Volume 25, Special Issue **Print ISSN: 1099 -9264**
Online ISSN: 1939-4675

TEXT CLASSIFICATION ANALYSIS BY MACHINE LEARNING JOB SEGMENTATION ALGORITHM

Duenpen Panurug, King Mongkut's University of Technology North Bangkok
Montean Rattanasiriwongwut, King Mongkut's University of Technology North Bangkok

ABSTRACT

The objective of this research was to analyze the classification of texts related to job qualifications. A text analysis algorithm (Text Classification Analysis) was applied. This research gathered the qualifications of job positions from the database of JOBBKK Company (as of March 9, 2020), which is the most popular recruitment agency in Thailand. Data on the job qualifications of 10,000 samples were used to classify workgroups by means of the 10-Fold Cross Validation test, which uses five algorithms: Decision Tree, naive Bayes, Learning base (Support Vector Machine), Random Decision Forest (Random Forest), and K-Nearest Neighbor. Performance measures were precision, recall, and each algorithm's F-measure value. The analysis results show that the Support Vector Machine (SVM) algorithm has a text recognition efficiency with a highest accuracy of 92.73%.

Key Words: Text Classification Analysis, Support Vector Machine, Random Decision Forest

INTRODUCTION

Nowadays, there is an increasing number of electronic documents on the Internet, and there is more storage in the form of databases. The text in this paper is unstructured. Unstructured text comes in the form of websites, e-mails, online bulletin boards, chat rooms, various social media, etc., in which the structured nature of information is 80%-90% of the total data (Kanimozhi et al., 2015). Today, people are accustomed to accessing or searching for information on the Internet. This is a change from the past, when people looked for information in textbooks and documents. In addition, users of Internet networks also exchange opinions, expressing views on subjects such as politics, marketing, education, or entertainment through websites.

The above message shows that the number of messages or comments on the Internet is both enormous and complicated Therefore, for the benefit of using messages or comments, Text Classification is therefore used in the analysis.

Today, many researchers present a wide variety of Internet-based text classification processes to analyze text, such as decision trees that help when choosing a field of study from an academic decision support system. 79.03% provide recommendations for choosing courses that matched the students' interests and abilities, with an accuracy of 61.11% (Fiarni et al., 2019). The text classification in this paper was based on 55,000 text classifications of group job titles. The performance comparison result was the most accurate at 96.25% (Mittal et al., 2020). The Learning base (Support Vector Machine) Journal classification for 251 subject articles tests the efficiency of the algorithm (Sheykhmousa et al., 2020). A Random Decision Forest using Algorithm K-Nearest Neighbor method for classifying job posting messages (Nasser & Alzaanin, 2020).

From the above research, it can be seen that there are several algorithms used to classify data, each of which has different fundamentals. Therefore, this research aims to find the best

algorithm efficiency for classifying the job qualification classes that are most closely related. By choosing algorithms with different essential characteristics, such as decision tree base, probability base (naive Bayes), learning base (Support Vector Machine), the Random Decision Forest, and K-Nearest Neighbor

This research structure consists of Part 2 - Related Theories and Research, Part 3 - Research Methodology, Part 4 - Results of Operations, and Part 5 - Summarizing Research Results.

RELATED THEORIES AND RESEARCH

The researcher studied the algorithm for classifying the data by looking at algorithms with different fundamental characteristics, namely, Decision Tree base, naive Bayes base, learning base (Support Vector Machine), Random Decision Forest base, and the base for finding the nearest neighbor (K-Nearest Neighbor), as follows:

Decision Tree

The Decision Tree Algorithm uses data to create a forecast model that is similar to a tree. Rules will be created for use in decision making. The decision tree is supervised learning, developing a classification model. A sample of pre-defined data, called a Training Set, can automatically predict a group of items that have not yet been categorized. The tree structure consists of nodes and branches, each represented by a feature of the data set to be learned and tested. Each branch of the tree is rendered in the test, and the Leaf Node represents a user-defined class. Criteria for selecting attributes to be used as tree nodes are derived from information gain calculations by considering characteristics with low information gain or low entropy, meaning that the feature has high categorization capability (Nasser & Alzaanin, 2020).

