

## Analysis of Naïve Bayes Algorithm Method for Outstanding Students at Yapendak Ajamu Private Junior High School

Putri Erwina<sup>1</sup>, Ibnu Rasyid Munthe<sup>2</sup>, Rahma Muti'ah<sup>3</sup>

<sup>1,2,3</sup>Universitas Labuhanbatu, Rantauprapat, Indonesia

[putrierwina09@gmail.com](mailto:putrierwina09@gmail.com), [ibnurasyidmunthe@gmail.com](mailto:ibnurasyidmunthe@gmail.com), [r.muthea5@gmail.com](mailto:r.muthea5@gmail.com)

### Abstract

Academic achievement is a change in skills or abilities that can be enhanced through learning situations. However, an issue arises at Yapendak Ajamu Private Junior High School, where student assessments are still manually inputted, resulting in inefficiency when transferring grades to paper, which are later re-entered into the e-Report system. The calculation of student scores is also done manually, and the criteria for determining outstanding students heavily rely on academic grades, while non-academic aspects are only considered as supporting data with unclear weighting. Consequently, the assessment lacks fairness in determining outstanding students. Moreover, the manual nature of the assessment and the fact that it is held solely by the homeroom teachers make it difficult to access. The Naïve Bayes algorithm method applies a classification system that includes academic grades, attitudes, attendance, and extracurricular activities. School is a place where students weigh knowledge for future needs, each school also has its own permissibility, both in terms of the best student creator school and a school that only has a few smart students, but it can be ascertained that every school wants all its students to have high intelligence, but intelligence is also not only created by the school but intelligence is also based on the students, Yapendak Ajamu Private Junior High School is a private school that has smart students Where this make Yapendak Ajamu Private Junior High School known to many people. These factors can be utilized by the school to determine outstanding students. Out of the 34 data training sessions processed in the Orange application, 30 students were predicted to be outstanding, while the remaining students were classified as not outstanding. The precision for predicting outstanding students is 1.000, while for predicting non-outstanding students, it is 0.104. Therefore, the conclusion drawn is that the grades of outstanding students are higher compared to those of non-outstanding students.

### Keywords

Academic; achievement; classification; junior high school; naïve bayes



## I. Introduction

School is a place where students acquire knowledge under the supervision of teachers. Currently, educational institutions are constantly striving to improve the quality of education, especially for junior high school students. In order to achieve this improvement in the quality of education, schools play a crucial role in motivating and developing the potential of students. One of the supporting activities for student development is the selection of outstanding students. Academic excellence among students is usually assessed based on their adherence to school rules, good attendance, consistently completing assignments, and achieving above-average grades. Most countries have a formal education system that is generally mandatory. In this system, students

progress through a series of teaching and learning activities in schools. The names for these schools vary by country, but generally include primary schools for children and continued to secondary schools for adolescents who have completed primary education.

The word school comes from Latin: *skhole*, *scola*, *scolae* or *skhola* which means: leisure time or leisure time, where at which time school is a leisure activity for children in the midst of their main activities, namely playing and spending time enjoying childhood and adolescence. Activities in their spare time are learning how to count, how to read letters and get to know about morals (ethics) and aesthetics (art). To accompany in *scola* activities, children are accompanied by experts and understand about children's psychology, so as to provide the greatest opportunity for children to create their own world through the various lessons above.

Currently, the word school changes its meaning to: is a building or institution for learning and teaching as well as a place to receive and teach lessons. The school is led by a principal. The principal is assisted by the deputy principal. The number of vice principals in each school is different, depending on the needs

At present, educational institutions are constantly working towards improving the quality of education, especially for junior high school students. In order to achieve this improvement, schools have an important role in motivating and developing the potential of students. One of the supporting activities for student development is the selection of outstanding students. Academic excellence is typically determined based on students' adherence to school rules, good attendance, consistent completion of assignments, and, importantly, their midterm and final exam grades to meet the average student grade.

Currently, the curriculum being used is the 2013 Curriculum, and the criteria used to determine outstanding students include knowledge, skills, attitudes, extracurricular activities, and attendance.

Researchers have identified issues with student data management. In the data input process, teachers still manually input student grades by writing them on sheets of paper containing the subjects taught. In the current digital era, it is highly inefficient to record student grades on paper, only to later enter them again into an electronic report (e-Rapor). The calculation of student scores is also done manually, which is not a wise use of paper.

Using the presumption of independent variables, Naive Bayes is a classification approach that relies only on probability. The relevance of Naive Bayes to the author's research is that it is used to assist researchers in classifying academic data for outstanding students at Yapendak Ajamu Private Junior High School.

