SHM PUBLISHER
SURYA HIJAU MANFAAT

Journal of Information System Exploration and Research

https://shmpublisher.com/index.php/joiser

p-ISSN 2964-1160 | e-ISSN 2963-6361

Techniques of Applied Machine Learning Being Utilized for the Purpose of Selecting and Placing Human Resources within the Public Sector

Panagiota Pampoukts[1*] , Spyridon Avdimiotis[2], Manolis Maragoudakis[3], Markos Avlonitis[4], Nikita Samantha[5], Praveen Hoogar[6], George Ruhago[7]

[1,3]*Department of Informatics, Ionian Unversity, Corfu, Greece*
[2,4,5]*International Hellenic University, Thermi, Thessaloniki, Greece*
[6]*Department of Anthropology, University of Karnatak: Dharwad, Karnatak, IN*
[7]*Department of Public Health and Social Sciences, University of Dar es Salaam, Dar es Saalam, TZ*

DOI: https://doi.org/10.52465/joiser.v1i1.91
Received 08 0ctober 2022; Accepted 07 December 2022; Available online 09 December 2022

Article Info

Keywords:
Selection;
Positioning;
Machine learning;
Assessment algorithm;
Classification

Abstract

In strategic human resource management, one of the most critical issues to focus on is the correct selection and placement of people. Within the confines of this framework, the reason for the study that was conducted was to explore the machine learning approaches that proved to be the most effective in assisting with the recruitment of personnel and the assessment of their positions. To accomplish this goal, a in a series of tests involving workers in the public sector, categorization algorithms were used. The purpose of these tests was to determine which employees would be the ideal fit in which workstations and to determine how workers should be distributed. For supporting the decision support system, an algorithm model was created. Used in the process of recruiting and evaluating potential workers based on the results of the tests that were given. The most important results of this study support the idea that using the People's Evaluation for Recruitment and Promotion Algorithm Model (EERPAM) would make hiring and promoting people in a company fairer.

1. Introduction

According to Becker and Huselid [1], the theory of strategic human resources (HR) management is predicated on the notion that human capital is a strategic asset that produces competitive advantages. Because of this, all HR procedures that are related in any way need to be established on defensible and accurate decision-making mechanisms to optimize the workers' potential for innovation and creativity.

[*] ***Corresponding Author:***

Panagiota Pampouktsi,
Department of Informatics,
Ionian University,
Corfu, Greece.
Email: ppthessaloniki@gmail.com

In this model, it is necessary to have an accurate and up-to-date overview of the requisite qualifications to correctly pick and distribute individuals. This is a requirement for the right selection of personnel. This is done in order to create long-term advantages for businesses [2], [3]. The true advantage that long-term successful firms possess is their ability to put the appropriate individuals in the appropriate roles [4]. The technology of computer science may be an instrumental strategic tool for efficient management. A mechanism for tracking resources, often known as an HRIS, can collect, store, manage, process, examine, retrieve, and disseminate all the data and information on an organization's human resources that was calculated in the processes that came before it [5]. In addition to technology and software, the Human Resources Information System (HRIS) is made up of employees, forms, rules, and procedures, but data is what matters most. The term "artificial intelligence," or AI, refers to the capacity of computers and other electronic devices to correctly interpret information received from the outside world [6], to gain knowledge from that information and to apply that knowledge to accomplish predetermined objectives and chores in a flexible and adaptable manner. Adaption [6], [7] conducted a poll that was connected to this topic and discovered that 78% of managers say they will believe AI recommendations in deciding.

It is extremely important for the strategic management of human resources. To have the capacity to accurately forecast the optimal way in which human resources may be matched to suitable jobs. This piece of study aims to foresee the right selection and placement of human resources in the public sector using machine-learning algorithms. That is a very new subject of research. The creation of a framework that is based on machine learning algorithms is the primary contribution made by this research. This framework will give a dependable tool for recruitment selection as well as assessment for optimal placement and promotion of personnel. This application will assist departments that deal with human resources in making choices that are accurate and impartial about the distribution of personnel.

The following is the structure of the paper: Studies that employed machine learning for selecting Section 2 of this text analyzes and discusses human resources. In Section 3, we provide the research approach that was used for the experiment. Many techniques for machine learning are described in Section 4 contrasted with one another, and measurements of model performance are discussed, with the goal of determining which approach yields the optimal results in terms of matching workers with the suitable roles [8]. In the fifth section, complete findings and discussion are shown, and the model that achieves the highest level of overall performance is suggested. Section 6 offers conclusions and suggestions for additional research.

2. Method

The study's initial objective was to find the best algorithm to help with selection and promotion decisions based on employees' performance reviews. A primary quantitative survey was performed with the help of a two-part questionnaire to achieve this end. First, we set out to define what qualities an Executive Director, Department Head, Entry-Level Employee (a role that anyone in the company could theoretically take on), and Mid-Level Employee (a role that requires some level of experience) should have (senior employees). The second stage was an attempt to catalog the skills and credentials of people currently employed. When developing the questionnaire, we followed the same guidelines as the Supreme Staff Selection Board, the Greek government agency in charge of hiring new public servants. In order to validate the questionnaire's ability to measure the targeted variables, researchers ran a series of tests on it [2], [9]. provided a framework for informing the research instrument's design strategy, which prioritized simplicity, accuracy, clarity, feasibility, and construct coherence. Validation in this instance was predicated on a plan that included a wide range of statistical tools and approaches [10]:

1) The Cronbach's alpha coefficient (which reached a value higher than 0.7),
2) Using techniques like EFA (Exploratory Factor Analysis) and
3) Confirmation for Factor Analysis (CFA).

2.1. Collecting Information and Having Qualified Staff

2.1.1. Information Gathering and Analysis

The dataset was compiled from responses to a survey administered to public sector workers in Greece and includes information about their education and experience. The employees in the sample came from a wide range of educational backgrounds and professional fields, including agriculture engineers, mechanical engineers, computer engineers, administration officers, etc. Information in question, in particular:

- Education at the university or technical institute degree level. Degree value: the grade received in the basic degree title.
- Master of Science Degree Holders: MSc Diploma, Relevant to Employment.
- PhD holders: PhD diploma, applicable to position.
- Have completed the National Academy of Public Administration program.
- The number of seminars attended in the past decade.
- Certification in Computer Competence.
- The sum of all an employee's years of professional experience.
- Extensive senior-level experience.
- Interview performance: the grade candidates received during the selection process.
- Committee membership: the percentage of employees who were active members of at least one committee.
- Number of published research articles by staff members; this relates to the preceding point.
- Where exactly do workers fall on the organizational chart? (Director, Head of Department, new or senior employee).

