Prediction of New Student Registration at Universitas Muhammadiyah Yogyakarta using the Naïve Bayes Classification Algorithm

Lusiana Ning Saputri*, Asroni, and Slamet Riyadi

Universitas Muhammadiyah Yogyakarta, Jln.Brawijaya, Tamantirto, Kasihan, Bantul, Yogyakarta 55183, Indonesia *Corresponding author: lusiana.ning.2012@ft.umy.ac.id

Abstract

All public and private universities perform new student admission each year, with the number of applicants reaching the thousands. Prospective new students who intend to continue and meet the higher education criteria must re-register according to the university's timetable after enrolling through their selected pathway and being pronounced passed. Due to the movement of prospective students to different departments or universities, the number of applications typically does not match the number of individuals who have re-registered. If the probability of a new student candidate departing can be discovered early, higher education management can take the necessary steps to retain the prospective student. Data mining and the Naïve Bayes algorithm were employed to analyze the data. The information was extracted from Universitas Muhammadiyah Yogyakarta's database of Information Technology freshmen applicants for 2016-2017. Microsoft Excel, RapidMiner, and SQL Server 2014 Management Studio were utilized.

Keywords: Data Mining, Naïve Bayes, RapidMiner, Re-registration Prediction.

1. Introduction

Universitas Muhammadiyah Yogyakarta (UMY) is located at South Ring Road, Kasihan, Bantul, the Special Region of Yogyakarta (DIY), with an A accreditation status under BAN-PT Decree No. 5237/SK/BAN-PT/Akred/PT/XII/2017. UMY conducts new student admission annually, with thousands of applicants registering. After enrolling through their selected pathway and being announced passed, students who decide to continue their studies and fulfill higher education standards must reregister at the time designated by the university. However, the number of applications frequently does not match the number of re-registered students because prospective students have transferred to other departments or universities.

The growth of information technology in systems and data processing is accelerating and becoming more precise. A server-based storage solution can effectively manage the thousands of new student data added annually. Using data mining, it is feasible to extract old data to generate new information for future needs. The number of registrants at a university can be utilized to predict the number of students who will re-register. In order for the university to anticipate future possibilities regarding the number of candidates. Suppose early detection of a potential new student's likely resignation is feasible. In that case, the higher education administration can take the required steps to retain the prospective student[1]. According to the Gartner Group, data mining is discovering significant links, patterns, and trends by analyzing massive data sets in a database using pattern recognition techniques such as statistics and mathematics[2]. Naïve Bayes is a straightforward probability classifier based on Bayes' theorem[3].

In the study of the Bayesiyan Classification Algorithm for Predicting New Student Registration at Stikes Hang Tuah Pekanbaru, it was found that from the results of registration data for new students who registered based on Pekanbaru City, Midwifery Department, Diploma 3 and Glombang 1, it can be seen that Posteriors who register are larger than those who do not register. So with these results it can be seen that the student will do the re-registration [4]. Previous research entitled Bayesian Classification Algorithm for Predicting New Student Registration at Stmik Widya Pratama, using the Naïve Bayes Algorithm in Data Mining Techniques, found that the number of registrants who registered was 658 people, and those who did not register were 255 people, out of a total of 913 people. who did the registration[5].

Naïve Bayes is a machine learning method that utilizes probability and statistical calculations proposed by the British scientist Thomas Bayes, namely predicting future probabilities based on previous experience. Bayes' theorem is an approximation to uncertainty as measured by probability[6].

The process of data mining is incoming data from various sources, this is integrated and placed in a data storage area or can be called a database. then taken and analyzed. After that it is passed to a data mining algorithm which produces output in the form of rules or some other type of 'pattern'. These are then interpreted to provide new discovery knowledge and potentially useful knowledge[7].

This study employs the Naïve Bayes classifier method to predict the data on reregistered students at UMY.

2. Method

2.1. Place and Time of Research

This study was performed in the Bureau of Information System (BSI) room of the AR Fachruddin B building at UMY from May 2019 to November 2020.

2.2. Tools and Materials

This study utilized additional tools to simplify research and data implementation. A. Research Tools

The research tools consisted of hardware and software.

- a. Hardware
 - 1. Personal Computer
 - 2. Intel i3-3217U 1.80 GHz with Nvidia GeForce 720M Cuda 2Gb
 - 3. 4 Gb Random Acces Memory (RAM)
- b. Software
 - 1. Operating system windows 8.1 64 bit
 - 2. RapidMiner
 - 3. SQL Server 2014 Management Studio
 - 4. Microsoft Office 2010

2.3. Research Flow

This study applied the Software Development Life Cycle (SDLC) waterfall model [8]. The flow is depicted in Figure 1. Waterfall model is very common model in software development. It needs results certainty in each steps so that going back to previous step is unnecessary except some condition applied it [9]. Thus, it suits well for this research work. **2.4. Literature Study**

The initial phase of this research was a literature study. It was undertaken to gather the necessary information for the research by reviewing relevant literature. The use of Naïve Bayes Algorithm are popular in the research that involve academics data. In 2017, Makhtar et. al. [10] used the algorithm to analyze on student performance in one of Malaysian university. The work also applied 10 fold cross-validation and resulted 73.4 accuracy.

The second work that used Naïve Bayes Algorithm was performed by Maitra et. al. [11] to analyze the authenticity of student feedback for faculty. They presented a proactive and outcome-based faculty feedback analysis model that employs the Nave Bayes Classifier to

sift and categorize the feedback provided by the students into invalid or valid categories based on the relative effect of the aforementioned quality features on the feedback measure.

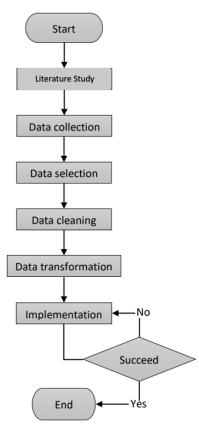


Figure 1 Research Flowchart

2.5. Data Collection

The second phase was data collection. It is crucial since it can influence the implementation process and the findings. The data were obtained from the BSI of UMY. 2.6. Data Selection

Thousands of general data required analysis and filtering to obtain more specific data before the existing database could be processed.

2.7. Data Cleaning

Data cleaning is the detection and correction of inaccurate and incomplete data. Data extracted from a database contains imperfect information, including missing, invalid, and duplicate information. This study possessed incorrect and unnecessary data attributes, necessitating data cleaning.

2.8. Data Transformation

During data transformation, the data were modified or combined into a format appropriate for data mining. This study employed Microsoft Excel to convert the data into CSV format. RapidMiner software could expedite the mining process through data transformation.

