IJCCS (Indonesian Journal of Computing and Cybernetics Systems)
Vol.x, No.x, July xxxx, pp. 1~5
ISSN (print): 1978-1520, ISSN (online): 2460-7258
DOI: 10.22146/ijccs.xxxx

Comparison Non-Parametric Machine Learning Algorithms for Prediction of Employee Talent

I Ketut Adi Wirayasa*[1], Arko Djajadi[2], H, andri Santoso[3], Eko Indrajit[3]
[1,2,3,4]Universitas Pradita; Scientia Business Park Tower I
Jl. Boulevard Gading Serpong
Blok O/1, Summarecon Serpong
Tangerang
[1,2,3,4]Department of Computer Science , Banten, Indonesia
e-mail: ***[1]ketut.adi.wirayasa@student.pradita.ac.id**, [2]arko.djajadi@pradita.ac.id, [3]h, andri.santoso@pradita.ac.id, [4]eko.indrajit@pradita.ac.id

Abstrak

Klasifikasi data ordinal merupakan bagian dari data kategorikal. Data ordinal terdiri dari fitur dengan nilai yang berdasarkan urutan atau ranking. Penggunaan metode machine learning di bagian manajemen Sumber Daya Manusia dimaksudkan untuk mendukung pengambilan keputusan yang didasarkan pada analisis data objektif dan bukan pada aspek subjektif. Tujuan dari penelitian ini adalah untuk menganalisis hubungan antar fitur dan apakah fitur yang digunakan sebagai faktor objektif dapat mengklasifikasi serta memprediksi karyawan tertentu bertalenta atau tidak. Penelitian ini menggunakan dataset publik yang disediakan oleh IBM analytics. Analisis pada dataset menggunakan uji statistika dan uji validitas confirmatory factor analysis, dimaksudkan untuk mengetahui hubungan atau korelasi antar fitur dalam merumuskan hypothesis testing sebelum membangun model non parametric machine learning dengan menggunakan komparasi dari empat algoritma yaitu Support Vector Machine, K-Nearest Neighbor, Decision Tree dan Artificial Neural Networks. Hasil pengujian dalam bentuk Confusion Matrix dan report classification dari setiap model. Evaluasi terbaik dihasilkan oleh algoritma Support Vector Machine dengan nilai Accuracy, Precision dan Recall yang sama yaitu sebesar 94.00%, Sensitivity 93.28%, tingkat False Positive rate 4.62%, tingkat False Negative rate 6.72%, dan nilai AUC-ROC curve 0.97 dengan kategori excellent dalam melakukan klasifikasi talent atau non-talent dari model prediksi employee talent.

Kata kunci*— non-parametric, machine learning, ordinal data, employee talent.*

Abstract

Classification of ordinal data is part of categorical data. Ordinal data consists of features with values based on order or ranking. The use of machine learning methods in Human Resources Management intends to support decision-making based on objective data analysis, not on subjective aspects. This study aims to analyze the relationship between features and whether the components used as objective factors can classify and predict certain talented employees or not. This study uses a public dataset provided by IBM analytics. Analysis of the dataset using statistical tests and confirmatory factor analysis validity tests, intended to determine the relationship or correlation between features in formulating hypothesis testing before building a model by comparing four algorithms, namely Support Vector Machine, K-Nearest Neighbor, Decision Tree, and Artificial Neural Networks. The test results are expressed in the Confusion Matrix and report classification of each model. The SVM algorithm produces the best evaluation. With the same Accuracy, Precision, and Recall values, 94.00%, Sensitivity

Received June 1st,2012; Revised June 25th, 2012; Accepted July 10th, 2012

93.28%, False Positive rate 4.62%, False Negative rate 6.72%, and AUC-ROC curve value 0.97 with an excellent category in performing classification of the employee talent prediction model.

Keywords— *non-parametric, machine learning, ordinal data, employee talent.*

1. INTRODUCTION

Data mining methods have been applied and have good prospects in the field of human resource management. The utilization of data mining tools has a positive impact in supporting management and policy development in organizations. Machine learning is one technique that can provide significant support for Human Resources Management (HRM) applications which are usually limited by interpretations and subjective decisions based on employee behavior [1]. By adopting technology, organizations will get many benefits by collecting, managing, and analyzing data. In terms of efficiency, competitive advantage, and better business competitiveness and leading to improvements in helping the decision-making process achieve the organizational goals that have been set before [1].

