International Journal of Multidisciplinary Research and Publications
ISSN (Online): 2581-6187

Implementation of Naïve Bayes for Selection Marketing Division

Sabar Rudiarto
Universitas Mercu Buana, Jl. South Meruya No.1, Indonesia
Email address: sabar.rudiarto@mercubuana.ac.id

Abstract**—***The Marketing Division is a job whose scope of work is the promotion of a product, good or service. The problem that always occurs in companies is that there is no department responsible for selecting reliable marketing employees. This problem has resulted in the employee recruitment process for the new marketing department still not being carried out professionally. This can happen because there is no standard method to support the assessment of the selection of prospective employees in the marketing division. An analysis needs to be carried out regarding the selection of employees in marketing division placements. By holding this analysis process for employees in new marketing department placements, it can be seen whether prospective employees in the marketing department have passed or not. Based on the existing problems, the data mining classification method is used to predict employee selection for the marketing division, namely by using the naïve Bayes method*.

Keywords**—** *Selection, Marketing, Data Mining, Naïve Bayes*.

I. INTRODUCTION

A company is a business entity that operates in the service sector and aims to obtain profits from the results of its operations. The aim of establishing a company is to generate maximum profits. In order to generate profits for the company, marketing is needed. Marketing is a very important element of a company in terms of achieving goals and maintaining the continuity of the company. A company can stand if it has good marketing. This marketing is carried out by a work department known as sales or marketing.

Marketing is a type of work that has the scope of work to market a product. The marketing process in the company has a significant impact on the success of the company's operations. Marketing plays an important role in marketing a company's products to maintain the continuity of the company's life. It can be said that a company can stand if it has reliable marketing. Therefore, HR management in the Marketing department must be done well. Starting from the stages of recruitment, selection, placement, development until the end of the career in question.

The selection process is the process of selecting people who have the abilities and competencies that suit the needs to fill job vacancies in a company (Garaika & Margahana, 2019). Employee selection is the main thing that a company must do to get employees who have excellent quality and are competent in doing all the company's work (Sari et al., 2015). Companies like to find that new marketing players only last for a short period of time. The main reason is a mistake in recruiting or selecting a new Marketing department where after working and carrying out product marketing activities it turns out that this Marketing department does not have the skills or qualifications needed to become a reliable Marketing person. The process of accepting new employees in the Marketing department is still not carried out professionally. This happens because there is no systematic standard method for assessing the selection of Marketing candidates. Therefore, the role of the Human Resources (HR) division is very necessary in the above problems, especially managers who carry out selection, which is needed from the start in the process of accepting new employees, especially in the Marketing department.

Research related to selecting placements in the marketing department by applying data mining methods. Data Mining is a process that can find something meaningful by classifying large data stored in repositories, using technology and statistical techniques (Iriadi and Nuraeni, 2016). Previous research, such as determining prospective employees at PTPN 12 Blater Tempurejo Jember City using the Naïve Bayes method, was research conducted by Nuryanti (2017). This research explains the creation of a platform for determining prospective employees using the Naive Bayes algorithm method, where the results obtained were Accuracy 63%, Precision 50%, and Recall 35% from 30 testing data.

II. LITERATURE REVIEW

1) Understanding Naïve Bayes

*Naive Bayes*is a method Classifying the data with results of 62% being feasible and 38% not being feasible is assisted by the RFM method as data analysis for each customer based on segmentation using the "usage rate" attribute in the data so that the processed data can be a basic reference in making decisions [4]. Simple Bayes Classification is a random classification based on Bayes theory. Bayes' Theorem will be combined with "Naïve" which means that each attribute/variable is independent. Naïve Bayes Classifier can be trained efficiently in supervised learning [5].

$$P(H \mid X) = \frac{P(H/X) + P(X/H)}{P(X)}$$

X : Data with unknown class
H: Hypothesis data is a specific class
P(H/X): Probability of hypothesis H based on condition X (posteriori probability)
P(H): Probability of hypothesis H (prior probability)
P(X/H) : Probability of X based on the conditions in the hypothesis HP(X) : Probability of X

Sabar Rudiarto, "Implementation of Naïve Bayes for Selection Marketing Division," *International Journal of Multidisciplinary Research and Publications (IJMRAP)*, Volume 7, Issue 3, pp. 45-48, 2024.