2.9. Implementation

Implementation is the primary procedure for discovering data. This study employed the Naïve Bayes algorithm and RapidMiner. RapidMiner was utilized to reprocess the data gathered from data mining to generate more precise analysis results. This study utilized training data containing Information Technology Department registrants in 2016, with the attributes of gender, province of origin, and registration pathway. Meanwhile, the data testing encompassed the same attributes of registrants in 2017. The prediction analysis applying RapidMiner and the Naïve Bayes algorithm yielded percentage-based findings.

3. Implementation and Discussion

3.1 Data Collection

This study utilized a data warehouse for UMY's students of the Information Technology Department Class of 2016. SQL Server 2014 Management was employed to access the data warehouse. Data warehouse access rights were acquired by making a copy of the BSI, along with a written approval letter from the BSI.

	5010147-5	Annual SQL Server Management	ni Shafio	* 17 gas
in the two Population (sing Task Marker Hag		Galler e	10.8	41-2-94-67-2
Construction (Construction (Co				
Bank	Microsoft Bereiter Bereiter Bereiter Bereiter Bereiter	SQL Server 2014	Next 1	
 B. Beckeller, LAR B. Beckeller, Debug Beckeller, Debug Beckell				
1 3 ALCOND CORNER -				

Figure 2 Entering the Database Server

In SQL Server 2014 Management, the database server must be connected as the initial step. A view to facilitating the retrieval of data was developed.

	William and					
10 No. 1 7.2.2.4 10 Internet 9 10 Internet 10 10 Internet 10 11 Internet 10 12 Internet 10 13 Internet 10 14 Internet 10 15 Internet 10 16 Internet 10 17 Internet 10 18 Internet 10 19 Internet 10 10 Internet 10	Cantala 7.50 Consent USANY CARDA 9 Consent a 9 December 1 Consent and Consent and Consent and Consent and		2 10 000 000 000 000 000 000 000 000 000	MARLE CHOICE MARLE CHOICE MARLE (PRO) MARLE MARL	ALL N.	
o D Astrony, o D Astrony, Stationard	day New Yorks	PARTICLE IN TRACTORY			nanikarahara	ALINE SAFTIN
	Recordering	Plane tone	ie a cuitoicune a		1	ALINE SAFTIA
1 Histology (1900) 11	Recordering	Provinsi non automotives non standard non standard		ในและสะ, สะ พ.เวเกม	พระการติด สาวาร	кана дантақ
1 Build (Holdard), Fill 1 Build (Holdard),	CT des Canaditi Cole des Delegiti Cole des Canaditi Parts Consello Tale	Providence and price canadian providence pro	ing a constraining a	maria in	HERETER CONTRACTOR	ALAK DATA
1 ImaxLuin, Heldaut, 15 ImaxLuin, Heldaut, 15 ImaxLuin, Heldaut, 15 ImaxLuin, Heldaut, 15 ImaxLuin, Heldaut, 16 ImaxLuin, Heldaut, 17 ImaxLuin, Heldaut, 18 ImaxLuin, Heldaut, 19 ImaxLuin, Heldaut, 10 ImaxLuin, Heldaut, 11 ImaxLuin, Heldaut, 12 ImaxLuin, Heldaut, 13 ImaxLuin, Heldaut, 14 ImaxLuin, Heldaut, 15 ImaxLuin, Heldaut, 16 ImaxLuin, Heldaut, 17 ImaxLuin, Heldaut, 18 ImaxLuin, Heldaut, 19 ImaxLuin, Heldaut, 10 ImaxLuin, Heldaut, 11 ImaxLuin, Heldaut, 12 ImaxLuin, Heldaut, 13 ImaxLuin, Heldaut, 14 ImaxLuin, Heldaut, 15 ImaxLuin, Heldaut, 16 ImaxLuin, Heldaut, 17 ImaxLuin, Heldaut, <td< td=""><td>CT des consections ma point the des Consections Consections Consections Consections Consections</td><td>Prove the set of an table transformer result of period material</td><td>ing a constraining a</td><td>Milantan Milantan Milantan</td><td>HEIREELAND OF HEIREE ALLENN MARKEN JR. STARLEN</td><td>ALAK DATA</td></td<>	CT des consections ma point the des Consections Consections Consections Consections Consections	Prove the set of an table transformer result of period material	ing a constraining a	Milantan Milantan Milantan	HEIREELAND OF HEIREE ALLENN MARKEN JR. STARLEN	ALAK DATA
10 Bealtan, Heldau, 12 Bealtan, Heldau, 13 Bealtan, 10 Bealtan, 11 Bealtan, 12 Bealtan, 13 Bealtan, 14 Bealtan, 15 Bealtan, 16 Bealtan, 17 Bealtan, 18 Bealtan, 19 Bealtan, 10 Bealtan, 11 Bealtan, 12 Bealtan, 13 Bealtan, 14 Bealtan, 15 Bealtan, 16 Bealtan, 17 Bealtan, 18 Bealtan, 19 Bealtan, 10 Bealtan, 11 Bealtan, 12 Bealtan, 13 Bealtan, 14 Bealtan, 15 Bealtan, 16 Bealtan, 17 Bealtan,	CT des Conselle Conse	Provinces and a closely freedor transfor statistical and the manager machine transformer free Carling free transformer free bases	HI HING AN CONSTRUCTION AND AND AND AND AND AND AND AND AND AN	ACOUNTRY DO	HEIRER HAR OF HEIRER ALL SAME	nuncjairtai