This study discusses the application of machine learning techniques in the HR department, which is carried out by analyzing datasets provided by IBM analytics. The selection of this dataset basically on the variables and attributes that reflect the employee database and have supporting variables and features owned by the organization, consisting of 35 variables and 1470 samples. It will use Four non-parametric algorithms, namely Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Decision Tree (DT), and Artificial Neural Networks (ANN). To choose these four algorithms based on: (1). The characteristics and types of data to be processed, (2). The number of variables and samples used, (3). What algorithm for classification and prediction, and (4). Each has advantages and disadvantages in generating models during training and data testing[2].

The objectives of this study are: (1). to analyze and compare the performance of machine learning non-parametric algorithms in conducting the classification and prediction process of employee talent based on ordinal category datasets, (2). The predictive models with concept of talent management using tested variables, (3). Determine whether the results of the comparison of non-parametric algorithms in classifying and predicting talented or non-talented employees can be used in objective decision-making. In addition, this research is helpful in: (1). Providing an alternative to developing concepts and application models in the talent management module (2). As a material for evaluating and testing relationships and relationships between variables based on hypothesis testing by previous researchers using machine learning methods and the Python programming language to study employee's talent prediction case.

This research contributes through empirical evaluation with an ordinal data analysis approach that uses non-parametric machine learning algorithms to predict employee talent. The research results conducted through the IBM HR Analytics dataset show that employee talent predictions using non-parametric machine learning algorithms are faster in data processing with a large number of variables, have a more accurate prediction accuracy level, and supporting variables show a significant correlation. From previous research hypotheses. It can be proven to be a supporting factor in the prediction of employee talent. This research also provides alternative solutions for organizations in utilizing existing staffing data for learning and developing employee career management systems that can be utilized by HR management in organizations.

2. METHODS

The organization's massive and employee information (big data) can be analyzed using machine learning technology. Previous researchers have researched the application of machine learning methods and algorithms in HRM and other applied sciences. Prediction of student activity level by comparing the SVM and DT algorithms using a dataset of 1530 samples [3], comparing the performance of the DT, SVM, KNN, and Naïve Bayes (NB) algorithms to the prediction of student alcohol consumption using a dataset of 1024 samples [4]. "Maintain, and Evaluate student's performance" using the DT algorithm, Linear Regression, Multiple Regression, and Logistic Regression [5]. The research on "Talent Identification in Soccer using a one-class SVM" identifies prospective soccer athletes [6]. The study predicts the right candidate for the right job by having the required qualities based on the applicant's resume. The analysis using approximately 500 samples through the DT algorithm, Naïve Bayes, and CART [7] are some examples of research that uses machine learning algorithms in the process.

The results of previous studies, machine learning algorithms in classifying and predicting produce a good level of accuracy and can be applied in the field of research to help make better decisions [7], [8], each algorithm has advantages, and disadvantages, which is lack of classification, and prediction [3]–[5], [8]. Classification and prediction results are influenced by several factors such as the number of training data samples used, data types and characteristics, selection of appropriate algorithms, and statistical methods [1], [8], [9]. No one algorithmic approach is superior to other methods for all problem cases or what is known as the "no free lunch" theory for the supervised machine learning method.

One of the statistical data processing is using non-parametric methods. The Wilcoxon Sum Rank test is a non-parametric statistical hypothesis that compares two related samples, matched samples, or repeated measurements of one model to assess whether the population means ratings differ [10]. The Mann-Whitney test is a non-parametric test used to determine the difference between the mean of two people equally distributed from two independent samples with an ordinal data form. The Kruskal Wallis test is a non-parametric test that assesses the difference between three or more groups of separate models that are not normally distributed (ordinal or ranked data) [11]. The Confirmatory Factor Analysis (CFA) test will strengthen the statistical tests' results. In proving the previous hypothesis, questioning whether there is a relationship or correlation between the dependent and independent variables measure to determine the construct validity of the sample in the survey [12], [13].

Receiver Operating Characteristic (ROC) curve in the Area Under Curve (AUC) in classifying the accuracy of the test results were used to provide comparison results between predictions and actual target values in the classification process [6], [14]. ROC describes model performance or comparison with a complete estimate of the classification threshold. The value in the ROC area varies between the 0 to 1 interval is shown in Table 1.

Table 1 AUC Value

AUC	Classification
0.90 - 1.00	Excellent
0.80 - 0.90	Good
0.70 - 0.80	Fair
0.60 - 0.70	Poor
< 0.60	Failure

In the work environment, employee job involvement relates to how a person manages his behavior at work and becomes part of the life cycle of an organization in achieving its goals. Employees engaged in work will feel that work will be more meaningful if they show better performance [15], [16]. Job satisfaction is essential to make an employee bring out his abilities to the fullest in his work [17].