Probability Base (Naive Bayes)

Naive Bayes uses the probability principle based on Bayes' Theorem by analyzing the relationship between each independent variable and the dependent variable to create probability conditions for each connection. Theoretically, the naive Bayes effect was predicted if all independent variables were independent of each other. Independent of any variable, Bayesian theory states that P(H) is the probability that an event H will occur and P(H|E) is the probability that an event H will arise when an event E occurs from a defined variable. Bayes' Theorem concept can predict events considered from the occurrence of events, as in Equation 1.

$$P\left(\frac{H}{E}\right) = \frac{P(H/E) \times P(H)}{P(E)} \qquad 1$$

Bayesian Logistic Regression Analysis is used to help study the relationship between the dependent variable and the independent variable to predict if an event will occur, or how likely it is to happen, with one variable set or several that are expected to affect the occurrence of that event in the logistic regression analysis. Parameters must be estimated using the method (Nasser, & Alzaanin, 2020).

Learning base (Support Vector Machine)

Support Vector Machine is a learning-based algorithm. The decision plane-based classification method was used to divide the data into two parts by using the principle of creating a centerline between the groups to have the distance between the boundaries of the two groups as wide as possible. The support vector machine uses a mapping function to convert data from the original input space to feature space and to create a kernel function on the feature to measure the similarity of data (Nasser & Alzaanin, 2020).

Random Decision Forest (Random Forest)

The Random Forest Algorithm is a type of regression tree decision created by taking a random sample from the training data. Data attributes are then built into a decision tree in which an unselected part of the sample will be used to test the decision tree. This is called Out-of-Bag (OOB), or Bagging. Independent results from each decision tree are taken into account for the most significant number of votes. The Random Forest algorithm does not need to have test data to estimate the error because the OOB data was already used to test the decision tree (Mittal et al., 2020).

K-Nearest Neighbor base

The K-Nearest Neighbor Algorithm is a method used to classify by deciding which type will be used to represent the condition or new cases by examining a certain number ("K" in the Nearest Neighbor Algorithm) of the same or closest cases or requirements. The sum of the number of conditions is obtained (Nasse & Alzaanin, 2020).

The researcher studied and synthesized the relevant research. The results are summarized as shown in Table 1.

Table 1
PRESENTS RESEARCH USING THE FIVE BASIC ALGORITHMS TO CLASSIFY MESSAGES.

Research	Year	Algorithm	Research Results
Comparison of job posting message classification with machine learning methods	2020	NB	Performance comparison results showed that the Random Forest algorithm message classification had high efficiency, accuracy 98.2%, Accuracy (Precision) 99.9%, Recall 98.2%, Overall efficiency (F-measure) 99.0%.
(Ibrahim et al., 2020)		SVM	
Subject: Machine Learning and Job Posting Classification: A Comparative Study		DT K-NN RF	
Comparison of job qualification characterization using machine learning techniques (Mittal et al., 2020)	2020	BNB	Classification of texts in this article has taken 55,000 messages. Words to classify groups of positions The performance comparison results showed that the LSVM algorithm precision (Precision) 96.25%.
Subject: Performance Comparisons of Machine Learning Classification Techniques for Job Titles Using Job Descriptions		MNB RF LSVM	
Classification of candidate profile data from Twitter using the naïve Bayes algorithm according to Methodology. DISC (Hartanto et al., 2019)	2019	NB	Job applicants' profiles were classified into four groups of opinions: Job Interests, Strengths, Dominance, Job Security
Subject: Job Seeker Profile Classification of Twitter Data Using the Naïve Bayes Classifier Algorithm Based on the DISC Method		W-IDF	When classifying opinions, the naive Bayes algorithm was found to be the most accurate.
A historical data analysis tool for job rankings. Sandanayake, et al., (2018)	2018	NB	Classification of the 30 job title data types tested for job ordering using naive Bayes and SVM algorithms found naive Bayes Precision (97.33%) and Support Vector Machine Precision.
Subject: Automated CV Analyzing and Ranking Tool to Select Candidates for Job Positions		SVM	

Classification of academic papers by Learning-based algorithm and Random Decision Forest. (Sheykhmousa et al., 2020) Subject: Support Vector Machine *vs*. Random Decision Forest for Remote Sensing Image Classification: A Meta-analysis and systematic review	2020	SVM RF	Journal classification for 251 academic articles to test the efficiency of the algorithm.