2.1.2. Simulation of Employee's Qualifications to Job Specification

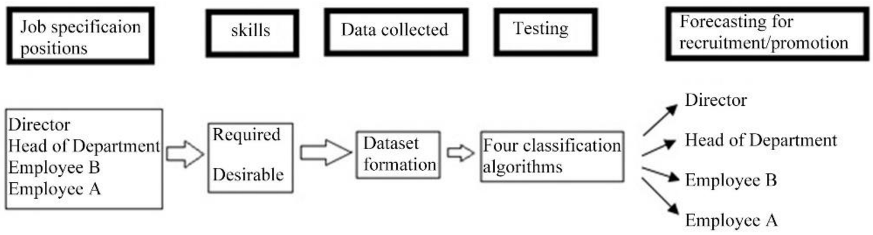

Figure 1. The conceptual experimental framework of the performed tests

The information that was gathered pertains to the primary classifications of jobs in the public sector, namely those of Manager, Division Chief, Staff Member A, and Staff Member B. A candidate may enter a corporation and take over the entry-level, most fundamental role known as Employee A. The positions that follow Employee A are Position B Employee, Department Head, and Director, respectively. According to the job criteria, the following are the abilities that are taken into consideration for these roles, whether they be needed or desirable:

1) Qualifications for the position of director include a bachelor's degree, the value of that degree, a master's degree, a doctoral degree (preferred), a national school for public administration (preferred), the number of seminars attended, fluency in European languages, an average evaluation score, years of experience, years in a position of authority, a score from an interview, evidence of group work and committee participation, a sample of research, and proof of computer literacy. 2) Manager: Bachelor's degree, total credits earned, Master's degree, Doctorate degree (preferred), National School for Public Administration (preferred), number of years of experience in the field.
2) For the position of Head of Department: Bachelor's degree, value of master's degree, PhD degree, national school for public administration, number of seminars, familiarity with European languages, mean score on performance reviews, years of experience, years in a position of authority (preferred), interview score, participation in teamwork.
3) The education level, degree value, number of master's and doctoral degrees held by Employee A (both of which are highly desirable), attendance at the national school for public administration (also highly desirable), number of seminars attended, proficiency in European languages (also highly desirable), years of experience (also highly desirable), and computer certification. 4) Qualifications for Staff Member B include a bachelor's degree, a Degree Value, a master's degree (preferred), and a Doctorate Degree (desirable).

4) The education level of Employee B, as well as the degree value, the master's degree (preferred), the doctorate (preferred), the national school for public administration (preferred), the number of seminars, the knowledge of European languages, the years of experience (preferred), and the computer certification. 5) The educational history of Employee A, including the greatest level of education that they have obtained, the value of their degrees, and whether they have attended any master's programs, doctorate programs, national schools for public administration, or seminars.

2.2. Preparation of a Data Base

2.2.1 Data Import

The database was constructed using the data imported from a spreadsheet created in Microsoft Excel. After conducting a first cycle of assessment in which the qualifications of workers were compared to the role specification, we selected 1010 cases that had a good match for them to serve as a genuine training set. The characteristics (traits) of the datasets were created by the needed and desired qualifications that constituted the necessary and preferred traits that defined employee profiles.

2.2.2 Questionnaire Evaluation

Questionnaire Evaluation as well as the Assignment of Weights to Selected Criteria In the first section of the survey, responses were solicited from employees about several different categories. The weighting factors were determined based on the means of their responses. We have received 196 complete responses, which is enough for conducting trustworthy statistical operations. The relative importance of the coefficients varied depending on a place (Director, Head of department and employees' positions). After that, each coefficient was given a standard value [11].

2.2.3 Total Score Calculation

A linear model was used to determine the total score for each employee. In this concept, workers are evaluated based on a combination of their individual skills and the role-specific weighting component. Then, all the data was exported to a CSV (Comma Separated Values) file so that it could be processed later.

2.3. The Actual Application of Machine Learning

For forecasting, the degree of compatibility between the capabilities of the workers and the job profile supervised machine learning schemata were used. Classifiers were trained on workers' data to provide accurate predictions about the degree to which people and positions are a good fit for one another [12]. The worker's information was encoded as a vector of feature values. The characteristics, in addition to the overall score that was accomplished by each worker, constituted the inputs. The kind of position, which may be any one of the four primary positions, was chosen as the output [12].

2.4. Classification of Machine Learning Algorithm

We narrowed the field of potential classification algorithms by selecting a subset of those that were accessible from the larger pool based on the following criteria:

1) Performance in forecasting, which refers to the capacity the ability of the system to precisely identify the position given the input parameters; 2) Models explain ability: this refers to the capacity of a human domain expert (often one working in the field of human resources) to comprehend the factors that contributed to the system's decision to provide a certain forecast rather than another [13]. 3) Using the criteria, this research identified four classification algorithms as the most effective among many other possible algorithms with different theoretical backgrounds, such as statistics, neurobiology, kernel functions, etc. Researchers in the field of machine learning also widely recognize these algorithms [14].

The J48 procedure induces decision trees. C4.5 was its original designation [15]. With the data provided, the C4.5 algorithm constructs a decision tree. Gained by analyzing the properties of the dataset that is currently available for training. Every time the process is repeated, the property or characteristics are selected whose values best distinguish the training examples based on their class name (training cycle) [16]. After there are no more qualities to investigate by the algorithm or when all the training instances have been segregated to a suitable level, it will terminate. In addition to this, J48 contains two distinct approaches to tree trimming: a) Subtree replacement is an operation that, when a particular subtree does not contribute to improved classification accuracy [17], makes a leaf the replacement for a node in a decision tree that corresponds to that node. The process of pruning begins at the uppermost leaves of the newly formed tree and works its way downward toward the tree's base. b) Subtree raising is the process of moving a node closer to the root of the tree, during which it may replace other nodes [18]. The vast majority of the time, the decision tree models are not significantly impacted by this kind of pruning [19].

1) Replaces a node in a decision tree with a leaf r. During each iteration of the algorithm, a decision tree is generated from a subset of the characteristics that are chosen at random [20]. Multiple decision trees are traversed using the input vector as a guide. The total number of iterations has already been decided upon. The overall error in categorization is equal to the meaning of the errors across all the iterations. To classify a new item from an input vector, multiple classification trees are created, and the output from each tree is used to create a final classification [21]. This allows Random Forest to classify a new object more accurately than the decision tree classifier [22]. The forest will choose the category that received the most positive comments and votes [23]. They are often able to attain great levels of performance. Additionally, despite the fact that the final trained model is capable of learning intricate connections, training results in easily understandable decision thresholds [15].
2) The Naive Bayes algorithm is a conditionally independent probabilistic classifier [24]. That is, the algorithm takes for granted that a feature's presence within the dataset does not depend on the presence of any other features outside of that feature's class value [25]. Although this assumption lacks empirical backing, it has been proven useful in solving a variety of classification problems. The ease with which it can be written, the speed with which it can produce predictions, and the fact that it can simplify challenges associated with predictive modeling all contribute to Naive Bayes' strong record of accomplishment of success [26]. This method just requires a little quantity of training data to establish the classification parameters that are essential. Because the variables in question are assumed independent of one another, it is not necessary to estimate the full covariance matrix; rather, one should simply determine the variable differentiation for each category [27].
3) Training support vector machines with Sequential Minimal Optimization (SMO) is an advanced method. A huge quadratic programming (QP) optimization issue is broken down into several smaller problems by it [28]. SMO works its way through each stage by solving the lowest possible optimization challenge [29]. Instead of running a complete library QP method, the inner loop of the algorithm is stated in a minimal amount of C code. This makes the process more efficient. Even if more optimization sub-problems are addressed, the overall QP issue is solved in a very short amount of time since each sub-problem can be handled in such a short amount of time (Platt, 1998). For training a support vector classifier, SMO employs the sequential minimum optimization approach, making use of polynomial or Gaussian kernels [30].

It is important to note that many alternative algorithms, including MLP, KStar, and AdaBoost, were tested on our dataset to determine their accuracy and ability to forecast outcomes, however the results were insufficient.

