UNIVERSITAS UTPADAKA SWASTIKA
Jl. KS Tubun No.11 Tangerang 15112 Banten, Telp. (021) 5589161-62 Fax. (021) 5589163
JISI Jurnal Ilmiah Sistem Informasi
Vol. 2 No. 02 (2024)
ISSN: 3046-711X

Detect Classification of Employees Tending to Move Work With the Naive Bayes Algorithm

Lukas Umbu Zogara

Teknologi Informasi, Utpadaka Swastika University, Tangerang, Indonesia, 15112
E-mail: [1],lukasumbuzogara68@gmail.com

Accepted: October 29, 2024 | Published: October 31, 2024

ABSTRACT

In recent years due to economic conditions and the uncertain situation of a country, many employees with a certain level of education, work experience and different countries in development and the level of income per capita of a country and several other factors cause many employees to move to a new place. the. Because there are many factors that cause employees to move careers and advances in information technology also make it difficult to predict what factors affect decision-making employees to move to work to new places. Therefore, it is necessary to know what factors and conditions or what are the employees so that they have a tendency to move to work so that the company can prevent, anticipate and immediately seek other solutions early if this condition must occur from its employees. Based on the problems described, this study discusses the classification of employees who have a tendency to change workplaces using the Naive Bayes algorithm. The goal is expected to identify the dominant factors in influencing employees to change workplaces. From the results of the research conducted by Naive Bayes, it was found that 3 dominant factors were influencing, namely the STEM area of expertise, the size of the company size and the level of education as well as the accuracy level of 80.79% and AUC 0.816.

KEYWORDS: Employees, Classification, Workplace, Algorithm, Naive Bayes.

1. Introduction

With the very rapid development of science and technology, it can help humans to solve all the problems and needs that are increasingly numerous and complex. In recent years due to economic conditions and the uncertain situation of a country, many employees with a certain level of education, work experience and different countries in development and the level of income per capita of a country and several other factors cause many employees to move to a new place. On the one hand, we also understand the human desire to have a better life, so employees must compete for jobs and better careers. This is what the authors adopt to be the subject of research, namely predicting the classification of employees who have a tendency to change jobs. Because there are many factors that cause employees to move careers and advances in information technology also make it difficult to predict what factors influence the decision-making employees to move to work to a new place.

Therefore, every company feels it is necessary to know what factors and conditions or what are the employees so that they have a tendency to move to work so that the company can prevent, anticipate and immediately seek other solutions early if this condition must occur from its employees. This is something a company should understand, because the progress of a company depends on the performance of its employees. So, if there are employees who are experienced, trained and have qualified skills for the development and progress of the company suddenly move to another place, of course this is undesirable to happen in the company.

Based on the problems described, this study discusses predicting the classification of employees who have a tendency to change workplaces with the Naive Bayes Algorithm. The purpose of holding it is to know the classification of employees who have a tendency to move to work and the dominant factors or attributes that determine the influence of employees to change workplaces.

2. Methodology

2.1. Naive Bayes Algorithm

The Naive Bayes algorithm is one of the algorithms contained in the classification technique. Naive Bayes is a classification using probability and statistical methods proposed by the British scientist Thomas Bayes, which predicts future opportunities based on previous experience, so it is known as the Bayes Theorem. The theorem is combined with Naive where it is assumed that the conditions between attributes are independent of each other. The Naive Bayes classification assumes that the presence or absence of certain features of a class has nothing to do with the characteristics of other classes

$$P(H|X) = \frac{P(X|H).P(H)}{P(X)}$$

Information:

- X : Data with an unknown class
- H : The data hypothesis X is a specific class
- P(H|X): Probability of hypothesis H based on condition X (posteriori probability)
- P(H) : The probability of the hypothesis H (prior probability)
- P(H|X): Probability of X based on hypothetical conditions
- H P (X): Probability X

By using the above equation, the data that has been obtained can be processed with the Naive Bayes algorithm for assessing the data to be classified.