Although talent management has a strategic role in a modern organization, it will not impact employees; much research has been performed on job satisfaction's mediating part [18]. Other research shows a close relationship between work-life balance, employee performance, job satisfaction, and work-life balance that can improve employee performance through employee job satisfaction [19]. Another hypothesis related to job involvement is closely related to improving employee performance and states that the higher a person's job involvement, the higher his employee performance [20]. There is an undoubtedly associated with the conceptual model of Talent Management, where there is a relationship between employee recognition and employee performance, and there is a relationship between the concept of talent management and employee performance [21].

Based on the results of previous studies, the formulation of hypotheses using the IBM analytics dataset resulted as the following:

a. H1: Is there a positive relationship between education and performance rating?
b. H2: Is there a positive relationship between environment satisfaction and performance rating?
c. H3: Is there a positive relationship between job involvement and performance rating?
d. H4: Is there a positive relationship between job level and performance rating?
e. H5: Is there a positive relationship between job satisfaction and performance rating?
f. H6: Is there a positive relationship between relationship satisfaction and performance rating?
g. H7: Is there a positive relationship between work-life balance and performance rating?
h. H8: Is there a positive, and convergent relationship, among other independent variables?

In this study, researchers used the performance rating variable as a target in the classification process. Other ordinal data such as education, environment satisfaction, job involvement, job level, job satisfaction, relationship satisfaction, and work-life balance variables were used as predictors..

2.1 Nonparametric Statistical Test

The ordinal data used for the experiment will go through statistical and CFA tests to strengthen hypothesis testing. Statistical tests were carried out on ordinal data using the Correlation Coefficient to determine the correlation or rank value relationship between 2 (two) variables [12]. After carrying out statistical tests and generating conclusions from hypothesis testing, the CFA validity test's analysis phase is carried out to test measurable and unmeasured variables. The CFA test carried out is only limited to testing variables by looking at the Keiser-Meyer-Olkin (KMO) test value and comparing the size of the sampling adequacy of each variable in a proportional measure. The primary variable efficiently (KMO >= 0.5), and Bartlett's test is a test of Sphericity that is used to determine whether there is a significant correlation between variables ($\alpha < 0.05$) [12], [23].

2.2 Data Testing

The pre-processing stages include data cleaning, which is carried out to ensure that no data is lost, null, or duplicated. Normalize the dataset (standardization) by assigning a value of 0 or 1. The following process is data selection by selecting the relevant data to use (ordinal data) and dividing the dataset into training and test data with a ratio of 90%: 10%, or 1323 samples, and 147 samples. Training and testing data are carried out using the selected algorithm model.

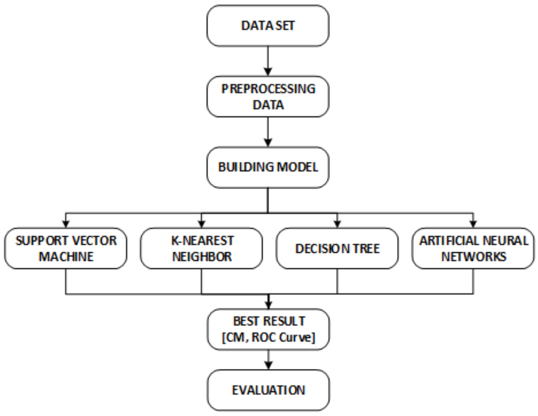

Figure 1 Research Methodology Proposal

The testing process uses training data from the formed model and further testing for evaluation. The research methodology proposal carried out at the training and model testing stages is shown in Figure 1.

The evaluation of the classification model carried out on data testing produces the value of the best model performance in predicting true or false objects displayed in the Confusion Matrix (CM) [24], report classification, and the ROC-AUC curve. CM consists of sections, namely True Positive (TP), False Negative (FN), False Positive (FP), and True Negative (TN), with the calculation parameters using the formula:

$$\text{Accuracy} = (TP + TN) / (TP+FP+FN+TN) \quad (1)$$

As shown in equation (1), the accuracy result explains that the model produces a correct prediction ratio for the classification of talent, and non-talent, from the entire sample. The accuracy is used to answer the question, "What percentage of the sample correctly predicts talent and non-talent?"

$$\text{Precision} = (TP) / (TP+FP) \quad (2)$$

As shown in equation (2), the precision results explain that the model produces a correct ratio of talent classification predictions compared to the overall sample results predicted by talent. The precision is used to answer the question, "What percentage of the correct sample of talent out of the total sample predicted talent??"

$$\text{Recall} = (TP) / (TP + FN) \quad (3)$$

The results of Recall or Sensitivity, as shown in equation (3), explain that the model produces a correct prediction ratio for talent classification compared to the entire sample of true (actual) talent. The Recall or Sensitivity is used to answer the question, "What percentage of the predicted sample is talent compared to the total sample that is talent?"

$$\text{Specificity} = (TN)/ (TN + FP) \quad (4)$$

As shown in equation (4), specificity results explain that the model produces a level of truth in predicting non-talents compared to the whole sample of non-talents. The Specificity is used to answer the question, "What percentage of the correct sample is non-talented compared to the total sample that is non-talented??"