A model that uses machine learning requires inputs before it can provide predictions. During our investigation, data on the credentials of 1010 workers were submitted. Our dataset included forty-eight (48) Employees A. As of this writing, there are 35 Directors and 721 Employees B. Likewise, there were 206 Department Heads working there. Figure 2 illustrates the proportion of each place in the training dataset that represents the class balance.

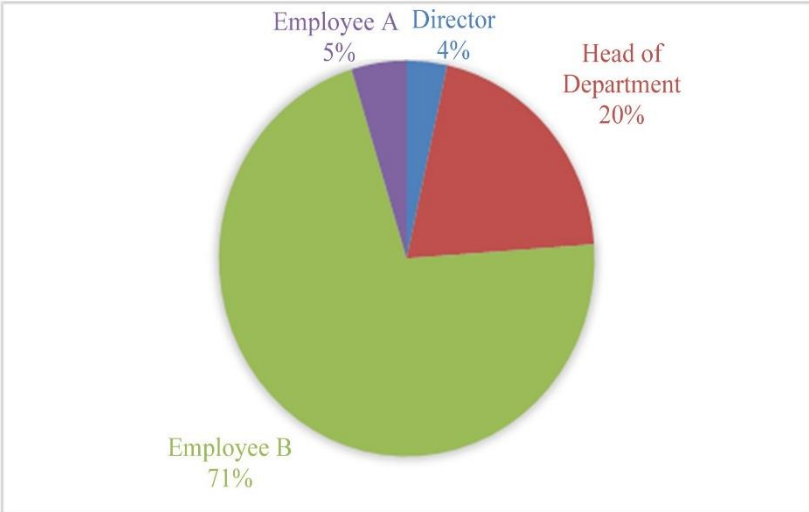

Figure 2. Chart showing the frequency of occurrence of class labels

Attribute categorization was built on decision trees. Bayesian algorithms (Naive Bayes) and support vector machines to create a new framework for human resource selection, placement, and management show in Figure 3.

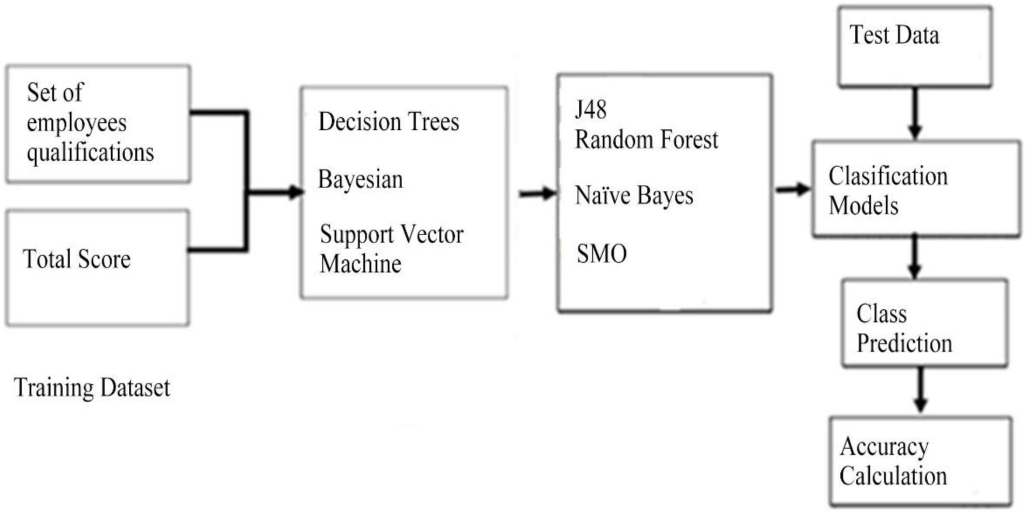

Figure 3. Analyzing the patterns that emerged from our experiment

The goal of a classifier is to develop a model for predicting the value of each class in a dataset it has never seen before (the test set), based on the available training data. To evaluate a model's capacity for generalization, it is necessary to calculate the typical level of danger posed by the whole collection of data points [31]. Since this cannot be done in real-world applications, we must count the risk using a test set to get an accurate estimate [32]. Selection of models based on their performance on a single test set after having been trained.

This happens when the model's behavior changes because of training in an attempt to precisely forecast all of the instances that it was trained on, but it fails when it is given new examples to learn from. This is an example of not adequately avoiding the risk of overfitting, which occurs when the model's behavior changes because of training. K-fold cross-validation, often known as CV, is one method that may be used in line with the recommendations for best practices in order to get a more accurate evaluation of the empirical risk [33].

Within the context of our primary experiments, we executed the required meta-learning modules in order to improve the performance of the algorithms. Because of this, we were forced to choose either the values that are preset as the default or the values that are automatically presented as the parameter values. The term "optimization" refers to the process of improving the algorithm classification's accuracy, estimated errors (particularly root mean square error), and overall performance [34]. It is important to point out; the default parameters of each method trumped any other permutation under some conditions, though. This is to be anticipated, as these parameters have been honed through experimentation with a huge collection of datasets provided by WEKA contributors. WEKA's Cross-Validation parameters selection meta-learner just marginally outperformed the competition for the J48 technique.

Fails to adequately guard against overfitting, which occurs when a trained model improperly modifies its output to precisely forecast all of the instances that it was trained on, but it fails when it is given new examples to learn. In accordance with the recommendations for best practices, k-fold cross-validation (CV) may be used to get a more precise assessment of the empirical risk [33].

3. Results and Discussion

The 10-fold cross-validation training strategy was the one that we used for the initial set of tests that we did. With the use of this technique, we won't have to go through the complete dataset each time; rather, we may train and test separately on n different subsets of the data that is readily accessible (each time the test set is one of the parts and the training set consists of the rest of the data). The average test risk may be a good approximation of the real generalization capability of the tested method, particularly when the complete cross-validation procedure is repeated (each time with a new data split), and n is adequately selected [35]. In addition, one may use the average test risk as a proxy for the tested algorithm's genuine generalization capabilities [33]. As part of our experiments, we randomly split the main sample into 10 smaller samples. Every 10 subsamples, only one was kept aside for model validation while the other 9 were used for model training [36]. Next, cross-validation was performed 10 times, with each of the 10 subsamples being used just once per round. The aggregated data from all 10 investigations was averaged. For the most part, we sent all of the training data into the following learning algorithms and let them figure out how to create a mapping between the inputs and the output class label that minimizes the prediction error. This allowed us to develop a model that was more accurate than our previous model.

3.1. Decision Trees

The primary applications of decision trees in our work are categorization and scenario planning [37]. They are shown by the rules IF-THEN-ELSE, beginning at the trunk of the tree and progressing all the way up to the branches and leaves. The aspects of the issue that have been analyzed are represented as nodes along the tree. Logical conditions are used to characterize individual tree aspects, which are referred to as nodes [38]. The names of the features are what give the nodes of a tree their identity, while the alternative values that a feature might have been what give the edges their names, and the various classes are what give the leaves their names.

3.1.1 The Classification Outcomes of the J48 Algorithm

The essential optimal parameters that are utilized for the J48 method that requires subtrees and pruned branches are shown in Table 1. The nature of our data led us to conclude that these parameters fulfilled our expectations; therefore, we may consider them satisfactory.