2.2. Research Methods

The research method used by the author in this study is an experimental research method, with the following stages consisting of 5 stages, namely the data collection stage, the data preprocessing stage, the feature selection stage, the data classification stage, and the classification result analysis stage. These stages are described in figure 1 below

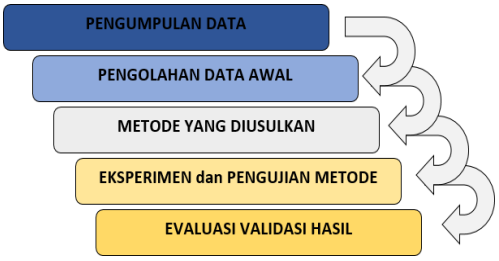


Figure 1. Research Methodology

The method used by the author in this study is an experimental research method, with the following stages:

a. Data collection

The data in this study were taken from the site www.kaggle.com, namely HR Analytics data 2019-2020: Job Change of Data Science. This data collection is done by doing a valid data filter. The author uses 19,158 data with 11 attributes as follows:

1. City _development _index,
2. Training hours,
3. Gender,
4. Relevant _experience,
5. Enrolled university,
6. Education level,
7. Major _ discipline,
8. Experience,
9. Company size,
10. Company type,
11. Last _new _ job.

b. Initial Data Processing

At the initial processing stage of the training dataset for data classification, the preprocessing stage is used so that missing data, symbols or punctuation marks can be removed. This process is the initial stage before testing the method, where the training dataset to be used is checked first for missing values. The preprocessing stage of the training dataset used by the author to use Rapid Miner to produce a training dataset in good condition

c. The Proposed Method
The method used in this research is the use of the Naive Bayes algorithm method

d. Experiment and Testing Methods
In testing the method proposed in this experiment, software is used as a tool, namely Rapid Miner. With the data used is the training data 19,158 data and with 11 attributes, where 1 attribute as a label and 10 attributes as regular.
In this study, the Attribute Label chosen was the Attribute Relevant Experience, in the sense that this test is that employees change jobs in accordance with work experience or field expertise

e. Evaluation of Results Validation
After knowing the results of the predictions in the classification process, the next step is testing. The resulting data is tested for accuracy by creating confusion matrix and the level of accuracy is calculated. After the accuracy is known, the resulting data can be known for its precision and recall values and visualized with the Area Under Curve (AUC). In addition, the level of accuracy of each curve can be tested again to see the level of presentation.

3. Results and discussion

3.1. Experiments and Research of the Naïve Bayes Algorithm Method
Data processing uses Rapid Miner tools for research the Naïve Bayes algorithm with 10 validations. The research process was carried out with the HRD train dataset that had been preprocessed beforehand, and the Naïve Bayes Algorithm research process was carried out as shown in figure 2 below.

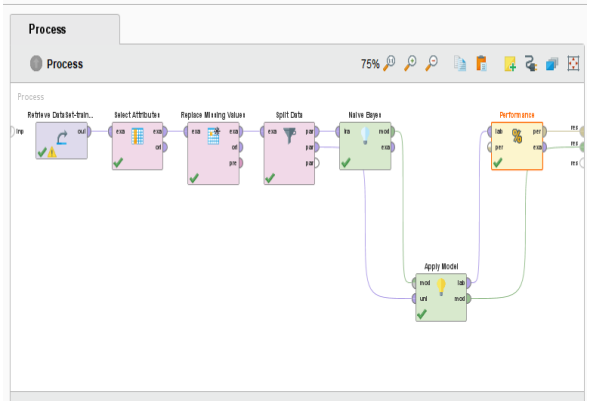


Figure 2. Research process of Naïve Bayes Algorithm with Rapid Miner

3.2 The results of the Naïve Bayes Algorithm research process can be seen in each regular attribute with the Relevant Experience attribute label.

In the following pictures is a Chart of 10 Attributes (regular) with Attributes (Label) Relevant-experience accordingly and incompatible.