III. Research Methodology

In this research, the data collection method used is Primary Data and Secondary Data:

Primary data

Data obtained directly from the research object through the process of observation and interviews.

Secondary Data

Data obtained from stakeholders relating to activity processes, organizational structure and documents used

In this case, it explains the steps taken to obtain a research methodology, which is a stage that must be implemented so that research can be carried out more focused and makes it easier to analyze existing problems.

The research stages are explained as follows:

Identification of problems

Identifying problems is a stage carried out by the writer regarding the problem to be researched. The method used is the SWOT method for analyzing system requirements.

Collecting data

Collect data by conducting interviews with stakeholders.

Data analysis

Analyzing and processing the data resulting from the interview and data collection process.

Study of literature

Study and collect sources in the form of books, magazines, e-books related to the concept in research.

IV. Results And Discussion

At this stage, the model is tested by calculating and obtaining rules for the algorithm applied, namely the Naïve Bayes method.

1. Testing the Naïve Bayes algorithm

The training data used in the naïve Bayes model is 359 data from Marketing candidates, divided into 2, namely 100 data that passed and 259 data that did not pass.

Prior probability is carried out in the form of the equation below:

Total data = 359
Pass Data = 100
Data Not Passed = 259
P(Passed) = 100 : 359 = 0.2785515
P(No) = 259 : 359 = 0.7214485

After obtaining the probability value for each hypothesis from the class, the next step is to calculate certain probability conditions (Probability X) using data based on the probability of each hypothesis (Probability H) or what is called Prior probability. Next, to find out the results of the calculation of the Prior Probability, a calculation is carried out by detailing the number of cases for each attribute of the data variable. The results of the calculation of the Prior Probability using the Naïve Bayes algorithm can be seen in the following table:

TABLE 1. Prior Probability Calculation

Node	Attribute		Juit's still a case	LOLOS	NOOOK	P(	X|Ci)
			(S)	(S1)	(S2)	LOLOS	QIDAK
ENTROPYQOTAL			359	100	259	0.2785515	0.7214485
1	Jennys Gender						
	P		235	50	185	0.5263158	0.7007576
	L		124	50	74	0.5263158	0.2803030
	S.Um		359				
2	TinBody height						
	<=151cm		41	4	37	0.0421053	0.1401515
	152-154cm		72	2	70	0.0210526	0.2651515
	155-157cm		101	41	60	0.4315789	0.2272727
	>= 158 cm		21	3	18	0.0315789	0.0681818
	<=162cm		21	2	19	0.0210526	0.0719697
	163-164cm		27	0	27	0.0000000	0.1022727
	165-167cm		65	41	24	0.4315789	0.0909091
	>=168cm		11	7	4	0.0736842	0.0151515
	S.Um		359				
3	Fallowt Body						
	<= 46 kg		10	0	10	0.0000000	0.0378788
	47 - 51 kg		149	30	119	0.3157895	0.4507576
	52 - 56 kg		47	15	32	0.1578947	0.1212121
	>=57kg		29	5	24	0.0526316	0.0909091
	<= 52 kg		4	0	4	0.0000000	0.0151515
	53-57kg		93	40	53	0.4210526	0.2007576
	58-61kg		16	4	12	0.0421053	0.0454545
	>=62kg		11	6	5	0.0631579	0.0189394
	S.Um		359				
4	Usia						
	<=20Known		46	8	38	0.0842105	0.1439394
	21- 22 years		237	73	164	0.7684211	0.6212121
	23-24Known		42	10	32	0.1052632	0.1212121
	>=25Known		34	9	25	0.0947368	0.0946970
	S.Um		359				
5	JARAK FROM RESIDENCE TO WORKPLACE						
	<=16km		97	33	64	0.3473684	0.2424242
	17-27km		74	27	47	0.2842105	0.1780303
	28-38km		172	38	134	0.4000000	0.5075758
	>=39km		16	2	14	0.0210526	0.0530303
	S.Um		359				
6	EXPERIENCEN						
	< 1 Year		308	87	221	0.9157895	0.8371212
		> = 1 Year	51	13	38	0.1368421	0.1439394
	S.Um		359				
7	TES CAPABILITY						
	<=61		56	7	49	0.0736842	0.1856061
	62-69		86	9	77	0.0947368	0.2916667
	70-76		189	69	120	0.7263158	0.4545455
	>=77		28	15	13	0.1578947	0.0492424
	S.Um		359				
8	TARGETING3 MONTHS						
	<=2Customerh		104	4	100	0.0421053	0.3787879
	3-10Customerh		58	3	55	0.0315789	0.2083333
	11-18Customerh		157	61	96	0.6421053	0.3636364
	>=19Customerh		40	32	8	0.3368421	0.0303030
	S.Um		359				