Table 2 CM - Talent and Non-Talent

Predicted & Observed	True Talent	True Non-Talent
Predictions Talent	True Positive (TP)	False Negative (FN)
Predictions Non-Talent	False Positive (FP)	True Negative (TN)

CM is used to represent the predictions and actual conditions of the data generated by the algorithm used. The performance results of the four algorithm models in CM [25]. True Positive is the real talent, True Negative is the actual non-talent, Positive Predictions is the talent prediction, and Negative Predictions is the non-talent prediction, as shown in Table 2. We use the accuracy for the evaluation process and determine the ratio of correct predictions (true positive and true negative) from the overall data. Meanwhile, AUC is used to show numbers that are directly related to the data. The AUC value describes the overall measurement results of the model's suitability with the indicator that the greater the AUC value, the better the variables studied predict events [25].

3. RESULTS and DISCUSSION

The research uses the Python programming language, where the input data comes from the IBM Analytics dataset, the dependent, and independent variables are ordinal type, using the nonparametric machine learning algorithm method SVM, KNN, DT, and ANN, through the analysis process of non-parametric statistical tests, and hypothesis testing.

3.1 Statistical Test Result

The results of statistical tests using the Mann Whitney U test, Wilcoxon Rank Sum, and Kruskal Wallis H test on the dataset based on statistical tests for all the independent variables with a p-value < 0.05. The conclusion of the hypothesis test on the correlation test results between the dependent and independent variables. There is a close correlation or relationship between the independent variables (education, environment satisfaction, job involvement, job level, job satisfaction, relationship satisfaction, work-life balance) and the dependent variable (performance ratings). Thus, the hypothesis test results state that there is a positive relationship between the independent variable and the dependent variable.

3.2. Hypothesis Testing

Hypothesis testing of the dependent variable performance rating as a target. And the independent variables are education, environment satisfaction, job involvement, job level, job satisfaction, job satisfaction, relationship satisfaction, and work-life balance as predictors by using statistical tests that have carrying to produce hypotheses:

a. H1: There is a positive relationship between education and performance rating.
b. H2: There is a positive relationship between environment satisfaction and performance rating.
c. H3: There is a positive relationship between job involvement and performance rating.
d. H4: There is a positive relationship between job level and performance rating
e. H5: There is a positive relationship between job satisfaction and performance rating.
f. H6: There is a positive relationship between relationship satisfaction and performance rating.
g. H7: There is a positive relationship between work-life balance and performance rating.
h. H8: There is a positive and convergent relationship between the job level, education, and other independent variables.

The KMO table and Bartlett's test show that the KMO value is 0.501, indicating a significant correlation between variables (>= 0.500). Likewise, Bartlett's Sphericity test, which has a value of 41.257 with a p-value of 0.011 < 0.05 (significant), is shown in Table 3, which means that the variable forming factors are pretty good and can be analyzed further.

Table 3 KMO and Bartlett's Test of Sphericity

KMO measure of sampling	0.501
Bartlett's Test of Sphericity, Chi-squared	41.257
, Sig.	0.011

3.3. Model Performance

Table 4 shows the accuracy results from training data and testing data from each model. Accuracy results show an increase after training using a model that was formed and tested using hyperparameter tuning.

Table 4 Accuracy – Training and Testing

No	Algorithms	Accuracy -Training	Accuracy - Testing
1	SVM	92.00%	94.00%
2	KNN	83.00%	84.00%
3	DT	81.00%	83.00%
4	ANN	91.00%	92.00%

3.3.1. ANN Algorithms Model Performance

Table 5 shows the number of testing data as many as 249 samples. The ANN model resulted in 117 examples of true positive and 112 actual negative samples, indicating that the prediction data following the talent classification is 117 samples. The non-talent classification prediction is 112 samples. While the real negative value of 13 or the prediction results of the non-talent category that do not match the actual are 13 samples. The true positive is seven or, and this result states that seven examples of predictive data with talent classification do not match. The final performance of the ANN model produces a precision value of 0.92, an accuracy level of 0.92 on the test results, and the ROC curve with an AUC value of 0.97 (excellent classification), as shown in Figure 2.

