choice using the metaheuristic algorithms for parameter tuning and to improve and enhance the model's functionality with the best performance. The experiments were divided into two parts:

1) A comparison between different classification algorithms to select the algorithm with the highest performance by evaluating the model in terms of Accuracy, Recall, Precision, and F1-score.
2) Apply metaheuristic algorithms for parameter tuning for the best algorithm selected from the previous part. This part also includes comparing the results of the different metaheuristic algorithms while providing recommendations for enhancing this application.

This paper's primary contribution is its recommendation of the metaheuristic and machine learning techniques for predicting the work preferences of employees on the chosen dataset. This is crucial when making decisions to manage labor and allocate resources as efficiently as possible. It can also help in creating surroundings that are specifically tailored for workers, increasing productivity, efficacy, and job satisfaction.

The remainder of the paper is organized as follows: Section II presents the Related Work to this study. Section III is the Background section containing root information about the concepts used in this research. Section IV presents the Methodology having a detailed overview of the experiments conducted for the research. Section V discusses the results obtained from conducting the experiments. Finally, the last section is the conclusion providing a summary of the paper and the final results.

II. RELATED WORK

Nowadays, employees may choose between working remotely and in the office. After the pandemic, according to some studies, more than 80% of employees prefer working at least half of their hours from home since they believe it to be just as productive and produces more work per hour [2]. As showed by Paula M. Caligiuri [5], after the pandemic, it was discovered that giving the option to work from home is one of the ways to create an FWA (Flexible work arrangement) where workers may pick where to work, with a high likelihood that it will last as long as they choose to accept this arrangement. Kumara et al. [7] also agreed that firms and employees might embrace a hybrid working style following the pandemic.

The findings show that workers increasingly favor working remotely for the following reasons: working from home encourages a healthy lifestyle as well as flexible schedules, lower travel expenses, and more, as demonstrated by Kumara et al. [7] in their work. Policies enabling workers to work from home can also boost productivity and save organizational costs. Working from home has several advantages for individuals and the company, according to Lingfeng BAO [3], including greater job satisfaction levels. Additionally, it was shown that remote employees saved a significant amount of money when working from home, improving their ability to manage work and life.

In contrast to the earlier research, Jayasutha and Arunachalam's study [4] revealed that although working remotely helped employees to become more independent and in charge of their work environment, doing so is not as beneficial as it may seem. The abrupt shift in the work environment, caused by a reduction in physical activity and dietary consumption, negatively influences both physical and mental health, according to the study. As a result, there was an increase in tension among the workers and their discomfort and agony. The employee's desire for employment after the pandemic will be impacted by their dissatisfaction with distant work and the epidemic. To some extent, Mehdi's research [2] has shown that a minority of workers opposed working remotely following the epidemic as their output dropped and their workload increased. The reasons for this include obstacles at home, taking care of kids or other family members, a poor internet connection, and an inappropriate work atmosphere. Moreover, it was shown that employment significantly impacted preferred work environments; for instance, around fifty-four percent of teachers prefer working outside. Although women tend to be less likely to prefer working from home during the pandemic, Paula M. Caligiuri's [5] research found that having other family members present while working remotely did not impact an employee's choice for that type of employment.

Several other papers studied remote work applications using machine learning algorithms. Kumara et al. [7] employed machine learning techniques such as Naive Bayes, Artificial Neural Networks, Random Forests, and Ensemble learning to predict the employee's preferred line of work. When tested with an Accuracy of 95.1%, ensemble learning had the greatest performance level. However, the model might be improved by recalculating the mean absolute error (MAE) and mean square error (MSE). Another researcher also used logistic regression for prediction; the model's Accuracy was close to 0.88, but no changes were made to raise the model's effectiveness [8]. Other studies used metaheuristic optimization for machine learning algorithms. This paper [9] followed a similar approach to test their hypothesis by using multiple machine learning models, including the Logistic Regression, Random Forest, and Decision Tree to predict the estimated slope stability for different datasets, but applied the FireFly Algorithm (FFA) for hyper-parameter tuning. The results were tested in terms of Accuracy, true positive rate, and true negative rate to influence the performance of the Logistic algorithm and Support Vector Machine. As for Junbo Qiu's paper [10], they used five metaheuristic algorithms, Particle Swarm Optimization (PSO), Grey Wolf Optimization (GWO), whale optimization algorithm (WOA), Butterfly Optimization Algorithm (BOA), and Sparrow Search Algorithm (SSA), to enhance the Extreme Learning Machines (ELM) predictability by optimizing the bias and weights of the model.

