

Application Of The Naïve Bayes Algorithm In Determining Sales Of The Month

Hendra Supendar^{1)*}, Rusdiansyah²⁾, Nining Suharyanti³⁾, Tuslaela⁴⁾

^{1,2,3)} Universitas Bina Sarana Informatika, Indonesia, ⁴⁾ Universitas Nusamandiri, Indonesia

¹⁾ hendra.hds@bsi.ac.id, ²⁾ rusdiansyah.rds@bsi.ac.id, ³⁾ nining.nni@bsi.ac.id,

⁴⁾ tuslaela.tll@nusamandiri.ac.id

Submitted : Mar 27, 2023 | **Accepted** : Apr 3, 2023 | **Published** : Apr 4, 2023

Abstract: One important factor for creating a healthy and growing company is the existence of sales rewards for employees to achieve sales targets every month. Assessing employees is not an easy thing when there are so many employees. This will make the assessment team have to look at the criteria carefully and carefully. Data manipulation can occur because it is difficult to make decisions with such large criteria and data without automated data mining. As a result, the company will not get competitive human resources. Sales targets are one of the keys to sales success because with sales targets, the sales prediction value can be used as a guide as a reference in determining product sales. One way to make better sales predictions is by utilizing data mining processing using the Naive Bayes algorithm. The Naive Bayes algorithm calculates the probability value of each of the attributes examined including attendance, sales targets and sales returns. Research with employee absence criteria, monthly sales and monthly sales invoice returns. From the results of the research that has been done, it can be concluded that the application of the Naive Bayes classifier method to the target data set for sales of goods achieves an optimization level of 95.78%, with attendance criteria greatly affecting employee performance so that product sales targets each month can be achieved.

Keywords: Sales Of The Month, absence , Sales Invoice, Sales Targets

INTRODUCTION

Today, with advances in technology, this data can be processed again into more useful data. In data mining, this information is very useful for increasing profits or assisting in preparing sales strategies. One of the uses of data mining with the Naive Bayes Classifier method is in monthly sales transaction data from sales people to find out available monthly sales transactions by predicting the probability of membership of a class. This information can be used to support the strategy of increasing sales targets every month (Muradiansyah & Siswanto, 2018). With this information a company can find out monthly sales transactions for a product contained in the company. So that the company can know and determine the target market in more detail. The development of the company follows the improvement in the quality of human resources incorporated in it. Quality human resources are characterized by the ability of each individual to complete the work that is their respective responsibility (Irawan, 2018). The company will also give appreciation to employees who excel in order to spur the performance of other employees (Sembiring, 2020). In data mining, this information is very useful for increasing profits or assisting in preparing marketing strategies (Ira Zulfa, 2020). One of the uses of data mining with the Naive Bayes Classifier method in general sales data is to determine the interest and interest of prospective buyers of available products by predicting the probability of membership of a class (Eko, 2018). This information can be used to support marketing strategies to make them more effective and efficient. With this information a company can determine the level of interest of buyers in a product

*name of corresponding author



contained in the company. So that the company can know and determine the target market in more detail (Suwarna, 2017). The focus of the problem in this research is how to apply the Naive Bayes Classifier method to sales data and sales transactions every month, which is then used to optimize the product marketing strategy so that it can be achieved maximally. This is because strong resources are needed for a good marketing strategy to deal with other business competitors and can also increase the company's sales turnover. The problem that has been experienced by the company for a long time in terms of sales targets is that employees still use conventional methods, only using word of mouth, still don't know how to optimize properly for marketing and sales targets, as well as lack of knowledge about good sales data processing techniques in overcoming the best solution to these problems. Assessing employees is not easy when there are so many employees. Assessing employees is not an easy thing when there are so many employees. This will make the assessment team have to look at the criteria carefully and carefully. Data manipulation can occur because it is difficult to make decisions with such large criteria and data without automated data mining. As a result, the company will not get competitive human resources (Iskandar, 2018). One way is to develop new research methods that can increase the accuracy of decision results, namely the use of data mining (Utami, 2020). The company has sales employees, who are always calculated every month for "Sales Of The Month" as an award for the best sales for that month, either in the form of absences, achieving sales targets and achieving billing targets. The purpose of this research is to find out information on determining sales of the month for the current month.