Table 5 CM Model ANN

Predicted & Observed	True Positive	True Negative	Class Precision
Predictions Positive	117	13	94.00%
Predictions Negative	7	112	90.00%
Class Recall	90.00%	94.00%	

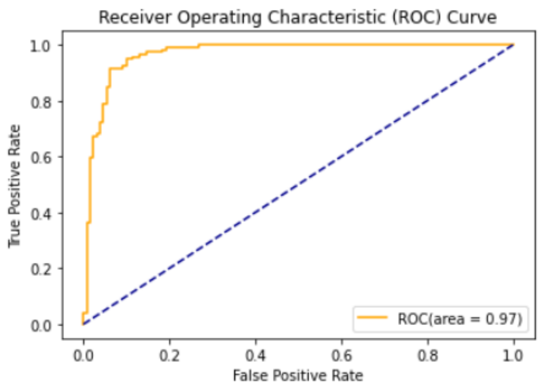

Figure 2 ROC Curve - Model ANN

3.3.2. DT Algorithm Model Performance

From the total testing data of 249 samples, the DT model yielded 118 samples of true positive and 89 actual negative samples. There is an indication that according to the talent classification, the prediction data is 118 samples, and the non-talent classification prediction is 89 samples, as shown in Table 6. While the actual negative value of 12 or the prediction results of the non-talent classification that do not match the real are 12 samples, and the true positive is 30. This result states that the predicted data with the talent classification that does not match the actual is 30 samples. The final performance of the DT model produces a precision value of 0.84, with an accuracy level of 0.83 on the test results, and the ROC curve with an AUC value of 0.85 (good classification), as shown in Figure 3.

Table 6 CM Model DT

Predicted & Observed	True Positive	True Negative	Class Precision
Predictions Positive	118	12	88.00%
Predictions Negative	30	89	80.00%
Class Recall	75.00%	91.00%	

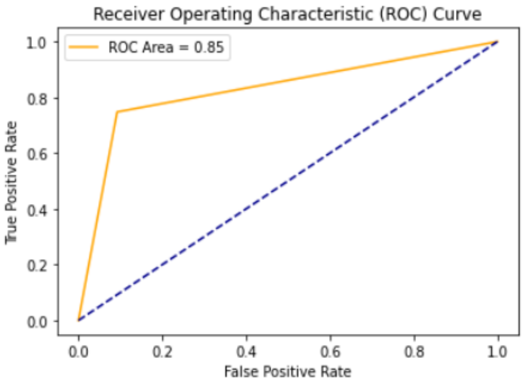

Figure 3 ROC Curve - Model DT

3.3.3. KNN Algorithm Model Performance

From the number of testing data as many as 249 samples, the KNN model resulted in 97 true positives and 113 actual negative examples. There is an indication that the prediction data according to the talent classification is 97 samples, and the non-talent classification prediction is 113 samples, as shown in Table 7. While the actual negative value of 33 or the prediction results of non-talent classifications does not match the real, there are 33 samples and six true positives. This result states that six examples of predictive data with talent classifications do not check the actual match. The final performance of the KNN model produces a precision value of 0.84 and an accuracy level of 0.83 on the test results, and the ROC curve with an AUC value of 0.91 (excellent classification) as shown in Figure 4.

Table 7 CM Model KNN

Predicted & Observed	True Positive	True Negative	Class Precision
Predictions Positive	97	33	77.00%
Predictions Negative	6	113	94.00%
Class Recall	95.00%	75.00%	

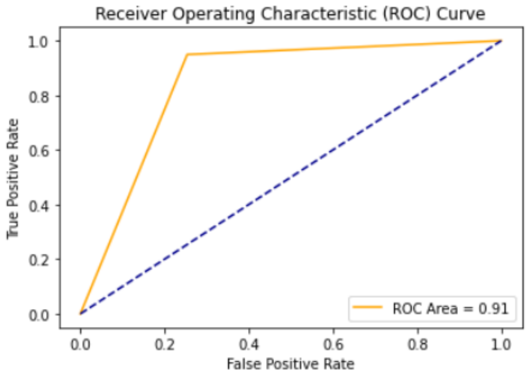

Figure 4 ROC Curve - Model KNN

3.3.4. SVM Algorithm Model Performance

The SVM model's testing data and 249 samples resulted in 124 actual positive samples and 111 actual negative samples. There is an indication that the prediction data according to the talent classification is 124 samples, and the non-talent classification prediction is 111 samples, as shown in Table 8. While the actual negative value of 6 or the prediction results of non-talent classifications does not match the existing, there are six samples and eight true positives. This result states that there are eight samples of predictive data with talent classifications that do not match. The final performance of the SVM model produces a precision value of 0.94, an accuracy level of 0.94 on the test results, and the ROC curve with an AUC value of 0.97 (excellent classification), as shown in Figure 5.