Algorithm description: DT=Decision Tree, NB=naive Bayes, RF=Random Forest, SVM=Support Vector Machine, LSVM=Linear SVM, K-NN=K-Nearest Neighbor, BNB=Bernoulli's naïve Bayes, MNB=Multinomial naïve Bayes, W-IDF=Weighted-Inverse Document Frequency

METHODS OF CONDUCTING RESEARCH

The study of the most suitable algorithm for text classification has the following conceptual frameworks and research procedures:

Text Classification Concept

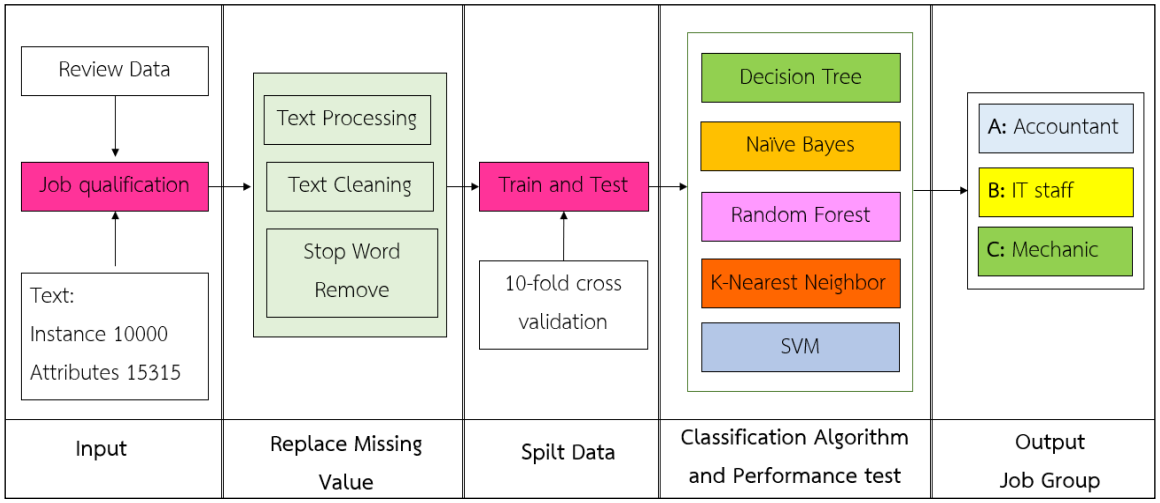

FIGURE 1
CONCEPTUAL FRAMEWORK OF TEXT CLASSIFICATION

Figure 1 shows the steps of the conceptual research framework on the analysis of text classification by evaluating the efficacy of job segmentation algorithms using machine learning methods. There are five steps to find the efficiency of text classification (1): Import study data and select 100,00 texts in the test; (2) Clean the texts, remove duplicate words, and replace missing values; (3) Divide the data into ten equal parts (10-fold cross-validation), (4) Take the model performance test data, (5) Display the results of the classification of workgroups in three classes - Class A: Accounting Workgroup, Class B: IT Group, Class C: Technician Group.

Instruments used in Research

The tool for this research is WEKA (Waikato Environment for Knowledge Analysis) version 3.7.4, a program that can be downloaded from the website under the control of the GPL License. WEKA, developed from Java, is popularly used for data mining jobs.

Algorithm Selection

Algorithm selection was a key factor in this research. The researcher studied algorithms to classify the messages, as follows: Decision Tree Base, naive Bayes, Learning Base, Support Vector Machine, Random Forest, K-Nearest Neighbor (Table 2).