Metaheuristic algorithms have gained significant recognition in several tasks across various fields as well due to their superior performance compared to other algorithms [11]–[14]. In the engineering domain, for instance, the researchers in this paper [15] proposed a PSO-SVM model that outperformed ANN, DNN, and ANFI algorithms in predicting damaged structures. Through rigorous testing involving 10 randomly damaged cases, the PSO-SVM model accurately predicted 8 damage levels and 10 locations, surpassing the limitations of other machine learning models that could only predict a maximum of 5 damage locations. These findings highlight the effectiveness and potential of metaheuristic algorithms in enhancing predictive capabilities in engineering applications. Another paper [16] proposed a new approach to predict the ground vibrations caused by blasting uses various criteria as input variables and applies gray wolf optimization (GWO), whale optimization algorithm (WOA), and Bayesian optimization algorithm to fine-tune the hyper-parameters of the extreme gradient boosting model, WOA-XGBoost model found the most effective compared to other machine learning models with the lowest RMSE (Root Mean Square Error) and MAE (Mean Absolute Error) values which confirmed the ability of the metaheuristic algorithm to enhance the performance of the model. The study of [17] used GWO to optimize thirteen hyperparameters and the model structures of Neural Basis Expansion Analysis for Time-Series (N-BEATS). This paper [18] proposed a hybrid GWO–PSO model with Random Forest algorithm to detect the intrusions on IoT network environment. It was applied on different datasets and reaches the highest performance compared with the decision tree, logistic regression, and Naive Bayesian algorithms with accuracy closed to 99.6% . Metaheuristics are widely known to be used in feature selection tasks as well; Sayed et al. [19] worked on three phases: the first phase is data pre-processing where the authors implemented SMOTE algorithm for oversampling. Next, Chaotic Dragonfly Algorithm (CDA) was implemented in feature selection phase to build and evaluate a support vector machine (SVM) classifier.Moreover, Gharehchopogh et al. [20] applied different chaotic maps on 24 UCI datasets to improve Vortex Search Algorithm (VSA) for feature selection. Another work [21] used the Gaussian Chaos Map-based initialization and neighbor search strategy with GWO for feature selection. They found that the proposed method obtained average accuracy improvements on Essays and Kaggle MBTI datasets having good convergence and scalability.

Most studies on remote work have solely looked at its analytical component. As far as we know, using parameter tuning with metaheuristic algorithms to predict the employees' work preferences is not found in other studies. This research used a predictive machine learning approach with metaheuristic algorithms and parameter tuning to study the employee's work preference to improve the model's predictions and provide new insights into preferences.

III. BACKGROUND

This section contains a general overview of the concepts covered in this work: the machine learning models used in this study and the metaheuristic algorithms.

A. MACHINE LEARNING MODELS

Logistic Regression, Decision Tree, Random forest, K-nearest neighbor, and Naive Bayes are well-known machine learning algorithms. The logistic regression models calculate the likelihood of a binary result. The logistic regression approach focuses on whether an event occurred rather than how frequently it occurred because it is focused on the statistics of choice in a situation [22]. The Naive Bayes Classifier is a type of probabilistic classification method that has its roots in Thomas Bayes' post-mortal Bayesian Theorem. The fundamental objective is to identify the optimal mapping between a new piece of data and a group of categories inside a certain problem domain [23]. In contrast, Decision Trees and Random Forests work in different ways. Decision Trees continuously divide input data into smaller groups by the class label; in the Decision tree, any path starting at the root is defined by a sequence of data elements up to a leaf node with a Boolean outcome (true or false) [24]. On the other hand, a random forest classifier uses bagging on samples, and a majority voting scheme for multiple decision trees. Random forest preserves many advantages of decision trees while producing better outcomes [25].

B. METAHEURISTIC ALGORITHMS

"Metaheuristic" refers to higher-level heuristics that have been put out as solutions to various optimization issues. Many metaheuristic algorithms have been successfully used recently to solve difficult problems. Because they produce the best/optimal answers even for very large problem sizes quickly, these algorithms are useful for addressing complicated problems. A wide range of optimization issues drawn the interest of Metaheuristic techniques starting with the complicated nature of these problems that are not very simple. Even though exact algorithms can offer the best results, they tend to have impractical computational times for huge data sizes. These problems can be substituted with the Metaheuristic algorithms as they provide practical and elegant solutions [26]. They are made to solve NP-Hard (non-deterministic polynomial-time hardness) optimization problem, which is classified based on the ability to find an optimal solution [27], in reasonable amounts of time [26].

Metaheuristic algorithms can be classified into two main categories:

1) Single solution-based Metaheuristic algorithms: beginning their optimization process with a single solution, and their solution is modified during the iteration.
2) Population (multiple) solution-based metaheuristic algorithm: these algorithms start their optimization process by generating a population of solutions. With each generation or iteration, the population of solutions changes.

Based on the algorithmic behavior, the metaheuristic algorithms are divided into four categories: evolution-based, swarm intelligence-based, physics-based, and human-related algorithms [28]. The following algorithms are the most popular metaheuristic algorithms [29], [30]:

1) PSO: Particle Swarm Optimization mimics how a flock of birds would migrate and separate themselves from one another as they searched for an ideal location in a multidimensional world [31].
2) SSA: Salp Swarm Algorithm replicates the swarming behavior of salps during seawater foraging. Salps typically form a swarm known as a salp chain in dense waters. The salp at the front of the chain is the leader in the SSA algorithm, while the remaining salps are referred to as followers [32].
3) GWO: The Grey Wolf Optimization is modeled after how grey wolves hunt.
4) WOA: By modeling the cooperative predation techniques of humpback whales, the whale optimization algorithm is an innovative mathematical model for addressing optimization issues [33].

IV. METHODOLOGY

The job market started using remote working during, post, and post-pandemic; this study shows the employee's work preferences by studying the relationship between the factors and their intended choice using metaheuristic algorithms for parameter tuning to enhance the model's performance to get better results.

To conduct this experiment, five machine learning models were constructed to be evaluated to see which model had the best performance. Using the evaluation measures, the next step is to merge the Metaheuristic Algorithms into this study. To put it more into a general overview, this study first include some pre-processing, changing the data into numerical data, and the unnecessary columns were dropped.