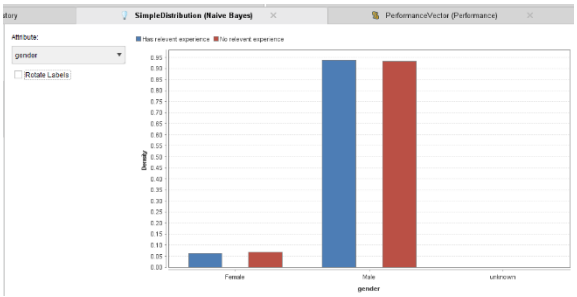


Figure 3. Chart of Attribute Gender

It can be seen from the results of the picture that the majority of employees who have the tendency to move to work are men

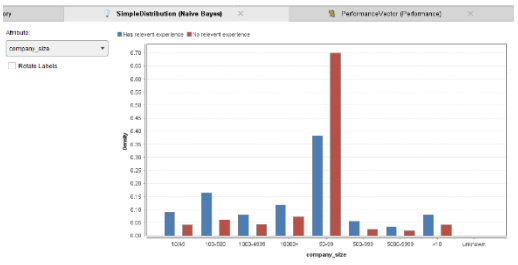


Figure 4. Chart of Attribute Company size

From the chart, that the companies that are interested in employees who have a tendency to move to work are middle-class companies (50-99)

UNIVERSITAS UTPADAKA SWASTIKA
Jl. KS Tubun No.11 Tangerang 15112 Banten, Telp. (021) 5589161-62 Fax. (021) 5589163
JISI
Jurnal Ilmiah Sistem Informasi
Vol. 2 No. 02 (2024)
ISSN: 3046-711X

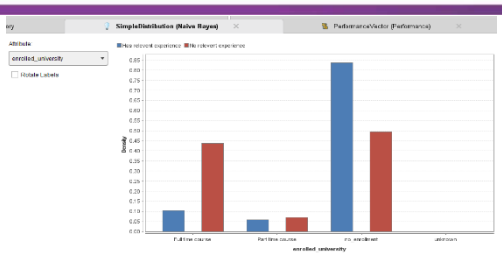


Figure 5. Chart of Attribute Enrolled University

It can be seen from the chart, that the majority of employees who have the tendency to move their jobs are not enrolled in tertiary institutions (no enrollment)

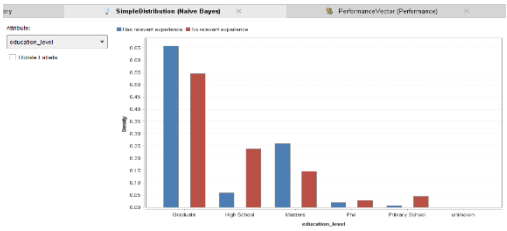


Figure 6. Chart of Attribute Education Level

It can be seen from the chart, that most of the employees who have the tendency to move to work are those who have a bachelor's level of education

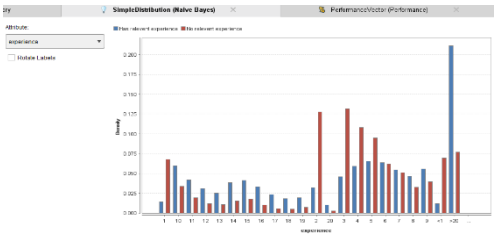


Figure 7. Chart of Attribute Experience

From the chart, it can be seen that most of the employees who have a tendency to move to work are those who have had a long work experience of up to 20 years.

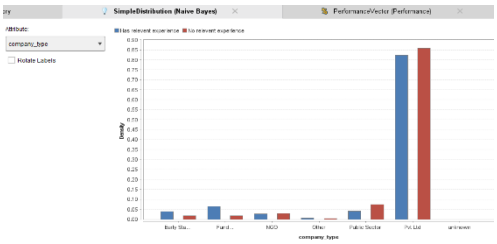


Figure 8. Chart of Attribute Company Type

From the chart, it can be seen from the test results that most of the employees moved to work to and from private companies (Pvt Ltd)

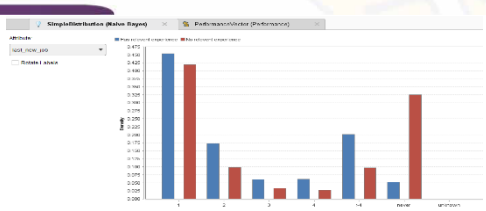


Figure 9. Chart of Attribute Last New Job

From the chart, you can see the results of the test that the majority of employees move to work to new places after they have worked for an average of 1 year in the new place.