Table 2
ALGORITHM SELECTION

Algorithm-based	Selected Algorithm
Tree based Classifier	Random Forest and Decision Tree (Ibrahim et al., 2020)
Probability based Classifier	Bayesian Logistic Regression and naive Bayes (Mittal et al., 2020)
Learning based Classifier	Support Vector Machine (Sheykhmousa et al., 2020)

Data Preparation and Analysis

This research gathered information about job qualifications from the database of JOBBKK Company (data as of March 9, 2020). 10,000 samples were selected. Examples of data such as "Award," "Export," "Operation," "Channel," Procurement," "Assessment," "Office," "Revenue," "Office Administration," and "Project," etc. were divided into 10 equal parts. The algorithm was tested for the efficiency of 5 types, as follows:

1. Decision Tree-based algorithm J48, developed from ID3, can be applied to discrete and continuous data. It can be seen that the capability of the algorithm of type J48 is different from ID3 because ID3 is only applicable to discrete data (Berson & Smith, 1997).
2. Naive Bayes is an algorithm that uses the principle of screening probabilities and answers by using five classification algorithms. In discriminating which data group had the highest precision, Mean Absolute Error (MAE) and Mean Squared Error (MSE) were the least. This will give the best grouping efficiency where the mean square error is a good valuation measure. This is because the mean squared error consists of variance and bias.
3. Learning base (Support Vector Machine) uses the SMO algorithm. The polynomial kernel is the best algorithm (Thakur & Biswas, 2020).
4. Random Forest Decision Tree Base. Independent results from each decision tree were taken into account for the most significant number of votes. The Random Forest algorithm does not require test data to estimate the error because the OOB data was already used to test the decision tree (Mittal et al., 2020)
5. The K-Nearest Neighbor base uses the IBK-type algorithm as the primary function. Weights, distances, and options can be set to determine k values using cross-validation (Shabani et al., 2020).

RESEARCH AND RESULTS

The research results on the analysis of text classification found the algorithm's efficiency for segmentation of job positions using machine learning methods based on 10,000 test datasets. The workgroups were classified into three groups: Class A=Accounting Group, Class B=IT Group, and Class C=Engineering Group, using five algorithms, as follows:

Decision Tree Base

Table 3
SUMMARY OF JOB CLASSIFICATION RESULTS

Correctly classified instances	8811	88.11%
Mean absolute error	0.1018	
Mean squared error	0.2599	

Table 3: Of 10000 samples, 8811 were grouped, representing 88.11%. The Mean Absolute Error (MAE) was 0.1018, which is relatively small. This shows that the grouping was

close to the actual value and had a Mean Squared Error (MSE) of 0.2599, which is relatively small. This indicates that the model is reasonably accurate.

Table 4
DETAILS OF THE GROUPING OF TASKS USING THE DECISION TREE BASE METHOD

Detailed Accuracy	A	B	C
TP Rate	0.911	0.899	0.825
FP Rate	0.037	0.101	0.035
Precision	0.945	0.774	0.914
Recall	0.911	0.899	0.825
F-Measure	0.928	0.832	0.867

Table 4. For classification responses (class A: Accounting), positive accuracy rate=0.911, positive error rate=0.037, accuracy=0.945, recall=0.911, balance=0.928, (class B: IT) positive accuracy rate=0.899, positive error rate=0.101, accuracy=0.774, recall=0.899, and balance=0.832, and (class C: Task Mechanic) positive accuracy rate=0.825, positive error rate=0.035, accuracy=0.914, recall=0.825, and balance=0.867.

Probability Base (Naive Bayes)

Table 5
SUMMARY OF JOB CLASSIFICATION RESULTS

Correctly classified instances	8697	86.97%
Mean absolute error	0.0876	
Mean squared error	0.2789	

Table 5. Of 10000 samples, 8697 grouped data, representing 86.97%, with a mean absolute error (MAE) of 0.0876, which is relatively small. This shows that the grouping is close to the actual value, and had a mean squared error (MSE) of 0.2789, which is relatively small. This indicates that the model is relatively accurate.