In prior probability there are two classes formed, namely:
Marketing Candidate Employee Class: Passed
Marketing Prospective Employee Class: No

The next stage is to use Prior Probability to determine the class of new case findings, by first calculating the Posterior

Sabar Rudiarto, "Implementation of Naïve Bayes for Selection Marketing Division," *International Journal of Multidisciplinary Research and Publications (IJMRAP)*, Volume 7, Issue 3, pp. 45-48, 2024.

Probability, this is done if a new case is found in data processing. The following is a posterior probability table for calculating new cases discovered:

TABLE 2. Posterior Probability Calculation

Data X		P(X|Ci)	
Atribut	Nilai	Lolos	Tidak
Jenis Kelamin	Perempuan	0.5263158	0.7007576
Tinggi Badan	155-157 cm	0.4315789	0.2272727
Berat Badan	47 - 51 kg	0.3157895	0.4507576
Usia	21- 22 Tahun	0.7684211	0.6212121
Jarak Dari Tempat Tinggal Ke Tempat Kerja	28-38 km	0.4000000	0.5075758
Pengalaman	< 1 Tahun	0.9157895	0.8371212
Tes Kapabilitas	70-76	0.7263157	0.4545454
Targeting 3 Bulan	11-18 Nasabah	0.6421052	0.3636363

Next, after knowing the probability value of each attribute relative to the probability of each class or which is formulated in the form of the equation P(X|Ci), the next step is to calculate the total probability of each class. The following is the equation for calculating the probability of each class:

P(X| Selection = Pass)

= 0.5263158 x 0.4315789 x 0.3157895 x 0.7684211 x 0.4000000 x 0.9157895 x 0.7263157 x 0.6421052

= 0.009416528

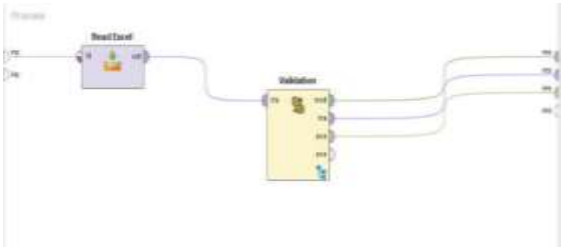
P(X|Selection = No)

= 0.7007576 x 0.2272727 x 0.4507576 x 0.6212121 x 0.5075758 x 0.8371212 x 0.4545454 x 0.3636363= 0.003132072

P(X|Selection= Passed) P(Passed) = 0.009416528 x 0.2785515 = 0.002622988

P(X|Selection= No) P(No) = 0.003132072 x 0.7214485 = 0.002259629

From the results of these calculations, it is known that the value of P(X|Passed) is greater than the value of P(X|No), so it can be concluded that this case will be classified as Passed. Experimental Results and Testing of the Naïve Bayes Algorithm Method using the K-Fold Cross Validation method using RapidMiner:



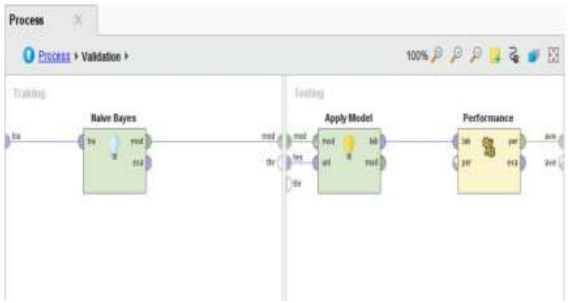


Figure 1. Testing the Naïve Bayes Algorithm Model

Figure 1 is a test of the naïve Bayes algorithm model using Rapid Minner software. Read Excel which in the picture is a tool for taking training data that will be made into a model. The data is then connected to Validation. In the validation process, tools are added for the model using Naïve Bayes and performance for the performance of the classification.