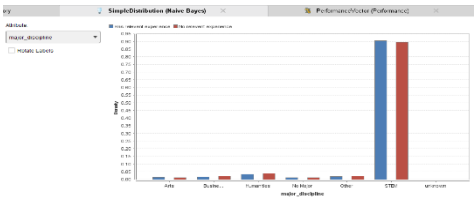


Figure 10. Chart of Attribute Major Discipline

From the chart, it can be seen that the results of the test are that most employees who have a tendency to move their jobs have disciplines of expertise in the fields of Science Technology, Engineering Mathematical (STEM)

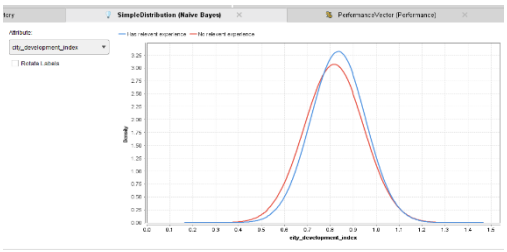


Figure 11. Chart of Attribute City Development Index

From the chart, it can be seen that the results of the test are that most employees who have a tendency to move their jobs are located or live-in cities that have a high average growth rate index (0.8-0.9)

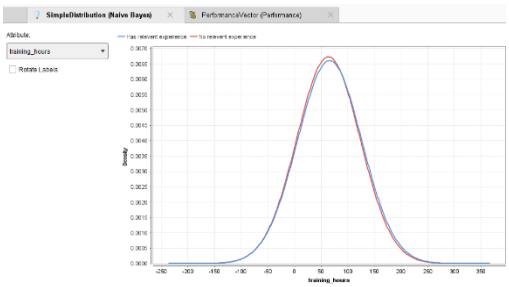
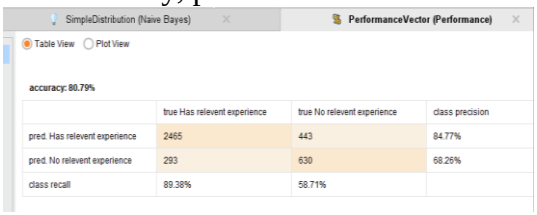


Figure 12. Chart of Attribute Training Hours

UNIVERSITAS UTPADAKA SWASTIKA
Jl. KS Tubun No.11 Tangerang 15112 Banten, Telp. (021) 5589161-62 Fax. (021) 5589163
JISI Jurnal Ilmiah Sistem Informasi
Vol. 2 No. 02 (2024)
ISSN: 3046-711X

From the chart, it can be seen that the results of the testing for most employees who have a tendency to change jobs are those who have undergone or attended training with a lot of material related to their fields, on average 60-80 hours of training.

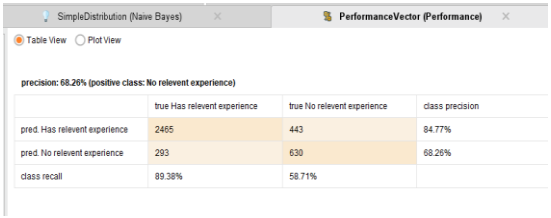
3.3. The results of the research process using the Naïve Bayes algorithm to classify HRD data in this research test resulted in accuracy, precision and recall


SimpleDistribution (Naive Bayes) | PerformanceVector (Performance)
Table View | Plot View
accuracy: 80.79%

	true Has relevent experience	true No relevent experience	class precision
pred. Has relevent experience	2465	443	84.77%
pred. No relevent experience	293	630	68.26%
class recall	89.38%	58.71%	

Figure 13. Accuracy results from the testing process Naive Bayes

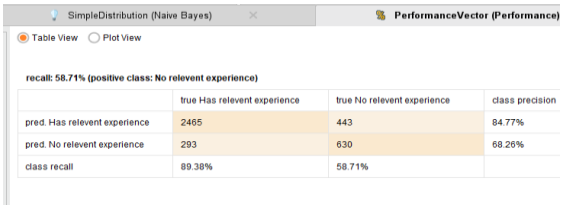
From the chart, it can be seen that the results of testing this research with the Naive Bayes algorithm are an accuracy level of 80.79%.