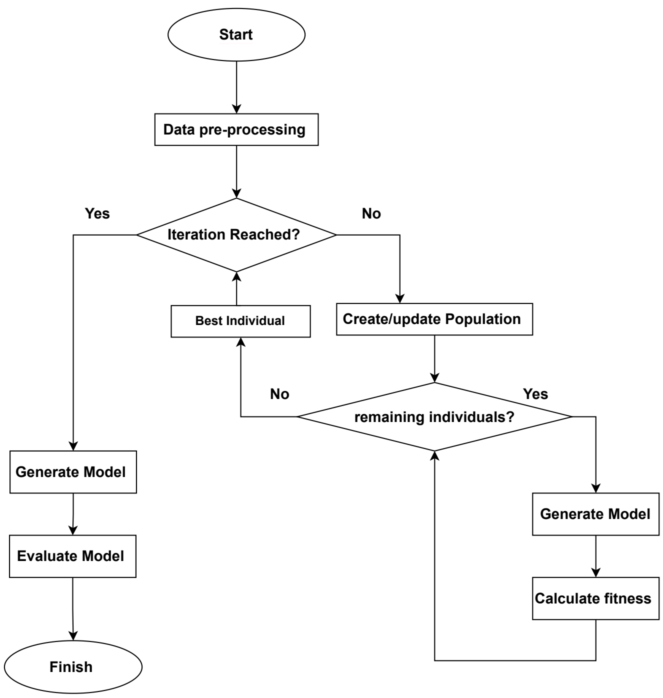

FIGURE 1: Flow Chart

Referring to the figure 1, the following steps are considered:

1) Pre-processing the data for the next steps. Data Preprocessing is responsible for encoding and transforming data for the machine to parse the data [34].
2) Looping through the iterations until the predefined number of iterations is reached.
3) For each iteration, the model creates a new population containing the individuals. An individual is a set of variables representing a proposed solution to the problem. In addition, the set of all individuals is called a population [35].

4) For each individual, the model is generated and the fitness value is calculated for the individual to check how good it is with respect to the problem's consideration. A fitness function is the objective function that shows how close a given individual is to achieving the best solution [35].
5) Once the number of iterations is reached, and according to the calculated fitness values for all the individuals for the iterations, the best individual with the best fitness value is recorded.
6) The model is regenerated for the best individual and is evaluated in terms of Accuracy, Precision, Recall, and f1-score measures. The Accuracy, Precision, Recall, and f1-score measures are discussed in the "Evaluation Measures" section.

The present study elucidates the methodology employed to conduct the research. Initially, an individual was constructed with four parameters, and subsequently, the genes in the individual were normalized to a range of 0 to 1 to conform to the requisite parameters. This is shown in Figure 2. The purpose of constructing the individual for the metaheuristic algorithm is to tune the hyper-parameters of the random forest, including the n_est, min_samples_leaf, max_depth, max_features hyper-parameters.

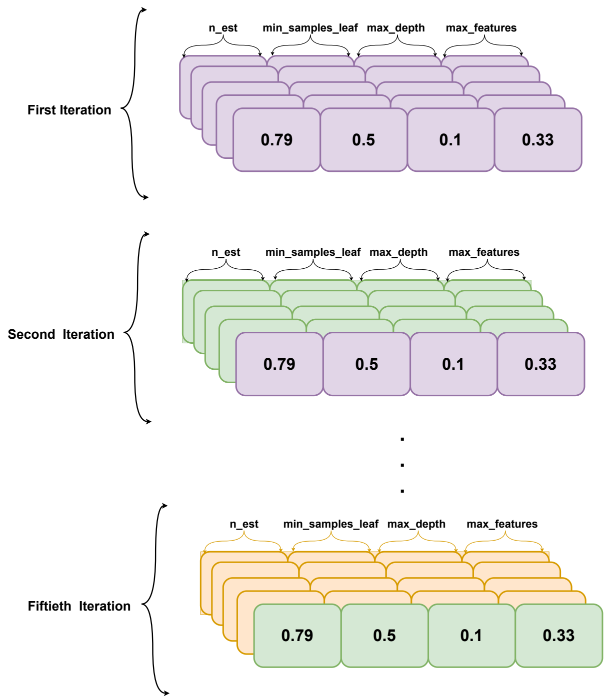

FIGURE 2: Individuals

After generating the individuals, the metaheuristic algorithm proceeds to optimize the parameters displayed in Figure 2. The algorithm then identifies the best individual within that iteration, subsequently compared to other optimized individuals in the search space during the next iteration. This cycle is repeated for up to 50 iterations to identify the best individual that enhances the model's performance and boosts evaluation measure rates.

The best individual is then used to identify the values of the parameters, enabling their inclusion in the random forest model. The dataset was divided into training and testing data, where the former was utilized to train the model for predictive purposes and the latter was used to assess model performance. The metaheuristic algorithms were employed to optimize the parameters and improve the model. Upon completion of this phase, the optimized parameters were processed using the random forest algorithm. The process can be observed in Figure

V. EXPERIMENTS AND RESULTS

This section breaks down the experiments that were done throughout the research. The Experimental setting investigates how the experiments were done. The dataset subsection describes the data. Next is the evaluation measures section explaining the measures used to evaluate the models. And finally, the subsection Experiments with parameter tuning contains an exposition of the results.

