A Combined Approach for Predicting Employees' Productivity based on Ensemble Machine Learning Methods

Ruba Obiedat and Sara Toubasi*
E-mail: r.obiedat@ju.edu.jo, tubasisara@gmail.com

King Abdullah II School for Information Technology, The University of Jordan, Amman 11942, Jordan

Keywords: MLP, J48, RBF, SVM, random forest, adaboost, bagging, productivity, accuracy

Received: November 24, 2021

Garment industrial sector is one of the most important business sectors in the world. It presents the lifeblood for many countries' economy. The demanding of garment merchandise in accretion year over year. There are many key factors affecting the performance of this sector including the employees' productivity. This research proposes a hybrid approach which aims to predict the productivity performance of garment employees by combining different classification algorithms including J48, random forest (RF), Radial Base Function network (RBF), Multilayer Perceptron (MLP), Naïve bayes (NB) and Support vector machine (SVM) with ensemble learning algorithms (Adaboost and bagging) on garment employees' productivity dataset. This work monitors three major evaluation metrics namely, accuracy, Mean Absolute Error (MAE) and Root Mean Square Error (RMSE). The results show that RF outperforms the other standard algorithms with accuracy of 0.983 and RSME of 0.1423. Applying Bagging and Adaboost with all standard classification algorithms on the dataset succeed in enhancing almost all classifiers' performance. Adaboost and bagging algorithms has been applied with all classification algorithms using different number of iterations starting from 1-100. The best result is achieved by applying Adaboost ensemble algorithm with J48 algorithm on its 20th iteration with an outstanding accuracy of 0.9916 and RSME of 0.0908.

Povzetek: .

1 Introduction

Machine learning (ML) is a branch of artificial intelligence that helps the computer to predict outcomes automatically by learning instantly from training data and previous experiences without any explicit programming. The idea of ML is trying to imitate the human's brain ability to solve problems and analyze it according to previous experiences. Thus, ML techniques are about using different algorithms on data to extract certain patterns that enhance decision-making process. There are various types of machine learning such as supervised learning, unsupervised learning, semi supervised learning and reinforcement learning, Zhang 2010 [1]. Each type of ML algorithm is used for solving specific kind of problems; some algorithms can be used for classification, other for regression while some are used for clustering. Choosing the suitable algorithm depends on the problem type and many other factors such as parametrization, time of learning, time of predicting, over fitting tendency and memory size, Mahesh 2019 [2]. All ML algorithms are useful techniques which assist people in various areas, such as data mining, image processing, and prediction analysis, Mona M. Jamjoom and 2021 [3].

ML algorithms could be used to solve different types of problems in various sectors depending on the type of algorithm. For instance, when the problem under study needs a prediction and analysis approach the suitable ML algorithm is the classification algorithms which help to predict the problem according to a given parameters. Classification algorithms are used in different domains such as medical sector, business sector, image recognition and many others. ML algorithms succeed in medical diagnosis specially when it is used for designing computer aided diagnosis (CADX) system which is a part of breast cancer detection on mammograms, Ozcift and Gulten 2011 [4]. Food image recognition system has been designed using ML algorithms for recording people's eating habits Taichi and Keiji 2009 [5]. Machine learning has been used also in finance sector like internet loan fraud prediction Fang et al. 2021 [6].

Ensemble learning (EL) is a machine learning mechanism that merges several base models in order to produce one optimal predictive model. EL has been used for increasing accuracy and consolidating the classification performance

Feng, Huang, and Ren 2018 [7]. In addition, ensemble learning algorithms contributed to the prediction in many sectors. Bagging and Boosting significantly improve predicting churn when applied on customer database of U.S wireless telecom company Lemmens and Croux 2006 [8].

Garment industry is a huge industry which employs millions of people and profits billions of dollars every year. The strength of garment economy makes the economic countries such as Bangladesh, India, China, and many countries focus on developing garment industrial sector Hearle 2016 [9]. Predicting risks and earning high profit, are the main goals of any industry. However, there are many types of risks affect the process in the garment industry sector. One of these risks is the description risk, which found to be the most critical risk type and can affect all other risks in the industry Chowdhury et al. 2019 [10].

According to many research, there are many key factors that affect the employees' productivity. Some of these factors include employee training, employee empowerment, and teamwork skillsHanaysha 2016 [11; Harfoushi and Obiedat 2011 [12]. In addition, the internal system in the manufactory has an effect on the productivity of employees. The effects include linking rewards to performance and initializing comfortable environment Evans and Davis 2015 [13; Harfoushi, Obiedat, and Khasawneh 2010 [14]. There are other key factors that have been found in previous research which studied a Bangladesh manufactory. It has been summarized into nine main key factors that are; working hours, wages and benefits, holidays, discrimination, harassment and abuse, workplace conditions, forced labor, welfare and employment relations Alam, Alias, and Azim 2018 [15]. Improving employees' productivity is one of the main goals of many manufactories especially those looking for stability and high standards productivity. Thus, the garment industries are one of the industrial sectors which are trying to find the easiest and fastest way to predict the productivity of employees in order to improve their performance.

2 Related work

This section discusses the main studies which focused on the usage of Machine Learning (ML) algorithms and ensemble learning algorithms in various sectors prediction issues.

Ensemble learning algorithms such as decision tree, adaBoost, Naïve Bayes, Random Forest and SVM were applied by study Bhatia, Arora, and Tomar 2016 [16] for presence of diabetic retinopathy, the results proved that the model could help in detecting symptoms earlier.

Outperformed results were found in a study conducted by Kruppa et al. 2013 [17] for credit risk prediction using framework of machine learning algorithms such as random forests (RF), k-nearest neighbors (KNN) and bagged k-nearest neighbors (BKN). Furthermore, a study by Balla, Rahayu, and Purnama 2021 [18] proved a promising result in predicting employee's productivity which is one of the most substantial factors in any organization. The study applied three classification algorithms namely, Neural Network (NN), Random Forest (RF) and Regressi Linier (RL). Random forest showed minimal values of correlation coefficient, MAE, and RMSE, which reflect that RF is very appropriate in predicting employee's productivity.