2. Evaluation and Validation of Results

Results of model testing that has been carried out by using Naïve Bayes to measure the level of accuracy and AUC (Area Under Curve).

a. Confusion Matrix Model

The Confusion matrix model uses an algorithm naïve Bayes classification, then enter the prepared testing data into the confusion matrix so that the results in Table 3 are obtained as follows:

TABLE 3. Confusion Matrix Naïve Bayes Classifier Algorithm

accuracy: 87.22% +/- 6.44% (mikro: 87.19%)

	true TIDAK	true LOLOS	class precision
pred. TIDAK	241	28	89.59%
pred. LOLOS	18	72	80.00%
class recall	93.05%	72.00%	

It is known that in Table 3, 241 were classified as Not in accordance with the predictions made using the Naïve Bayes Algorithm method, then 28 data were predicted to be No but the results turned out to be Pass, 72 data of the Pass class were predicted to be suitable, and 18 data were predicted to be Pass but turned out to be No. Based on Table, it shows that the level of accuracy using the Naïve Bayes algorithm is 87.22%, and can be calculated to find the values of accuracy, sensitivity, specificity, ppv, and npv in the equation below:

$$acc = \frac{tp + tn}{tp + tn + fp + fn} \quad acc = \frac{241 + 72}{241 + 72 + 28 + 18}$$

$$sensitivity = \frac{tp}{tp + fn} \quad sensitivity = \frac{241}{241 + 28}$$

$$specitivity = \frac{tn}{tn + fp} \quad specitivity = \frac{72}{72 + 18}$$

$$ppv = \frac{tp}{tp + fp} \qquad ppv = \frac{241}{241 + 18}$$

$$npv = \frac{tn}{tn + fn} \qquad npv = \frac{72}{72 + 28}$$

Sabar Rudiarto, "Implementation of Naïve Bayes for Selection Marketing Division," *International Journal of Multidisciplinary Research and Publications (IJMRAP)*, Volume 7, Issue 3, pp. 45-48, 2024.

Tabel 4. Hasil perhitungan algoritma *Naïve Bayes*

	Nilai (%)
Accuracy	87,22
Sensitivity	89,59
Specitivity	80
PPV	93,05
NPV	72

a. ROC Curve Evaluation

The calculation results are visualized with The ROC curve can be seen in Figure 2 which expresses the confusion matrix from Table. The horizontal line is false positive and the vertical line is true positive.

Figure 2. AUC value in the ROC graph of the Naïve Bayes Classifier Algorithm

From Figure there is a ROC graph with an AUC (Area Under Curve) value of 0.920 where the results can be stated as Excellent Classification.

V. CONCLUSION

Based on the research explained above, it can be seen that the data testing carried out using the applied model and the %Performance obtained with an accuracy value of 100% can be categorized as an excellent Naïve Bayes algorithm model.

REFFERENCES

[1] S. Apriliyani, P. Piksi, and G. Bandung, "Use Of Electronic Medical Records For Purpose * Correspondent Author : Sinta Apriliyani," vol. 1, no. 10, pp. 1399–1410, 2021.

[2] Fitriani Astika, "Implementation of Electronic Medical Records (EMR) at 'X' Hospital Pekanbaru in 2019," J. Hosp. Manag. Heal.Sci., vol. 1, no. 1, pp. 43–53, 2020, [Online]. Available: http://journal.al-matani.com/index.php/jhmhs/article/view/26.

[3] R. Ordila, R. Wahyuni, Y. Irawan, and M. Yulia Sari, "Application Of Data Mining For Grouping Patient Medical Record Data Based On Type Of Disease Using A Clustering Algorithm (Case Study: PT.Inecda Poly Clinic)," J. Computer Science ., vol. 9, no. 2, pp. 148–153, 2020, doi: 10.33060/jik/2020/vol9.iss2.181.

[4] I. Ranggadara, G. Wang, and ER Kaburuan, "Applying Customer Loyalty Classification with RFM and Naïve Bayes for Better Decision Making," Proc. - 2019 Int. Semin. Appl. Technol. Inf. Commun.Ind. 4.0 Retrosp. Prospect. Challenges, iSemantic 2019, no. September, pp. 564–568, 2019, doi: 10.1109/ISEMANTIC.2019.8884262.