A. EXPERIMENTAL SETTINGS

A laptop with an Intel core i5-10210U CPU @ 1.60GHz 2.10 GHz/ 8 GB RAM was used to conduct this experiment with the Python 3.10 software. The packages used in this case are: Matplotlib used for plotting and generating quality graphs [36], Pandas for data manipulation [37], Numpy for arrays [38], Seaborn for graphic visualization [39], and Sklearn for fetching machine learning algorithms [40]. The code executing settings are to be shown in Table 1. Thirty runs are considered for the experiments, with 50 iterations and 20 individuals for each run. The code terminates after applying the 50 iterations for all the runs. The ranges of the different parameters are considered with the range values of [100, 200], [20, 80], [1, 20], and [0.1, 1] for the n-estimators, Maximum depth, Minimum sample leaf, and Max_features, respectively [41]. The initial values of the hyper-parameters are initiated randomly. EvoloPy framework [42] is customized by adding a fitness function for hyper-parameter tuning and model generation.

TABLE 1: Experimental Settings

Item	Value
Number of Runs	30
Number of Iterations	50
Population size	20
n-estimators	Max = 200, Min = 100
Maximum depth	Max = 80, Min = 20
Minimum sample leaf	Max = 20, Min = 1
Max_features	Max = 1, Min = 0.1

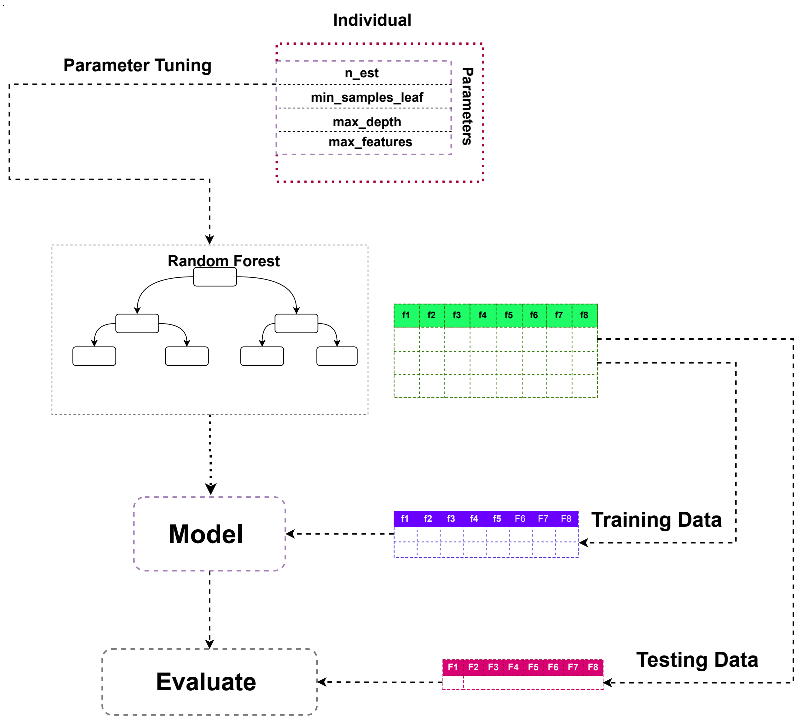

FIGURE 3: Methodology of parameter tuning of the Random Forest using Metaheuristic algorithms

B. DATASET

This research used a public dataset available on [1] Kaggle under the title (Predict if people prefer working from home vs working from office post-Covid-19), it consists of 208 rows and 19 features ID, Name, Age, Occupation, Gender, Same office location, Kids, Remote Work (RM) save money, RM Quality time, RM Better Sleep, Calmer Stressed, RM Professional Growth, RM Lazy, RM Lazy, RM productive, Digital Connection Sufficient, RM Better Work Life Balance, RM Improved Skill-set, RM Job Opportunities, and Target.

The dataset is based on a survey conducted during the pandemic, the purpose is to know the employee's points of view on their work location preferences and whether they prefer working from home or from the office. Table 2 provides an overview of the features utilized in this study. Each feature represents a question that was asked of the employees via a survey. For instance, the feature "same office location" corresponds to the question "Do you reside in the base location as your office?" and is answered with either a yes or a no. While some features are binary in nature, others are answered using a scale. For example, the feature "RM productive" is answered on a scale of 1-5 in response to the question "On a scale of 1-5, are you more productive with working remotely? 1 indicates not being productive at all, while 5 indicates being highly productive while working remotely."

According to Figure 4, which studies the correlation between the features, the two highest features affecting their choices are the "Better Work-Life Balance" and "Better Sleep"; so whether their sleeping improvement, employees prefer to work remotely yet if not, they would rather go back to their offices. The better Work-Life balance was highly correlated with their choices; some employees felt that bringing their work home created an imbalance between their home life and work life, making them eager to return to their offices. In opposition, having remote work job opportunities or not did not really affect their choices.

[1]https://www.kaggle.com/datasets/anninasimon/predict-if-people-prefer-wfh-verses-wfo-data

C. EVALUATION MEASURES

Evaluation metrics are crucial to machine learning since they are frequently used as objectives to optimize while building learning models and to compare various learning algorithms [43].