Table 6
DETAILS OF THE GROUPING OF TASKS USING THE NAIVE BAYES BASE METHOD

Detailed Accuracy	A	B	C
TP Rate	0.923	0.693	0.958
FP Rate	0.027	0.031	0.133
Precision	0.960	0.895	0.764
Recall	0.923	0.693	0.958
F-Measure	0.941	0.781	0.850

Table 6. Classification answers (class A: Accounting) positive accuracy rate=0.923, positive error rate=0.027, accuracy=0.960, recall=0.923, and balance=0.941 (class B: IT jobs) positive accuracy rate=0.693, positive error rate=0.031, accuracy=0.895, recall=0.693, and balance=0.781, and (class C: Task Mechanic) positive accuracy rate=0.958, positive error rate=0.133, accuracy=0.764, recall=0.958, and balance=0.850.

Learning base (Support Vector Machine)

Table 7 SUMMARY OF JOB CLASSIFICATION RESULTS		
Correctly classified instances	9273	92.73 %
Mean absolute error	0.2424	
Mean squared error	0.3057	

Table 7. 10000 samples, 9273 samples were grouped, representing 92.73%. The mean absolute error (MAE) was 0.2424, which was relatively low. This shows that the grouping is close to the actual value. And had a mean squared error (MSE) of 0.3057, which is an acceptable value. This shows that the model is relatively accurate.

Table 8 DETAILS OF THE GROUPING OF TASKS USING THE BASE METHOD. LEARNING (SUPPORT VECTOR MACHINE)			
Detailed Accuracy	A	B	C
TP Rate	0.957	0.899	0.913
FP Rate	0.045	0.037	0.029
Precision	0.937	0.904	0.934
Recall	0.957	0.899	0.913
F-Measure	0.947	0.902	0.924

Table 8. Classification answers (class A: Accounting) positive accuracy rate=0.957, positive error rate=0.045, accuracy=0.937, recall=0.957, and balance=0.947 (class B: IT jobs), positive accuracy rate=0.899, positive error rate=0.037, accuracy=0.904, recall=0.899, and balance=0.902, and (class C: Task Mechanic) positive accuracy rate=0.913, positive error rate=0.029, accuracy=0.934, recall=0.913, and balance=0.924.

Random Decision Forest

Table 9 SUMMARY OF JOB CLASSIFICATION RESULTS		
Correctly classified instances	9239	92.39 %
Mean absolute error	0.2039	
Mean squared error	0.2612	

Table 9.10000 samples, 9239 samples were grouped, representing 92.39%. The mean absolute error (MAE) was 0.2039, which was relatively low. This shows that the grouping is close to the actual value and had a mean squared error (MSE) of 0.2612, which is relatively small. This indicates that the model is rather accurate.

Table 10 DETAILS OF THE GROUPING OF WORK BY THE METHOD. RANDOM DECISION FOREST			
Detailed Accuracy	A	B	C
TP Rate	0.967	0.861	0.923
FP Rate	0.059	0.023	0.036
Precision	0.920	0.935	0.921
Recall	0.967	0.861	0.923
F-Measure	0.943	0.896	0.922

Table 10. Classification responses (class A: Accounting) positive accuracy rate=0.967, positive error rate=0.059, accuracy=0.920, recall=0.967, and balance=0.943, (class B: IT) positive accuracy rate=0.861, positive error rate=0.023, accuracy=0.935, recall=0.861, and balance=0.896, and (class C: Task Mechanic) positive accuracy rate=0.923, positive error rate=0.036, accuracy=0.921, recall=0.923, and balance=0.922.

K-Nearest Neighbor base

Table 11 **SUMMARY OF JOB CLASSIFICATION RESULTS**		
Correctly classified instances	8280	82.8%
Mean absolute error	0.1183	
Mean squared error	0.3136	

Table 11. 10000 samples, 8280 samples were grouped, representing 82.8%. The mean absolute error (MAE) was 0.1183, which was relatively small.

Table 12 **DETAILS OF JOB GROUPING, FIND THE NEAREST NEIGHBOR (K-NEAREST NEIGHBOR)**			
Detailed Accuracy	**A**	**B**	**C**
TP Rate	0.915	0.709	0.819
FP Rate	0.138	0.057	0.071
Precision	0.823	0.826	0.838
Recall	0.915	0.709	0.819
F-Measure	0.867	0.763	0.828

Table 12. Classification answers (class A: Accounting) positive accuracy rate=0.915, positive error rate=0.138, accuracy=0.823, recall=0.915, and balance=0.867, (class B: IT jobs), positive accuracy rate=0.709, positive error rate=0.057, accuracy=0.826, recall=0.709, and balance=0.763, and (class C: Task Mechanic) positive accuracy rate=0.819, positive error rate=0.071, accuracy=0.838, recall=0.819, and balance=0.828.