LITERATURE REVIEW

Previous studies that the authors use as references include the following:

1. Research conducted by Odi Nurdiawan, et al, 2018. Research with the Naive Bayes Classifier Algorithm also works on numeric data types which can facilitate the analysis process. The process in this method is the process of analyzing patterns of pre-existing sales data (Learning Phase) based on attributes, namely type, time, size tested and process of analysis. The results of the process resulted in an accuracy rate of 97.22% (Eko, 2018).
2. Research conducted by Herry Derajad Wijaya, et al, 2020. Research to obtain accuracy values for drug sales data, especially types of vitamins which are often the choice of customers who need these medicines by using a data mining classification algorithm, namely the Naive Bayes algorithm. This study uses the Rapidminer version 8 tool as a medium for testing the data to be processed to obtain accuracy and ROC values. The accuracy value shows at a value of 88.00% (Wijaya, 2020).
3. Research conducted by Anugrah Angga Ronaldi, et al, 2020. Research on CV Mitra Artha Sejati's sales data in 2017, 2018, 2019, and 2020. The results of prediction calculations using the Naive Bayes algorithm produce a prediction accuracy rate of 94.59% with class precision, namely "Yes" 100.00%, "No" 94.44%, and for class recall namely "Yes" 33.33%, and "No" 100.00% (Ronaldi, 2021).

Data mining is a process that uses statistical, mathematical, artificial intelligence, and machine learning techniques to extract and identify useful information and related knowledge from various databases (Handoko, 2018).

The Naive Bayes algorithm is an algorithm that studies the probability of an object with certain characteristics belonging to a certain group/class, for classification problems. It is based on Bayes' probability theorem. In Bayes' Theorem, if there are two separate events (say X and H), then Bayes' Theorem is formulated as follows (Sriyano, 2021):

$$P(H|X) = \frac{P(X|H)}{P(X)} \cdot P(H) \quad (1)$$

Information :

X = Data with unknown class

H = Hypothesis data X is a specific class

P(H|X) = Probability of hypothesis H based on condition x (posteriori prob.)

*name of corresponding author



Naive Bayes is one of the data mining methods used in classification problems based on the application of the Bayes theorem. Naive Bayes will calculate the posterior probability for each occurrence value of the target attribute in each data sample (Mutiara, 2020). Furthermore, Naive Bayes will classify the data sample into the class that has the highest posterior probability value.

$P(H)$ = Probability of the hypothesis H (prior prob.)

$P(X|H)$ = Probability of X under these conditions

$P(X)$ = Probability of X

The basic idea of Bayes' rule is that the result of the hypothesis (H) can be estimated based on some observed evidence (E). There are several important things from the Bayesian rule, viz (Prayoga, 2018):

1. An initial/priori probability H or $P(H)$ is the probability of a hypothesis before the evidence is observed
2. An ultimate probability H or $P(H|E)$ is the probability of a hypothesis after the evidence has been observed

Rapid Miner is a machine learning data mining, text mining and predictive analytics environment. RapidMiner is software that is open (open source). RapidMiner is a solution for analyzing data mining, text mining and predictive analysis (Irnanda, 2020). RapidMiner uses various descriptive and predictive techniques to provide insight to users so they can make the best decisions (Ardiansyah, 2018). RapidMiner has approximately 500 data mining operators, including operators for input, output, data preprocessing and visualization (Irnanda, 2021).

METHOD

The stages in this research include research steps. The framework in this research is described as follows:

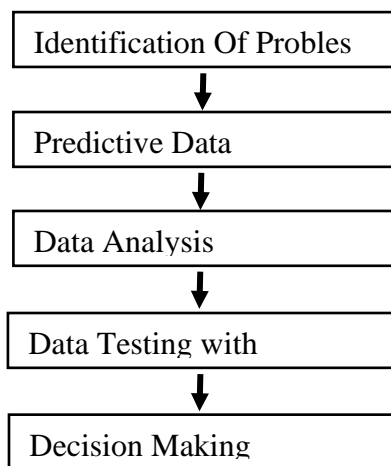


Figure 1: Research Framework

The data used for data analysis in research on employee attendance data, target billing data and sales data. The results of the research conducted can be seen in the image below..

Table 1. Attendance Data for November 2022

Name	Entry Amount	Mark
Tedi	25	100
Yandi	25	100
Arum	25	100
Asniar	24	98

*name of corresponding author



Mery	23	96
Kolidi	24	98
Bayu	25	100
Sardi	22	94
Aan	25	100
Mahyudi	24	98
Dikas	24	98
Jefry	25	100
Iyamo	25	100
Januar	24	98
Yulinus	21	92

In table 1, Absence Display Data Sales working days in November are determined by the company at less than 2 if per day of absence for 1 month, some employees reach 100% with full entry requirements for 1 month.