Decision tree classification algorithms utilized by Attygalle and Abhayawardana 2021 [19] for investigating and visualizing employee productivity and any other social phenomenon with evidence. Moreover, decision tree methods and data mining tools employed by Ďurica, Frnda, and Svabova 2019 [20] to build a model for predicting financial difficulties of polish companies. The results presented prediction power around 98% and more. In addition, Mahoto et al. 2021 [21] had used three machine learning algorithms (Multiclass Random Forest, Multiclass Logistic Regression, Multiclass one-vs-all) in order to help business workers to set product pricing and discounts depending on customer behavior, the model showed outstanding results in product price prediction. On the other hand, prediction model has been built by study Sorostinean, Gellert, and Pirvu 2021 [22] using decision tree methods and data mining tools for investigating the effect of decision tree methods and ensemble learning for improving performance prediction in assembly assistance system. The results demonstrated that the gradient boosted decision trees was the best through all the decision tree-based methods.

Some studies evaluated worker 's performance of textile company by using ML and ensemble learning algorithm, such as study as Saad 2020. [23] which applied different Machine learning algorithms including, decision tree and bagging algorithm to achieve the highest accuracy. The CHAID model produced high-level specificity and sensitivity.

Four different ML algorithms including, support vector machine, optimized support vector machine (using genetic algorithm), random forest, XGBoost and Deep Learning were used by El Hassani, El Mazgualdi, and Masrour 2019 [24] for predicting the overall equipment effectiveness (OEE) which is a performance measurement of manufacturing industry. Deep learning and random forest with cross validation manifest the best results for predicting OEE. Additionally, an approach built in study De Lucia, Pazienza, and Bartlett 2020 [25] of ML and logistic regression used for financial performance prediction by focusing on predicting the accuracy of main financial indicators such as Return of Equity (ROE) and Return of Assets (ROA). The ML algorithms were performed perfectly for predicting ROE and ROA.

All studies and research work mentioned above focused on combining two or more classifiers and how this integration of different techniques and algorithms can help in prediction. This research focuses on combining classification algorithms with bagging and Adaboost. In addition, the iterations from 1 to 100 are recorded to study how these combinations influence the accuracy, RMSE, and MAE values of predicting employees' productivity. Detailed comparisons between our study and the studies mentioned above shown in Table 1.

3 Classification algorithms

3.1 Decision tree

A decision tree (DT) is a popular classification technique. DT aims to build a model that predict the value of target variable. It represents the decision and the possible outcomes by building a flow chart structure with nodes, and leaves. The node without incoming edges is called root, but the node with outgoing edge is called internal or tested node, while the other nodes are called decision nodes. Decision tree chooses the best node by calculating the uncertainty of an attribute which called information gain for each node. The node with the highest gain is chosen as rooted node and the rest nodes are used again for information gain calculation. The algorithm goes through all the possible nodes to calculate the value of attribute x and the cut-off value Ihya et al. 2019 [26]. The decision tree flow chart shown in Figure 1.

The J48 is an execution of the C4.5 decision tree algorithm. J48 creates the decision tree by classifying new instances from the attribute values of training dataset. The time it comes through the training set, it admits the attributes which are responsible for classifying the various instances most accurately. All the

possible feature's values with ambiguity equal zero are assigned to the concern branch by terminating it Uma Mahesh et al. 2021 [27].

3.2 Random Forest

Random forest classification depends on creating number of trees based on the binary recursive partitioning trees by generating random variables. The tree consists of two types of nodes; the root node that involves the entire predictor area, and the terminal node that represents the last part of the predictor area. The splitting criteria depends on the value of predictor variable. When the predictor variable is smaller than the split, the point goes to the left and the rest go to the right El Hassani, El Mazgualdi, and Masrour 2019 [24]. Below equation represents the classifier where $\ominus\, i$ represents the number of independent vectors distributed identically so that every tree has a vote for most popular class of input X, De Lucia, Pazienza, and Bartlett 2020 [25].

$$Space = h\,(X, \ominus\, i);\quad i = 1, 2, 3, ., nT \qquad (1)$$

3.3 Naïve bayes

Nave bayes is a probabilistic classifier which simplifies learning by defining the features as independent given class. Each class describes by feature vector. Despite of the simplicity of Naive Bayesian classifier, it is doing well, and it used very often because it outperformed more complicated classification methods. Bayes theorem work on calculating the posterior probability, $P\,P(c|x)$, from $P(c), P(x), and\ P(x|c)$, the equation below shows the simple form of Bayes theorem, where $X = (X1, \dots Xn)$ is a value of predictor, and C is a class Narayanan, Arora, and Bhatia 2013 [28].

$$P(X|C) = P\,(C|X) * P(X)\,/\,P\,(C) \qquad (2)$$

3.4 Multilayer perceptron

Multilayer perceptron (MLP) classifier is a feedforward neural network. MLP structure consists of three layers: input, hidden and output layer. The minimum number of layers is 3 layers as shown in Figure 2 which consists of input layer, hidden layer, and output layer.

Reference	Author	Year	Dataset	ML algorithm	Evaluation measurements
16	Bhatia, K., S. Arora, and R. Tomar	2016	Retinal image processing algorithms	- DT. - AdaBoost. - NB - RF - SVM	**Accuracy 94%**
17	Kruppa, J., et al	2013	Payment histories of installment credits	- RF - KNN - Bagged KNN.	**AUC 95.5%**
18	Balla, I., S. Rahayu, and J.J. Purnama	2021	Garment worker productivity dataset	- NN. - RF. - RL.	**MAE 0.0787**
19	Attygalle, D. and G. Abhayawardana	2021	Employees of a non-government organization in Sri Lanka	- Classification trees	**Accuracy 74%**
20	Mahoto, N., et al	2019	Financial and economic indicators of Polish companies	- Prediction models Based on DT	**Accuracy 98.2%**
21	Mahoto, N., et al	2021	Customers and products dataset	- Multi class RF. - Multi class LR	**Mean 90%**
22	Sorostinean, R., A. Gellert, and B.-C. Pirvu	2021	Trainees and manufacturing workers.	- classification and regression trees. - RF - Gradient Adaboost DT	**Accuracy 65.11%**
23	Saad, H	2020	Data collected from Libyan Textile Company	- RF. - Boosted trees - Interactive tree (CART and CHAID). -	**Accuracy 99.16%**
24	El Hassani, I., C. El Mazgualdi, and T. Masrour	2019	Cable production industry dataset	- SVM - Optimized SVM - RF - XGBoost. - Deep Learning	**MAE 6.16**
25	De Lucia, C., P. Pazienza, and M. Bartlett	2020	ASSET4/EIKON database.	Machine learning model (Rf, SVM, KNN, ANN, Ridge regression)	**Mean dependent var 4.55**

Table 1: Related work comparison.