The role of the Naive Bayes method is to analyze and predict student grade data by classifying the different classes of student grades to identify and observe outstanding students. The data used in this analysis includes student grades, attitudes and ethics, participation in school organizations, and attendance.

Knowledge Discovery in Databases (KDD) is a multi-step procedure that begins with data cleansing and ends with knowledge display. Data mining is one of these steps.

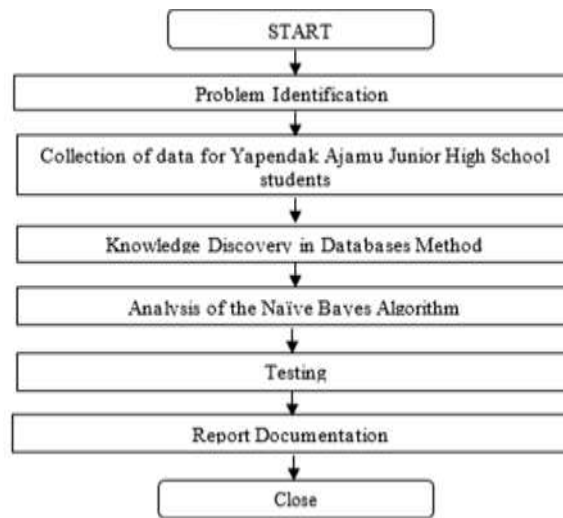
## **II. Research Method**

Because each characteristic is treated as if it were separate in the Naive Bayes method, it is easy to implement. The Naive Bayes algorithm may be thought of as a subset of classification methods. The Naive Bayes method is based on the Bayesian Theorem, which states that it is possible to estimate future probabilities using information about the past. The Naive method, which considers attributes to have no interdependence, is paired

with this idea. The Naïve Bayes algorithm assumes that specific characteristics of a class are not related to other classes and has a very low error rate and simple probability calculation compared to other classification methods.

Knowledge Discovery in Databases (KDD) is defined as the extraction of potential information. The process of Knowledge Discovery in databases involves extracting patterns or trends from data and accurately transforming them into easily understandable information.

The sequence of work steps in the research is as follows:



### III. Result and Discussion

Data mining is a step in performing Knowledge Discovery in Databases (KDD). Knowledge discovery is a process that consists of data cleaning, data integration, data selection, data transformation, data mining, pattern evaluation, and knowledge presentation.

In Data Mining, there are three parts: Association, Classification, and Clustering. Association is a process used to discover relationships present in the attribute values of a dataset. Classification is a technique used to predict the class or properties of each data instance. Clustering involves grouping data without relying on specific data classes, but instead grouping objects according to relevant topics. Prediction shares a similar definition with classification; however, in classification, data is based on the behavior or values expected in the future, which are observed from patterns/values in the past. One of the main methods of data mining is Supervised Learning, where the algorithm learns based on the values of the target indicator linked to predictor indicators.

One of the analysis methods found in data mining is Classification. Classification itself is defined as a process used to discover models or functions that describe and differentiate data into classes, aiming to find patterns that have more value in relatively large to very large datasets. Classification involves examining the characteristics of objects and assigning them to one of the predefined classes. In classification, there is only one attribute among many attributes that can be a possibility, called the target attribute, while the other attributes present are referred to as predictor attributes. Each possible value that the target attribute can take indicates the class that is predicted based on the values of the predictor attributes.

Determining relevant data and uncovering hidden patterns within a dataset is the goal of Knowledge Discovery in Databases (KDD). This information is contained within a large database that was previously unknown and potentially holds valuable insights.

The steps of Knowledge Discovery in databases (KDD) are as follows:

1. The first step in data mining is called "Data Selection," and it involves picking which data points will be utilised.

<b>Respondent</b>	<b>Mark</b>	<b>Extracurricular</b>	<b>Personality</b>	<b>Presence</b>	<b>Label</b>
Ahmad Adriansyah	82	B	B	237	Achievement
Aldi Irvan Syahputra	86	B	A	239	Achievement
Aroma Nusantara Sianipar	91	A	A	217	Achievement
Baihakki	85	A	A	240	Achievement
Bima Wirangga Siregar	80	B	B	232	Achievement
Dewi Sri Rizky Br. Tambunan	92	B	A	239	Achievement
Dina Nur Safitri	91	C	C	236	Achievement
Erwin Siregar	90	C	C	211	Achievement
Fitriani	81	B	B	223	Achievement
Frekdi Swandi Zebua	90	B	A	231	Achievement
Friska Afriana Br. Simamora	84	C	B	232	Achievement
Gusti Ananda Sabriansyah	85	C	B	229	Achievement
Hanna Yanti Simbolon	78	C	C	215	Underachieving
Hotma Ruli Tua Siagian	84	B	B	240	Achievement
Khairani	92	A	A	222	Achievement
Ledi Sri Natalia Pakpahan	90	A	A	227	Achievement
Lukas Rio Rizki Banjarnahor	83	C	B	237	Achievement
Melky Martin Simatupang	82	B	B	232	Achievement
Muhammad Basir Maulana Hsb	90	B	B	238	Achievement
Muhammad Faisal Nasution	79	C	C	209	Underachieving
Nuraini	88	B	B	180	Achievement
Putri Yosia Butar-Butar	94	B	B	237	Achievement
Rahma Dani	89	C	B	231	Achievement
Rahmad Zanuardi Harahap	90	C	B	221	Achievement
Rakes Mitu Wantober Sihombing	83	B	B	235	Achievement
Ratih Suryani	87	B	B	214	Achievement
Riska Putriana	91	A	A	240	Achievement
Rossy Pratiwi Simanungkalit	90	B	B	236	Achievement
Rosverawati Batubara	81	C	C	219	Achievement
Rut Meilisa Br. Simangunsong	84	C	C	197	Achievement
Selly Novita Sari	92	A	A	238	Achievement
Siska Silvia	93	A	A	234	Achievement
Teguh Prasetyo	78	C	C	240	Underachieving
Yuwanda Muhammad Fahri	75	C	C	221	Underachieving

2. Preprocessing consists of two stages. The first stage is the cleaning process, which involves checking for duplicate or inconsistent data and correcting any errors, such as typos. The second stage is integration, which is performed on attributes that identify unique entities.
3. Third, transformation is the action of changing data into a form that can be processed by data mining.

**Table 1**

P(N)	Amount	Percentage
90-100	14	5.6
80-89	16	8
70-79	4	0.4

P(E)	Amount	Percentage
A	7	2.1
B	14	8.4
C	13	1.3

P(K)	Amount	Percentage
A	10	3.5
B	16	8.8
C	8	0.8

P(KD)	Amount
<220	7
>220	27

4. Interpretation/Evaluation. This stage involves identifying interesting relationships within the identified knowledge base and generating distinctive patterns and predictive models that are evaluated to assess whether the existing study meets the desired objectives.
5. Knowledge. Users will be shown the created patterns. At this point, the freshly minted information may be used as a reference for decision-making and is readily accessible to all parties involved.

### 3.1 Discussions

Data processing involves all activities to prepare the data that will be manually completed. The data is processed from raw data collected during data collection. The data processed in Microsoft Excel includes academic grades, extracurricular activities, moral values, personality, and attendance of Yapendak Ajamu Private Junior High School students. With this data, the author will classify student grades using the Naïve Bayes algorithm.

Explanation:

X = data with an unknown class or label

C = hypothesis that the data belongs to a specific class.

$P(C|X)$  = probability of the hypothesis given the condition (posterior probability)

$P(c)$  = probability of the hypothesis

$P(X|C)$  = probability based on the condition in the hypothesis

P(x) = probability

P(c | x) = Value

Calculate the total occurrences:

$$90-100 = 14/34$$

$$80-89 = 16/34$$

$$70-79 = 4/34$$

$$P | C \text{ "n"} = 14/34 \times 16/34 \times 4/34$$

$$= 0,4117647059 \times 0,4705882353 \times 0,1176470588$$

$$= 0,0227966619$$

Extracurricular

Calculate the total occurrences :

$$A = 7/34$$

$$B = 14/34$$

$$C = 13/34$$

$$P | C \text{ "e"} = 7/34 \times 14/34 \times 13/34$$

$$= 0,2058823529 \times 0,4117647059 \times 0,3823529412$$

$$= 0,0324140037$$

Personality

Calculate the total occurrences :

$$A = 10/34$$

$$B = 16/34$$

$$C = 8/34$$

$$P | C \text{ "k"} = 10/34 \times 16/34 \times 8/34$$

$$= 0,2941176471 \times 0,4705882353 \times 0,2352941176$$

$$= 0,0325666599$$

Presence

Calculate the total occurrences :

