

Application of Naïve Bayes Algorithm for Measuring the Suitability of the Work Position in Ministry of Home Affairs

Penny Dwi Harnaning
Universitas Gunadarma
Program Magister Manajemen
Sistem Informasi

Muhammad Subali, PhD
Sekolah Tinggi Teknik Multimedia
Cendekia Abditama
Teknik Informatika

Karmilasari
DR, SKom, MM
Universitas Gunadarma
Fakultas Ilmu Komputer dan
Teknologi Informasi

ABSTRACT

The Ministry of Home Affairs a ministry which is under and responsible to the President through the Minister, who has the task of holding affairs in the field of internal government to assist the President in organizing the governance of the country. In carrying out governmental tasks of domestic government, the Ministry of Home Affairs requires qualified State Civil Apparatuses and their placement must be in accordance with their competencies. Until now there are still problems with the placement of employees in positions that are not in line with their competencies. Data mining in this study was implemented to measure the level of suitability of the employees to the positions they occupy by involving large amounts of data, the technique used for classification is the Naïve Bayes algorithm which is used to determine the extent of suitability between employees and the occupied positions. The attributes used consist of three attributes, namely the level of education, education and training, and rank / classification. The object of this study are 4202 employee profile data consisting of 305 employees who hold the position of Administrator (III. A), 16 employees who occupy the position of Administrator (III. B), 806 employees who occupy Supervisory positions (IV. A), 82 employees occupying the Supervisory position (IV. B), and 2993 employees who hold the Implementing position. Accuracy of suitability between employees and the position they occupy is based on testing of the results of the classification of the Naïve Bayes algorithm using 90% training data and 10% testing data is 83%.

General Terms

Data Mining, Naïve Bayes Algorithm, Employee Placement.

Keywords

Data Mining, Naïve Bayes Algorithm, Ministry of Home Affairs, Employee Placement.

1. INTRODUCTION

Human resources are valuable assets with an important role for an organization or company to achieve their goals. If human resources are not managed properly and correctly, then it is not possible for an organization or company to run properly and optimally as expected.

To achieve organizational goals and avoid mistakes in managing human resources, then in terms of the placement of, the principle of "The right man in the right place" applies which means that in placing employees in a position in terms of expertise, both in terms of education and training that have been followed by these employees.

The Personnel Bureau as the staffing officer in the Ministry of

Home Affairs has tried to place civil servants in the right position in an organizational unit. However, based on the data seen, there are still irregularities found in the implementation of the placement of civil servants, including there are still civil servants who are placed in organizational units that are not in accordance with their education, training followed, and rank of the civil servants.

Due to the placement of civil servants who are not based on their latest education, training, and rank, there are still civil servants who are placed based on vacant positions in organizational units regardless of their latest educational background, there are still civil servants who do not have leadership education and training III and IV occupying administrative positions with the administrator (echelon III. A and III. B) and supervisor (echelon IV. A and IV. B), and still there are civil servants with high rank occupying implementary positions.

To measure the suitability of civil servants with the positions they occupy, competency assessments can be carried out by conducting an assessment of each civil servant who occupies a position. However, the implementation of this competency assessment is still not going well because of the large number of civil servants within the Ministry of Home Affairs, particularly civil servants with administrative positions. It is necessary to process data on civil servants in various ways, one of which is to utilize data mining.

2. LITERATURE REVIEW

2.1 Definition of Data Mining

Data mining is an iterative and intensive search (discovery) activity that aims to extract knowledge and data sets that did not / did not yet carry important meanings. The intended knowledge can be in the form of patterns, correlation, changes, anomalies, structures, formulas, rules, or models that emerge from the data. The dataset, has various attributes / features and with or without class / label / target. Each attribute / feature can be of categorical / nominal or numeric type. Data mining helps in making decisions in an era where all decisions are based on evidence which is widely obtained from data. The process starts from taking data from existing sources (databases or data warehouses) which can be in the form of documents, files, images, articles, photos, and so forth.

2.2 Data Pre-processing

Data processing is the process of transforming raw data into an understandable format. Raw data is usually incomplete (missing values), inconsistent, noisy (contain errors) and there

are extreme values (outliers). So before data mining is performed, these conditions must be solved first. Data Pre-processing consists of several stages, as follows:

1. Data Cleaning.
The process of cleaning up data that is incomplete, error, outlier and inconsistent.
2. Data Integration.
Merging data from various data sources (databases, files, etc.) into one data warehouse (data warehouse).
3. Data Reduction.
Data mining process with big data will take a long time. So the data must be reduced, both in terms of dimensions (number of variables) and volume (number of cases / objects). In this process feature selection can be performed or compressed data into fewer dimensions with Principal Component Analysis techniques.
4. Data Transformation.
Data changes to a form suitable for subsequent data mining processes. This process includes normalization, aggregation and smoothing.

2.3 Naïve Bayes

Naïve Bayes is a probability / opportunity based classification method. This method calculates a set of probabilities by adding up the frequency and combination of values from a given dataset. This method uses the Bayes Theorem and assumes independent or non-interrelated attributes given by values to class variables.