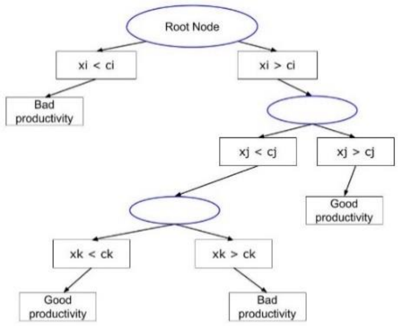

Figure 1: Decision tree flowchart.

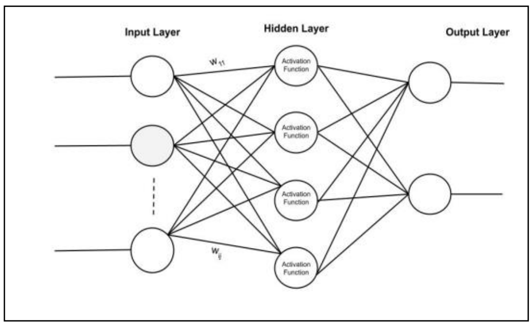

Figure 2: Three-layer multilayer perceptron neural network.

The input layer handout the input to the next layers. Thresholds and weights should be calculated for each hidden node and output node. Input nodes and output nodes has linear activation functions, but the hidden

nodes has nonlinear activation functions which are called sigmoid function Nazzal, El-Emary, and Najim 2008 [29]. Each signal passes among a node in a sequence layer that has the original input multiplied by weights with thresholds added then it passes among activation function.

The input to the jth hidden unit, $net_p(j)$, is expressed in equation (3). The N input units are represented by the index K, $Whi\ (J, K)$ denotes the weight connecting the K_{th} input unit to the J_{th} hidden unit Delashmit and Manry 2005 [30].

$$net_p(j) = \sum_{k=1}^{N+1} w_{hi}\ (j,k).x_p(k) \qquad 1 \le j \le N_h \qquad (3)$$

The output activation for the P_{th} training pattern, $O_p(j)$, being expressed by equation (4)

$$O_p(j) = f(net_p(j)) \qquad (4)$$

The nonlinear activation is typically chosen to be the sigmoidal function

$$f(net_p(j)) = \frac{1}{1+e^{-net\ p(j)}} \qquad (5)$$

3.5 Radial Base Function

Radial Base Function classifier or (RBF) is a feed forward network algorithm that has minimum 3 layers which are input layer, hidden layer, and output layer. In RBF the hidden layer weights are absent, also the activation function/sigmoid function is not used to calculate the hidden-units' outputs, rather than each output Z_j is acquire the input X to an n-dimensional parameter vector μ_j associated with the j_{th} hidden unitLeung, Lo, and Wang 2001 [31].

The equation below shows the response of characteristics of j_{th} hidden unit, (j= 1,2, …. J).

$$Z_j = k\left[\frac{||X-\mu_j||}{\sigma_j^2}\right] \qquad (6)$$

3.6 Support vector machine

Support vector machine (SVM) is a supervised learning algorithm that depends on implicitly mapping the sample vectors into a high dimensional, nonlinear feature space which is called kernel trick. The samples separate into a kernel using a similarity function called the optimal separating hyperplane (OSH). It minimizes the risk of misclassifying and maximizes the distance between two parallel plans. Each training data labeled as data points of the following form Cao 2019 [32]:

$$M = \{(x_1, y_1), (x_2, y_2), \ldots\ldots, (x_n, y_n)\} \qquad (7)$$

Where $y = 1/-1$, is a constant that refers to the class to which that point belongs, n=number of data sample, and x_n is a p-dimensional real vector.

SVM classifier works first on mapping the input vectors to decision value then executes the classification using proper threshold value.

4 Ensemble learning algorithms

Ensemble methods aim to enhance the predictive performance for a given classification algorithms. Bagging and Adaboost present the two most popular ensemble algorithms.

4.1 Bagging

Bootstrap Aggregating-Bagging algorithm is a homogeneous weak learner that generates sampling instances from the training set to produce an aggregated predictor which is acquired using majority voting rule. Bagging works very well for overfit models, because it works on decreasing the variance mean squared error (MSE) for a given operation such as decision trees or another algorithm by choosing a variable and arranging them into linear model. The dataset is signified by $L_{i=}(Y_i, X_i)(i = 1, \ldots\ldots, n)$ X_i is p-dimensional explanatory variable for i_{th} instant and Y_i is the real valued response Yaman and Subasi 2019 [33]. The Pseudocode of Bagging is shown in Figure 3.

Algorithm 1 Pseudocode for Bagging

Initialize the parameters

a. DInitialize the parameters

b. $= \oslash$ .the ensemble

c. L, the number of classifiers to train

For k = 1, . . . , L

a.Take a bootstrap sampleS_k from **Z**

b.Build a classifier D_k using as the training set

c.Add the classifier to the current ensemble, $D = D \cup D_K$.