$$< 220 = 7/34$$

$$> 220 = 27/34$$

$$P | C \text{ "kd"} = 7/34 \times 27/34$$

$$= 0,2058823529 \times 0,7941176471$$

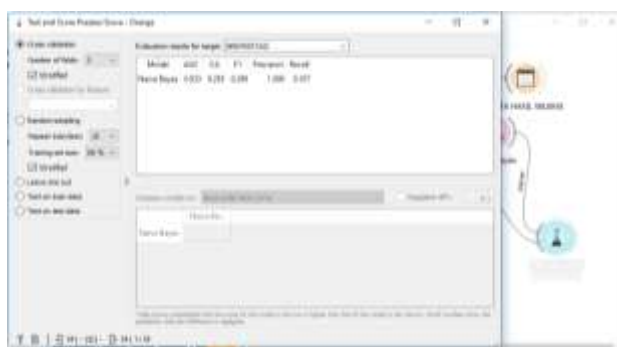
$$= 0,1634948097$$



Figure 1. Interface orange



**Figure 2.** Data mining process



**Figure 3.** Result for target achievement



**Figure 4.** Result for target underachieving

## IV. Conclusion

Based on the research findings and discussions, it can be concluded that the Naïve Bayes algorithm can be used as one of the methods for classification in predicting outstanding students at Yapendak Ajamu Private Junior High School. The parameters used in this study include student grades, personality, extracurricular activities, and attendance. Based on the test results, the AUC (Area Under the Curve) value is 0.945, demonstrating that the Nave Bayes algorithm is effective at grouping pupils into appropriate categories. Out of the 34 data training sessions processed in the Orange application, 30 students were predicted to be outstanding, while the remaining students were classified as not outstanding. The precision for predicting outstanding students is 1.000, while for predicting non-outstanding students, it is 0.104. Therefore, the conclusion drawn is that the grades of outstanding students are higher compared to those of non-outstanding students.

Researchers hope that Yapendak Ajamu Private Junior High School. will continue to maintain students' grades in schools through better education, so that the nation's successors have good intelligence, it is also hoped that the Yapendak Ajamu Private Junior

High School and its teaching staff will continue to work competently and always provide innovative, creative and good learning so that Yapendak Ajamu Private Junior High School continue to create smart and achievement students.

## References

- A. F. Watratan, A. P. B, D. Moeis, S. Informasi, and S. P. Makassar, "Implementation of the Naive Bayes Algorithm to Predict the Spread of Covid-19 in Indonesia," *J. Appl. Comput. Sci. Technol.*, vol. 1, no. 1, pp. 7–14, 2020.
- A. K. Clustering and G. Belajar, "Aplikasi Data Mining Menggunakan Algoritma K-Means Clustering untuk Mengelompokkan Mahasiswa Berdasarkan Gaya Belajar," vol. 12, 2022, doi: 10.34010/jati.v12i1.
- B. A. B. Ii and S. Sebelumnya, "BAB II TINJAUAN PUSTAKA 2.1 Studi Sebelumnya," pp. 7–20, 2019.
- B. Siregar, F. A. Pangruruk, S. T. Siridion, and U. Matana, "Pengenalan Data Science dan Profesi Data Scientist di SMA Pramita Tangerang," vol. 1, no. 2, pp. 87–96, 2022.
- D. Yunita, R. Amalia, and P. Rosyani, "Analisa Prestasi Siswa Berdasarkan Kedisiplinan, Nilai Hasil Belajar, Sosial Ekonomi dan Aktivitas Organisasi Menggunakan Algoritma Naïve Bayes," vol. 3, no. 4, pp. 2–5, 2018.
- E. W. Ambarsari, "Pemodelan Reward Rule Game Streamer Indonesia Tingkat Amatir dengan Orange Data Mining," *STRING (Satuan Tulisan Ris. dan Inov. Teknol.*, vol. 4, no. 1, p. 9, 2019, doi: 10.30998/string.v4i1.3606.
- H. D. Wijaya and S. Dwiasnati, "Implementasi Data Mining dengan Algoritma Naïve Bayes pada Penjualan Obat," vol. 7, no. 1, pp. 1–7, 2020.
- H. Hozairi, A. Anwari, and S. Alim, "Implementasi Orange Data Mining Untuk Klasifikasi Kelulusan Mahasiswa Dengan Model K-Nearest Neighbor, Decision Tree Serta Naive Bayes," *Netw. Eng. Res. Oper.*, vol. 6, no. 2, p. 133, 2021, doi: 10.21107/nero.v6i2.237.
- H. Putri, A. I. Purnamasari, A. R. Dikananda, O. Nurdiawan, and S. Anwar, "Penerima Manfaat Bantuan Non Tunai Kartu Keluarga Sejahtera Menggunakan Metode NAÏVE BAYES dan KNN," vol. 3, no. 3, pp. 331–337, 2021, doi: 10.47065/bits.v3i3.1093.
- I. A. Nikmatun, U. Diponegoro, I. Waspada, and U. Diponegoro, "IMPLEMENTASI DATA MINING UNTUK KLASIFIKASI MASA STUDI MAHASISWA MENGGUNAKAN ALGORITMA K-NEAREST NEIGHBOR," vol. 10, no. 2, pp. 421–432, 2019.
- I. Technology and C. Science, "No Title," vol. 4, pp. 20–29, 2021.
- J. Informasi, S. Oktarian, and S. Defit, "Klasterisasi Penentuan Minat Siswa dalam Pemilihan Sekolah Menggunakan Metode Algoritma K-Means Clustering," vol. 2, 2020, doi: 10.37034/jidt.v2i3.65.
- J. Prima, J. Sistem, I. Komputer, and V. No, "PENERAPAN METODE NAÏVE BAYES DALAM MEMREDIKSI PRESTASI SISWA DI SMA NEGERI 1 PANOMBEIAN PANEL," vol. 4, no. 2, 2021. "ENENG SHERLIN AGUSTIN\_Si20".
- M. Berprestasi, "IMPLEMENTASI NAIVE BAYES UNTUK SISTEM PREDIKSI," vol. 8, no. 2, 2022.
- M. I. A. PRAMUDYA, "Penerapan Market Basket Analysis Menggunakan Proses Knowledge Discovery in Database (Kdd) Sebagai Strategi Penjualan ...," *Repository.Radenintan.Ac.Id*, 2022, [Online]. Available:



[http://repository.radenintan.ac.id/id/eprint/21682%0Ahttp://repository.radenintan.ac.id/21682/1/SKRIPSI 1-2.pdf](http://repository.radenintan.ac.id/id/eprint/21682%0Ahttp://repository.radenintan.ac.id/21682/1/SKRIPSI%201-2.pdf)

- M. K. M. Nasution, "Sains Data," no. March, 2020, doi: 10.13140/RG.2.2.21816.49924/2.
- N. Yustira, D. Witarsyah, and E. Sutoyo, "IMPLEMENTASI ALGORITMA NAÏVE BAYES CLASSIFICATION UNTUK KLASIFIKASI KELULUSAN MAHASISWA TEPAT WAKTU ( STUDI KASUS : PROGRAM STUDI SISTEM INFORMASI UNIVERSITAS TELKOM )".
- P. Studi, T. Informatika, and U. Mulawarman, "KNN vs Naive Bayes Untuk Deteksi Dini Putus Kuliah Pada Profil Akademik Mahasiswa," vol. 3, no. 2, 2019.
- R. F. Rohman, P. Studi, T. Informatika, and F. Teknik, "KRITERIA KETUNTASAN MINIMAL SISWA MENGGUNAKAN METODE NAÏVE BAYES," 2020.
- R. Sistem, "Klasifikasi Berita Menggunakan Algoritma Naive Bayes Classifier Dengan," vol. 1, no. 10, pp. 3–8, 2021.
- R. Wirawan, P. S. Informatika, F. Komunikasi, D. A. N. Informatika, and U. M. Surakarta, "Analisa terhadap tingkat keaktifan siswa pada kegiatan pembelajaran sekolah dengan metode naïve bayes," 2022.
- R. Y. Hayuningtyas, "Penerapan Algoritma Naïve Bayes untuk Rekomendasi Pakaian Wanita," *J. Inform.*, vol. 6, no. 1, pp. 18–22, 2019, doi: 10.31311/ji.v6i1.4685.
- S. Indah Nurhafida and F. Sembiring, "Analisis Text Clustering Masyarakat Di Twiter Mengenai Mcdonald'Sxpbs Menggunakan Orange Data Mining," *SISMATIK (Seminar Nas. Sist. Inf. dan Manaj. Inform.*, pp. 28–35, 2021.
- S. M. Gaib et al., "Analisis Efektivitas Pembelajaran Daring Google Classroom Menggunakan Metode Naïve Bayes," vol. 5, no. 2, pp. 36–40, 2021.
- Y. Astuti, I. R. Wulandari, I. Sulistyowati, B. A. M, F. Sains, and U. K. Immanuel, "Algoritma Naive Bayes untuk Mengklasifikasikan Kepribadian Siswa SMP Berdasarkan Tipologi," vol. 10, pp. 480–489, 2021.