The Bayes Rule / Theorem is as follows:

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

Where:

X = Data with unknown classes

H = Data hypothesis is a specific class

P(H|X) = Hypothesis H probability is based on condition X (posterior probability)

P(H) = Probability of Hypothesis H (prior probability)

P(X|H) = Probability of X based on conditions on the hypothesis H

P(X) = Probability of X.

There are two assumptions in the Naïve Bayes method, namely:

1. All attributes are priority / equally important.
2. All attributes are statistically independent (the value of one attribute is not related to the value of another attribute)

These assumptions are mostly not always true, but in practice even if the assumptions are not fulfilled, this method still produces good results. Because in Naïve Bayes it is assumed that all attributes are mutually independent, the equation can change into:

$$P(C|F_1 \dots F_n) = \frac{P(C)P(F_1 \dots F_n|C)}{P(F_1 \dots F_n)}$$

$$P(F_1 \dots F_n)$$

Where:

C = Class

$F_1 \dots F_n$ = Characteristics of the required clues
For the classification

$P(C|F_1 \dots F_n)$ = The probability of class C when the condition characteristic of $F_1 \dots F_n$

P(C) = Class C probability (prior probability)

$P(F_1 \dots F_n|C)$ = Probability of the characteristic $F_1 \dots F_n$ in line with class C conditions

$P(F_1 \dots F_n)$ = Probability characteristic of $F_1 \dots F_n$

2.4 Steps of the Naïve Bayes algorithm

The steps of the Naïve Bayes algorithm are as follows:

1. Prepare a dataset.
2. Calculate the number of classes in the training data.
3. Calculate the same number of cases with the same class.
4. Multiply all results according to the testing data that the class will look for.
5. Compare results per class, the highest value is set as a new class.

2.5 Microsoft Excel

MS Excel is a premium application made by the world's largest software company, Microsoft. This application is the accounting and mathematical office application used by all people of the world today, for reasons of exceptional convenience and features. A very important feature for accountants, statisticians, researchers or research students is the statistical and mathematical function or formulas, so Microsoft's MS Excel application is very useful for database system users for data mining and reporting analysis.

3. RESEARCH METHOD

The research method in this study consisted of several stages carried out, as shown in the following chart:

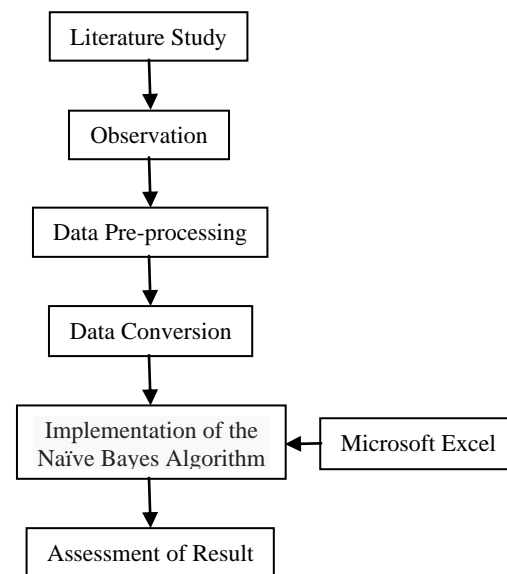


Fig 1: Block diagram of research method

Caption:

1. Literature Study

At this stage, researchers search for written sources related to the problem to be studied, such as determining the research theme, problem formulation, methods, and objectives of the study, before conducting direct research on the research object.

2. Observation

At this stage, the researcher conducted interviews with several sources related to the implementation and decision making in the placement of civil servants within the Ministry of Home Affairs, gathering information in the form of regulations relating to the placement of civil servants in administrative positions, taking data from sources.

3. Data Pre-processing

At this stage, the researcher conducted the process of changing the data of the raw data taken through the Ministry of Home Affairs Management Information System database into an understandable format. The data processing stage consists of the following 4 (four) steps:

a. Data Cleaning

In this process, from the data obtained through the Personnel Management Information System, sorting is done, namely by removing data that is not used. The sorting of data is carried out on data of civil servants with daily executive positions (PLH), job executing positions (PLT), and special staff of the Minister of the Interior, because daily executive positions (PLH) and job implementing positions (PLT) are not definitive positions of employees civil service, while the special staff position of the Minister of the Interior is not an administrative position for research. From the results of data cleaning, the amount of data obtained for this study was 4202 data.

b. Data Transformation

In this process, the data that has been cleaned will be changed into consist of various names into several kinds of names that have the same meaning.

c. Data Integration

In this process, the data that has been cleaned and changed will be merged with other data that are interrelated to produce an attribute that can be used for the process of measuring the suitability of positions with employees.

d. Data Reduction

In this process, from the data will be reduced of the attributes that are not used for the process of measuring the suitability of the position with employees, such as NIP attributes, Employee Names, and Components will be removed.