[5] ASR Sinaga and D. Simanjuntak, "Expert System for Detecting Toddler Malnutrition Using the Naïve Bayes Classifier Method," J. Inkofar, vol. 1, no. 2, pp. 54–60, 2020, doi: 10.46846/jurnalinkofar.v1i2.110.

[6] A. Febrianto, S. Achmadi, and AP Sasmito, "Application of the K-Means Method for Clustering Visitors to the Itn Malang Library," J. Mhs. Tech.Inform., vol. 5, no. 1, pp. 61–70, 2021.

[7] R. Rusliani, "Cash Receipts and Disbursements Application Program Using PHP at the Noor Hidayaturrahim Mosque in Banjarmasin. This study discusses the application of the accounting information system for cash receipts and disbursements at the Noor Hidayaturrahim Mosque in," no. 36, pp. 1–10, 2020.

[8] F. Technology, AND Informatics, and U. Dinamika, "Designing Website User Interfaces onCV. Alpha Raya Technic," 2021.

[9] MI Puspita, I. Ranggadara, and I. Prihandi, "Zachman framework for designing information system logistics management," Int. J. Recent Technol. Eng., vol. 8, no. 3, pp. 4030–4034, 2019, doi: 10.35940/ijrte.C5377.098319.

[10] AF Sallaby and I. Kanedi, "Designing a Doctor's Schedule Information System Using the CodeIgniter Framework," J. Media Infotama, vol. 16, no. 1, pp. 48–53, 2020, doi: 10.37676/jmi.v16i1.1121.

[11] RT Waruwu and A. Sindar, "Expert System Determines Types of Child Developmental Disorders Using the Certainty Factor Method," Sci. Comput. Sci. Informatics J., vol. 1, no. 2, pp. 1–4, 2020.

[12] M. Sari, S. Defit, and GW Nurcahyo, "Expert System for Detecting Diseases in Children Using the Forward Chaining Method," J. Sistem Inf. and Technol., vol. 2, pp. 5–9, 2020, doi: 10.37034/jsisfotek.v2i4.114.

[13] A. Apriliani et al., "Forecasting Sales Trends of Restaurant Menus Using Sales Trend Forecast of a Restaurant Menus Using Single Moving," J. Teknol. Inf. and Computer Science., vol. 7, no. 6, pp. 1161–1168, 2020, doi: 10.25126/jtiik.202072732.

[14] RR Rizky and ZH Hakim, "Expert System for Determining Hypertension in Pregnant Women at Adjidarmo Rangkasbitung Regional Hospital, Banten Province," J. Sisfokom (Inf. and Computer Systems), vol. 9, no. 1, pp. 30–34, 2020, doi: 10.32736/sisfokom.v9i1.781.

[15] Priyono and N. Nuris, "TOPSIS Method Decision Support System for Diagnosing Dengue Fever," Inti Nusa Mandiri Journal, vol. 15, no. 1, pp. 51–58, 2020.

[16] EA Saputra and S. Nurmiati, "Web-Based Decision Support System for Determining Symptoms of Major Diseases at Salak Bogor Hospital," J. Ilmu Komput., pp. 1–4, 2019, [Online]. Available: https://jurnal.pranataindonesia.ac.id/index.php/jik/article/view/58.

[17] Y. Yuvidarmayunata, "Web-Based Expert System Using Backward Chaining Method to Determine Appropriate Nutrition for Pregnant Women," INTECOMS J. Inf. Technol. Comput. Sci., vol. 1, no. 2, pp. 231–239, 2018, doi: 10.31539/intecoms.v1i2.302.

[18] AW Ganda Anggara, Gede Pramayu, "Building an expert system using Bayes' theorem to diagnose lung disease," Semin.Nas. Technol. Inf. and Multimed. 2016, pp. 79–84, 2016.

[19] TF Ramadhani, I. Fitri, and ETE Handayani, "Web-Based ISPA Disease Diagnosis Expert System Using Forward Chaining Method," JOINTECS (Journal Inf. Technol. Comput. Sci., vol. 5, no. 2, p. 81, 2020 , doi: 10.31328/jointecs.v5i2.1243.

[20] A. Gunawan, S. Defit, and S. Sumijan, "Expert System for Identifying Gynecological Diseases Using the Android-Based Forward Chaining Method," J. Sistem Inf. and Technol., vol. 2, no. 1, pp. 15–23, 2020, doi: 10.37034/jsisfotek.v2i1.30.