Table 8 CM Model SVM

Predicted & Observed	True Positive	True Negative	Class Precision
Predictions Positive	124	6	95.00%
Predictions Negative	8	111	94.00%
Class Recall	93.00%	95.00%	

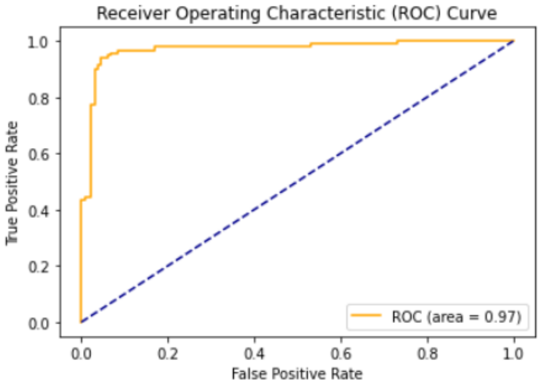

Figure 5 ROC Curve - Model SVM

3.3.5. Model Comparison

Evaluation of the model performance resulted in the SVM algorithm, which has the highest accuracy of 94.00%, compared to other algorithm models. That confirms that SVM has a more accurate level of accuracy in making predictions for the classification of talent, and non-talent as shown in Table 9.

Table 9 Algorithm Model Comparison

DESCRIPTION	SVM	KNN	DT	ANN
Accuracy	94.00%	84.00%	83.00%	92.00%
AUC	0.97	0.91	0.85	0.97
Precision	94.00%	86.00%	84.00%	92.00%
Recall	94.00%	85.00%	83.00%	92.00%
Sensitivity	93.28%	94.96%	74.79%	94.12%
Specificity	95.38%	74.62%	90.77%	90.00%
PPV	94.87%	77.40%	88.12%	89.60%
NPV	93.94%	94.17%	79.73%	94.35%
TPR	93.28%	94.96%	74.79%	94.12%

SVM has a precision value of 94.00% and recalls 94.00%, higher than the other models. In other words, SVM is better at predicting a positive sample of talent but is non-talented, rather than indicating that a sample that is expected to be non-talented but is a talent. Furthermore, SVM also has a specificity value of 95.38% higher than other algorithm models. That means that the SVM model produces a low false-positive rate or 4.62% on the test result. The resulting prediction model has an error in predicting a sample that is non-talented but stated to be quite a low talent compared to the results from other models, as shown in Table 10.

Table 10 Talent, and Non-talent Prediction

PREDICTION		SVM	KNN	DT	ANN
Talent	True	95.38%	74.62%	90.77%	90.00%
	False	4.62%	25.38%	9.23%	10.00%
Non Talent	True	93.28%	94.96%	74.79%	94.12%
	False	6.72%	5.04%	25.21%	5.88%

SVM also has an AUC value of 0.97 (excellent classification), although this value is the same as the ANN algorithm. However, SVM is superior in specificity value and has a lower false-positive rate, as shown in Figures 6 and 7.