4. Data Conversion

At this stage, researchers conducted a process of comparing data on the employee profile attributes with the data on the job condition attributes. The attributes used such as Education, Training, and Rank, derived from data tables in the Ministry of Home Affairs Civil Service Information System that illustrates the real conditions of civil servants. While the source of attribute position requirements as comparative data to obtain a dataset used in the classification process is obtained from regulations that still applies. The results of the comparison of each employee's profile attributes with the position

requirements attribute, obtained four attributes that will be used in classification techniques, namely Education Level, Training, Rank, and Classification.

5. Implementation of the Naïve Bayes Algorithm

At this stage, the researcher conducted the processing of data that has been preprocessed to measure the suitability of positions with employees using the Naïve Bayes algorithm.

6. Implementation of Naïve Bayes Algorithm using Microsoft Office

At this stage, the data is processed using the Naïve Bayes algorithm with the help of Microsoft Excel tools regarding the suitability of the employees and their positions.

7. Assessment of Results

At this stage the results will be calculated using the Naïve Bayes algorithm formula, both manually and using Microsoft Office tools which will be explained in the Results and Discussion chapter.

4. RESULTS AND DISCUSSION

4.1 Implementation of the Naïve Bayes Algorithm

The Naïve Bayes algorithm implementation consists of the following steps:

1. Prepare a dataset. The dataset used in processing this data uses 4202 training data, and 420 testing data.
2. Calculate the number of classes in the training data. Classes in training data consist of two categories, namely suitable and unsuitable: Number of suitable classes = 3379. The number of unsuitable classes = 823. Then, $P(C = \text{"Suitable"}) = 3379/4202 = 80\%$. $P(C = \text{"Unsuitable"}) = 823/4202 = 20\%$.
3. Calculation of the same number of cases with the same class. $P(X | C)$
 - a. First data Processing Testing (No. 1, with the title Administrator (III. A)): $P(\text{Educational level} | \text{Classification})$. $P(\text{Education Level} = \text{"S2"} | \text{Classification} = \text{"Suitable"}) = 703/3379 = 21\%$. $P(\text{Education Level} = \text{"S2"} | \text{Classification} = \text{"Unsuitable"}) = 495/823 = 60\%$. $P(\text{Education and Training} | \text{Classification})$. $P(\text{Education and Training} = \text{"Technical Training"} | \text{Classification} = \text{"Suitable"}) = 610/3379 = 18\%$. $P(\text{Education and Training} = \text{"Technical Training"} | \text{Classification} = \text{"Unsuitable"}) = 125/823 = 15\%$. $P(\text{Rank} | \text{Classification})$. $P(\text{Rank} = \text{"Stylist. I (III / d)"} | \text{Classification} = \text{"Suitable"}) = 596/3379 = 18\%$. $P(\text{Rank} = \text{"Stylist. I (III / d)"} | \text{Classification} = \text{"Unsuitable"}) = 179/823 = 22\%$.
 - b. Secondary Testing Data Processing (No. 2 with Supervisory position (IV. A)): $P(\text{Educational level} | \text{Classification})$. $P(\text{Educational level} = \text{"S2"} | \text{Classification} = \text{"Suitable"}) = 703/3379 = 21\%$. $P(\text{Education Level} = \text{"S2"} | \text{Classification} = \text{"Unsuitable"}) = 495/823 = 60\%$. $P(\text{Education and Training} | \text{Classification})$. $P(\text{Education and Training} = \text{"None"} | \text{Classification} = \text{"Suitable"}) = 2131/3379 = 63\%$. $P(\text{Education and Training} = \text{"None"} | \text{Classification} = \text{"Unsuitable"}) = 436/823 = 53\%$. $P(\text{Rank} | \text{Classification})$. $P(\text{Rank} = \text{"Junior Administrator Tk. I (III / b)"} | \text{Classification} = \text{"Suitable"}) = 899/3379 = 27\%$. $P(\text{Rank} = \text{"Junior Administrator Tk. I (III / b)"} | \text{Classification} = \text{"Unsuitable"}) = 65/823 = 8\%$.