1 Bisclash, Holdad, 15 Bisclash, 16 Bisclash, 17 Bisclash, 18 Bisclash, 19 Bisclash, 10 Bisclash, 11 Bisclash, 12 Bisclash, 13 Bisclash, 14 Bisclash, 15 Bisclash, 16 Bisclash, 17 Bisclash, 18 Bisclash, 19 Bisclash, 10 Bisclash, 11 Bisclash, 12 Bisclash, 13 Bisclash, 14 Bisclash, 15 Bisclash, 16 Bisclash, 17 Bisclash, 18 Bisclash, 19 Bisclash, 10 Bisclash, 11 Bisclash, 12 Bisclash, 13 Bisclash, 14 Bisclash,	CT description Control des Control Control des Control	Provinces only provided theory reacting of policies only the second of the second theory of the second the second of the the second of the second of the second of the the second of the second of the second of the the second of the second of the the second of the second of the second of the the second of the second of the second of the the second of the second of the second of the the second of the second of the second of the the second of the second of the second of the the second of the second of the second of the the second of the second of the second of the the second of the second of the second of the the second of the	and an observation and a second secon	AN JULIES AND	HEREETANG OF HEREET, BALLAN MARK AT JR. T. Sant Her. T. Sant Her. T. Sant Her.	alan sama(
0 max.Ltdl, HEIGAM, 10 max.Ltdl, HEIGAM, 10 max.Ltdl, HEIGAM, 11 max.Ltdl, HEIGAM, 12 max.Ltdl, HEIGAM, 13 max.Ltdl, HEIGAM, 14 max.Ltdl, HEIGAM, 15 max.Ltdl, HEIGAM, 16 max.Ltdl, HEIGAM, 17 max.Ltdl, HEIGAM, 18 max.Ltdl, HEIGAM, 10 max.Ltdl, HEIGAM, 10 max.Ltdl, HEIGAM, 11 max.Ltdl, HEIGAM, 12 max.Ltdl, HEIGAM, 13 max.Ltdl, HEIGAM, 14 max.Ltdl, HEIGAM, 15 max.Ltdl, HEIGAM, 16 max.Ltdl, HEIGAM, 17 max.Htdl, HEIGAM, 18 max.Htdl, HEIGAM, 19 max.Htdl, HEIGAM, 10 max.Htdl, HEIGAM, 11 max.Htdl, HEIGAM,	CT des-Carolite Carolite ps-Self-self-self- des-Carolite Carolite Controls of the resource of the resou	Die gewannen of the second of	and an observation and a second secon	Micautian Micautian Micautian Micautian Micau Micau Micau Micau Micau Micau Micautian	Hannes Lakor, or provide an ison and at its. There has There has There has	aune puttin
1 Bisclash, Holdad, 15 Bisclash, 16 Bisclash, 17 Bisclash, 18 Bisclash, 19 Bisclash, 10 Bisclash, 11 Bisclash, 12 Bisclash, 13 Bisclash, 14 Bisclash, 15 Bisclash, 16 Bisclash, 17 Bisclash, 18 Bisclash, 19 Bisclash, 10 Bisclash, 11 Bisclash, 12 Bisclash, 13 Bisclash, 14 Bisclash, 15 Bisclash, 16 Bisclash, 17 Bisclash, 18 Bisclash, 19 Bisclash, 10 Bisclash, 11 Bisclash, 12 Bisclash, 13 Bisclash, 14 Bisclash,	Construction of the second sec	Preserved outry of also calculated reach, as period hard the reach of the biotecome	eril 1	MULTER DATA	HEIREESING OF HEIREES ALLEMAN MARKE JE JK. 17 Tanis Arts. 17 Tanis Arts. 17 Tanis Arts. 17 Tanis Arts. 18 Tanis Arts.	ALUNC SAITTAN
1 Barr, 1243, MERABAK, 13 1 2 Barr, 1243, MERABAK, 13 1 2 Barr, 1243, MERABAK, 13 1 3 Barr, 1243, MERABAK, 13 1 3 Barr, 1243, MERABAK, 13 1 4 Barr, 1243, MERABAK, 13 1 5 Barr, 1243, MERABAK, 13 1 6 Barr, 1243, MERABAK, 13 1 7 Barr, 1243, MERABAK, 13 1 8 Barr, 1243, MERABAK, 13 1 9 Barr, 1243, MERABAK, 14 1 10 Barr, 1243, MERABAK, 14 1 11 Barr, 1243, MERABAK, 14 1 12 Barr, 1244, MERABAK, 1243, MERABAK, 14 1 13 Barr, 1244, MERABAK, 1243, MERABAK, 14 1 14 Barr, 1244, MERABAK, 1243, MERABAK	Torona and a second sec	P water trees and in an calculation of each or type and the second of the second of type calculation of the second calculation of the second calcula	101 1000 1010 101 1000 1010 1000 1 1000 1010 1000 1 1000 1010 1000 1 1000 1010 1000 1 1000 1010 1000 1 1000 1010 1000 1 1000 1000 1000 1 1000 1000 1000	- AR(207)(B), Dov - AR(207)(B) - Prillip - Prillip - Prillip - Prillip - Prillip - Prillip - Prillip - Prillip	4004,00 (K, MONE) 44-244 4044,00 (K, MONE) 17 Salat Mar. 17 Salat Mar. 17 Salat Mar. 17 Salat Mar. 17 Salat Mar. 17 Salat Mar. 17 Salat Mar.	ALIAC SAITA
0 344.2104/1912ALM 1 0 344.2104/1912ALM 1 0 344.2007 1 0	Annual Control of Cont	P date total and ph de Celebra Pender des ph de Celebra Pender de Unit Material de Celebra de Celeb	41) 44 Autor Annual Annual 54 Autor Annual Annual 5 Annual		None 11 and O' Works's AL Same None 11 and These Nex. These Nex.	ALIAN SAITAN