Returen *D*

Classification phase

Run $D1, \ldots, DL$ **on the input x**
The class with the maximum number if votes is chosen as the label x

Algorithm 2 Pseudocode for AdaBoost

Given $(x_1, y_1), , (x_m, y_m)$ $where$ $x_i \in X, y_i \in Y = -1, +1$

Initialize $D_1(i) = 1/m$

$For(t) = 1 \quad For, \ldots, T:$

Train weak learner using distribution D_t

Get weak hypothesis(h_t) : $(X \rightarrow \{-1, +1\})$ with error$\epsilon_t = P_r \widetilde{D}_t h_t(x_i) \neq y_i$

choose $\alpha t = 1/2ln\,((1 - \epsilon t)/\epsilon t)$

4.2 Adaboost

Boosting is referred to Adaptive Boosting, it is a homogenous learner who produces a series of classifiers aiming to improve the accuracy of the classifier. Depending on each classifier performance, the training set will be chosen. The incorrectly classified sample will be selected more often than the correctly classified samples. Consequently, a new classifier produced by boosting algorithm which performs well on new dataset. Using the weighted majority vote, boosting will influence the classifier. Training sets prepared as $(x_1, y_1), \ldots.(x_n, y_n).x_i \in X$, while X symbolize instance space, and training set members are labeled with $y_i \in y = \{-1, +1\}$. All weights given to training set equal 1/m Bühlmann 2012 [34] . Adaboost calling weak learning algorithm repeatedly according to T which presents the

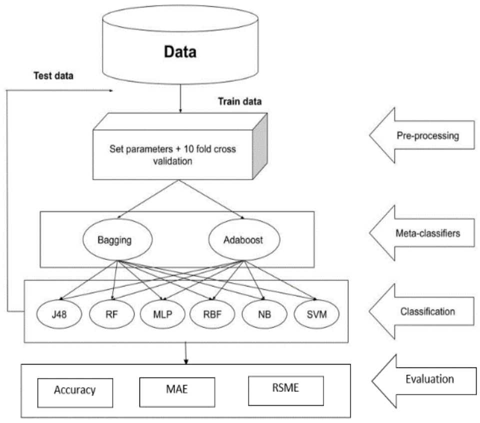

Figure 5: Process model.

times of iterations. The Pseudocode of Adaboost is shown in Figure 4.

5 Methodology

This section describes in detail the research process of the proposed work and the used datasets (Garment employee productivity), each of which will be discussed in detail in the following subsections.

5.1 Research process

This research follows a four main stages methodology framework. First, it applies six classification algorithms namely, J48, Multilayer Perceptron, Random Forest, Radial base Function, naïve bayes and Support vector machine. After that, it uses Bagging algorithm with every classification algorithm. Followed by applying Adaboost ensemble algorithm with every classification algorithm as well. All results are calculated using 10 folds cross-validation and fixed parameters of every classification algorithm.

Finally, the results are evaluated using the accuracy, MAE and RMSE measurements. Figure 5 below presents the main stages.

5.2 Dataset

This research used Garment employee productivity dataset. Garment employee productivity dataset contains 1197 instances divided into two classes: 747 "good" and 450" bad".

The data was collected and prepared by Imran, Rahim, and Ahmed 2021 [35]. The original Garment employee productivity contains 15 attributes between integer and real type as shown in Table 2.

5.3 Evaluation and measurements

Evaluation metrics are various measurements that provide a complete image about machine learning prediction performance. This study used three measurements namely, Accuracy, MAE, and RMSE.

Accuracy

No.	Attribute	Description
1	date	Date in MM-DD-YYYY
2	day	Day of the Week
3	quarter	A portion of the month. A month was divided into four quarters
4	department	Associated department with the instance
5	team_no	Associated team number with the instance
6	no_of_workers	Number of workers in each team
7	no_of_style_change	Number of changes in the style of a particular product
8	targeted_productivity	Targeted productivity set by the Authority for each team for each day.
9	smv	Standard Minute Value, it is the allocated time for a task
10	wip	Work in progress. Includes the number of unfinished items for products
11	over_time	Represents the amount of overtime by each team in minutes
12	incentive	Represents the amount of financial incentive (in BDT) that enables or motivates a particular course of action
13	idle_time	The amount of time when the production was interrupted due to several reasons
14	idle_men	The number of workers who were idle due to production interruption
15	actual_productivity	The actual % of productivity that was delivered by the workers. It ranges from 0-1.

Table 2: Attributes information.

Accuracy is a measurement which gives an indication about machine learning prediction if it works effectively or not.

$$\text{Accuracy} = \frac{\text{Number of correct predictions}}{\text{Total number of predictions}} \quad (8)$$

It also could be calculated by positive and negative predictions as the following equation:

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}} \quad (9)$$

Where TP= True Positives, TN= True Negative, FP = False Positive, FN = False Negative.

Mean Absolute Error Value

MAE is the absolute value of the individual prediction error, while the prediction error is the predicted error subtracted from the actual error of the instance. The calculations of MAE shown in equation (10) Vujović [36].

$$MAE = \frac{1}{n}\sum_{j=1}^{n}\left|p_{ij} - \sum\nolimits_{j=1}^{n}\left|p_{ij} - T_j\right|\right| \quad (10)$$

Where $P(ij)$ is the predicted value by the individual model i of record j, Tj is the target value of record j.

Root Mean Square Error

RMSE is called also standard error (SE), is an error which gives a full picture of error distribution Chai and Draxler 2014 [37], the equation of RMSE as shown below

$$RMSE = \sqrt{\frac{\sum_{j=1}^{n} {P_{ij}}^2}{n}} \qquad (11)$$

5.4 Experiments and results

This research concentrated on achieving the highest accuracy with minimal values of MAE and RMSE for predicting employees' productivity. Firstly, all classification algorithms have been applied on Garment employee productivity dataset and the accuracy, MAE and RSME values have been recorded as shown in Table 3. The results show that all the classification algorithms have achieved a high accuracy exceeding 80%. The highest accuracy was 0.983 using RF classification while the J48 has achieved the lowest MAE and RSME with 0.0259, 0.1241 respectively. Bagging and Adaboost have been applied with all classification algorithms on the dataset. Both ensemble algorithms succeed in enhancing almost all classifiers' performance, but Adaboost has outperformed Bagging algorithms, the results presented in Table 4 & 5. In order to gain higher accuracy and lower MAE and RMSE values; Adaboost and bagging algorithms has been applied with all classification algorithms using different number of iterations starting from 1-100. When Adaboost was combined with classification algorithms using different numbers of iterations the results of MLP, NB, and RF didn't show any changes. However, the other classification algorithms including J48, RBF and SVM shows variation in their performance. J48 achieves outstanding results on 20 iterations, with accuracy of 0.9916 and a low MAE and RSME of 0.0083 and 0.0908 respectively, the results shown in Table 6. Additionally, the results of RBF and SVM have been improved. Bagging with classification algorithms have been applied using different number of iterations as well. The results prove that J48 and MLP has achieved an outstanding result on the 90 iterations, while RF on first iteration, NB on 10 iterations, but SVM and RBF on 20 iterations, bagging with classification algorithms using different number on iterations are displayed in Table 7. Figures 6 and 7 show a summary and visualized representation of the MAE results of Bagging and Boosting using different numbers of iterations.