- c. Third Testing Data Processing (No. 3, with the position of Executor): $P(\text{Educational level} \mid \text{Classification})$. $P(\text{Education Level} = \text{"Senior High School"} \mid \text{Classification} = \text{"Suitable"}) = 479/3379 = 14\%$. $P(\text{Education Level} = \text{"Senior High School"} \mid \text{Classification} = \text{"Unsuitable"}) = 1/823 = 0\%$. $P(\text{Education and Training} \mid \text{Classification})$. $P(\text{Education and Training} = \text{"None"} \mid \text{Classification} = \text{"Suitable"}) = 2131/3379 = 63\%$. $P(\text{Education and Training} = \text{"None"} \mid \text{Classification} = \text{"Unsuitable"}) = 436/823 = 53\%$. $P(\text{Rank} \mid \text{Classification})$. $P(\text{Rank} = \text{"Junior Administrator Tk. I (III / b)"} \mid \text{Classification} = \text{"Suitable"}) = 899/3379 = 27\%$. $P(\text{Rank} = \text{"Junior Administrator Tk. I (III / b)"} \mid \text{Classification} = \text{"Unsuitable"}) = 65/823 = 8\%$.
- d. Processing of Testing Data Fourteenth thirteen (No. 1943, with the title of Administrator (III. B)): $P(\text{Educational level} \mid \text{Classification})$. $P(\text{Educational level} = \text{"S2"} \mid \text{Classification} = \text{"Suitable"}) = 703/3379 = 21\%$. $P(\text{Education Level} = \text{"S2"} \mid \text{Classification} = \text{"Unsuitable"}) = 495/823 = 60\%$. $P(\text{Education and Training} \mid \text{Classification})$. $P(\text{Education and Training} = \text{"Leadership Training Tk. IV"} \mid \text{Classification} = \text{"Suitable"}) = 524/3379 = 16\%$. $P(\text{Education and Training} = \text{"Leadership Training Tk. IV"} \mid \text{Classification} = \text{"Unsuitable"}) = 104/823 = 13\%$. $P(\text{Rank} \mid \text{Classification})$. $P(\text{Rank} = \text{"Leadership trainee (IV / a)"} \mid \text{Classification} = \text{"Suitable"}) = 253/3379 = 7\%$. $P(\text{Rank} = \text{"Leadership trainee (IV / a)"} \mid \text{Classification} = \text{"Unsuitable"}) = 140/823 = 17\%$.
- e. Processing of Testing Data of the four hundred and twenty (No. 2913, with the title of Supervisor (IV. B)): $P(\text{Educational level} \mid \text{Classification})$. $P(\text{Educational level} = \text{"S2"} \mid \text{Classification} = \text{"Suitable"}) = 703/3379 = 21\%$. $P(\text{Education Level} = \text{"S2"} \mid \text{Classification} = \text{"Unsuitable"}) = 495/823 = 60\%$. $P(\text{Education and Training} \mid \text{Classification})$. $P(\text{Education and Training} = \text{"Technical Training"} \mid \text{Classification} = \text{"Suitable"}) = 610/3379 = 18\%$. $P(\text{Education and Training} = \text{"Technical Training"} \mid \text{Classification} = \text{"Unsuitable"}) = 125/823 = 15\%$. $P(\text{Rank} \mid \text{Classification})$. $P(\text{Rank} = \text{"Administrator Tk. I (III / d)"} \mid \text{Classification} = \text{"Suitable"}) = 596/3379 = 18\%$. $P(\text{Rank} = \text{"Administrator Tk. I (III / d)"} \mid \text{Classification} = \text{"Unsuitable"}) = 179/823 = 22\%$.
4. Multiply all Suitable results with the testing data the class will look for.

$P(X \mid C)$

- a. First Data Processing Testing (No. 1, with the title Administrator (III. A)): $P(X \mid \text{Classification} = \text{"Suitable"}) = 0.21 * 0.18 * 0.18 * 0.80 = 0.53\%$. $P(X \mid \text{Classification} = \text{"Unsuitable"}) = 0.60 * 0.15 * 0.22 * 0.20 = 0.39\%$.
- b. Secondary Testing Data Processing (No. 2 with Supervisory position (IV. A)): $P(X \mid \text{Classification} = \text{"Suitable"}) = 0.21 * 0.63 * 0.27 * 0.80 = 2.81\%$. $P(X \mid \text{Classification} = \text{"Unsuitable"}) = 0.60 * 0.53 * 0.08 * 0.20 = 0.49\%$.
- c. Third Testing Data Processing (No. 3, with the position of Executor): $P(X \mid \text{Classification} = \text{"Suitable"}) = 0.14 * 0.63 * 0.27 * 0.80 = 1.91\%$. $P(X \mid \text{Classification} = \text{"Unsuitable"}) = 0.00 * 0.53 * 0.08 * 0.20 = 0\%$.
- d. Four Hundred and thirteenth Processing of Testing

Data (No. 1943, with the title of Administrator (III. B)): $(X \mid \text{Classification} = \text{"Suitable"}) = 0.21 * 0.16 * 0.07 * 0.80 = 0.19\%$. $P(X \mid \text{Classification} = \text{"Unsuitable"}) = 0.60 * 0.13 * 0.17 * 0.20 = 0.25\%$.

- e. Four hundred and twentieth Processing of Testing Data (No. 2913, with the title of Supervisor (IV. B)): $(X \mid \text{Classification} = \text{"Suitable"}) = 0.21 * 0.18 * 0.18 * 0.80 = 0.53\%$. $P(X \mid \text{Classification} = \text{"Unsuitable"}) = 0.60 * 0.15 * 0.22 * 0.20 = 0.39\%$.
5. Comparing results per class with the highest value is set as a new class.
- a. First Data Processing Testing (No. 1, with the title of Administrator (III. A)): Have the highest value on the corresponding probability of 0.53% means they suitable with the class.
- b. Secondary Testing Data Processing (No. 2 with Supervisory position (IV. A)): Having the highest value on the corresponding probability of 2.81%, means it consider in the suitable class.
- c. Third Testing Data Processing (No. 3, with the position of Executor): Have the highest value on the corresponding probability of 1.91%, suitable class.
- d. Four Hundred and Thirteenth Processing of Data Testing (No. 1943, with the title of Administrator (III. B)): Having the highest value on the probability is not appropriate at 0.25%, considers nsuitable in class.
- e. Four Hudrend and Twetieth Processing of Testing Data (No. 2913, with the title of Supervisor (IV. B)): Have the highest value on the corresponding probability of 0.53%, means classified in suitable class.