Table 2. Closing Sales Data for November 2022

CLOSING DATA OF BILLING TARGETS FOR NOVEMBER 2022				
Sales name	TARGET	REALIZATION	PERCENTAGE	MARK
Tedi	6.383.007.555	3.606.256.279	56,50%	600
Yandi	1.173.911.464	841.227.471	71,66%	800
Arum	264.974.723	272.216.926	102,73%	1400
Asniar	175.103.500	114.143.835	65,19%	700
Mery	167.692.825	139.949.382	83,46%	1000
Kolidi	281.519.560	259.558.421	92,20%	1100
Bayu	232.282.116	193.311.941	83,22%	1000
Sardi	309.558.635	249.950.125	80,74%	1000
Aan	1.208.945.589	1.157.899.709	95,78%	1200
Mahyudi	237.806.812	196.641.380	82,69%	1000
Dikas	719.179.564	466.145.430	64,82%	600
Jefry	171.614.246	160.513.402	93,53%	1200
Iyamo	196.024.742	157.383.822	80,29%	1000
Januar	328.112.927	307.982.509	93,86%	1200
Yulinus	183.081.207	151.431.610	82,71%	1000

In table 2, target billing data for the end of November 2022 with an explanation of determining the value as follows: Description: < 50% =0, 51% - 65 % = 600, 65% - 70% =700, 70% - 75%= 800, 75% - 80%=900, 80% - 90% = 1000, 90% - 93%= 1100, 93% - 95%=1200, 95% - 100%=1300, >100% =1400.

Table 3. Closing Data for Billing Refunds for November 2022

SALES CLOSING DATA FOR November 2022				
NAMA SALES	TARGET	REALISASI	PERSENTASE	NILAI
Tedi	2.476.719.085	2.562.213.503	103,45%	1000
Yandi	1.541.032.841	1.561.753.194	101,34%	1000
Arum	262.886.769	127.840.515	48,63%	0
Asniar	182.717.894	95.125.687	52,06%	0

*name of corresponding author



Mery	135.562.500	125.023.952	92,23%	800
Kolidi	215.986.803	190.112.669	88,02%	800
Bayu	239.327.994	212.320.074	88,72%	800
Sardi	204.217.074	166.172.712	81,37%	800
Aan	960.438.222	1.049.380.067	109,26%	1000
Mahyudi	140.714.000	114.665.990	81,49%	800
Dikas	377.351.069	431.429.426	114,33%	1100
Jefry	136.562.000	88.418.500	64,75%	0
Iyamo	140.877.600	153.195.300	108,74%	1000
Januar	208.410.800	218.218.900	104,71%	1000
Yulinus	165.374.568	176.714.000	106,86%	1000

In table 3, sales data for the end of November 2022 with an explanation of determining the value as follows: Description: < 70% = 0.71% - 100% = 800, 101% - 110% = 1,000, > 110% = 1,100.

RESULT

Figure 2. Display of the results of the Test Test

RESU LT	PREDICT ION	CONFIDE NCE ABSENCE	CONFIDE NCE SALES	CONFIDE NCE AR	NAM E	ABSEN CE	ACH IE SAL E	ACHI EV AR
ENOU GH	ENOUGH	0,999	0,001	0	Tedi	100	1000	600
GOOD	GOOD	0,039	0,961	0	Yandi	100	1000	800
ENOU GH	ENOUGH	1	0	0	Arum	100	0	1400
ENOU GH	ENOUGH	1	0	0	Asnia r	98	0	700
ENOU GH	GOOD	0.415	0,585	0	Mery	96	800	1000
GOOD	GOOD	0.010	0.990	0	Kolidi	98	800	1100
GOOD	GOOD	0.009	0.991	0	Bayu	100	800	1000
GOOD	GOOD	0.000	1.000	0	Sardi	94	800	1000
VERY GOOD	VERY GOOD	0.000	0.000	1.000	Aan	100	1000	1200
GOOD	GOOD	0.009	0.991	0	Mahy udi	98	800	1000
ENOU GH	ENOUGH	1.000	0.000	0	Dikas	98	1100	600
ENOU GH	ENOUGH	1.000	0.000	0	Jefry	100	0	1200
GOOD	GOOD	0.007	0.993	0	Iyamo	100	1000	1000
GOOD	GOOD	0.018	0.982	0	Januar	98	1000	1200
GOOD	GOOD	0.000	1.000	0	Yulin us	92	1000	1000

In Figure 2. t can be seen that the percentage for Correctly Classified Instance is very good on behalf of Aan, from the attendance data, the sales target is 109.26% and the accuracy rate is 95.78%.. Test results with rapidminer software, test results with rapidminer, the best sales employee was Aan with Very Good criteria according to calculations with Excel.

*name of corresponding author



DISCUSSIONS

Based on the results of the study predicting sales targets at the end of the month with total data of 25 working days of sales target transactions by employees, for Correctly Classified Instances, totaling 16 data, which has an accuracy of 95.78%, the average sales target accuracy rate is above 90% and according with previous research on sales targets conducted by Odi Nurdiawan, et al, 2018. an accuracy rate of 97.22% is produced for predicting sales targets. Research conducted by Anugrah Angga Ronaldi, et al, 2020. Research on sales data produces a prediction accuracy rate of up to 94.59%.