The classification performance of a classifier in relation to certain test data is summarized by a confusion matrix. The four matrix cells are referred to as True Positives (TP), False Positives (FP), True Negatives (TN), and False Negatives (FN) [44]. This study used four main measures: Accuracy, Precision, Recall, and F1 Score:

1) Accuracy: Accuracy is defined as the capacity of a test to accurately distinguish between the employees' work preference. It calculates the proportion of true positives and true negatives in all the evaluated instances resulting in an estimation of the Accuracy of a test. The Accuracy measure is mathematically described as the following [45]:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

TABLE 2: Dataset Description with the features' description

Feature	Question	Answer Type
Same Office Location	Do you reside in the base location as your office?	Yes or No
Kids	Do you have Kids	Yes or No
RM save money	Do you feel you were able to save money with remote work?	Yes or No
RM quality time	Do you feel that remote work has given you more quality time with family/friends?	Yes or No
RM better sleep	Has your sleep cycle improved with work from home?	Yes or No
Calmer Stressed	Are you calmer or more stressed than usual since remote work began?	Calmer or Stressed
RM professional growth	On a scale of 1-5, Do you feel WFH has affected your professional growth adversely?	5-yes it's affecting me badly, 1 No it doesn't affect me
RM lazy	On a scale of 1-5, Do you feel WFH has made you lazy?	5-extremely lazy, 1-Nope, not lazy at all
RM productive	On a scale of 1-5, are you more productive with working remotely?	1-not productive at all,5-extremely productive
digital connect sufficient	Do you feel digital connect is sufficient?	yes or no
RM better work life balance	On a scale of 1-5, do you feel you have a better work-life balance with remote work?	1-No not at all, 5-Yes
RM improved skillset	On a scale of 1-5,how much has your skillset improved in the last two years?	1:Not improved, 5:Improved drastically
RM job opportunities	Do you think there are more job opportunities with remote work?	Yes or No
Target	In the future which of the following do you think is more suited for you?	WFH (Work from Home) or WFO (Work from Office)

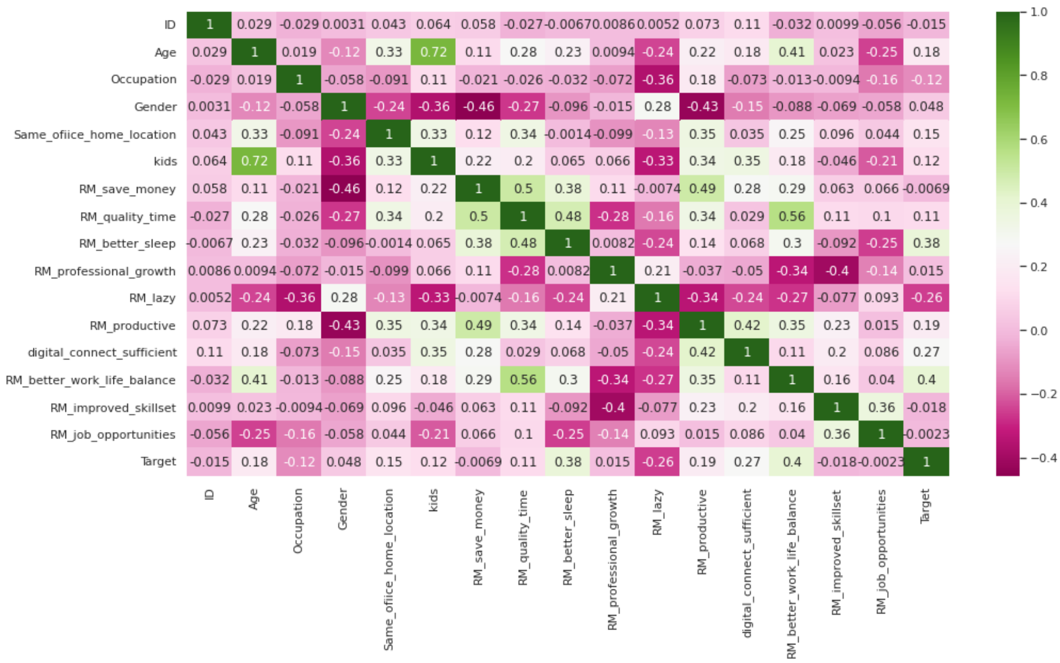

FIGURE 4: The correlation between the different features and the label of the dataset

2) Precision: The positive instances in a positive class that are accurately predicted from the total anticipated instances are measured by Precision [46]:

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

3) Recall is computed as the ratio of positive instances to the total number of positive instances in the search space. In essence, it provides an estimate of the completeness of the search results, indicating the proportion of positive instances that have been successfully predicted. [47]. The mathematical presentation of the Recall is shown as the following equation [48]:

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

4) F1-Score: Also known as the f-measure, drawn more attention in the classification context, particularly to assess unbalanced classification issues across a range of applications, including machine learning, computer vision, data analytics, and natural language processing. It may be easily understood as the harmonic mean of the two degrees of freedom in the confusion matrix, including the Precision and Recall [49]:

$$F1 - score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{4}$$

D. EXPERIMENTS WITH DIFFERENT ALGORITHMS

This research used five algorithms: Logistic Regression, K-Nearest Neighbor, Naive Bayes, Decision Tree, and Random Forest. These models were used to predict the employees' choice of work. As shown in Table 3, the algorithms were measured in terms of Accuracy, Precision, Recall, and F1-score. The random forest algorithm had the best model performance compared to the other algorithms. The Naive Bayes model did not perform as well as the other ones, as presented in the table. The Random forest and the Decision tree were the highest-performing models, and since the random forest is made of multiple decision trees, it was higher than the decision tree.

E. EXPERIMENTS WITH PARAMETER TUNING

As shown in table 4, the SSA algorithm increased the performance of the random forest in terms of Accuracy, Recall, Precision, and F1-score. The PSO was next in enhancing the model's performance but had the lowest Recall compared to the other algorithms. The GWO performed poorly in this case compared to the others, yet its Recall was higher than the PSO and WOA. Figure 5 shows the convergence curve for each metaheuristic algorithm, indicating the fitness value for each iteration, which is considered as the complement of the Accuracy measure. It is observed from the figure that the SSA and WOA, were both close to achieving higher performance, but the SSA had a lower fitness value and it had the lead; even though the WOA is stabilized first.