4.2 Implementation of the Naive Bayes Algorithm Using Microsoft Excel

The Naive Bayes algorithm implementation consists of the following steps:

1. Prepare a dataset. The dataset used in processing this data uses 4202 training data, and 420 testing data.
2. Calculate the number of classes in the training data. Classes in training data consist of two categories, namely appropriate and not appropriate, so the probability to be suitable and not yet appropriate is as follows: Number of suitable classes= 3379. The number of unsuitable classes = 823. Then, $P(C = \text{"Suitable"}) = \text{COUNTIF}(\text{"FS4":\text{"FS4205"}, \text{"I10"})/\text{COUNTA}(\text{"FS4":\text{"FS4205"}}) = 80\%$. $P(C = \text{"Unsuitable"}) = \text{COUNTIF}(\text{"FS4":\text{"FS4205"}, \text{"J10"})/\text{COUNTA}(\text{"FS4":\text{"FS4205"}}) = 20\%$.
3. Calculation of the same number of cases with the same class. $P(X \mid C)$
 - a. First Data Processing Testing (No. 1, with the title Administrator (III. A)): $P(\text{Educational level} \mid \text{Classification} = \text{"S2"} \mid \text{Classification} = \text{"Suitable"}) = \text{VLOOKUP}(\text{"J51"}, \text{"SH11":\text{"SL18,4"}}) = 21\%$.

$P(\text{Educational level} = \text{"S2"} \mid \text{Classification} = \text{"Unsuitable"}) = \text{VLOOKUP}(\text{"J51"}, \text{"SH11":\text{"SL18,5"}}) =$