CONCLUSION

From the research that has been done, it can be concluded that the application of the naive Bayes classifier method to the sales target data set achieves an optimization rate of 95.78%. The problem is that the target is not reached at the end of each month due to absent employees who do not enter so that the targets set are not achieved.

REFERENCES

- Ardiansyah, D. (2018). Algoritma c4.5 untuk klasifikasi calon peserta lomba cerdas cermat siswa smp dengan menggunakan aplikasi rapid miner. *Jurnal Inkofar*, 1(2), 5–12.
- Eko, A. (2018). Penerapan Data Mining Pada Penjualan Barang Menggunakan Metode Naive Bayes Classifier Untuk Optimasi Strategi Pemasaran. *Jurnal Teknologi Informasi Dan Komunikasi*, (April), 100–110.
- Handoko, K. (2018). Data Mining Pada Jumlah Penumpang Menggunakan Metode Clustering. *SNISTEK*, (1), 97–102.
- Irawan, B. (2018). Organisasi formal dan informal: tinjauan konsep, perbandingan, dan studi kasus. *Administrative Reform*, 6(4), 195–220.
- Irnanda, K. F. (2020). Penerapan Klasifikasi C4 . 5 Dalam Meningkatkan Kecakapan Berbahasa Inggris dalam Masyarakat. *SAINTEKS*, 304–308.
- Irnanda, K. F. (2021). Analisa Klasifikasi C4 . 5 Terhadap Faktor Penyebab Menurunnya Prestasi Belajar Mahasiswa Pada Masa Pandemi. *MEDIA INFORMATIKA BUDIDARMA*, 5, 327–331. <https://doi.org/10.30865/mib.v5i1.2763>
- Iskandar, D. (2018). Strategi peningkatan kinerja perusahaan melalui pengelolaan sumber daya manusia dan kepuasan kerja dan dampaknya terhadap produktivitas karyawan. *JIBEKA*, 12, 23–31.
- Ira Zulfa. (2020). Implementasi data mining untuk menentukan strategi penjualan buku bekas dengan pola pembelian konsumen menggunakan metode Apriori (studi kasus : Kota Medan). *TEKNIKA: JURNAL SAINS DAN TEKNOLOGI*, 16(1), 69–82.
- Murdiansyah, A. O., & Siswanto, I. (2018). Algoritma Naive Baiyes Classsifier pada Aplikasi Data Mining Berbasis Web. *SKANIKA*, 1(1), 284–290.
- Mutiara, E.-. (2020). Algoritma Klasifikasi Naive Bayes Berbasis Particle Swarm Optimization Untuk Prediksi Penyakit Tuberculosis (Tb). *Swabumi*, 8(1), 46–58. <https://doi.org/10.31294/swabumi.v8i1.7668>
- Prayoga, N. D. (2018). Sistem Diagnosis Penyakit Hati Menggunakan Metode Naive Bayes. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer (J-PTIHK)*, 2(8), 2666–2671.
- Ronaldi, A. A. (2021). Implementasi Data Mining Untuk Prediksi Penjualan Pestisida Pada Cv MitraA Artha Sejati Menggunakan ALgoritma Naive Bayes. *KESATRIA(Jurnal Penerapan Sistem Informasi Dan Manajemen*, 2(1), 39–46. Retrieved from <http://eprosiding.ars.ac.id/index.php/pti>

*name of corresponding author



- Sembiring, H. (2020). Pengaruh Motivasi Dan Lingkungan Kerja Terhadap Kinerja Karyawan Pada Bank Sinarmas Medan. *Jurakunman*, 13(1).
- Sriyano, C. S. (2021). Pendeteksian Berita Hoax Menggunakan Naive Bayes Multinomial Pada Twitter dengan Fitur Pembobotan TF-IDF. *Ie-Proceeding of Engineering*, 8(2), 3396–3405.
- Suwarna, A. (2017). Pengaruh Kualitas Jasa Terhadap Loyalitas Pelanggan Indosat Im3 Prabayar Di Desa Sangkanhurip. *Manajemen Ekonomi dan Akuntansi*, 1(2), 41–63.
- Utami, S. F. (2020). Penerapan Data Mining Algoritma Decision Tree Berbasis PSO. *Sainteks*, 677–681.
- Wijaya, H. D. (2020). Implementasi Data Mining dengan Algoritma Naive Bayes pada Penjualan Obat. *Jurnal Informatika*, 7(1), 1–7. <https://doi.org/10.31311/ji.v7i1.6203>

*name of corresponding author



This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.