Results from the J48 classification are shown in Figure 4, and they include the accuracy and recall of training (learning). This algorithm led to the formation of four classes that correspond to the four primary categories of employment rolls. The best findings that were recovered were for the job of Director, which reached 1.0 for recall. Head of Department and Employee B both had excellent calculations (above 0.95) for the post.

The recall performance of Employee A was good, coming in at 0.85. The accuracy of the genuine predicted values was rather good, and in every instance, it was at least 0.94. According to the output of the J48 classifier, there were 979 instances that were correctly classified and 31 instances that were incorrectly classified. The incorrectly classified instances concerned employees at the there was a gradational void between B-level workers and A-level workers and the department head.

The RMSE and MAE both measured in at 0.1173 (root mean squared error) and 0.0244 (mean absolute error). For the purposes of prediction, having a high recall is very helpful. According to our investigations, it was extremely high for the positions of Management Team: Director, Department Head, and Employee B, but it was adequate for Employee A. This method is believed to be particularly beneficial for the purposes of classification and prediction since it has a high level of accuracy. A human specialist has deemed this algorithm appropriate for use in the interpretation of the forecasting process.

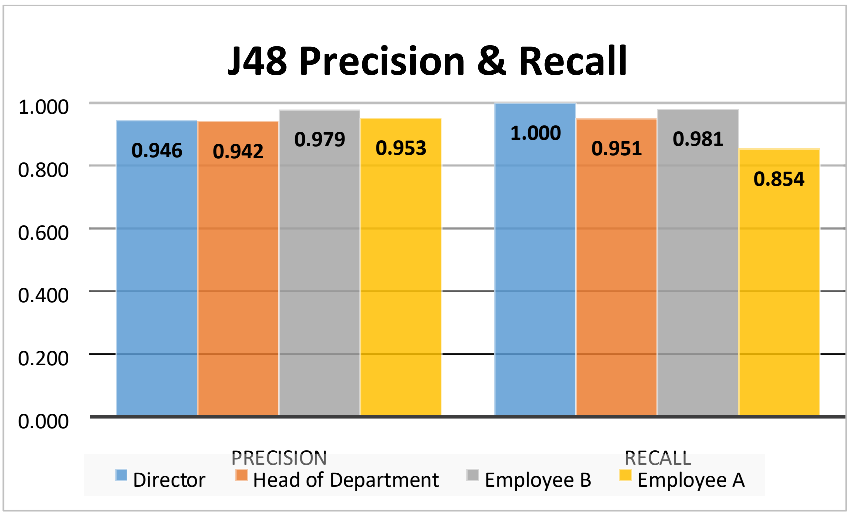

Figure 4. Precision and recall for J48 classification algorithm

Table 1. J48 algorithm parameter values

Binary Splits	**False**
Confidence Factor	0.1
Minimum instances per leaf	3
Reduced Error Pruning	False
Subtree Raising	True
Pruned True	True
Laplace Smoothing	False

Figure 5 shows the J48 tree, which indicates that overall score and years of experience in authority are the two most essential factors in determining an applicant's suitability for a job. The system prioritized the overall score above all other factors. The algorithm predicted that a total score of >0.6825 would be able to differentiate between the Director position and the Employee A position (0.2474), While Employee B's prior management experience helps set him apart from the department head, so does Employee A's. In cases where the number of years in authority is greater than zero, the total score proved to be a reliable criterion, while in cases where the number of years in authority is zero, the total score proved to be unsafe. When the number of authority years is greater than two, the total score and authority years together may be able to tell the two groups apart (positions). Overall performance evaluations are crucial in determining who gets the job. The developed model achieved a 96.93% accuracy rate.

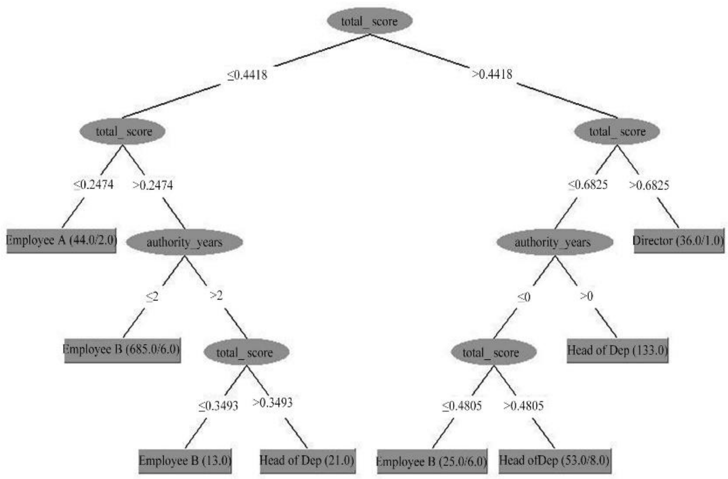

Figure 5. The J48 classifier made the tree

3.1.2 Algorithm for Random Classification Outcomes

The Random Forest algorithm's baseline settings are tabulated in Table 2 (the WEKA defaults were deemed adequate in this case).

Table 2. Random Forest parameter values

Deepest Possible	**Unlimited**
Extent of characteristics	0
Quantity of trees to be produced	100
Seed	1

The training results for Random Forest classification (precision and recall) are shown in Figure 6. The recall was also high for all classes (positions), but only for the Head of department and Employee B, while the other two classes had recalls of about 0.80. Due to low recall values, its predictive power was diminished for two classes. Because of its effective differentiation of roles, this classifier successfully classified 960 examples. Fifty cases were misclassified, including seven directors, eighteen department heads, sixteen B employees, and nine employees. Accuracy of the built model was 95.04%, with a mean absolute error of 0.0559 and a root mean squared error of 0.1465.

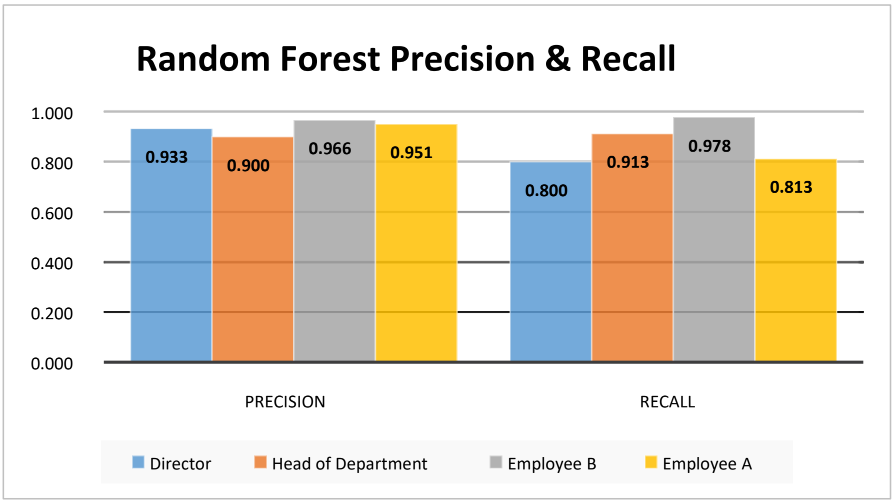

Figure 6. Visual representation of the random forest classification algorithm's recall and accuracy

3.2. Applications of Bayesian Algorithms

Classification using Bayes' theorem is a time-tested and popular machine learning strategy [39]. The theory behind it is to determine the likelihood of an occurrence by examining the likelihood of a previous occurrence. The purpose of a Bayesian classifier is not to calculate the probability of each conceivable label, but rather to select the class that is most likely to be correct [40].