6 Comparison and discussion

This study focuses on finding the best approach for predicting employees' productivity. After reviewing all previous work and their results shown in Table 1, it can be noticed that only one study used the same garment employee productivity dataset [18]. Study [18] had followed a typical ML approach as it applied standard ML algorithms (Neural Network (NN), Random Forest (RF) and Regressi Linier (RL)) without any ensemble algorithms or following any other hybrid approach that can help in improving their results. On the other hand, other studies such as [16, 22] used the ML algorithm with ensemble algorithms, but the results showed higher values of MAE or lower accuracy. Moreover, only one study done by [23] combined the ensemble algorithm (Bagging) with four different decision tree algorithms to predict the worker performance of Libyan Textile Company. The accuracy result was very close to our study results, which is 99.1%. However, study [23] used a different dataset that

Algorithm	J48	RF	MLP	RBF	NB	SVM
Accuracy	0.950	**0.983**	0.981	0.834	0.855	0.936
MAE	**0.0259**	0.0972	0.151	0.1345	0.2758	0.0643
RMSE	**0.1241**	0.1423	0.210	0.1737	0.3371	0.2536

Table 3: Classification algorithms.

Bagging						
Algorithm	J48	RF	MLP	RBF	NB	SVM
Accuracy	0.983	0.983	**0.986**	0.877	0.861	0.877
MAE	**0.0271**	0.1229	0.0392	0.2124	0.2758	0.0689
RMSE	0.116	0.1664	**0.113**	0.3033	0.3371	0.2289

Table 4: Bagging with classification algorithms.

Boosting						
Algorithm	J48	RF	MLP	RBF	NB	SVM
Accuracy	**0.991**	0.986	0.981	0.873	0.855	0.960
MAE	**0.01**	0.1051	0.0216	0.1478	0.1795	0.045
RMSE	**0.097**	0.1528	0.1394	0.301	0.3377	0.179

Table 5: Boosting with classification algorithms.

	Boosting											
Class-ifier	**Num Iteration**	1	10	20	30	40	50	60	70	80	90	100
J48	**Accuracy**	0.9825	0.9908	**0.9916**	0.9900	0.9916	0.9908	0.9900	0.9900	0.9900	0.9900	0.9900
	MAE	0.0259	0.0101	**0.0083**	0.0100	0.0090	0.0093	0.0100	0.0100	0.0100	0.0100	0.0100
	RMSE	0.1241	0.0970	**0.0908**	0.1001	0.0924	0.0959	0.1001	0.1001	0.1001	0.1001	0.1001
MLP	**Accuracy**	0.9808	0.9808	0.9808	0.9808	0.9808	0.9808	0.9808	0.9808	0.9808	0.9808	0.9808
	MAE	0.0256	0.0216	0.0216	0.0216	0.0216	0.0216	0.0216	0.0216	0.0216	0.0216	0.0216
	RMSE	0.1201	0.1394	0.1394	0.1394	0.1394	0.1394	0.1394	0.1394	0.1394	0.1394	0.1394
Random forest	**Accuracy**	0.9858	0.9858	0.9858	0.9858	0.9858	0.9858	0.9858	0.9858	0.9858	0.9858	0.9858
	MAE	0.1051	0.1051	0.1051	0.1051	0.1051	0.1051	0.1051	0.1051	0.1051	0.1051	0.1051
	RMSE	0.1528	0.1528	0.1528	0.1528	0.1528	0.1528	0.1528	0.1528	0.1528	0.1528	0.1528
Naïve bayes	**Accuracy**	0.8546	0.8546	0.8546	0.8546	0.8546	0.8546	0.8546	0.8546	0.8546	0.8546	0.8546
	MAE	0.2758	0.2758	0.2758	0.2758	0.2758	0.2758	0.2758	0.2758	0.2758	0.2758	0.2758
	RMSE	0.3371	0.3371	0.3371	0.3371	0.3371	0.3371	0.3371	0.3371	0.3371	0.3371	0.3371
RBF	**Accuracy**	0.8730	**0.8780**	0.8772	0.8772	0.8772	0.8772	0.8772	0.8772	0.8772	0.8772	0.8772
	MAE	0.2096	0.1478	0.1453	0.1452	0.1452	0.1452	0.1452	0.1452	0.1452	0.1452	0.1452
	RMSE	0.3302	0.3010	0.2984	0.2982	0.2982	0.2982	0.2982	0.2982	0.2982	0.2982	0.2982
SVM	**Accuracy**	0.9348	0.9599	0.9683	0.9708	**0.9758**	0.9741	0.9741	0.9724	0.9716	0.9724	0.9724
	MAE	0.0652	0.0446	0.0336	0.0301	0.0263	0.0268	0.0265	0.0272	0.0283	0.0272	0.0272
	RMSE	0.2553	0.1789	0.1569	0.1537	0.1460	0.1485	0.1499	0.1542	0.1573	0.1572	0.1572

Table 6: Boosting with Classification algorithms using different number of iterations.