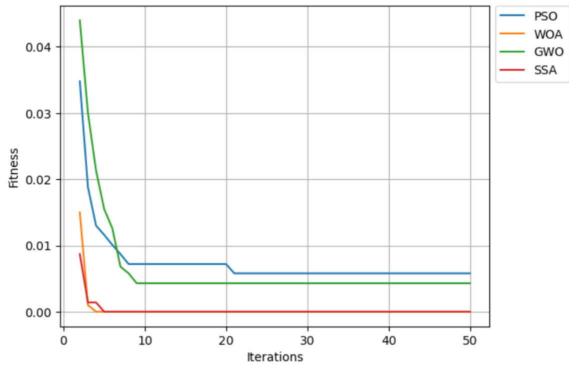

FIGURE 5: Convergence curves for each metaheuristic algorithm for tuning the parameters of the Random Forest

In summary, the random forest achieved the best results compared to the other algorithms, including K-NN, Decision Tree, Logistic Regression, and Naive Bayes. Tuning the hyper-parameters of the random forest using different metaheuristic algorithms, shows that the SSA metaheuristic algorithm can suggest the best values of the hyper-parameters of the Random Forest compared to the other metaheuristic algorithms

VI. CONCLUSION

With Covid-19 taking over the world, everyone was forced to deal with remote working. With such a change, it is important for employers and employees to have an insight into what their employee's work preferences are and how they can provide a comfortable work environment. This paper used machine learning algorithms to investigate the reasons behind their choices and their preferences. The Random Forest algorithm achieved the best performance in comparison to the other algorithms. Yet this model was enhanced by using metaheuristic algorithms for parameter tuning, having SSA and WOA for increasing the performance of the model.

By predicting employees' choices between remote and office work, this study provides an adaptable model that can be applied to other decision-making scenarios in corporate environments. Predictive modeling is beneficial in scheduling, labor management, and resource allocation optimization. Additionally, it can help with creating individualized work environments for employees, customizing rewards, and locating chances for career advancement. Through the application of machine learning techniques, firms can improve decision-making procedures, resulting in increased production, efficiency, and job satisfaction.

For future references, the model will be used on a bigger dataset, in addition, there will be other metaheuristic algorithms applied to the model and study their effects on the model. Also, more techniques would be applied like feature selection.

AVAILABILITY OF DATA AND MATERIAL

Data is available based on a reasonable request.

TABLE 3: Experiments with different Meta-algorithms

Algorithm	Avg Accuracy	Avg Recall	Avg Precision	Avg F1 Score
Random Forest	0.9423	0.9773	0.9011	0.9375
K-NN	0.7391	0.5862	0.7391	0.6538
Decision Tree	0.9057	0.8965	0.8816	0.8889
Logistic Regression	0.7619	0.6071	0.8095	0.6938
Naive Bayes	0.6825	0.5714	0.6666	0.6153

TABLE 4: Experiments with different algorithms

RF-Algorithm	Accuracy	Recall	Precision	F1 Score
Random Forest	0.9423	0.9773	0.9011	0.9375
RF-PSO	0.9691	0.992	0.94	0.9647
RF-GWO	0.9647	0.9954	0.9279	0.9599
RF-WOA	0.9667	0.9931	0.9341	0.962
RF-SSA	0.9739	1	0.9437	0.9705

CONFLICTS OF INTEREST

The authors have declared that there is no conflicts of interest. Non-financial competing interests.

COMPLIANCE WITH ETHICAL STANDARDS

This article does not contain any studies with human participants or animals performed by any of the authors.

FUNDING

No funding is received.

CREDIT AUTHOR STATEMENT

All authors contributed equally to this paper.