SimpleDistribution (Naive Bayes) | PerformanceVector (Performance)
Table View | Plot View
precision: 68.26% (positive class: No relevent experience)

	true Has relevent experience	true No relevent experience	class precision
pred. Has relevent experience	2465	443	84.77%
pred. No relevent experience	293	630	68.26%
class recall	89.38%	58.71%	

Figure 14. Precision results from the testing process Naive Bayes

From the chart, it can be seen that the results of testing this research with the Naive Bayes algorithm are the value of the precision level of 68.29%


SimpleDistribution (Naive Bayes) | PerformanceVector (Performance)
Table View | Plot View
recall: 58.71% (positive class: No relevent experience)

	true Has relevent experience	true No relevent experience	class precision
pred. Has relevent experience	2465	443	84.77%
pred. No relevent experience	293	630	68.26%
class recall	89.38%	58.71%	

Figure 15. Recall results from the testing process Naive Bayes

From the chart, it can be seen that the results of testing this research using the Naive Bayes algorithm are 58.71% recall rate.

3.4. The results of the research process with the Naïve Bayes algorithm to classify HRD data in this research produce AUC and ROC graphs in figure below.

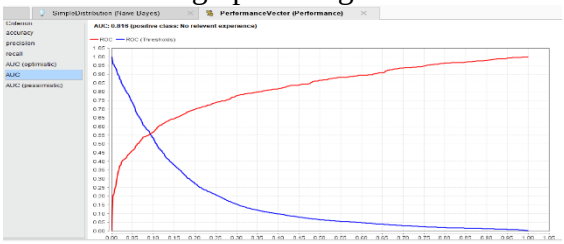


Figure 16. AUC results from the testing process Naive Bayes

From the chart, it can be seen that the results of testing the Naive Bayes AUC algorithm are 0.816

3.5. The results of the research process with the Naïve Bayes algorithm to classify HRD data in this research produce Performance Vector in figure below

PerformanceVector

```
PerformanceVector:
accuracy: 80.68%
ConfusionMatrix:
True:   Has relevent experience No relevent experience
Has relevent experience:        2462    444
No relevent experience: 296     629
precision: 68.00% (positive class: No relevent experience)
ConfusionMatrix:
True:   Has relevent experience No relevent experience
Has relevent experience:        2462    444
No relevent experience: 296     629
recall: 58.62% (positive class: No relevent experience)
ConfusionMatrix:
True:   Has relevent experience No relevent experience
Has relevent experience:        2462    444
No relevent experience: 296     629
AUC (optimistic): 0.816 (positive class: No relevent experience)
AUC: 0.816 (positive class: No relevent experience)
AUC (pessimistic): 0.816 (positive class: No relevent experience)
```

Figure 17. Performance Vector results from the testing process Naive Bayes

From the results of the image, it can be seen that the results of the vector performance of this research test using the Naive Bayes algorithm are in accordance with the relevant experience 2462 and not according to 629. And it is known that the results of the accuracy of the recall precision and AUC are also known.

3.6. The results of the testing process using the Naïve Bayes algorithm for classifying HRD data in this research resulted in a Simple Distribution in the figure below.