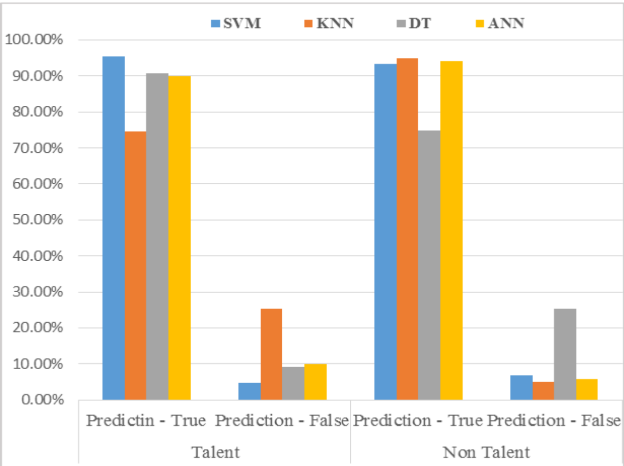

Figure 6 Graph – Comparative of Model Predictions

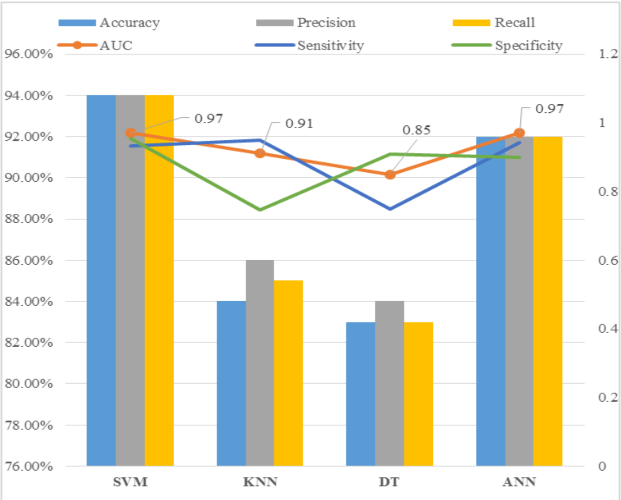

Figure 7 Graph - Model Performance Comparison

4. CONCLUSIONS

The ordinal data has different characteristics in handling. Machine Learning Algorithm is one of the tools that can extract ordinal data into information for decision-making. Using a comparison of four non-parametric machine learning models, namely SVM, KNN, DT, and ANN, the ordinal data went through the stages of non-parametric statistical tests on the dataset used in this study CFA validity tests in formulating hypothesis testing.

The results of hypothesis testing on the dataset state a correlation or relationship between the dependent and independent variables. The existence of a variable that mediates the relationship between the dependent variable and the independent variable. We can conclude that the ordinal data used in the dataset can be analyzed using an algorithm model to classify and predict. From the results of training and testing of accuracy based on CM, and the ROC-AUC curve is the SVM algorithm, where the model produces an accuracy of 94.00%, AUC of 0.97, and also have FPR and FNR values of 4.62% and 6.72% with a minimal difference with a low error rate..

We recommended for further research, prediction models and talent or non-talent classification analysis used as a guide and initial process in developing methods for classifying talented or non-talented employees using ordinal data. Prediction models and analysis of talent or non-talent classification can also be used as tools in the preparation of deep learning-based application systems for the concept of talent management. The use of more datasets or updated data is a recommendation by using feature engineering techniques. It can identify data characteristics, and the addition of new features from the sample dataset will improve prediction results and accuracy..

REFERENCES

[1] F. Fallucchi, M. Coladangelo, R. Giuliano, and E. William De Luca, "Predicting Employee Attrition Using Machine Learning Techniques," *Computers*, vol. 9, no. 4, p. 86, Nov. 2020, doi: 10.3390/computers9040086.

[2] A. Dandekar, D. Basu, and S. Bressan, "Differentially Private Non-parametric Machine Learning as a Service," *Differ. Priv. Non-parametric Mach. Learn. as a Serv.*, vol. 11706 LNCS, pp. 189–204, 2019, doi: 10.1007/978-3-030-27615-7_14.

[3] S. Wiyono, "Perbandingan Algoritma Machine Learning SVM dan Decision Tree untuk Prediksi Keaktifan Mahasiswa," *Sinkron*, vol. 3, no. 1, pp. 105–108, 2018.

[4] N. Sagala and H. Tampubolon, "Komparasi Kinerja Algoritma Data Mining pada Dataset Konsumsi Alkohol Siswa," *Khazanah Inform. J. Ilmu Komput. dan Inform.*, vol. 4, no. 2, p. 98, Dec. 2018, doi: 10.23917/khif.v4i2.7061.

[5] M. G M, G. M. Mujtaba, and M. Rahmath, "Maintain and Evaluate students' performance Using Machine Learning," *Int. J. Comput. Trends Technol.*, vol. 68, no. 6, pp. 57–63, 2020, doi: 10.14445/22312803/ijctt-v68i6p110.

[6] S. Jauhiainen, S. Aÿrämö, H. Forsman, and J. P. Kauppi, "Talent identification in soccer using a one-class support vector machine," *Int. J. Comput. Sci. Sport*, vol. 18, no. 3, pp. 125–136, 2019, doi: 10.2478/ijcss-2019-0021.

[7] A. S. & S. Bhardwaj, "Resume Ranking And Performance Appraisal Using Predictive Mining And Machine Learning : Talent Management System," *Int. J. Manag. Appl. Sci. ISSN 2394-7926*, vol. 4, no. 6, 2018.

[8] M. Moustafa Reda, M. Nassef, and A. Salah, "Factors Affecting Classification Algorithms Recommendation: A Survey," in *8th International Conference on Soft Computing, Artificial Intelligence and Applications*, Jun. 2019, pp. 83–99, doi: 10.5121/csit.2019.90707.

[9] A. Çetinkaya and Ö. K. Baykan, "Prediction of middle school students' programming talent using artificial neural networks," *Eng. Sci. Technol. an Int. J.*, vol. 23, no. 6, pp. 1301–1307, 2020, doi: 10.1016/j.jestch.2020.07.005.

[10] C. N. Refugio, "An Empirical Study on Wilcoxon Signed Rank Test An Empirical Study on Wilcoxon Signed Rank Test Ana Marie Durango," *J.*, no. December, p. 12, 2018, doi: 10.13140/RG.2.2.13996.51840.