Figure 3 View Data from Training Data of the Information Technology Department Class of 2016

A Martine Cost	10100							
and the second se	William			_			_	
THE REAL PROPERTY AND	in anona a	an-mailed 1	-					
Vani +	TY CANADA	,	10 DOM	ARTMENT .	Eal	IT CRMANU	CHOKE .	
in factors lines	C? Without	well.	101111111111111111111111111111111111111	skeed.	-	Childhee	6 A	
T INLAURED,	MOLEVES,			CARDONNE (B		214,000,000		
de Aldrein, ille e m	COMMON D		See Division	71.0		COURSE PARTY		
ALLANDAR	Distant	10x 14	a fam. lat's	ine South	in the	Die .		0.
C meansurem.	T PARAMETERS	1 10	and in					
() the Aller and The The State	7 decimin	CLARKERS BL PR	Chickley Thissester,	and Constants	United, My CANN	N. M. A. PT.	NULAME OF AU	VECT des CORNELINE DATTIN.
BAR BODD THR	10m 2074	#THENT WORK CO	F (INFORTMENT)					
the second second second		451.0045.005						
the flass, Philad	184.5.188	ARE CAMPAGED.	Mini Calenaky Diversion 21	DEADLER FROM T	ON .			- (AMR) CROCK SHOULAND AND
D des COA/TH,	the fills	ATMENT COLUMN	Buccassalar, to-strict co	DEADLER FROM T	ION ALEMAKT NEED	riseled meaning	rin:	 CAMPLE CHERT PREMIMER HILL
an DALENDAR	184.5.188	ATABLE DUALS	Ale Calification (Second Second Secon	DEADLER FROM T	NAME OF ME	Haddel in Andre Britishisk	NAME (16,16)	 CAMPLE CREEK President CREEK
An USAC PRICAT	de traj	Alarah (M.a.) Hasalari Hasalari	Au calain, there co pages (select nilling Autocal	DEADLER FROM T	CON RELEVANCE TO LEA NAME OF MIL. Tangan Aran T.	Regardan	NAME OF JR.	(AVABL_CECT Transition) and
an DAL PROVINCE	the fills	HARD DUAL HARDON	An CANADA DIGO C parate Constant States An Constant Sector States Sector States	1 1 1 1 1 1 1 1	NAME OF A DESCRIPTION O	nejariuk nejariuk hutu	Hani Natabi (16,06) 10 Tanàn Inter- 10 Tanàn Inter-	CANNEL CHOILE THE AREA HE
An DAN PRIMA An DA	de traj	HARD DATE	Au calain, there co pages (select nilling Autocal	1 1 1 1 1 1 1 1	CON RELEVANCE TO LEA NAME OF MIL. Tangan Aran T.	nejariuk nejariuk hutu	NAME OF JR.	- (ANNUE, CHEEK THE ANNUE OF
An USAN PROBAT An USAN PROBATION AN USAN PROBATIONA	da mij	HARD DUAL HARDON	An CANADA DIGO C parate Constant States An Constant Sector States Sector States	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	NAME OF A DESCRIPTION O	folder och seine Belgiseften Profes Profes Profes Profes	Hani Natabi (16,06) 10 Tanàn Inter- 10 Tanàn Inter-	- (ANNUE, CHEEK THE ANNUEL HEE
metilaxi, motelaxi metilaxi, motelaxi metilaxi, motelaxi metilaxi,	dia CAN dia Dilip CANAFACE Transition Transition Transition	HARD DATE	an callence constant pages (constanting) Applied Applied Technologieses For Forwards Woor Applied opticities	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	ION Radionation toolage Nation, CS., ML, Tamper Arabit, D, Tamper Arabit, D, Tamper Arabit, D, Tamper Arabit, D,	nebel conset nebelset totals totals totals totals	National Controls Of France States Of France States Of France States	- ANNE CERT DEVICE THE
Anti LAN, PRIARE Anti LAN, PRIARE Anti LAN, PRIARE Anti LAN, PRIARE Anti LANC, Anti	dia CAN dia Tala CANANCE Therei Ta Therei Ta Therei Ta	Hard Dr. A.	dia catalah penteri pagan (wener holper Apatasah Referensi Megener Referensi Megener Referensi Megener Referensi Megener Referensi Megener	10000000000000000000000000000000000000	ION RAMMANE MALA NAME OF MALA Surger Area II. Surger Area II. Surger Area II. Surger Area II.	nethel of heads netheline heads head	NAME OF STR. 10 Yound ONLY 10 Yound ONLY 10 Yound ONLY 10 Yound ONLY 10 Yound ONLY	- Laware Local Constraint and
Analysis,	An China An Think Canning Canning Canning Can Canning Can Canning Can Canning Can Can Can Can Can Can Can Can Can Can	Analysis of the second se	A BALE CARALLER CHARGE CONTROL OF STREAM CONTROL OF STREAM PROJECT CONTROL OF STREAM CONTROL OF STREAM CONTROL OF STREAM CONTROL OF STREAM CONTROL OF STREAM CONTROL OF STREAM	10000000000000000000000000000000000000	ION Representation Name of March Strappolation (C. Strappolation (C. Strappolation (C. Strappolation (C.	NAME OF DAMAGE NO. CONTRACT NO. CONTRACT	Figure Named (191,191, 10 Transit Anton, 10 Transit Anton, 10 Transit Anton, 10 Transit Anton, 10 Transit Anton,	- Lander Lader Landeration
minilation minilat	An CANA	Nacional Contract Nacional Secondaria Second	 an casalan product or product cards training Antibated Antibated Retrain Higginson 	10000000000000000000000000000000000000	NAME AND ADDRESS OF AD	Haddal at seals Helitak Hadda Hadda Hadda Hadda Hadda	Figure	(WWR, PELT SHARANE WE
Analysis,	An China An Think Canning The Canning The	Hand Care	All Catalog Deletics party (2014 https:// Participaries Partic	10000000000000000000000000000000000000	NAME OF CONTRACT OF A DESCRIPTION OF A D	Hadad os Adad Halitadia Intelse St St St St St St St St St St St St St	Mathef, 19 (M). 1 Constant office 2 Constant office 2 Constant office 3 Constant offi	ELANARE, INC. I NUMBER OF
An A Construction And A Construction	de case de case de case case case case case case case case	Hardon F. Hardone RC RC RC RC RC RC RC RC RC RC RC RC RC	All Casality Detect 20 people (2010 Higher Anti-teach References Angeres All Insue People Lak- media (2010) All Insue People Lak- media (2010) All Insue People All Insu	10000000000000000000000000000000000000	NAME CONTRACT DESIGN NAME OF MALE Nonpolation () Nonpolation () Nonpolation () Nonpolation () Nonpolation () Nonpolation () Nonpolation () Nonpolation () Nonpolation ()	negaariak negaariak heats 21 22 22 23 23 23 24 24 24 24 24 24 24 24 24 24 24 24 24	Partiel Natable (15, 16) 11 Tabata Anton 12 Tabata Anton	ELANANG, DEGLA DEMAANDE HEE
Bit BUN (PRO),	de cital de cital de table (AMMAL) (AMMAL) (Sama Cital (Sama Cital)) (Sama Cital (Sama Cital)) (Sama Cital) (Sama Cita	001 2000 1 2-1 1402 0 00 2-1 1402 0 00 01 01 01 01 01 01 01 01 01	All Casality Deletics Party Cardinal Statements Party Cardinal Statements Party S	10000000000000000000000000000000000000	ON Anne (Ar.M. Serger Lean C. Serger Lean C.	ne jariak osalak ne jariak indis indis indis indis indis indis jariak jariak	Addati (M. M. Addati (M. M. 10 Tanahaman 10 Tanahaman	- INNEL OF LATER AND NO
mit.mit/max, mit.mit/mit/mit/mit/mit/mit/mit/mit/mit/mit/	de case de case de case case case case case case case case	Hardon F. Hardone RC RC RC RC RC RC RC RC RC RC RC RC RC	All Casality Detect 20 people (2010 Higher Anti-teach References Angeres All Insue People Lak- media (2010) All Insue People Lak- media (2010) All Insue People All Insu	10000000000000000000000000000000000000	NAME CONTRACT DESIGN NAME OF MALE Nonpolation () Nonpolation () Nonpolation () Nonpolation () Nonpolation () Nonpolation () Nonpolation () Nonpolation () Nonpolation ()	nd Sairtak nd Sairtak Sairtak Sairtak Sairtak Sairtak Sairtak Sairtak Sairtak Sairtak Sairtak Sairtak Sairtak Sairtak	Partiel Natable (15, 16) 11 Tabata Anton 12 Tabata Anton	ELANANG, DEGLA DEMAANDE HEE

Figure 4 View Data from Testing Data of the Information Technology Department Class of 2017

The data comprised prospective students of the Information Technology Department Class of 2016 who had and had not re-registered, totaling 805 registrants. In addition, the 2017 Class of the Information Technology Department possessed 1,293 prospective new students. These were pure data from the database of UMY.