- 60%. P(Education and Training | Classification). P(Education and Training = "Technical Training" | Classification = "Suitable") = VLOOKUP(\$K51,\$H\$23:\$L\$27,4,) = 18%. P(Education and Training = "Technical Training" | Classification = "Unsuitable") = VLOOKUP(\$K51,\$H\$23:\$L\$27,5,) = 15%. P(Rank | Classification). P(Rank = "Administrator Tk. I (III / d)" | Classification = "Suitable") = VLOOKUP(\$L51,\$H\$32:\$L\$45,4,) = 18%. P(Rank = "Administrator Tk. I (III / d)" | Classification = "Unsuitable") = VLOOKUP(\$L51,\$H\$32:\$L\$45,5,) = 22%.
- b. Secondary Testing Data Processing (No. 2 with Supervisory position (IV. A)): P(Educational level | Classification). P(Educational level = "S2" | Classification = "Suitable") = VLOOKUP(\$J52,\$H\$11:\$L\$18,4,) = 21%. P(Educational level = "S2" | Classification = "Unsuitable") = VLOOKUP(\$J52,\$H\$11:\$L\$18,5,) = 60%. P(Education and Training | Classification). P(Education and Training = "None" | Classification = "Suitable") = VLOOKUP(\$K52,\$H\$23:\$L\$27,4,) = 63%. P(Education and Training = "None" | Classification = "Unsuitable") = VLOOKUP(\$K52,\$H\$23:\$L\$27,5,) = 53%. P(Rank | Classification). P(Rank = "Junior Administrator Tk. I (III / b)" | Classification = "Suitable") = VLOOKUP(\$L52,\$H\$32:\$L\$45,4,) = 27%. P(Rank = "Junior Administrator Tk. I (III / b)" | Classification = "Unsuitable") = VLOOKUP(\$L52,\$H\$32:\$L\$45,5,) = 8%.
- c. Third Testing Data Processing (No. 3, with the position of Executor): P(Educational level | Classification). P(Education Level = "Senior High School" | Classification = "Suitable") = VLOOKUP(\$J53,\$H\$11:\$L\$18,4,) = 14%. P(Education Level = "Senior High School" | Classification = "Unsuitable") = VLOOKUP(\$J53,\$H\$11:\$L\$18,5,) = 0%. P(Education and Training | Classification). P(Education and Training = "None" | Classification = "Suitable") = VLOOKUP(\$K53,\$H\$23:\$L\$27,4,) = 63%. P(Education and Training = "None" | Classification = "Unsuitable") = VLOOKUP(\$K53,\$H\$23:\$L\$27,5,) = 53%. P(Rank | Classification). P(Rank = "Junior Administrator Tk. I (III / b)" | Classification = "Suitable") = VLOOKUP(\$L53,\$H\$32:\$L\$45,4,) = 27%. P(Rank = "Junior Administrator Tk. I (III / b)" | Classification = "Unsuitable") = VLOOKUP(\$K53,\$H\$23:\$L\$27,5,) = 8%.
- d. Four hundred and thirteenth Processing of Testing Data (No. 1943, with the title of Administrator (III. B)): P(Educational level | Classification). P(Educational level = "S2" | Classification = "Suitable") = VLOOKUP(\$J463,\$H\$11:\$L\$18,4,) = 21%. P(Educational level = "S2" | Classification = "Unsuitable") = VLOOKUP(\$L463,\$H\$32:\$L\$45,4,) = 60%. P(Education and Training | Classification). P(Education and Training = "Tk. IV Leadership Training" | Classification = "Suitable") = VLOOKUP(\$K463,\$H\$23:\$L\$27,4,) = 16%. P(Education and Training = "Leadership Training Tk. IV" | Classification = "Unsuitable") = VLOOKUP(\$L463,\$H\$32:\$L\$45,4,) = 13%. P(Rank | Classification). P(Rank = "Guide (IV / a)" | Classification = "Suitable") = VLOOKUP(\$L463,\$H\$32:\$L\$45,4,) = 7%. P(Rank = "Guide (IV / a)" | Classification = "Unsuitable") = VLOOKUP(\$L463,\$H\$32:\$L\$45,5,) = 17%.
- e. Four hundred and Twentieth Processing of Testing (No. 2913, with the title of Supervisor (IV. B)): P(Educational level | Classification). P(Educational level = "S2" | Classification = "Suitable") = VLOOKUP(\$J470,\$H\$11:\$L\$18,4,) = 21%. P(Educational level = "S2" | Classification = "Unsuitable") = VLOOKUP(\$J470,\$H\$11:\$L\$18,5,) = 60%. P(Education and Training | Classification). P(Education and Training = "Technical Training" | Classification = "Suitable") = VLOOKUP(\$K470,\$H\$23:\$L\$27,4,) = 18%. P(Education and Training = "Technical Training" | Classification = "Unsuitable") = VLOOKUP(\$K470,\$H\$23:\$L\$27,5,) = 15%. P(Rank | Classification). P(Rank = "Administrator Tk. I (III / d)" | Classification = "Suitable") = VLOOKUP(\$L470,\$H\$32:\$L\$45,4,) = 18%. P(Rank = "Administrator Tk. I (III / d)" | Classification = "Unsuitable") = VLOOKUP(\$L470,\$H\$32:\$L\$45,5,) = 22%.
4. Multiply all results suitable with the testing data that the class will look for. **P(X | C)**
- a. Data Processing Testing First (No. 1, with the title Administrator (III. A)): P(X | Classification = "Suitable") = VLOOKUP(\$J51,\$H\$11:\$L\$18,4,)*VLOOKUP(\$K51,\$H\$23:\$L\$27,4,)*VLOOKUP(\$L51,\$H\$32:\$L\$45,4,)*K5 = 0.53%. P(X | Classification = "Unsuitable") = VLOOKUP(\$J51,\$H\$11:\$L\$18,5,)*VLOOKUP(\$K51,\$H\$23:\$L\$27,5,)*VLOOKUP(\$L51,\$H\$32:\$L\$45,5,)*K6 = 0.39%.
- b. Secondary Testing Data Processing (No. 2 with Supervisory position (IV. A)): P(X | Classification = "Suitable") = VLOOKUP(\$J52,\$H\$11:\$L\$18,4,)*VLOOKUP(\$K52,\$H\$23:\$L\$27,4,)*VLOOKUP(\$L52,\$H\$32:\$L\$45,4,)*K5 = 2.81%. P(X | Classification = "Unsuitable") = VLOOKUP(\$J52,\$H\$11:\$L\$18,5,)*VLOOKUP(\$K52,\$H\$23:\$L\$27,5,)*VLOOKUP(\$L52,\$H\$32:\$L\$45,5,)*K6 = 0.49%.
- c. Third Testing Data Processing (No. 3, with the position of Executor): P(X | Classification = "Suitable") = VLOOKUP(\$J53,\$H\$11:\$L\$18,4,)*VLOOKUP(\$K53,\$H\$23:\$L\$27,4,)*VLOOKUP(\$L53,\$H\$32:\$L\$45,4,)*K5 = 1.91%. P(X | Classification = "Unsuitable") = VLOOKUP(\$J53,\$H\$11:\$L\$18,5,)*VLOOKUP(\$K53,\$H\$23:\$L\$27,5,)*VLOOKUP(\$L53,\$H\$32:\$L\$45,5,)*K6 = 0%.
- d. Four hundred and thirteenth Processing of Testing Data (No. 1943, with the title Administrator (III. B)): P(X | Classification = "Suitable") = VLOOKUP(\$J463,\$H\$11:\$L\$18,4,)*VLOOKUP(\$L463,\$H\$32:\$L\$45,4,)*VLOOKUP(\$K463,\$H\$23:\$L\$27,4,)*K5 = 0.16%. P(X | Classification = "Unsuitable") = VLOOKUP(\$L463,\$H\$32:\$L\$45,5,)*VLOOKUP(\$K463,\$H\$23:\$L\$27,5,)*K6 = 0.17%.