Bagging												
Classifier	**num iterations**	1	10	20	30	40	50	60	70	80	90	100
J48	**Accuracy**	0.9816	0.9833	**0.9850**	0.9858	0.9866	0.9866	0.9866	0.9866	0.9866	**0.9875**	0.9858
	MAE	0.0252	0.0271	0.0275	0.0274	0.0271	0.0272	0.0272	0.0273	0.0272	0.0272	0.0272
	RMSE	0.1301	0.1160	0.1135	0.1131	0.1119	0.1124	0.1122	0.1126	0.1123	0.1118	0.1117
MLP	**Accuracy**	0.9724	0.9858	0.9858	0.9883	0.9883	0.9875	0.9875	0.9866	0.9875	**0.9891**	0.9883
	MAE	0.0359	0.0392	0.0393	0.0393	0.0390	0.0389	0.0389	0.0393	0.0395	0.0395	0.0394
	RMSE	0.1485	0.1130	0.1115	0.1113	0.1109	0.1101	0.1101	0.1105	0.1103	0.1103	0.1100
Random forest	**Accuracy**	**0.9841**	0.9841	0.9841	0.9841	0.9841	0.9841	0.9841	0.9841	0.9841	0.9841	0.9841
	MAE	0.1216	0.1229	0.1230	0.1232	0.1230	0.1231	0.1233	0.1232	0.1232	0.1232	0.1232
	RMSE	0.1710	0.1664	0.1666	0.1667	0.1665	0.1665	0.1666	0.1665	0.1664	0.1664	0.1664
Naïve bayes	**Accuracy**	0.8446	**0.8613**	0.8613	0.8605	0.8580	0.8580	0.8563	0.8580	0.8580	0.8580	0.8571
	MAE	0.2756	0.2758	0.2768	0.2770	0.2771	0.2772	0.2772	0.2770	0.2770	0.2770	0.2770
	RMSE	0.3389	0.3371	0.3376	0.3376	0.3378	0.3379	0.3379	0.3378	0.3378	0.3378	0.3378
RBF	**Accuracy**	0.9791	0.9833	**0.9841**	0.9841	0.9841	0.9841	0.9841	0.9841	0.9841	0.9841	0.9841
	MAE	0.2115	0.2124	0.2132	0.2132	0.2132	0.2132	0.2138	0.2138	0.2138	0.2138	0.2138
	RMSE	0.3342	0.3033	0.3009	0.3009	0.3009	0.3009	0.3017	0.3017	0.3017	0.3017	0.3017
SVM	**Accuracy**	0.9348	0.8772	**0.9365**	0.9365	0.9365	0.7700	0.8000	0.8100	0.8100	0.7900	0.7800
	MAE	0.0652	0.0689	0.0695	0.0695	0.0695	0.0699	0.0701	0.0704	0.0707	0.0708	0.0709
	RMSE	0.2553	0.2289	0.2293	0.2283	0.2283	0.2289	0.2283	0.2287	0.2291	0.2294	0.2292

Table 7: Bagging with Classification algorithms using different number of iterations.

contains 12 attributes and only 121 instants, it presents only a small dataset comparing to the garment employee productivity dataset utilized by this study (15 attributes with 1197 instances). Furthermore, study [23] focused only on applying decision tree algorithms with ensemble algorithms, while our study applied six different ML algorithms including J48, RF, MLP, RBF and SVM combined with Bagging and Boosting ensembles. Additionally, by comparing our work with the rest of studies mentioned in the related work, to the best of our best knowledge, no one had followed the same approach in this field by combining different ML algorithms with ensemble learning (Bagging and Adaboost) using various number of iterations. Also, this study highlighted that the number of iterations on some algorithms made a serious change on accuracy such as MLP while other algorithms don't show any changes, which made an indicator that the number of iterations affect the results and made a great addition to our study.

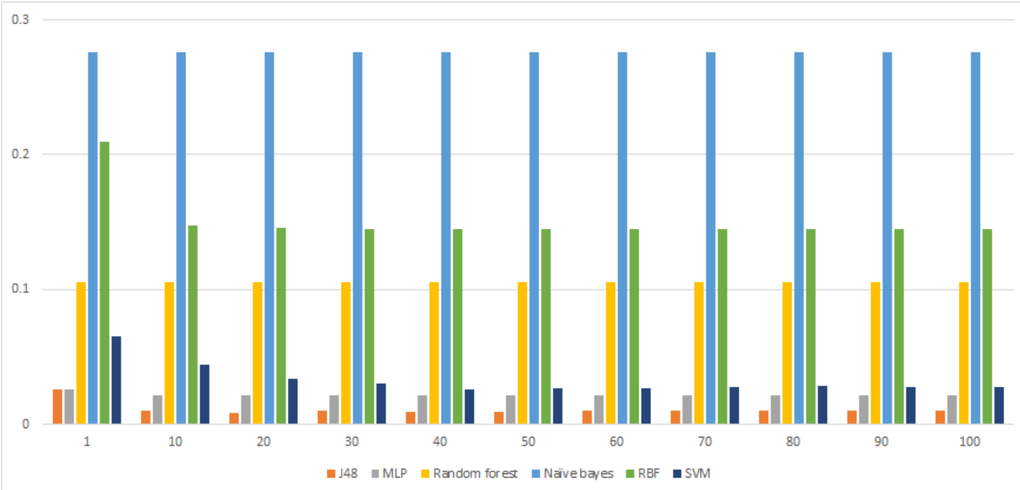

Figure 6: MAE of boosting with classification algorithms.