[11] E. A. Bedoya-Marrugo, L. E. Vargas-Ortiz, C. A. Severiche-Sierra, and D. D. Sierra-Calderon, "Kruskal-Wallis Test for the Identification of Factors that Influence the Perception of Accidents in Workers in the Construction Sector," *Int. J. Appl. Eng. Res.*, vol. 12, no. 17, pp. 6730–6734, 2017, [Online]. Available: http://www.ripublication.com.

[12] R. P. Sarmento and V. Costa, "Confirmatory Factor Analysis -- A Case study," *ResearchGate*, p. 39, 2019, [Online]. Available: http://arxiv.org/abs/1905.05598.

[13] X. Lee, B. Yang, and W. Li, "The influence factors of job satisfaction and its relationship with turnover intention: Taking early-career employees as an example," *An. Psicol.*, vol. 33, no. 3, p. 697, Jul. 2017, doi: 10.6018/analesps.33.3.238551.

[14] D. A. Kristiyanti and M. Wahyudi, "Feature selection based on Genetic algorithm, particle swarm optimization and principal component analysis for opinion mining cosmetic product review," in *2017 5th International Conference on Cyber and IT Service Management (CITSM)*, Aug. 2017, pp. 1–6, doi: 10.1109/CITSM.2017.8089278.

[15] S. Fernández-Salinero, Á. G. Collantes, F. R. Cifuentes, and G. Topa, "Is job involvement enough for achieving job satisfaction? The role of skills use and group identification," *Int. J. Environ. Res. Public Health*, vol. 17, no. 12, pp. 1–11, 2020, doi: 10.3390/ijerph17124193.

[16] A. Yuspahruddin, A. Eliyana, A. D. Buchdadi, Hamidah, T. Sariwulan, and K. Muhaziroh, "The effect of employee involvement on job satisfaction," *Syst. Rev. Pharm.*, vol. 11, no. 7, pp. 490–498, 2020, doi: 10.31838/srp.2020.7.72.

[17] R. Setiawan, A. Eliyana, T. Suryani, and J. Christopher, "Creating job satisfaction in a strict organization," *Opcion*, vol. 36, no. SpecialEdition27, pp. 376–385, 2020.

[18] D. R. Wickramaaratchi and G. D. N. Perera, "The Impact of Talent Management on Employee Performance: The Mediating Role of Job Satisfaction of Generation Y Management Trainees in the Selected Public Banks in Sri Lanka," *Sri Lankan J. Hum. Resour. Manag.*, vol. 10, no. 1, p. 21, Jun. 2020, doi: 10.4038/sljhrm.v10i1.5648.

[19] M. D. V. S. Mendis, "The Impact of Work Life Balance on Employee Performance with Reference to Telecommunication Industry in Sri Lanka: A Mediation Model," *ResearchGate*, vol. 12, no. January 2017, p. 30, 2018, doi: 10.4038/kjhrm.v12i1A2.

[20] A. Kusmaningtyas and R. Nugroho, "Effect of Job Involvement on Employee Performance through Work Engagement at Bank Jatim," *Univers. J. Manag.*, vol. 9, no. 2, pp. 29–37, 2021, doi: 10.13189/ujm.2021.090201.

[21] M. N. El, "Talent Management , Employee Recognition And Performance In The Research Institutions," *Sciendo*, vol. 14, no. 14, pp. 127–140, 2019, doi: 10.2478/sbe-2019-0010.

[22] F. Firmansyah, Rozanah Katrina Herda, Angga Damayanto, and Fajar Sidik, "Confirmatory Factor Analysis To Know the Influencing Factors of Elementary School Students' Self-Concept in Jetis Sub District, Bantul Regency," *JISAE J. Indones. Student Assess. Eval.*, vol. 6, no. 2, pp. 196–202, 2020, doi: 10.21009/jisae.062.010.

[23] L. L. Chan and N. Idris, "Validity and Reliability of The Instrument Using Exploratory Factor Analysis and Cronbach ' s a lpha," *IJARBS*, vol. 7, no. 10, pp. 400–410, 2017, doi: 10.6007/IJARBSS/v7-i10/3387.

[24] A. N. Noercholis, "Comparative Analysis of 5 Algorithm Based Particle Swarm Optimization (Pso) for Prediction of Graduate Time Graduation," *Matics*, vol. 12, no. 1, p. 1, 2020, doi: 10.18860/mat.v12i1.8216.

[25] T. T. Maskoen and D. Purnama, "Area Under the Curve dan Akurasi Cystatin C untuk Diagnosis Acute Kidney Injury pada Pasien Politrauma," *Maj. Kedokt. Bandung*, vol. 50, no. 4, pp. 259–264, 2018, doi: 10.15395/mkb.v50n4.1342.