Summary of Efficiency Results in Group Classification

This study examined the results of group classification based on precision, mean absolute error, and mean squared error. The overall results are shown in Table 13.

Table 13 **THE RESULTS OF THE ALGORITHM FOR FINDING THE EFFICIENCY OF CLASSIFYING WORK GROUPS USING FIVE METHODS.**			
Algorithm	**Precision**	**MAE**	**MSE**
DT	88.11%	0.1018	0.2599
NB	86.97%	0.0876	0.2789
RF	92.39%	0.2039	0.2612
SVM	92.73%	0.2424	0.3057
K-NN	82.8%	0.1183	0.3136

Table 13. The Support Vector Machine (SVM) method had the most accurate text classification efficiency of 92.73%, with mean absolute and mean squared errors of 0.2424 and 0.3057, respectively.

CONCLUSION AND DISCUSSION

This research was to determine the efficiency of precision in classifying workgroups. The researcher has compiled the qualifications for the position from the database of JOBBKK Company (as of March 9, 2020), which is the most popular recruitment agency in Thailand. Data on job qualification of 10,000 samples were used to classify the workgroup using the 10-Fold Cross Validation test, which uses five algorithms: Decision Tree, naive Bayes, Probability Base Learning (Support Vector Machine), Random Forest, and K-Nearest Neighbor. Performance metrics were precision, recall, and F-measure of each algorithm. The study found that Support Vector Machine (SVM) workgroups were classified with 92.73% accuracy, followed by Random Forest (RF) 92.39%, Decision Tree (DT) 88.11%, naive Bayes (NB) 86.97%, and K-Nearest Neighbor (K-NN) 82.8%. This research is divided into class A: accounting work, class B: IT work, and class C: technician work of decision support providing advice (RS).

For study, Sanz-Garcia, et al., (2012). Study of the popular online hotel reservations today. But the problem is that customers are more likely to book at the last minute. This situation has a significant impact on further operations. This research, therefore, proposes a new model to predict bookings to prevent the effect of last-minute bookings. The proposed model is a combination of the Genetic Algorithm and k-nearest neighbors. The results showed that the improved model significantly improved booking data. Compared to traditional methods, And can improve a helpful demand forecasting calendar. This study is consistent with Song et al. (2015) studying artificial muscle activators used in medical robots for human rehabilitation. using a power-to-weight ratio and inherent safety characteristics. But the problem was that it was difficult to obtain a precise model in analyzing this model. Due to the compression of air and flexible materials as the core, this actuator model is neither linear nor time-variant. Three approaches to modeling were proposed back Propagation algorithm, Genetic Algorithm, and a combination of BP and Modified Genetic Algorithm. An integrated approach has better performance.

Same as Angskun & Angskun (2014). This research presents an individual tourist attraction recommendation system by cluster technique. The K-Means technique uses Hatigan's law (Hartigan’s Rule), and Analytical Hierarchy (AHP) were combined to create a tourist attraction ranking model. According to preferences and conditions governing to be most suitable Ranked for tourists the rating assessment uses statistics to measure the correlation between variables that have a rating scale: The Spearman Correlation Coefficient, data obtained from 400 tests, the order of 50 attractions, tourists. Four hundred people the process of combining grouping techniques and an analytical hierarchy process consisted of four steps: 1) building a clustering model 2) calculating the priority of each criterion 3) calculating the priority of the criterion. Each group 4) Prioritize the attractions the model evaluation results revealed that the sequence obtained from the model is correct. This study is Precise as the tourists have arranged which no matter how much the number of attractions increases, it does not affect the model's accuracy. The Spearman correlation coefficient was closer to 1 for every number of attractions tested. This study is consistent with da Silva, et al., (2016). This research optimizes parameters, selects the best algorithms and techniques. Calculate the matrix to calculate the predictive rating (rating) for the target audience. The algorithms used in the research were collaborative filtering algorithms using Pearson correlation, Euclidean, Spearman, Tanimoto, and Loglikelihood and optimization using the Genetic Algorithm (GA) algorithm, which is a movie dataset (MovieLens), a program that uses MATLAB. Performance evaluation is based on error metrics, averaging the root mean square of other error values (Root Mean Square). Error: RMSE) is to consider the difference between the predicted value and the actual value. Suppose the obtained value is equal to zero. In that case, it indicates that there is no error at all in the research focusing on adjusting the efficiency of the recommended system to be more accurate. Research has shown excellent results and low-cost computation by combining a collaborative filtering