3.2.1. Classification Outcomes of the Naïve Bayes Algorithm

The Nave Bayes algorithm's default settings are shown in Table 3.

Table 3. Values of the Nave Bayes Parameter

Parameter	Action values
Implement a Kernel Estimator	No
Make use of discretization in the presence of a supervisor	No
Optimal thresholding	No

Training results for a selected NB algorithm are shown for classification (precision and recall) in Figure 7. Almost all classes (positions) have a precision greater than 0.85, with the exception of Employee a. (0.712). Employee B was the most productive, bringing in the most money (0.982). Director's recall was 1.0, which is very good. Employee A was the only one who scored below a recall value of 0.9. (0.875). Because of its excellent recall values, this algorithm also has satisfactory prediction capability. With 941 cases accurately categorized, this classification method successfully differentiated between positions. The mean absolute error (MAE) was 0.0401 and the root mean squared error (RMSE) was 0.1692, while accuracy of classification estimations reached 92.9960%.

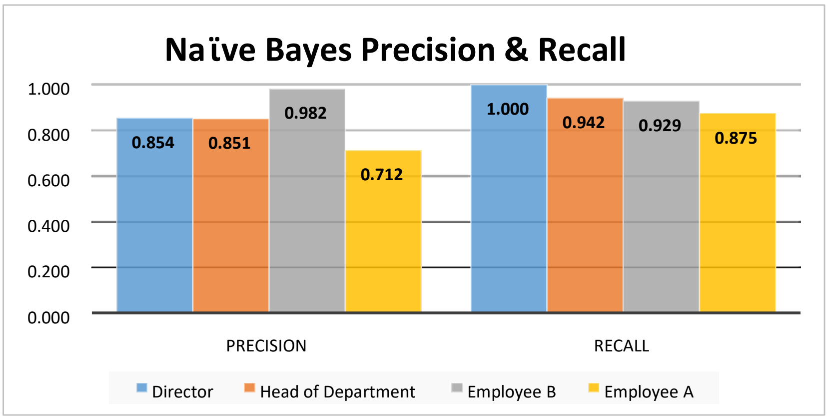

Figure 7. Naive Bayes (NB) classification method accuracy and recall

3.3. SVD, or Support Vector Machines

Boser first proposed supporting vector machines (SVM) for addressing classification and regression challenges [41]. Classification issues may give rise to models with variable decision margins, depending on the parameters that are used in the analysis. It is possible to have a margin that is either linear or non-linear [42]. According to Scholkopf a linear support vector machine, also known as an SVM, is a hyperplane that divides a training set into two groups [43], [44]. The objective of this type of SVM is to increase the margin in the region of characteristics, which is measured by the distance of the hyperplane from the nearest positive or negative examples [45]. Because margins optimization deals with the appropriate location of the margins, complexity in the margins does not influence generalization. Support vector machines lower the empirical risk for problems involving classification and regression [33].

3.3.1. The Results of Classification Carried Out Using the SMO Algorithm

The main parameters of the SMO algorithm are shown in Table 4, along with their default defaults (default values of WEKA). Figure 8 presents the classification training results obtained with the SMO algorithm. These findings include recall and precision. Again, accuracy was rather good (varying from

0.82 for Employee A to 0.93 for the Director and Employee B), and recall was more than 0.70 for all employees (it was low reaching only 0.48). The performance that SMO turned in about Employee A's capacity for remembering and prediction was below average. 934 occurrences were effectively categorized, suggesting that the categorization procedure was successful. The accuracy of the created model was 92.47%, with a root mean squared error of 0.3221 and a mean absolute error of 0.2565.

Table 4. SMO parameter values

Parameters	Value
Complexity parameter	1.0
Round-off error	1.0E−12
Filter Type	Normalize training data
Kernel	Poly Kernel (exp. 1.0)
Random seed for cross validation	1
Tolerance parameter	0.001

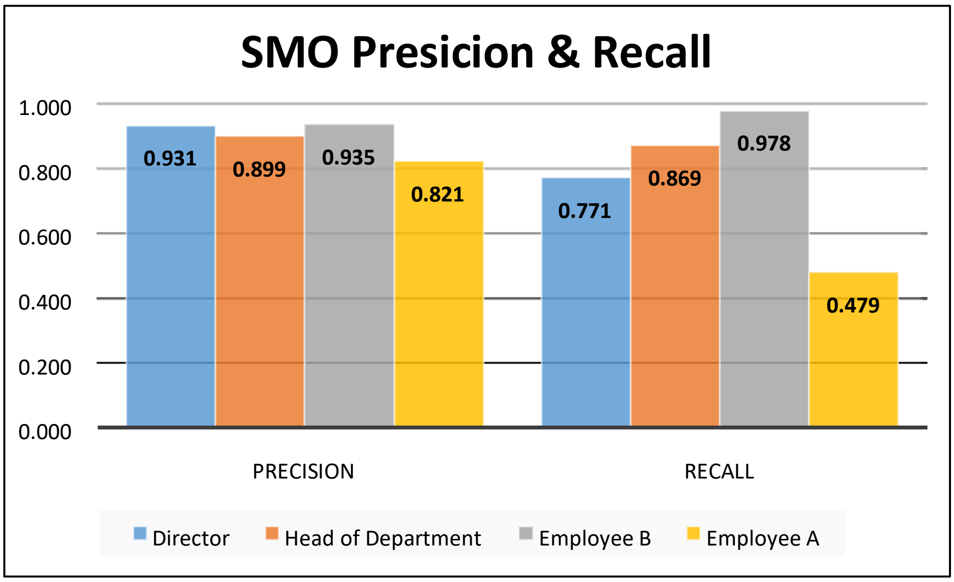

Figure 8. Precision and Recall for SMO classification algorithm

3.4. Summary of Experimental Findings and Analysis

This study makes use of supervised machine learning techniques, such as decision trees, support vector machines, and Bayesian algorithms, to look into the future and predict how human resources should be distributed in the best possible way. A general-purpose ability model-learning framework was used to ascertain which place was the most suitable for each method. This framework combined both observable factors and hidden patterns ingrained in the employees' official qualifications to assist learning from many perspectives. No matter the algorithm used to categorize data, there were only four groups that adequately reflected the whole range of possible roles in an organization with a cross-validation sample size of 10, the mean accuracy of classification was 97.03% for J48, 95.04% for Random Forest, 93.16% for Nave Bayes, and 92.41% for SMO. It was determined that algorithm J48 had the least amount of error while estimating errors. The accuracy measures of MAE = 0.0244 and RMSE = 0.1173 were obtained. Given these results, the J48 algorithm has the potential to outperform competing prediction methods. SMO was found to have the greatest MAE and RMSE (0.2561 and 0.3221 respectively). While SVM has been shown to beat base classifiers in a few earlier studies, this was not the case for us. The longer training times and reduced performance after trying out various kernels and regularization parameters both pointed to the fact that an excessive amount of the training data was converted into support vectors. The best classifier for a given job is itself task-dependent, and the no free lunch theorem states that no classifier system is inherently superior to the others.