- $\$K463, \$H\$23: \$L\$27,4,)*VLOOKUP(\$L463, \$H\$32: \$L\$45,4,)*K\$5 = 0.19\%$. $P(X | \text{Classification} = \text{"Unsuitable"}) = VLOOKUP(\$J463, \$H\$11: \$L\$18,5,)*VLOOKUP(\$K463, \$H\$23: \$L\$27,5,)*VLOOKUP(\$L463, \$H\$32: \$L\$45,5,)*K\$6 = 0.25\%$.
- e. Four hundred and twentieth Processing of Testing Data (No. 2913, with the title Supervisor (IV. B)): $P(X | \text{Classification} = \text{"Suitable"}) = VLOOKUP(\$J470, \$H\$11: \$L\$18,4,)*VLOOKUP(\$K470, \$H\$23: \$L\$27,4,)*VLOOKUP(\$L470, \$H\$32: \$L\$45,4,)*K\$5 = 0.53\%$. $P(X | \text{Classification} = \text{"Unsuitable"}) = VLOOKUP(\$J470, \$H\$11: \$L\$18,5,)*VLOOKUP(\$K470, \$H\$23: \$L\$27,5,)*VLOOKUP(\$L470, \$H\$32: \$L\$45,5,)*K\$6 = 0.39\%$.
5. Compare results per class, the highest value is set as a new class.
- a. First Data Processing Testing (No. 1, with the title Administrator (III. A)): $= IF (P51 > Q51, \text{"Suitable"}, \text{"Unsuitable"}) = \text{Suitable}$.
- b. Secondary Testing Data Processing (No. 2 with Supervisory position (IV. A)): $= IF (P52 > Q52, \text{"Suitable"}, \text{"Unsuitable"}) = \text{Suitable}$.
- c. Third Testing Data Processing (No. 3, with the position of Executor): $= IF (P53 > Q53, \text{"Suitable"}, \text{"Unsuitable"}) = \text{Suitable}$.
- d. Four Hundred and Thirteenth Testing Data Processing (No. 1943, with the title Administrator (III. B)): $= IF (P463 > Q463, \text{"Suitable"}, \text{"Unsuitable"}) = \text{Unsuitable}$.
- e. Four hundred and twentieth Processing of Testing Data (No. 2913, with the title Supervisor (IV. B)): $= IF (P470 > Q470, \text{"Suitable"}, \text{"Unsuitable"}) = \text{Suitable}$.

From the results of data processing with the Naïve Bayes algorithm using Microsoft Excel with 90% training data and 10% testing data, the accuracy of the suitability level between employees with positions occupied is based on the measurement of the position gap of 83% which is divided into several classifications as shown in the table following:

Table 1. Confusion Table

CLASS PREDICTION	KLASIFIKASI	
	SUITABLE	UNSUITABLE
SUITABLE	310	58
UNSUITABLE	13	39

Accuracy = 83%

Table Description:

- a. First class classification (Suitable class prediction and Suitable classification): $= COUNTIFS(\$O\$51: \$O\$470, \$O474, \$M\$51: \$M\$470, P\$473) = 310$.
- b. The second classification (Unsuitable class prediction and Suitable classification): $= COUNTIFS(\$O\$51: \$O\$470, \$O475, \$M\$51: \$M\$470, P$

$\$473) = 13$.

- c. The third classification (Suitable class prediction and Unsuitable classification): $= COUNTIFS(\$O\$51: \$O\$470, \$O474, \$M\$51: \$M\$470, Q\$473) = 58$.
- d. The fourth classification (Unsuitable class prediction and Unsuitable classification): $= COUNTIFS(\$O\$51: \$O\$470, \$O475, \$M\$51: \$M\$470, Q\$473) = 39$.
- e. Accuracy Value $= (P474 + Q475) / SUM(P474: Q475) = 83\%$.

5. CONCLUSIONS AND SUGGESTIONS

5.1 Conclusions

The conclusions that can be drawn from the results of research conducted in this study are:

1. This study uses the data of civil servants who occupied Administration positions last year with a total of 4244 employees. Some of the employee data have noise in the data cleaning process, 42 employees have noise data, so the data that can be used was 4202 data.
2. Based on the results of comparison of data on educational level attributes, there are 16 employees with elementary education level and 16 employees with junior secondary education level who do not match their placement in occupied positions.
3. Based on the results of comparison of data on the attributes of education and training, 102 employees were obtained through Leadership Training Tk. IV, 124 employees with Technical Training and 394 employees who have no education or training at all who do not fit the placement in the occupied position.
4. Based on the results of comparison of data on rank, 6 employees with young guide rank (IV / c), 184 employees with rank of guide I (IV / b), 35 employees with rank of Patron (IV / a), 1 employee with rank of administrator (III / d) and 1 employee with rank of administrator (III / c) that is not yet suitable for placement in the occupied position.
5. Based on the results of data processing using the Naïve Bayes algorithm using 90% training data and 10% testing data, the accuracy of the level of conformity between employees and the position occupied based on the measurement of the position gap of 83%.