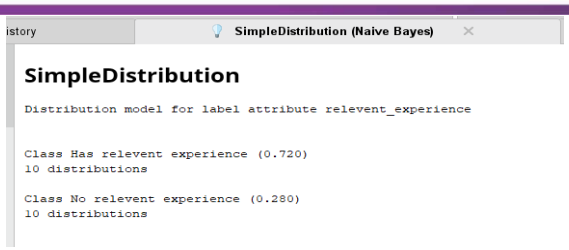


Figure18. Simple Distribution results from the testing process Naive bayes

From the chart, it can be seen that the results of testing with the Naive Bayes algorithm produce Simple Distribution showing that the Has Relevant Experience Class is 0.720 (10 distributions) and the Class No Relevant Experience is 0.280 (10 distributions).

3.7. The results of the testing process using the Naïve Bayes algorithm for classifying HRD data in this research produce a list of factors that affect employees to change places of work as follows

Table 1. List of Factors That Affect Employees To Change Employment

Column Weights

Influence of columns on the predictions of this specific model

Column	Weight
major_discipline	0.205
company_size	0.138
education_level	0.114
enrolled_university	0.103
last_new_job	0.087
experience	0.066
city_development_index	0.044

Based on the list, it can be seen that the results of research with the Naive Bayes algorithm, of the 10 attributes or factors that influence employees to change their place of work, there are 7 dominant factors in the following order:

1. Major discipline 0.205
2. Company size 0.138
3. Education level 0.114

3.8. The following is a comparison of the results of research using the data mining algorithm method with the same HRD dataset on several datamining algorithm methods are Naive Bayes, Generalized Linear Model, Deep Learning and Decision Tree

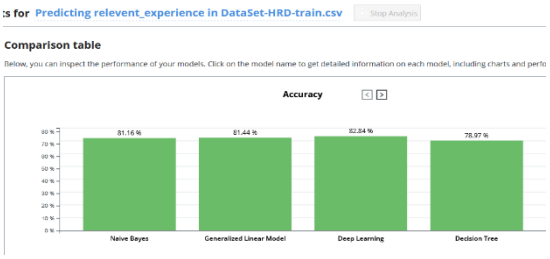


Figure 19. Comparison of the results of Accuracy from research with several data mining algorithms

From the chart results, it can be seen that the accuracy value of Naive Bayes is 81.16% and each of the other data mining algorithm research, the average is 81.10%.

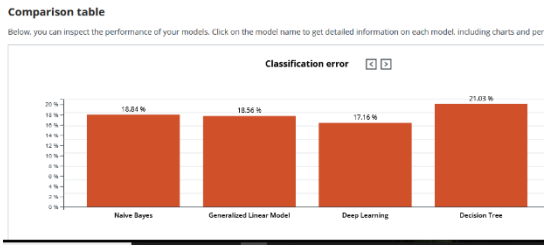


Figure 20. Comparison of the results of Classification Error from research with several datamining algorithms

From the chart results, it can be seen that the Classification error value of Naive Bayes is 18.84% and each of the other data mining algorithms research, the average is 18.88%.

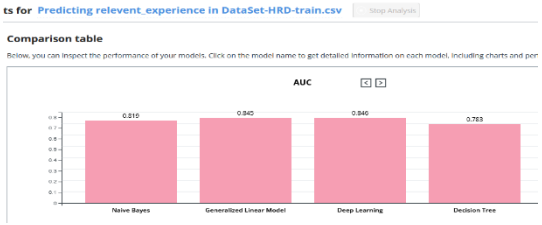


Figure 21. Comparison of the results of AUC from research with several datamining algorithms

From the chart results, it can be seen that the AUC value of Naive Bayes is 0.819 and each of the other data mining algorithms research, the average is 0.823.

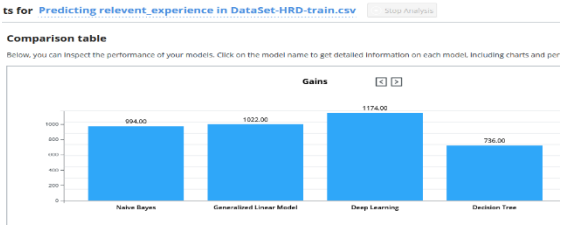


Figure 22. Comparison of the results of Gains from research with several datamining algorithms

From the chart results, it can be seen that the Gains value of Naive Bayes is 994 and each of the other data mining algorithms research, the average is 982.