Due to its consolidation of the previously mentioned recall and accuracy metrics, the F-Measure gives a comprehensive evaluation of the models. Its F-Measure is the harmonic mean of its recall and accuracy measures [46]. Results from the F-Measure classification using the primary machine learning methods are shown in Figure 9. The Algorithm, with the Random Forest and Naive Bayes methods coming in a close second and third, respectively, maximizes accuracy and recall. However, SMO has the lowest recall values of all the algorithms and so gives the worst outcomes. Based on our findings, J48 provides classification that is more precise and, by extension, more precise calculations of appropriate job positions. Random Forest comes in second. Comparing the parameters of the various rest algorithms, I found them all to be suitable (but with lower pairing accuracy).

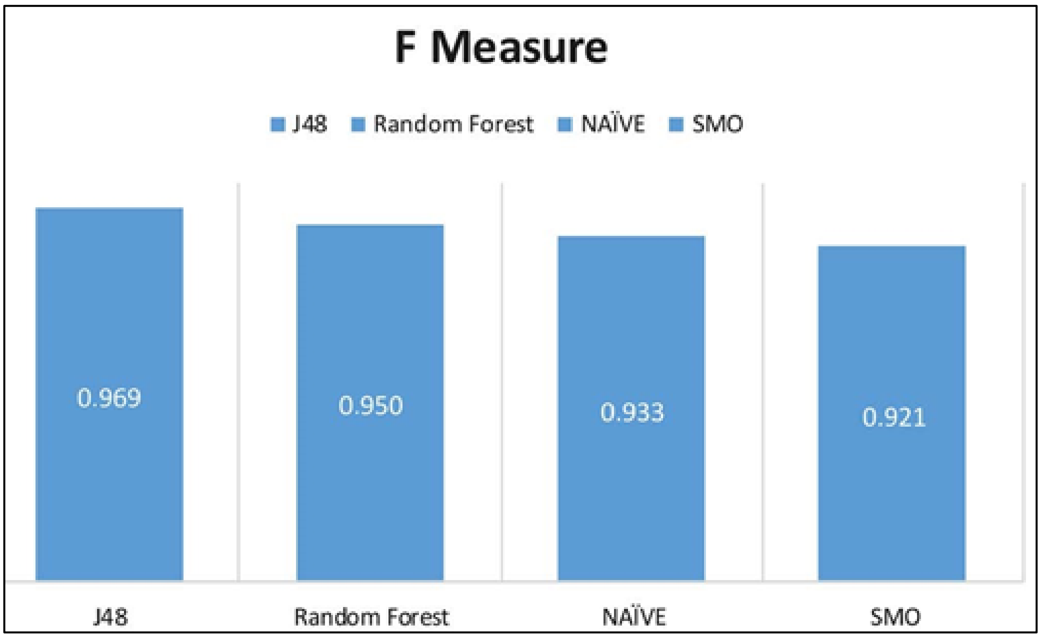

Figure 9. The F-measure for the J48, Random Forest, Naive Bayes, and SMO algorithms

A variety of different approaches, as was said, were also investigated. We put a lot of effort into developing neural networks, particularly the Multilayer Perceptron (MLP). Although the other algorithms performed better, the F-Measure value lagged at 0.8822, making it the worst performer. This was mostly because of the especially low recall values for the instance of the Head of Department post (0.625), which was a major problem. When compared to other algorithms, its mean recall of 0.844 indicates that it has relatively weak prediction accuracy. In comparison to Naive Bayes and SMO, the accuracy was satisfactory (around 0.93), but it lagged behind J48 and was on par with Random Forest. However, we did not provide extensive details about Multilayer Perceptron's performance because it lagged tree-based algorithms in our experiment in terms of predictive ability. Another difficulty is that MLP makes predictions in a "black box" fashion, meaning that after making a forecast, no information is offered to a human expert to help them determine the underlying reasons that were used to make the prediction.

In similar experiments, Using decision trees and the CHAID classifier algorithm developed for machine learning, Chien and Chen attempted to establish some principles for the objective selection of the most qualified candidates for employment in the high-technology sector [47]. They confidently (63%-96%) classified five groups of jobs (job or work description) according to shared characteristics and requirements. Successful classification of HR strategies for selecting new employees using the Naive Bayes algorithm is the subject of another recent article [48]. The work seems to have rather low levels of effectiveness and precision. Using HR data and job titles, Varshney et al. (2014), working with data from IBM salespeople, reported accuracy of roughly 80%. (The main attributes). Depending on the algorithm and the parameters, [49] found an accuracy anywhere from 60% to 80%. Up until now, every single study has been funded by the commercial sector.

Our research indicates an individual's total score, which accounts for all his or her qualifications. In combination with the requirements of the job description, is the best predictor of whether they will be successful in either being selected for a position of authority (higher) or being recruited. Total employee score in relation to prior experience in such positions is the most essential attribute for promotion considerations [50]. Is the sum score based on employees' qualifications related to the current work used for selection and recruitment? To conserve materials and aid the decision-support mechanism, we propose a method of quickly and accurately assigning personnel to specific positions.

Our experimental results showed that all the classifiers we tested had a precision of over 0.9, but that J48 had the highest F-measure (which combines precision and good pairing) of the bunch at over 0.90. For J48, we saw a similarly high rate of accuracy (97%) in our classifications. J48 also significantly reduces the time required to construct the model (only 0.01 seconds). As a result, we recommend using the J48 method to construct the Employees' Evaluation for Recruitment and Promotion Algorithm Model (EERPAM), which is the recommended solution. J48 exhibited very promising categorization findings and great prediction capacity about the goals of personnel selection and reallocation. We successfully categorized the 1010 workers into four classes (job roles) using high precision, pairing, and accuracy, and we made predictions for future people selection based on those forecasts. When compared to prior studies on the use of machine learning for personnel selection, ours proved to be more accurate and predictive[51][49]. Our proposed learning-based methodology has real-world application since it can forecast the best candidates for open positions and then place them in those positions based on their formal qualifications and the requirements of the job description.

4. Conclusion

This study was carried out in the public sector as part of an innovative strategy for the selection and placement of government workers. Each of the four was given bounds by the proposed machine-learning model. Primary locations, which resulted in a classification prediction accuracy of more than 90 percent, allowing for improved candidate placement and compatibility. The most significant conclusion that emerged from our investigation was the fact that machine learning can precisely predict which workers would be assigned to which positions. As a result, HR departments will have a reliable tool for assessing employees and making fair choices about where to place them and whether to promote them. The proposed strategy can serve as a blueprint for personnel selection, placement (vertical and horizontal positioning), and promotion to higher ranks (vertical positioning). One limitation of our study was the relatively small size of the sample we used (1010 total occurrences), which was made up of graduate staff from universities and technical institutions. From this vantage point, the supplied model is probably not suited for workers with a secondary or lower level of graduate education. In terms of future avenues for study, we plan to assess the efficacy of the suggested model by comparing the results of our projected classes to actual employee data. This will help us identify the most qualified applicants for open positions and evaluate their suitability considering the job requirements.

REFERENCES

[1] B. E. Becker and M. A. Huselid, “Strategic human resources management: Where do we go from here?,” *J. Manage.*, vol. 32, no. 6, pp. 898–925, 2006, doi: 10.1177/0149206306293668.