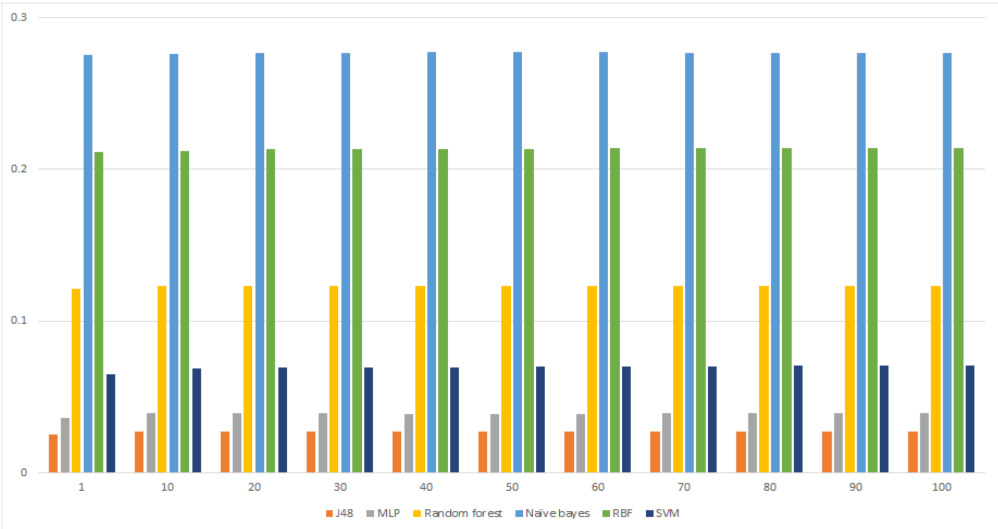

Figure 7: MAE of bagging with classification algorithms.

7 Conclusion

The employees' productivity plays an essential role in the manufacturing sector. Thus, many studies highlight the employees' productivity subject. This study focused on predicting garment employee productivity using different machine learning algorithms such as J48, RF, SVM, NB, and RBF with and without ensemble learning algorithms including, bagging and Adaboost. Our proposed approach succeeds in enhancing almost all classifiers' performance. J48 was the superior comparing with all other applied algorithms. The best results were obtained by J48 combined with Adaboost on 20^{th} iterations with 0.9916 accuracy, 0.0083 MAE and 0.0908 RSME. Consequently, J48 with Adaboost algorithm found to be the best for garment employee productivity prediction.

References

[1] Zhang, Yagang. 2010. New advances in machine learning (BoD–Books on Demand). https://books.google.com.qa/books?hl=en&lr=&id=2nQJEAAAQBAJ&oi=fnd&pg=PR7&dq=+machine+learning+&ots=fH16V9SEos&sig=lWb2Shc_S0awsEigIzvs0YuYXfg&redir_esc=y#v=onepage&q=machine%20learning&f=false

[2] Mahesh, Batta. 2019. Machine Learning Algorithms - A Review DOI: 10.21275/ART20203995

[3] Mona M. Jamjoom, Eatedal A. Alabdulkareem*, and Myriam Hadjouni , Faten K. Karim and Maha A. Qarh. 2021. 'Early Prediction for At-Risk Students in an Introductory

[4] Programming Course Based on Student Self-Efficacy', Informatica
DOI: https://doi.org/10.31449/inf.v45i6.3528
https://www.informatica.si/index.php/informatica/article/view/3528/1621

[5] Ozcift, Akin, and Arif Gulten. 2011. 'Classifier ensemble construction with rotation forest to improve medical diagnosis performance of machine learning algorithms', Computer methods and programs in biomedicine, 104: 443-51
DOI: https://doi.org/10.1016/j.cmpb.2011.03.018.