3.2 Data Selection

Data Selection is the process of selecting data for analysis from a database based on research requirements. The data possessed the attribute names sex, province name, and non-subsidized pathway, derived from data collected using View Data in SQL Server Management. Sex is the gender, the province name represents the province of origin, and the non-subsidized pathway refers to the registration pathway prospective new students took. To simplify research, the labels of these attributes were altered; specifically, sex became gender, province name became province, and non-subsidized pathway became registration pathway.

3.3 Data Cleaning

Collecting data necessitated a cleaning procedure to prevent data duplication before it could be mined for insights. After the data had been thoroughly cleaned, they were saved in a new Microsoft Excel CSV-formatted dataset. The data extracted from the Information Technology Department contained null data; thus, it must be eliminated because the data mining system employing the Naïve Bayes method could not analyze null data.

In training data, 246 records were discovered to include the same data. However, there were 347 records in testing data, resulting in the registrant registering several times and entering incorrect data.

0	- C ŵ	V a marchine										
6	Concernance and second				(TREADED)	REAR AND AN A STREET AND A ST	PTTURES/	CBNS		* *	n III 6	
		ALLEMANTIZON = III III	We Level									1
	1000	L	1.1.1		-	Carlon and an an						
7		contraction and		1112	8.11.9	- Y - S - D - In h	10 T - A -					-
+			1.2.1									
٥.	intention of Transie	should finance	HER NAME	LUF TRUMU	ad tertio	AME OF DESIGNATION						2
	ADD-LOWN	Diff Number Drive		es Auto Davassian	THE	Dr Tanis referance						100
	VIRGINIZ	INTER OF MALE PORT AND IN		ra Auto Developer	Fallet.	Sylami remain						
	100000	1978 Auto Salesar		in American Statements	6404	Tribers emerging						
	markets 2	20 di Grap Sana And		en Agen Tananaken	1089	S r Taxan Internation						
	engine las	2010 Taxof Futures	1.000	in Auto Servation	PRIME	ST THINK MANAGER						
	10000037	THE ADDRESS		part of the second	Tid.ast.	51 Max viteration						
	with size in	In M. and Talky Schemes	1.0000	pa dope the reader	**	8 1 Ma						
	THE OWNER OF THE OWNER	the literature		to Anth Danageration	1.07	a the standard and						
	TANK ADD TO .	19th local Tarty Artist	1.0000	ro Aven Tananien	197	entena nomena						
	10022000	27% AS Supplement	1.040	ATTACANT AND A DOLLAR DO	ORT	ST Sector other states						
	NUMBER OF T	27 K-Harra Shainak Sylatioutemena	2.244	in identical fragmations	FIELD	St lass strains						
	Watercom.	2016 made landywiften	1040	all taking in particular	112.45	21 Tells I Dirights						
	10011061	JUST & Dana Harton	1200	all latences (lagranged the	87	STN 10 10 DAMAGE						
	10000101044	2010 Katherian Babrie	1240	ali belimane 'rispakelie	PURDA	21 lass storate						
	meconic	20 Killion Human Banashari	1.7.00	at taiming frankatia	1.07	Different enternalities						
	401010400	215 Tr-Deter-Duberneti	1744	in latences Topologia	17	117e-ci reteralta						
	108122112	2010 President Australia	0.0	and inclusions "Imposituation	PROK.	11 Terris principle						
	100003107	2018 Ella son Voelanery	0.044	all taken to a second at the	PROF	STRON PROVIDE						
	DOMESTICAL.	20 In Mari, Ball at Prototol	1.044	at latenasi regatatia	100	A 1 Marcal Volumental at						
	MMATCH	20 hatrid with	1.040	in minute Tigetalla	647	ATTACK MANAGES						
	minertel?	DOM: Taxing Dive Calcimic	1.044	an spontage ingual after	87	En fansk Administra						
	106102318	274 Realized heat		enterna l'apalana	PADE	STIBUS PARTAILS						
	1010-0001	2010 Avrian Dates (Frankrist	1.044	on latinuina Tispiakatia	17	S17ault relevable						
	10011471	10 H Aver Roome		ATTRACT BOUNDED	Sum-	S174-ck trimates						
	10010012	Bill Almacitant Mate	124	shieger' available	ar	ST Tates edenation						
	10041103	37 Rotannal For Senani	1044	situlation in the second state	181	5 Class nameka						
	0004/1012	20 K. Kulaning Plat Notice	1044	and a second and and	1281	51 Wash Internation						2
		March & Article & Partners	1.00	of the second strength strength of the second strength st	-	And the second s						

Figure 5 Conditional Formatting 1 in Training Data

Conditional formatting is a feature used to automatically modify cell or range formats in Microsoft Excel or Google Spreadsheets based on specified criteria. This study employed criteria based on the number of FullName column cells containing identical data.