[2] P. Pampouktsi, S. Avdimiotis, M. Maragoudakis, and M. Avlonitis, “Applied Machine Learning Techniques on Selection and Positioning of Human Resources in the Public Sector,” *Open J. Bus. Manag.*, vol. 09, no. 02, pp. 536–556, 2021, doi: 10.4236/ojbm.2021.92030.

[3] E. Memili, H. “Chevy” Fang, B. Koc, Ö. Yildirim-Öktem, and S. Sonmez, “Sustainability practices of family firms: The interplay between family ownership and long-term orientation,” *J. Sustain. Tour.*, vol. 26, no. 1, pp. 9–28, 2018.

[4] J. Collins, “Good to Great - (Why Some Companies Make the Leap and others Don’t),” *NHRD Netw. J.*, vol. 2, no. 7, pp. 102–105, 2009, doi: 10.1177/0974173920090719.

[5] S. R. M. Zeebaree, H. M. Shukur, and B. K. Hussan, “Human resource management systems for enterprise organizations: A review,” *Period. Eng. Nat. Sci.*, vol. 7, no. 2, pp. 660–669, 2019, doi: 10.21533/pen.v7i2.428.

[6] A. Kaplan and M. Haenlein, “Siri, Siri, in my hand: Who’s the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence,” *Bus. Horiz.*, vol. 62, no. 1, pp. 15–25, 2019, doi: 10.1016/j.bushor.2018.08.004.

[7] Vegard Kolbjørnsrud, Richard Amico, and Robert J. Thomas, “How Artificial Intelligence Will Redefine Management,” *Harv. Bus. Rev.*, pp. 2–7, 2016.

[8] G. Carleo *et al.*, “Machine learning and the physical sciences,” *Rev. Mod. Phys.*, vol. 91, no. 4, p. 45002, Dec. 2019, doi: 10.1103/RevModPhys.91.045002.

[9] M. J. García de Yébenes Prous, F. Rodríguez Salvanés, and L. Carmona Ortells, “Validation of questionnaires,” *Reumatol. Clin.*, vol. 5, no. 4, pp. 171–177, 2009, doi: 10.1016/j.reuma.2008.09.007.

[10] S. Sovey, K. Osman, and M. E. E. MohdMatore, “Exploratory and Confirmatory Factor Analysis for Disposition Levels of Computational Thinking Instrument Among Secondary School Students,” *Eur. J. Educ. Res.*, vol. 10, no. 3, pp. 1075–1088, 2022, [Online]. Available: https://www.researchgate.net/profile/Suntonrapot-Damrongpanit/publication/356662582_Effects_of_Mindset_Democratic_Parenting_Teaching_and_School_Environment_on_Global_Citizenship_of_Ninth-grade_Students/links/61a6dda685c5ea51abc0f7b6/Effects-of-Mindset-Dem.

[11] Y. S. Abu Sultan, M. J. Al Shobaki, S. S. Abu Naser, and S. A. El Talla, “The Style of Leadership and Its Role in Determining the Pattern of,” *Int. J. Acad. Manag. Sci. Res.*, vol. 2, no. 6, pp. 26–42, 2018.

[12] F. Fallucchi, M. Coladangelo, R. Giuliano, and E. W. De Luca, “Predicting employee attrition using machine learning techniques,” *Computers*, vol. 9, no. 4, pp. 1–17, 2020, doi: 10.3390/computers9040086.

[13] M. De-Arteaga, R. Fogliato, and A. Chouldechova, “A Case for Humans-in-the-Loop: Decisions in the Presence of Erroneous Algorithmic Scores,” in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 2020, pp. 1–12, doi: 10.1145/3313831.3376638.

[14] R. Hou and Y. Xia, “Review on the new development of vibration-based damage identification for civil engineering structures: 2010–2019,” *J. Sound Vib.*, vol. 491, p. 115741, 2021, doi: https://doi.org/10.1016/j.jsv.2020.115741.

[15] J. R. Quinlan, “Bagging, boosting, and C4.5,” *Proc. Natl. Conf. Artif. Intell.*, vol. 1, no. Quinlan 1993, pp. 725–730, 1996.

[16] A. Joshuva, R. S. Kumar, S. Sivakumar, G. Deenadayalan, and R. Vishnuvardhan, “An insight on VMD for diagnosing wind turbine blade faults using C4.5 as feature selection and discriminating through multilayer perceptron,” *Alexandria Eng. J.*, vol. 59, no. 5, pp. 3863–3879, 2020, doi: https://doi.org/10.1016/j.aej.2020.06.041.

[17] L. J. Rong, “Distance learning quality assessment of universities based on interval monotonic decision tree algorithm,” *Comput. Electr. Eng.*, vol. 102, p. 108116, 2022, doi: https://doi.org/10.1016/j.compeleceng.2022.108116.

[18] P. Xu *et al.*, “Contact Sequence Planning for Hexapod Robots in Sparse Foothold Environment Based on Monte-Carlo Tree,” *IEEE Robot. Autom. Lett.*, vol. 7, no. 2, pp. 826–833, 2022, doi: 10.1109/LRA.2021.3133610.

[19] I. H. Witten and E. Frank, “Data mining: practical machine learning tools and techniques with Java implementations,” *Acm Sigmod Rec.*, vol. 31, no. 1, pp. 76–77, 2011.

[20] N. Kappelhof *et al.*, “Evolutionary algorithms and decision trees for predicting poor outcome after endovascular treatment for acute ischemic stroke,” *Comput. Biol. Med.*, vol. 133, p. 104414, 2021, doi: https://doi.org/10.1016/j.compbiomed.2021.104414.

[21] L. Vanfretti and V. S. N. Arava, “Decision tree-based classification of multiple operating conditions for power system voltage stability assessment,” *Int. J. Electr. Power Energy Syst.*, vol. 123, p. 106251, 2020, doi: https://doi.org/10.1016/j.ijepes.2020.106251.

[22] Q. Guo *et al.*, “Urban Tree Classification Based on Object-Oriented Approach and Random Forest Algorithm Using Unmanned Aerial Vehicle (UAV) Multispectral Imagery,” *Remote Sens.*, vol. 14, no. 16, 2022, doi: 10.3390/rs14163885.

[23] N. Octaviani Faomasi Daeli, “Sentiment Analysis on Movie Reviews Using Information Gain and K-Nearest Neighbor,” *Open Access J Data Sci Appl*, vol. 3, no. 1, pp. 1–007, 2020, doi: 10.34818/JDSA.2020.3.22.

[24] S. Ruan, H. Li, C. Li, and K. Song, “Class-Specific Deep Feature Weighting for Naïve Bayes Text Classifiers,” *IEEE Access*, vol. 8, pp. 20151–20159, 2020, doi: 10.1109/ACCESS.2020.2968984.

[25] N. Arora and P. D. Kaur, “A Bolasso based consistent feature selection enabled random forest classification algorithm: An application to credit risk assessment,” *Appl. Soft Comput.*, vol. 86, p. 105936, 2020, doi: https://doi.org/10.1016/j.asoc.2019.105936.

[26] Z. Haiyun and X. Yizhe, “Sports performance prediction model based on integrated learning algorithm and cloud computing Hadoop platform,” *Microprocess. Microsyst.*, vol. 79, p. 103322, 2020, doi: https://doi.org/10.1016/j.micpro.2020.103322.

[27] N. Shrestha, “Factor Analysis as a Tool for Survey Analysis,” *Am. J. Appl. Math. Stat.*, vol. 9, no.