https://www.sciencedirect.com/science/article/pii/S0169260711000836
[6] Taichi, Joutou, and Yanai Keiji. 2009. "A food image recognition system with Multiple Kernel Learning." In 2009 16th IEEE International Conference on Image Processing (ICIP), 285-88. DOI: 10.1109/ICIP.2009.5413400
[7] Fang, Weiwei, Xin Li, Ping Zhou, Jingwen Yan, Dazhi Jiang, and Teng Zhou. 2021. 'Deep Learning Anti-Fraud Model for Internet Loan: Where We Are Going', IEEE Access, 9: 9777 DOI: 10.1109/ACCESS.2021.3051079
[8] Feng, Wei, Wenjiang Huang, and Jinchang Ren. 2018. 'Class imbalance ensemble learning based on the margin theory', Applied Sciences, 8: 815 DOI: https://doi.org/10.3390/app8050815
[9] Lemmens, Aurélie, and Christophe Croux. 2006. 'Bagging and boosting classification trees to predict churn', Journal of Marketing Research, 43: 276-86 DOI: https://doi.org/10.1509/jmkr.43.2.276
[10] Hearle, Chris. 2016. 'Skills, employment and productivity in the garments and construction sectors in Bangladesh and elsewhere', London: OPM. https://assets.publishing.service.gov.uk/media/5977616f40f0b649a7000022/Skills_productivity_and_employment.pdf
[11] Chowdhury, Nighat Afroz, Syed Mithun Ali, Zuhayer Mahtab, Towfique Rahman, Golam Kabir, and Sanjoy Kumar Paul. 2019. 'A structural model for investigating the driving and dependence power of supply chain risks in the readymade garment industry', Journal of Retailing and Consumer Services, 51: 102-13 DOI: https://doi.org/10.1016/j.jretconser.2019.05.024. https://www.sciencedirect.com/science/article/pii/S0969698918311822
[12] Hanaysha, Jalal. 2016. 'Testing the effects of employee empowerment, teamwork, and employee training on employee productivity in higher education sector', International Journal of Learning and Development, 6: 164-78 DOI: DO - 10.5296/ijld.v6i1.9200
[13] Harfoushi, Osama, and Ruba Obiedat. 2011. 'E-Training acceptance factors in business organizations', International Journal of Emerging Technologies in Learning (iJET), 6: 15-18 DOI: doi:10.3991/ijet.v6i2.1443
[14] Evans, W Randy, and Walter D Davis. 2015. 'High-performance work systems as an initiator of employee proactivity and flexible work processes', Organization Management Journal, 12: 64-74 DOI: https://doi.org/10.1080/15416518.2014.1001055
[15] Harfoushi, Osama, Ruba Obiedat, and Sahar Khasawneh. 2010. 'E-learning adoption inside Jordanian organizations from change management perspective', International Journal of Emerging Technologies in Learning (iJET), 5: 49-60 DOI: doi:10.3991/ijet.v5i2.1260
[16] Alam, Mohammad, Rosima Alias, and Mohammad Azim. 2018. 'Social Compliance Factors (SCF) Affecting Employee Productivity (EP): An Empirical Study on RMG Industry in Bangladesh', 10: 87-96. https://www.researchgate.net/publication/326733299
[17] Bhatia, Karan, Shikhar Arora, and Ravi Tomar. 2016. "Diagnosis of diabetic retinopathy using machine learning classification algorithm." In 2016 2nd international conference on next generation computing technologies (NGCT), 347-51. IEEE DOI: 10.1109/NGCT.2016.7877439.
[18] Kruppa, Jochen, Alexandra Schwarz, Gerhard Arminger, and Andreas Ziegler. 2013. 'Consumer credit risk: Individual probability estimates using machine learning', Expert systems with applications, 40: 5125-31 DOI: https://doi.org/10.1016/j.eswa.2013.03.019. https://www.sciencedirect.com/science/article/pii/S0957417413001693
[19] Balla, Imanuel, Sri Rahayu, and Jajang Jaya Purnama. 2021. 'GARMENT EMPLOYEE PRODUCTIVITY PREDICTION USING RANDOM FOREST', Jurnal Techno Nusa Mandiri, 18: 49-54 DOI: https://doi.org/10.33480/techno.v18i1.2210
[20] Attygalle, Dilhari, and Geethanadee Abhayawardana. 2021. 'Employee Productivity Modelling on a Work From Home Scenario During the Covid-19 Pandemic: A Case Study Using Classification Trees', Journal of Business and Management Sciences, 9: 92-100 DOI: 10.12691/jbms-9-3-1
[21] Ďurica, Marek, Jaroslav Frnda, and Lucia Svabova. 2019. 'Decision tree based model of business failure prediction for Polish companies', Oeconomia Copernicana, 10: 453-69 DOI: 10.24136/oc.2019.022
[22] Mahoto, Naeem, Rabia Iftikhar, Asadullah Shaikh, Yousef Asiri, Abdullah Alghamdi, and Khairan Rajab. 2021. 'An Intelligent Business Model for Product Price Prediction Using Machine Learning Approach', 30: 147-59 DOI: 10.32604/iasc.2021.018944
[23] Sorostinean, Radu, Arpad Gellert, and Bogdan-Constantin Pirvu. 2021. 'Assembly Assistance System with Decision Trees and Ensemble Learning', Sensors, 21: 3580 DOI: https://doi.org/10.3390/s21113580
[24] Saad, Hamza. 2020. 'Use Bagging Algorithm to Improve Prediction Accuracy for Evaluation of Worker Performances at a Production Company', arXiv preprint arXiv:2011.12343 DOI: 10.4172/2169-0316.1000257
[25] El Hassani, Ibtissam, Choumicha El Mazgualdi, and Tawfik Masrour. 2019. 'Artificial intelligence and machine learning to predict and improve efficiency in manufacturing industry', arXiv e-prints: arXiv: 1901.02256
[26] De Lucia, Caterina, Pasquale Pazienza, and Mark Bartlett. 2020. 'Does good ESG lead to better financial performances by firms? Machine learning and logistic regression models of public enterprises in Europe', Sustainability, 12: 5317 DOI: https://doi.org/10.3390/su12135317
[27] Ihya, Rachida, Abdelwahed Namir, Sanaa El Filali, Mohammed Ait Daoud, and Fatima Zahra Guerss. 2019. "J48 algorithms of machine learning for

predicting user's the acceptance of an E-orientation systems." In Proceedings of the 4th International Conference on Smart City Applications, 1-8. DOI: 10.1145/3368756.3368995

[28] Uma Mahesh, Janni, K. Naganjaneyulu, P. Likitha, and K. Aishwarya. 2021. Analysis of J48 Algorithm in Classification-Ebola Virus
DOI: 10.13140/RG.2.2.17135.76961

[29] Narayanan, Vivek, Ishan Arora, and Arjun Bhatia. 2013. "Fast and accurate sentiment classification using an enhanced Naive Bayes model." In International Conference on Intelligent Data Engineering and Automated Learning, 194-201. Springer DOI: 10.1007/978-3-642-41278-3_24

[30] Nazzal, Jamal, Ibrahim El-Emary, and Salam Najim. 2008. 'Multilayer Perceptron Neural Network (MLPs) For Analyzing the Properties of Jordan Oil Shale', World Applied Sciences Journal, 5.
http://www.idosi.org/wasj/wasj5(5)/5.pdf

[31] Delashmit, Walter H, and Michael T Manry. 2005. "Recent developments in multilayer perceptron neural networks." In Proceedings of the seventh Annual Memphis Area Engineering and Science Conference, MAESC.
https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.318.4243&rep=rep1&type=pdf

[32] Leung, Henry, Titus Lo, and Sichun Wang. 2001. 'Prediction of noisy chaotic time series using an optimal radial basis function neural network', IEEE Transactions on Neural Networks, 12: 1163-72 DOI: 10.1109/72.950144.

[33] Cao, Wangcheng. 2019. 'Application of the Support Vector Machine Algorithm based Gesture

[34] Recognition in Human-computer Interaction', informatica Informatica 43 (2019) 123–127 123 DOI: https://doi.org/10.31449/inf.v43i1.2602

[35] Yaman, Emine, and Abdulhamit Subasi. 2019. 'Comparison of bagging and boosting ensemble machine learning methods for automated EMG signal classification', BioMed research international, 2019 DOI: https://doi.org/10.1155/2019/9152506

[36] Bühlmann, Peter. 2012. 'Bagging, Boosting and Ensemble Methods', Handbook of Computational Statistics DOI: 10.1007/978-3-642-21551-3_33

[37] Imran, Abdullah Al, Md Shamsur Rahim, and Tanvir Ahmed. 2021. 'Mining the Productivity Data of Garment Industry', International Journal of Business Intelligence and Data Mining, 1 DOI: 10.1504/IJBIDM.2021.10028084

[38] Vujović, Željko Đ. 'Classification Model Evaluation Metrics'

[39] Chai, T., and R. R. Draxler. 2014. 'Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature', Geosci. Model Dev., 7: 1247-50 DOI: 10.5194/gmd-7-1247-2014.
https://gmd.copernicus.org/articles/7/1247/2014/