	CAMARIA									1144	-	÷
1		the mant forme liese fasts date		and to be a second of the second						- 12	a line	ł
-		10% - 3 % 5 12 12- 00		- + + z = <u>A</u> - 4	811-8	$\tau: T \in \{0, \cdots, N^{(n)}\}$	- 1	1				*
											-	
6		TARJORA - FIRANON	1.800	- RAME, OF, PROVANCE	- 411,007.01	ANK OLDERATING						6
	10010304	and instances Rear		1 Nation April Desemint	AND	Sillank ethmake						
	VALUES	2010 Californi Fatti Aulifa		3 Kerps Ash Developer	PM24	21741k ottroaks						
	10022003	2016 maria filament		1 Regis Aust Devision	71055	21 Talue otherpase						
	0840110	20 K. Ornar Simo And		Charge Ant Davasies	187	2 (Teloli Inforceba						
	TORNEON	214 Decthorage		1 Sample Auth Danisation	Think	27 Monte triberial Ara						
	10000120	27.97 Abeyune		C Tranges Aust Davasation	710/4	TTTM of URITALS						
	1001340	37-91 Add Taping Derivative		T Karys Aret Davesalen	-	215k4 stored to						
	108410072	278:31e8x17x83		1 hange dark Devenime	187	Ellipsik bilanakka						
	WHERE AND A	STR met Talukter		1 Kept Aut Draken	445	a the examples						
	NENDER	216 Hack Dyards Systematical		difactorianese toppicate	Falls	21 laint startable						
	1000045	2745 Availab Butting A Public		C. Daniel Inclusion Vegetiane	PALM	3 (Spinik Schutzpika						
	10411061	2010 Project President		1 Danet Industry Vepelant	47	E-Claim etternalite						
	100.0104	2110 Hatvanar Ragine		I Denthictory Vepticity	7404	El Talvé referrative						
	108-0308	214 InterNation Reportion		1 Dentitalmene Vignilate	087	5-Tanii mmaka						
	1010-0008	39 61 TV Diviter Submonali		-Connectatives Vegetata	07.	317M-Withmake						
	10100-0	2010 Fatchis Aptaci		- Canolidation Taplitate	PM04.	3118-ID INDIGAL						
	100000107	2010 2 Grant Vancenty		Citamic Informal Vagatilation	P10.00	D'I'M of all and a second second						
	100010-007	2019 Han Ball at Plating		1. Daniel receive vignatiana	1.87	21758183.08/14959						
	100000000	IN M. Passer Works.		* Stand I division Topyleine's	287	817M of Alternative						
	INCOMPANY.	37% Tasks fre Lature		1 Daniel Infolme Veryleigter	10	Erfleini ethniatka						
	VALUED	22-9 - Fasting test		C Canot I pincese Viggenatio	2404	3 (Section education						
	WITHHE	216 Anna Salai Vicanaria		 Danstiemene Vagenate 	10	21 face attends						
	tig risets	2010 And Ballys		1. Creative transitions	diame.	217 Auto Informative						
	and based	20 M. Advantition of Values		1 Daniel Informer Vegatiana	10	STRAM MONANA						
	#396-111X	If it inverteel for honey		7. Cantol Information Consultation	110	STRUG HERMAN						
	VIRENDOM	2010 Autor Safe Statema		1 Danshistmens Veryalista	FARTH	117anii kilonakka						
	and the set	2010 Link Dark Service		1 Denti latente Viguelata	07	31 Nami Islandka						

Figure 6 Conditional Formatting 2 in Training Data

	-> C &		CATA MANNE T 25										- D. A.				
	- C 2		D A https://docs.	feedigrau,) t		0,6/1565	-72c-85el	mp4Nik2;	POdDRug Y25	QUeleQv0a	Clied 18g	85 ··	- © 2			0.6	1
-	DATA TESTING TI 2017	* 5 4	And the second se											-	a then		
-	The Edit View Insert If:	max Date	Tank Address Male Access	why latestas										-		•	
-			10+ DAMAG 10		1.4.10		A										
	In sec.		and manyor				*										
	TALING C																
		<u> </u>															
	ALL DRIVE	-	1 Intel (1 PROVIDE 1 Terrar for Development	T All Safer 1													
	to Talka heata		8 Nergen-Acut Devasion	Pada													
	CHARGE BET HORS		1 Terge Add Developer	100													
	Contraction score		1 Nerges Acet Development														
	in Lines Buildingson		< Hargo-Anal Davasters	Pada													
	CONTRACTOR OF		1 Terrary Aug Development														
	tate Internet, 112		University Investory														
	the Durah sub-Search		< Nergen-Acute Desumation	10													
	the floor framework		E Terran Audi Deventern	-													
1	Table 1 Security		E Nergen-Fault Development	PAGE													
	mail also big		 Nergen-Acab Devasionen 	087													
1	topia Pala		I Terrary Aug Development	PAGE													
1	an lints logoant		2 Genetications Vegetarie	PAGE													
1	auto Anno Feature		1 Depth Minard Vigoriants	PAGE .													
٦,	Manual Roat Design		1 Desit Minera Yapataria	70													
0	Trans line in		I Denti Minare Vegetare	PADA													
	ant Pargetty		Countrainers Viguesters														
	BURAN ALVER ADVE		1 Geneticativas Vagataria	17													
	ing Roha Ninalauma		Case Time	ar													
	OWNER ADDRESS (CANAD		1 Depit stimes Yigalarts	01													
	ARCH DV: PRAIL.070		1 Deal Minks Vapataria	17													
	utmat Page		1 Denti Minace Vigotere	-													
	to film Norson		1 des Ting	087													
	Telep Televille		1 Dent Minera Viguraria	10													
	had introducers		1 Denti Minera Vigalaria	arr													
	ONE TROUBLE		1 10000	10													
	Ne Mere		1 Dest Mines Vigitaria	7101													
	totals lidge from		E Danit Minara Vigalaria	PAGE													
	enge (Hyer-knistje		1 Destroyment Vigward														
	Chie TREventeux		1 March 1	en off													
			1. Janua Tangath														

Figure 7 Conditional Formatting 1 di Testing Data

Conditional formatting is a feature to automatically change cell or range formats in Microsoft Excel or Google Spreadsheets under determined criteria. This study utilized criteria following the number of FullName column cells containing identical data.

First Xnov suffered for	extensets Instit 1940												-	8
0x7A 1651MG 112017 - 0	ianyi 🗙 📓 Galta Teatensi 11 Ji	Hi-lim X	+											
) → @ @	0 A https://doc	geogik.com/1	preschived	6/6/1586-	(Zo-85e81)	(Mulazije	status (2)	QUellarda	Columb	1-0 (B	m)	- D ☆	16 E	θ
DATA TESTING TI 2017														
	former Dana Hora Addison Halla Acces I In al Algunar - Mathematica - M						W						_	_
	a f d an annar - a		<u>_</u> ~ 0		T - H - A		1.0					-	Conditional format rules	
112.20.00	THE THERE IS PRIME	AL 144										_	 Assessment with the 	
Parlana Regard	E Verse Aug Description	100												
the light letter	E Terren And Devening	200										_		
INCOME AND INCOME.	Chargen And Descention	10												
Performant Name	- Harpen Analy Descentions													
Revises Buildings	1 Harry And Tenanters	Page 1												
STANDARCHAR	1 Territo Add. Development	-			_									
This Island, 51	1 Terrary Aug Deventory	-			-									
NOT TA AN ALL THAT	1 Names Aug Description													
Hate Day Spring	E Terren Anti Deveniter													
Partie 1 March 10	E Terren And Development	PHON I												
Date All Disk	Companying Street Ten services	187												
regate Parts	E Territo And Development	Page 1												
Las farta Argeneri	E Desh shree Tayra's	PHER.												
Table Army Party	1 Desh Mines Taylor	200												
Palanush Roat Centure	1 Death Minister Transfer In	PLCN .												
Proper lists	I Dent Monte Vignation	PAGE 1												
Dava Parquete	Classic Minana Vigotaria	-												
Distant Arrists	1 Deput Million Vignatio													
Densel'in ardoma (Clianic)														
HANDA DALIMAN (1970)	1 Desh shney Yigalara 1 Desh shney Yigalara	0												
monal Teal	1 Sect Sheet Tapers													
North Test	1 deal birries Toporaria													
Party faile Internet	1 Description Transferie													
The climiters														
the contact	1 Seech Million Vigolate	200												
Tankin Villa fut	1 Seech Million Vigolana	Page 1												
	I Deph Strees Yourse													
Henry Otye: Anthja	1 Dech Miners Tigerarb	PHONE IN COLUMN AND ADDRESS ADDRES ADDRESS ADDRESS ADD												
KINE TREAMINE A	1 Marchae													
DRIVE SPACE DRIVING WE	1 Dente Minace Togetaria													
Rates Frinal, Hearth	E Sand Minas Tigata's	ALC:												
	1 Sand Minaso Vigolate													
Pratici Artica Sart	E Sand Minaw Yigalara	See.												
and Products	· Dept Mines Tighters	PHON												
Petro Caracifront	 Desk strain righters 	PHONE INC.												