1, pp. 4–11, 2021, doi: 10.12691/ajams-9-1-2.
[28] M. Sornalakshmi *et al.*, “Hybrid method for mining rules based on enhanced Apriori algorithm with sequential minimal optimization in healthcare industry,” *Neural Comput. Appl.*, vol. 34, no. 13, pp. 10597–10610, 2022, doi: 10.1007/s00521-020-04862-2.
[29] K. P. Lagaza, A. K. Kashyap, and A. Pandey, “Spider Monkey Optimization Algorithm Based Collision-Free Navigation and Path Optimization for a Mobile Robot in the Static Environment,” in *Advances in Mechanical Engineering*, 2020, pp. 1459–1473.
[30] S. S. Keerthi, S. K. Shevade, C. Bhattacharyya, and K. R. K. Murthy, “Improvements to Platt’s SMO algorithm for SVM classifier design,” *Neural Comput.*, vol. 13, no. 3, pp. 637–649, 2001, doi: 10.1162/089976601300014493.
[31] Z. Pan, Z. Meng, Z. Chen, W. Gao, and Y. Shi, “A two-stage method based on extreme learning machine for predicting the remaining useful life of rolling-element bearings,” *Mech. Syst. Signal Process.*, vol. 144, p. 106899, 2020, doi: https://doi.org/10.1016/j.ymssp.2020.106899.
[32] K. Gorman and S. Bedrick, “We Need to Talk about Standard Splits,” in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, Jul. 2019, pp. 2786–2791, doi: 10.18653/v1/P19-1267.
[33] P. B. Keenan and P. Jankowski, “Spatial Decision Support Systems: Three decades on,” *Decis. Support Syst.*, vol. 116, pp. 64–76, 2019, doi: 10.1016/j.dss.2018.10.010.
[34] G. Li, F. Li, C. Xu, and X. Fang, “A spatial-temporal layer-wise relevance propagation method for improving interpretability and prediction accuracy of LSTM building energy prediction,” *Energy Build.*, vol. 271, p. 112317, 2022, doi: https://doi.org/10.1016/j.enbuild.2022.112317.
[35] Y. Xu and R. Goodacre, “On Splitting Training and Validation Set: A Comparative Study of Cross-Validation, Bootstrap and Systematic Sampling for Estimating the Generalization Performance of Supervised Learning,” *J. Anal. Test.*, vol. 2, no. 3, pp. 249–262, 2018, doi: 10.1007/s41664-018-0068-2.
[36] M. Roccetti, G. Delnevo, L. Casini, and S. Mirri, “An alternative approach to dimension reduction for pareto distributed data: a case study,” *J. Big Data*, vol. 8, no. 1, p. 39, 2021, doi: 10.1186/s40537-021-00428-8.
[37] Y. Nieto, V. Gacía-Díaz, C. Montenegro, C. C. González, and R. González Crespo, “Usage of Machine Learning for Strategic Decision Making at Higher Educational Institutions,” *IEEE Access*, vol. 7, pp. 75007–75017, 2019, doi: 10.1109/ACCESS.2019.2919343.
[38] L. Xiaoxue, B. Xuesong, W. Longhe, R. Bingyuan, L. Shuhan, and L. Lin, “Review and Trend Analysis of Knowledge Graphs for Crop Pest and Diseases,” *IEEE Access*, vol. 7, pp. 62251–62264, 2019, doi: 10.1109/ACCESS.2019.2915987.
[39] N. Bindra and M. Sood, “Detecting DDoS Attacks Using Machine Learning Techniques and Contemporary Intrusion Detection Dataset,” *Autom. Control Comput. Sci.*, vol. 53, no. 5, pp. 419–428, 2019, doi: 10.3103/S0146411619050043.
[40] M. Chatterjee and A. S. Namin, “A fuzzy Dempster–Shafer classifier for detecting Web spams,” *J. Inf. Secur. Appl.*, vol. 59, p. 102793, 2021, doi: https://doi.org/10.1016/j.jisa.2021.102793.
[41] H. Faris, M. Habib, M. Faris, M. Alomari, and A. Alomari, “Medical speciality classification system based on binary particle swarms and ensemble of one vs. rest support vector machines,” *J. Biomed. Inform.*, vol. 109, p. 103525, 2020, doi: https://doi.org/10.1016/j.jbi.2020.103525.
[42] B. E. Boser, V. N. Vapnik, and I. M. Guyon, “Training Algorithm Margin for Optimal Classifiers,” *Perception*, pp. 144–152, 1992.
[43] Y. Zhang, B. Yu, and H. M. Gu, “Face recognition using curvelet-based two-dimensional principle component analysis,” *Int. J. Pattern Recognit. Artif. Intell.*, vol. 26, no. 3, pp. 1–13, 2012, doi: 10.1142/S0218001412560095.
[44] T. L. Nikmah, M. Z. Ammar, Y. R. Allatif, R. M. P. Husna, P. A. Kurniasari, and A. S. Bahri, “Comparison of LSTM , SVM , and Naive Bayes for Classifying Sexual Harassment Tweets,” *J. Soft Comput. Explor.*, vol. 3dend, no. 2, pp. 131–137, 2022.
[45] K. Tan, W. S. Lee, H. Gan, and S. Wang, “Recognising blueberry fruit of different maturity using histogram oriented gradients and colour features in outdoor scenes,” *Biosyst. Eng.*, vol. 176, pp. 59–72, 2018, doi: https://doi.org/10.1016/j.biosystemseng.2018.08.011.
[46] J. S. Mapa, A. Sison, and R. P. Medina, “A Modified C4.5 Classification Algorithm: With the Discretization Method in Calculating the Goodness Score Equivalent,” *ICETAS 2019 - 2019 6th IEEE Int. Conf. Eng. Technol. Appl. Sci.*, pp. 4–7, 2019, doi: 10.1109/ICETAS48360.2019.9117309.
[47] C. F. Chien and L. F. Chen, “Data mining to improve personnel selection and enhance human

capital: A case study in high-technology industry," *Expert Syst. Appl.*, vol. 34, no. 1, pp. 280–290, 2008, doi: 10.1016/j.eswa.2006.09.003.

[48] B. Prasetiyo, Alamsyah, and M. A. Muslim, "Analysis of Building Energy Efficiency Dataset Using Naive Bayes Classification Classifier," *J. Phys. Conf. Ser.*, vol. 1321, no. 3, p. 32016, Oct. 2019.

[49] A. Azar, M. V. Sebt, P. Ahmadi, and A. Rajaeian, "A model for personnel selection with a data mining approach: A case study in a commercial bank," *SA J. Hum. Resour. Manag.*, vol. 11, no. 1, pp. 1–10, 2013, doi: 10.4102/sajhrm.v11i1.449.

[50] S.-O. Gerdt, E. Wagner, and G. Schewe, "The relationship between sustainability and customer satisfaction in hospitality: An explorative investigation using eWOM as a data source," *Tour. Manag.*, vol. 74, pp. 155–172, 2019, doi: https://doi.org/10.1016/j.tourman.2019.02.010.

[51] K. R. Varshney, V. Chenthamarakshan, S. W. Fancher, J. Wang, D. Fang, and A. Mojsilović, "Predicting employee expertise for talent management in the enterprise," *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, pp. 1729–1738, 2014, doi: 10.1145/2623330.2623337.