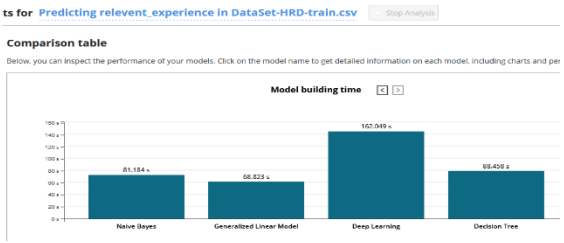


Figure 23. Comparison of the results of Model Building Time from research with several datamining algorithms

From the chart results, it can be seen that the Model Building Time value from Naive Bayes is 81.184s and each of the other data mining algorithms research, the average is 100.128s.

4. Conclusion

Based on research to predict the classification of employees who have a tendency to move to work using the Naive Bayes Algorithm method, the following points are obtained.

1. In this research with 10 (attribute) factors that influence employees to change workplaces, research using the Naive Bayes algorithm shows that there are 3 factors that are very dominant in influencing the tendency of employees to change workplaces, namely the skill factor with discipline in the fields of expertise in Science Technology, Engineering. Mathematical (STEM) with the highest score of 0.205, followed by the company size factor with a value of 0.138 and the next factor is the level of employee education with a value of 0.114.
2. The results of the research process with the Naïve Bayes algorithm to classify HRD data in this research produce Performance Vector with an accuracy value of 80.79% and Area Under Curve (AUC) 0.816.
3. From the chart, it can be seen that the results of research with the Naive Bayes algorithm produce Simple Distribution showing that the Has Relevant Experience Class is 0.720 (10 distributions) and the Class No Relevant Experience is 0.280 (10 distributions

References

[1] A. S. De Oliveira Góes and R. C. L. De Oliveira, "A Process for Human Resource Performance Evaluation Using Computational Intelligence: An Approach Using a Combination of Rule-Based Classifiers and Supervised Learning Algorithms," in IEEE Access, vol. 8, pp. 39403-39419, 2020, doi: 10.1109/ACCESS.2020.2975485.

[2] Liangjun Yu, Shengfeng Gan,Yu Chen,And Meizhang He, “Correlation-Based Weight Adjusted Naive Bayes”. IEEE, March 2020.Doi:10.1109/ACCESS.2020. 2973331

[3] S. Wang, J. Ren and R. Bai, "A Regularized Attribute Weighting Framework for Naive Bayes," in IEEE Access, vol. 8, pp. 225639-225649, 2020, doi: 10.1109/ ACCESS. 2020.3044946.

[4] A. Radaideh, F. Dweiri and M. Obaidat, "A Novel Approach to Predict the Real Time Sentimental Analysis by Naive Bayes & RNN Algorithm during the COVID Pandemic in UAE," 2020 International Conference on Communications, Computing, Cybersecurity, and Informatics (CCCI), 2020, pp. 1-5, doi: 10.1109/ CCCI49893.2020.9256587.

[5] Z. Xue, J. Wei and W. Guo, "A Real-Time Naive Bayes Classifier Accelerator on FPGA," in IEEE Access, vol. 8, pp. 40755-40766, 2020, doi: 10.1109/ACCESS. 2020.2976879.

[6] C. K. Aridas, S. Karlos, V. G. Kanas, N. Fazakis and S. B. Kotsiantis, "Uncertainty Based Under-Sampling for Learning Naive Bayes Classifiers Under Imbalanced Data Sets," in IEEE Access, vol. 8, pp. 2122-2133, 2020, doi: 10.1109/ACCESS.2019.2961784.

[7] F. Alzamzami, M. Hoda and A. E. Saddik, "Light Gradient Boosting Machine for General Sentiment Classification on Short Texts: A Comparative Evaluation," in IEEE Access, vol. 8, pp. 101840-101858, 2020, doi: 10.1109/ACCESS.2020.2997330.