Figure 8 Conditional Formatting 2 di Testing Data

Sorting duplicate training data resulted in a reduction of 706 records and the production of 1,169 records of testing data.

		+ X 10217 AM0284											-	
.)	-> C 1	0	https://do	- google.co	m/spreadshe	etu, izy'i UMSE	CMPP/05/201	ofa-2ah/Kollp?	ISGQEESWEW;	hateyi (B	s) 0	Ω ±∎	000	•
		MAHASIDINA TI 2016 -			edi ya acce	0.00							é Stars	
	~ 0.7	105 - 1 5 4, 4	121- Aul		• B Z	6 ≜ ≤	B 11 - E -	Υ - 10 - 10 - 10	• F E 7	- I ·		Conditional format rules	×	
	Jania, Kalamia	Proini	Jaria Datar									 Add another rule 		
	-	Java Targah	PHEAM			-						-		
		Kaimantan Tergah	PMDK			•								
		Kaimantan Terpah	PMDK											
	1	External Uters	007											
		Java Teur	PTELAN											
	1	Tuno Trougen David (1771)	PMDK											
	1	Daniel Islands Topolate	CHT											
	1	Emplo	PMDK											
	1	Langung	087											
		Etabelia:	097											
		Earliet	0.01											
	4	Jeve Torgah	10											
	9	Sumalary Selaten	PHEIABI											
	1	Kalmantan Timur	087											
		Name Terrygons Daniel (NTE)	087											
	1	Java Tenut	087											
	1	Lampung	P1424											
		ceve Troph	**											
		European I and	PTELON											
	1	Kalmarter lived	097											
		Sulevenities	PMDK											
	1	Java Tegah	C01											
		SumdersUters	C01											
		Kalmantan Selatan	PMOK											
		Denth latimena Topyskata	CIRT											
		Jeve Torgah	CBT											
		Sunders/Jims	PROAM									~		
		The set The set of the												

Figure 9 Attributes of Gender, Province, and Registration Pathway of Training Data

	MG712017 - Bao	9 X 🖬 🖬	TA TRAFFIC	672916-64	-× +												
-) → @	₽	0	🖨 Marci	Non people	kam'ipre	ectrimento, Adj	/1546-YZo-I	5+670)4Ns	625/5404	w72:00+	kQv0xDijk	dt#gid+0	(8%)	(9 ☆	16. C	θ
		0 M () Annual		Accessibility	Control I was a											a 🖬	-
		5 (5 (6 (m)))	and an a	B	$z + \underline{\lambda}$	4 B II -	E-1-1	- 5	P 8 1	- 2 -					-	Conditional format rules	
																A DECEMBER OF	
Janua, Parlamon	Pulle	Jacob, Dallar Prices	·														
c	-	THER.	-														
C		10	-														
C		11	-														
5		PHON.	-														
5		10	-														
h		10															
h	A.																
		10															
P	*	PMCH.															
		ORT PAGE	_														
2		Photo: No.	-														
-	-	PADA PADA	-														
C		PAGE .	-														
C		(THEN	-														
C			-														
£		18	-														
3		04															
3.		07															
3.		100															
		187															
2		10															
2		CET INCO															
2		PHONE INCOME.	-														
r	-	PAGE .	-														
-	-	100	-														
5		107	-														
5		PAGE .	-														
5	-	GIT	-														
		Read of Lot of L															
		PAGE 1	-														
1.		(ALC N															

Figure 10 Attributes of Gender, Province, and Registration Pathway of Testing Data

Attributes used to be processed in the application were first initialized.

Table 1.	Gender Initialization
Gender	Initialization
0 (Female)	F
1 (Male)	М

Table 2 Gender Initializ	ation
Province of Origin	Initialization
Sumatera, Riau, Bangka Belitung	А
Jawa, Bali, Madura	В
Kalimantan	С
Sulawesi	D
NTB, NTT, Papua	E
Overseas	F

3.4 Data Transformation

This study processed the data through the SQL Server 2014 Management Studio database. Then, they were converted into a CSV file for data processing in the RapidMiner software. Data attribute names sex, province name, and non-subsidized pathway were changed to gender, province, and registration pathway.

) → C*	ŵ 0	A https://doc	geoglass	m/spreadsities	es,is/10MSE	QNE4820	oGs-2atel Kollpl	35GIJEDWAN	harry: (88	0 6	14	7 W	000	
	U MAHASIDINA TI 2016 - 1 View Inset Farmat Data											9	é State	4
	105-155.0.0					a na lata	i - li - Ye				~	Conditional formal rules		
					-		A . 11 . 7					Conditional format rules	×	
A		e												
Jania Kalami	Proint	Jaria Datar										 Add another rule 		
1	Jeve Torgah	PHUAN												
1	Kalevantan Tergah	PMDK				1					_			
	Kalevanian Troppi	P1424				-								
1	T-description	0.07												
	carve Terror	PTELAN												
1	Name Terggers Sweet (NTE)	PMDK												
1	Depait Islimente Vogenkatio	007												
1	Employ	PMDK												
1	Lampung	007												
	Walatio	0.01												
	Tarian	CBT												
1	Java Traph	-												
1	Turnelword Hoden	PTEM												
1	Kalerantas Titur	CMT												
8	Name Torogam David (NTE)	C81												
4	Java Tenut	001												
4	Lampung	PMDK												
4	Jeve Tergah	10												
0	Sumday Bank	PHUAN												
1	Kalmanian Band	081												
	Euleves/Uses	PIECK												
4	Java Bright	C87												
	Sundary/Jack	087												
1	Kalmanter Selater	PMDK												
1	Denth Islimena Yoppicata	0.01												
1	Jeve Tergah	0.81												
1	Sumaional/Item	PTRAM												
	in a Branch	ALC: 1.8												

Figure 11 Data Transformation

3.5 Implementation

This study utilized probability (prediction) based on the Naïve Bayes algorithm. This method could better predict possibilities by comparing past experiences.