REFERENCES

[1] M. Gibbs, F. Mengel, and C. Siemroth, "Work from home & productivity: Evidence from personnel & analytics data on it professionals," University of Chicago, Becker Friedman Institute for Economics Working Paper, no. 2021-56, 2021.
[2] T. Mehdi and R. Morissette, "Working from home: Productivity and preferences," 2021.
[3] L. Bao, T. Li, X. Xia, K. Zhu, H. Li, and X. Yang, "How does working from home affect developer productivity?–a case study of baidu during covid-19 pandemic," arXiv preprint arXiv:2005.13167, 2020.
[4] V. Jayasutha and T. Arunachalam, "Deep learning with statistical analysis for stress prediction of remote working it employees in covid-19 pandemic," in 2022 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS). IEEE, 2022, pp. 179–184.
[5] P. Caligiuri, H. De Cieri et al., "Predictors of employees' preference for working from home post-pandemic," Business and Economic Research, vol. 11, no. 2, pp. 1–9, 2021.
[6] O. Biran and K. R. McKeown, "Human-centric justification of machine learning predictions." in IJCAI, vol. 2017, 2017, pp. 1461–1467.
[7] B. Kumara, G. Herath, P. Wijeratne, and K. Banujan, "Work from home after covid-19: Machine learning-based approach to predict employee's choice," in 2022 International Conference on Decision Aid Sciences and Applications (DASA). IEEE, 2022, pp. 147–150.
[8] A. SIMON, "Wfh verses wfo prediction - logistic regression," 2021. [Online]. Available: https://www.kaggle.com/code/anninasimon/wfh-verses-wfo-prediction-logistic-regression.htm
[9] C. Qi and X. Tang, "Slope stability prediction using integrated metaheuristic and machine learning approaches: a comparative study," Computers & Industrial Engineering, vol. 118, pp. 112–122, 2018.
[10] J. Qiu, X. Yin, Y. Pan, X. Wang, and M. Zhang, "Prediction of uniaxial compressive strength in rocks based on extreme learning machine improved with metaheuristic algorithm," Mathematics, vol. 10, no. 19, p. 3490, 2022.
[11] A. H. Sadeghi, E. A. Bani, A. Fallahi, and R. Handfield, "Grey wolf optimizer and whale optimization algorithm for stochastic inventory management of reusable products in a two-level supply chain," IEEE Access, 2023.
[12] S. Lim, D. Choi, S. H. Lee, C. Kang, and J.-W. Park, "Frequency stability enhancement of low-inertia large-scale power system based on grey wolf optimization," IEEE Access, vol. 10, pp. 11 657–11 668, 2022.
[13] M. E. Bento, D. Dotta, R. Kuiava, and R. A. Ramos, "A procedure to design fault-tolerant wide-area damping controllers," IEEE access, vol. 6, pp. 23 383–23 405, 2018.
[14] H.-P. Hsu and C.-N. Wang, "Hybridizing whale optimization algorithm with particle swarm optimization for scheduling a dual-command storage/retrieval machine," IEEE Access, vol. 11, pp. 21 264–21 282, 2023.
[15] T. Cuong-Le, T. Nghia-Nguyen, S. Khatir, P. Trong-Nguyen, S. Mirjalili, and K. D. Nguyen, "An efficient approach for damage identification based on improved machine learning using pso-svm," Engineering with Computers, pp. 1–16, 2021.
[16] Y. Qiu, J. Zhou, M. Khandelwal, H. Yang, P. Yang, and C. Li, "Performance evaluation of hybrid woa-xgboost, gwo-xgboost and bo-xgboost models to predict blast-induced ground vibration," Engineering with Computers, pp. 1–18, 2021.
[17] H. Lin and C. Wang, "Digwo-n-beats: An evolutionary time series prediction method for situation prediction," Information Sciences, p. 120316, 2024.
[18] P. K. Keserwani, M. C. Govil, E. S. Pilli, and P. Govil, "A smart anomaly-based intrusion detection system for the internet of things (iot) network using gwo–pso–rf model," Journal of Reliable Intelligent Environments, vol. 7, pp. 3–21, 2021.
[19] G. I. Sayed, A. Tharwat, and A. E. Hassanien, "Chaotic dragonfly algorithm: an improved metaheuristic algorithm for feature selection," Applied Intelligence, vol. 49, no. 1, pp. 188–205, 2019.
[20] F. S. Gharehchopogh, I. Maleki, and Z. A. Dizaji, "Chaotic vortex search algorithm: metaheuristic algorithm for feature selection," Evolutionary Intelligence, vol. 15, no. 3, pp. 1777–1808, 2022.
[21] H. Lin, C. Wang, and Q. Hao, "A novel personality detection method based on high-dimensional psycholinguistic features and improved distributed gray wolf optimizer for feature selection," Information Processing & Management, vol. 60, no. 2, p. 103217, 2023.
[22] E. Y. Boateng and D. A. Abaye, "A review of the logistic regression model with emphasis on medical research," Journal of data analysis and information processing, vol. 7, no. 4, pp. 190–207, 2019.
[23] F.-J. Yang, "An implementation of naive bayes classifier," in 2018 International conference on computational science and computational intelligence (CSCI). IEEE, 2018, pp. 301–306.
[24] C. S. Lee, P. Y. S. Cheang, and M. Moslehpour, "Predictive analytics in business analytics: decision tree," Advances in Decision Sciences, vol. 26, no. 1, pp. 1–29, 2022.
[25] Y. Qi, "Random forest for bioinformatics," in Ensemble machine learning. Springer, 2012, pp. 307–323.
[26] T. Dokeroglu, E. Sevinc, T. Kucukyilmaz, and A. Cosar, "A survey on new generation metaheuristic algorithms," Computers & Industrial Engineering, vol. 137, p. 106040, 2019.
[27] A. Alridha, A. M. Salman, and A. S. Al-Jilawi, "The applications of np-hardness optimizations problem," in Journal of Physics: Conference Series, vol. 1818, no. 1. IOP Publishing, 2021, p. 012179.
[28] P. Agrawal, H. F. Abutarboush, T. Ganesh, and A. W. Mohamed, "Metaheuristic algorithms on feature selection: A survey of one decade of research (2009-2019)," IEEE Access, vol. 9, pp. 26 766–26 791, 2021.
[29] R. Qaddoura, H. Faris, I. Aljarah, and P. A. Castillo, "Evocluster: an open-source nature-inspired optimization clustering framework," SN Computer Science, vol. 2, no. 3, pp. 1–12, 2021.
[30] ——, "Evocluster: An open-source nature-inspired optimization clustering