5.2 Suggestions

For further development, suggestions that can be made as input for future research are:

1. Based on the use of existing data further research can use more updated, because the data used in this study are dynamic at any time, i.e. changes occur which include the movement of employees in positions, there are employees who move work areas, pensions, died, resigns, and the addition of new employees is possible every year.
2. Based on the results of comparison of data on the attributes of the education level, it is expected that employees who have not met the qualification of position occupied with the level of education, can be given the opportunity to pursue a program equivalent to High

School (SMA) in accordance with the requirements of the position occupied.

3. Based on the results of comparison of data on the attributes of education and training, it is expected that employees who have not met the placement of positions occupied with education and training can be given the opportunity to attend education and training in accordance with the position occupied.
4. Based on the results of comparison of data on rank attributes. , it is expected that employees who do not fit the rank with the latest level of education qualification will be given to continue their education in accordance with the rank.
5. Research using data mining in addition to the Naïve Bayes algorithm, as a comparison to test the accuracy of training data and data testing to measure the suitability of employee placement.

6. REFERENCES

- [1] Adinugroho, Sigit dan Yunita Arum Sari. 2018. Implementasi Data Mining Menggunakan Weka. Malang: UB Press.
- [2] Awang, Y. Mesak dan Ferdinand R. Anigomang. 2019. Penerapan Metode Analisis Hirarki Proses (AHP) Sistem Penunjang Keputusan (SPK) Penempatan Pegawai Kantor Kecamatan Alor Barat Laut dengan Menggunakan Aplikasi Expertchoice. Seminar Nasional Teknologi Informasi dan Komunikasi STI&K (SeNTIK). ISSN : 2581-2327.
- [3] Buulolo, Efori. 2020. Data Mining Untuk Perguruan Tinggi. Yogyakarta: Deepublish.
- [4] Dewi, Santika. Esmeralda C. Djamal, dan Rezki Yuniarti. 2017. Optimalisasi Penempatan Guru Sekolah Dasar di Kecamatan Cikajang Kabupaten Garut Menggunakan Algoritma Genetika. Seminar Nasional Komputer dan Informatika (SENASKI). ISBN: 978-602-60250-1-2.
- [5] Fitriaty dan Ninik Vurwanti. 2015. Analisis Penempatan Pegawai pada Badan Kepegawaian Pendidikan dan Pelatihan Daerah di Kabupaten Sarolangun Tahun 2013 – 2015. eJournal Pemerintahan Integratif. ISSN 2337-8670.
- [6] Karnoto, Edi. 2018. “Penerapan Data Mining Untuk Memprediksi Tingkat Kesesuaian Penempatan Pegawai Terhadap Jabatan Yang Diduduki pada Badan Penyelenggara Pengelolaan Kepegawaian Menggunakan Metode Decision Tree dengan Algoritma C4.5”. Magister Sistem Informasi. Universitas Gunadarma. Jakarta.
- [7] Muflikhah, Laili. Dian Eka Ratnawati, dan Rekyan Regasari Mardi Putri. 2018. Data Mining. Malang: UB Press.
- [8] Pramana, Setia dkk. 2018. Data Mining dengan R, Konsep Serta Implementasi. Bogor: In Media.
- [9] Perpres RI No. 11. (2015). Peraturan Presiden Republik Indonesia Nomor 11 Tahun 2015 tentang Kementerian Dalam Negeri.
- [10] Pramudhita, N. Agung. Hadi Suyono, dan Erni Yudaningtyas. Penggunaan Algoritma Multi Criteria Decision Making dengan Metode Topsis dalam Penempatan Karyawan.
- [11] P. R. N. 1., 2017. Peraturan Pemerintah Republik Indonesia Nomor 11 Tahun 2017 Tentang Manajemen Pegawai Negeri Sipil. Jakarta: s.n.
- [12] Silaban, Maryati Sri dan Arief Rifa’i H. 2017. Analisis Kesesuaian Penempatan Pegawai pada Dinas Kependudukan dan Pencatatan Sipil Kota Pekanbaru. PUBLIKa. Vol 3, No. 1 Hal. 135-153.
- [13] Sutrisno, Edy. 2017. Manajemen Sumber Daya Manusia. Jakarta: Kencana.
- [14] U. R. N. 5., 2014. Undang-Undang Republik Indonesia Nomor 5 Tahun 2014 Tentang Aparatur Sipil Negara. Jakarta: s.n.
- [15] Warsito, Setyaning Windry. 2018. Implementasi Analisis Jabatan dalam Penempatan Pegawai di Bagian Kepegawaian Sekretariat DPRD Kabupaten Bandung. LPPM STIE Muhammadiyah Bandung. E-ISSN: 2621-5306 P-ISSN: 2541-5255.