algorithm and a Genetic Algorithm (GA). Loglikelihood and Tanimoto are the best, while GA parameter adjustments become more and more accurate.

REFERENCES

Angskun, T., & Angskun, J., (2014). A personalized system for travel attraction Recommendation using a clustering technique and an analytic hierarchy Process. *Suranaree Journal of Science and Technology, 8*(2), 87-109.

Berson, A., & Smith, S.J. (1997). Data warehousing, data mining, and OLAP, McGraw-Hill, New York, An evolutionary approach for combining results of recommender systems techniques based on collaborative filtering**.** *Expert Systems with Applications 53*(1), 204-218.

Fiarni, C., Sipayung, E.M., & Tumundo, P.B. (2019). Academic decision support system for choosing information systems sub majors programs using decision tree algorithm. *Journal of* TT *Information Systems Engineering and Business Intelligence*, *5*(1), 57-66.

Hartanto, A.D., Utami, E., Adi, S., & Hudnanto, H.S. (2019). Job seeker profile classification of twitter data using the naïve Bayes classifier algorithm based on the DISC method. In *2019 4th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE)* 533-536. IEEE.

Kanimozhi, K.V., & Venkatesan M. (2015). "Unstructured Data Analysis - A Survey." *International Journal of Advanced Research in Computer and Communication Engineering*, 4(3), 223-225.

Larose, D.T. (2005). *Discovering knowledge in data*: *An introduction to data mining*. John wiley and sons, Hoboken, N.J., 222.

Mittal, S., Gupta, S., Shamma, A., Sahni, I., & Thakur, N. (2020). Performance comparisons of machine learning classification techniques for job titles using job descriptions.

Nasser, I., & Alzaanin, A.H. (2020). Machine learning and job posting classification: A comparative study. *International Journal of Engineering and Information Systems (IJEAIS) ISSN*, 6-14.

Sandanayake, T.C., Limesha, G.A.I., Madhumali, T.S.S., Mihirani, W.P.I., & Peiris, M.S.A. (2018). Automated CV analyzing and ranking tool to select candidates for job positions. In *Proceedings of the 6th International Conference on Information Technology: IOT and Smart City* 13-18.

Sanz-Garcia (2012). Application of Genetic Algorithms to Optimize a Truncated Mean k-Nearest Neighbours Regressor for Hotel Reservation Forecasting. *Hybrid Artificial Intelligent Systems. HAIS*, 7208, 79-90.

Shabani, S., Samadianfard, S., Sattari, M.T., Mosavi, A., Shamshirband, S., & Várkonyi-Kóczy, A.R. (2020). Modeling pan evaporation using gaussian process regression, k-nearest neighbors, random forest, and support vector machines, Comparative analysis. *Atmosphere*, *11*(1), 66.

Sheykhmousa, M., Mahdianpari, M., Ghanbari, H., Mohammadimanesh, F., Ghamisi, P., & Homayouni, S. (2020). Support vector machine *vs*. random forest for remote sensing image classification: A meta-analysis and systematic review. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*.

Song. (2015). Modeling of pneumatic artificial muscle using a hybrid artificial neural network approach. *Mechatronics, 31*(2), 124-131.

Thakur, D., & Biswas, S. (2020). Smartphone based human activity monitoring and recognition using ML and DL: a comprehensive survey. *Journal of Ambient Intelligence and Humanized Computing*, 1-12.