framework in python," in International conference on the applications of evolutionary computation (part of EvoStar). Springer, 2020, pp. 20–36.
[31] M. Imran, R. Hashim, and N. E. A. Khalid, "An overview of particle swarm optimization variants," Procedia Engineering, vol. 53, pp. 491–496, 2013, malaysian Technical Universities Conference on Engineering & Technology 2012, MUCET 2012. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1877705813001823
[32] A. E. Hegazy, M. Makhlouf, and G. S. El-Tawel, "Improved salp swarm algorithm for feature selection," Journal of King Saud University - Computer and Information Sciences, vol. 32, no. 3, pp. 335–344, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1319157818303288
[33] G. Sun, Y. Shang, K. Yuan, and H. Gao, "An improved whale optimization algorithm based on nonlinear parameters and feedback mechanism," International Journal of Computational Intelligence Systems, vol. 15, no. 1, pp. 1–17, 2022.
[34] K. Maharana, S. Mondal, and B. Nemade, "A review: Data pre-processing and data augmentation techniques," Global Transitions Proceedings, vol. 3, no. 1, pp. 91–99, 2022.
[35] A. E. Eiben and J. E. Smith, Introduction to evolutionary computing. Springer, 2015.
[36] N. Ari and M. Ustazhanov, "Matplotlib in python," in 2014 11th International Conference on Electronics, Computer and Computation (ICECCO). IEEE, 2014, pp. 1–6.
[37] W. McKinney et al., "pandas: a foundational python library for data analysis and statistics," Python for high performance and scientific computing, vol. 14, no. 9, pp. 1–9, 2011.
[38] C. R. Harris, K. J. Millman, S. J. Van Der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith et al., "Array programming with numpy," Nature, vol. 585, no. 7825, pp. 357–362, 2020.
[39] A. H. Sial, S. Y. S. Rashdi, and A. H. Khan, "Comparative analysis of data visualization libraries matplotlib and seaborn in python," International Journal, vol. 10, no. 1, 2021.
[40] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg et al., "Scikit-learn: Machine learning in python," the Journal of machine Learning research, vol. 12, pp. 2825–2830, 2011.
[41] F. Hidayat and T. M. S. Astsauri, "Applied random forest for parameter sensitivity of low salinity water injection (lswi) implementation on carbonate reservoir," Alexandria Engineering Journal, vol. 61, no. 3, pp. 2408–2417, 2022.
[42] H. Faris, I. Aljarah, S. Mirjalili, P. A. Castillo, and J. J. M. Guervós, "Evolopy: An open-source nature-inspired optimization framework in python." IJCCI (ECTA), vol. 1, pp. 171–177, 2016.
[43] J. Huang and C. X. Ling, "Constructing new and better evaluation measures for machine learning." in IJCAI, 2007, pp. 859–864.
[44] K. M. Ting, Confusion Matrix. Boston, MA: Springer US, 2010, pp. 209–209.
[45] P. Baratloo, "1: simple definition and calculation of accuracy, sensitivity and specificity," Emerg (Tehran), no. 3, p. 48.
[46] M. Hossin and M. N. Sulaiman, "A review on evaluation metrics for data classification evaluations," International journal of data mining & knowledge management process, vol. 5, no. 2, p. 1, 2015.
[47] W. H. Walters, "Google scholar search performance: Comparative recall and precision," portal: Libraries and the Academy, vol. 9, no. 1, pp. 5–24, 2009.
[48] S. Seo, Y. Kim, H.-J. Han, W. C. Son, Z.-Y. Hong, I. Sohn, J. Shim, and C. Hwang, "Predicting successes and failures of clinical trials with outer product–based convolutional neural network," Frontiers in Pharmacology, vol. 12, p. 670670, 2021.
[49] D. J. Hand, P. Christen, and N. Kirielle, "F*: an interpretable transformation of the f-measure," Machine Learning, vol. 110, no. 3, pp. 451–456, 2021.

SALMA ALMOHTASIB an Information Sciences student at Al Hussein Technical University, is set to graduate in 2024. With a keen interest in Artificial Intelligence, she combines her academic pursuits with a passion for staying at the forefront of technological advancements.

RANEEM QADDOURA Dr. Raneem Qaddoura is an Assistant Professor at Al Hussein Technical University, with a combined experience of 17 years in both academia and industry. She holds a Ph.D. in computer science with a focus on machine learning and data mining. Throughout her career, Dr. Qaddoura has held several important positions such as Assistant Professor, Instructor, Project Management Office (PMO) Manager, Technical Team Leader, Senior Web Developer, and Application Developer. Her extensive experience has led to the publication of over 40 papers in prestigious journals, book chapters, and conferences indexed in leading databases. She is also an active research member of the Evolutionary and Machine learning group (Evo-ML.com), which focuses on evolutionary algorithms, machine learning, and their applications for solving fundamental problems in different areas. Her current research involves evolutionary computation, deep learning, data classification, and clustering.

ISLAM ALOMARI is An assistant Instructor, Department of Data Science and Artificial Intelligence at Al Hussein Technical University, she has bachelor's degree in computer information systems and holding a master's degree in data science, she has five years of teaching experience and has a number of research in various fields: data streams curation, data visualization and evolutionary computation.

ABDELAZIM G. HUSSIEN received the bachelor's degree from the Mathematics and Computer Science Department, Faculty of Science, Menoufia University, Egypt, in 2013, and the master's degree from the Faculty of Science, Menoufia University, in 2018. He is currently a Teaching Assistant with the Faculty of Science, Fayoum, Egypt. He was a PhD student in Faculty of Science, Menoufia University from 2018 – 2021. He is now researcher in Linkoping University, Sweden. He has around 80 papers in prestigious journals. He developed many optimization algorithms in literature such as Snake Optimizer, Fick's Law Algorithm, Pied Kingfisher Optimizer, etc. His research interest include optimization, swarm intelligence, cheminformatics, simulation & modeling.

SHUANG WANG received Ph.D. degree from University of Chinese Academy of Sciences in 2020. From August 2020 to August 2022, she was an associate professor with School of Information Engineering, Sanming University, China. She is currently an associate professor with New Engineering Industry College, Putian University, China. Her primary research interest include swarm intelligence, metaheuristics, and artificial intelligence.