Using the RapidMiner software, training and testing data were divided and evaluated separately. The Information Technology Department yielded 706 training and 1,169 testing data records.

After selection, cleaning, and initialization, the following training data in CSV format were prepared for data mining.

CHANN	HARMEN N	Data 10	17945 11207	× 1	BATA TRA	INNEST 2011	× 🖬	AMERIC NUM	MARKE 2	🔹 Tak	quare Web	×	- IN 19	encircular	H-C- X	10 PTT PA	NOADURAN-	6- X
)→	C 🚖	0	A https:/	the a goo	gk.com/sp	wadsheets/S	/1HPC-Put	lu NOS TadAver	Notemine	47521ac31	fo41/DrAd	lidtfgi:	(75) -			ł		Θ
	ARU MIRHADISWA 1																	
***	in the lost for	nati Data Toota	NOTION MADE	No. and Surger	Lot off, edit	10000.000											_	_
	9	した時間に	- 24		1 2 4 1	4.00.00	· · · ·	8-8-1	- D B Y	- 2 -								-
Sector Mar	T Paula	Annie, Talla	Thread and	-														_
	4	10468	-															-
	4	OBT	-															-
		087	100															
		197	144															
		10	-															
	4	Page 1	-															
		CBT																
		Page .																
		PHON																
		PHEN.																
		1987	-															
		1000	-															
		CHE																
		15																
		1007																
		0484	-															
		Dam's																
			-															
	-	100.00	-															
	-	-																
	-		-															
			-															
		-																
		-	-															
		200	-															
			-															
		CONT.																
		100	100															
		0.87	-															
		PROPERTY.	-															
		car	-															
		PMDK.	100															
		Carrow Carrow	100															_

Figure 12 Training Data in CSV Format

No. 121. 764	113017 0 11 0 Nast Runal Data no. 5 5 5 5 5 1 Nast Nast Nast Nast Nast Nast Nast Nast Nast	A https://do	and the local	<u>⊥</u> = =	1 - 1 #-	L + 1 + 1	Rating.		uwqvad	oleanta.	(6%)	0	Ŷ	а 1 * ж	0 0000 0 0000
10 10 10 1 10 10 10 10 10	1001 Normal 200 10 - 1 5 5 5 1 10 10 - 1 5 7 1 10 10 - 1 10 - 1 1	0- 1660 AL - 1	1.2.8	<u>1</u> = 0	11 - 1 H -			7 - 1 -							
an june				_				7 - 1 -							
1 1	2200	- min	s												
1 1	2200	- 100000													
	2200	2	-												
	1000	-													
	*	-													
1															
3 #															
	1.00	-													
	7828														
h #	10	-													
P 4	/NOL	-													
P 8	- 197	-													
	197	-													
* *	-	-													
5 B	10	100													
h #	. ar	-													
P #		-													
		10													
		-													
		-													
2	79.01														
<u> </u>	-														
	1904														
	100														
1 1		-													
1 1	10														
2 1	796401	-													
1 8	/NOK														
	170,001	-													
1. 8	1983	10													
P 8	10	100													
1. 8	10	-													
1. 8	ar	-													
1	R.	100													
· ·	PIOL	-													

Figure 13 Testing Data in CSV Format

3.6 RapidMiner Software Testing

In this study, the re-registration attribute served as the label. The projections of prospective new students who would re-register based on Information Technology Department information were examined. The data must be in CSV format to be processed by the Rapidminer software.

After the data were converted to CSV format, importing was performed. It was accomplished by including the read CSV operator in the view process (Import - Data - Read CSV), followed by renaming training and testing data through drag-and-drop, as displayed in the following figures.

	<new proce<="" th=""><th>sa*> - RapidMiner 5.3.015 @ SAPUTRI</th><th></th><th>- 0 -</th></new>	sa*> - RapidMiner 5.3.015 @ SAPUTRI		- 0 -
Die ER Docess Jools view Holp				
🗋 📦 🖩 🖬 🖉 🔊 🛝	4 🕨 🖩 🔳 🛐 🗟 👁			
Constants 10 10 0 10 Constant 10 10 10 Constant 1	Process 7 100. 1 0 • • · · · · · · · · · · · · · · · · ·	بې کې کې ۲۰ تا ۲۰ کې 	ipetoch	Control 11
A Real Docesson A Real Accesson A REAL ACCESSO			🚵 4 hidden expert param	alas
er ing der 8 der Lacal Repository (Lasiana Ving Tapatri)			O Here X 11 C	met X
			Process	
	A Potens II 10 Lts II		Smaph	
	St 2 potential problems			
	Message The mandatory parameter "car file" to undefined The mandatory parameter "car file" to undefined.	Tees Leaden	every process. Acti	which is the outer most operator of ivate Windows PC settings to activate Windows.

Figure 14 Drag and Drop and the Read CSV Operator

•	<new process<="" th=""><th>#> - RapidMiner 5.1015 @ SAPUTRI</th><th></th><th>- 5</th></new>	#> - RapidMiner 5.1015 @ SAPUTRI		- 5
			Rifeendes II @(crist)	
Constant II II a B		a - 🗆 🛊 🔟 a .		
Process Control (37)	A . H E MARCHINE .	0.11 V 11 0 0 0		
4 KAMU (52)	Their Process		Bata training (Read CSV)	
Expository Access (5) Import (37)			A Import Configuration Weard	
Des (19)	No. Contaction and		cor fie	
- FastCostwitrFormat	1		esium separatan	
- Read SAB			😥 use quelles	
Final All,	Eula Toning		avotes character	
Read Aller Read Offer Read Outstance			itip commente	
Repositorios X 2 4 1			🕑 parse numbers	
			decimal character	
G DB G Local Reportery (Locara King Tapatr)			O Help II Convent II	
			A Read CSV (tapithiner Core)	
	A Problems 11 Diag 21			
	St. 2 polential problems		Synopsis	
	Vessage The mandatory parameter "cavifie" is undefined.	Fixes Lacation	This operator is used to read CSV files.	
	The mandatory parameter 'confile' is undefined.	🕎 Set mandatory parameter 🦂 Data Testing	Description Activate Windows Go to PC settings to activate	

Figure 15 Importing Training File

The Import Configuration Wizard must be selected to bring up the Data Import Wizard form of the first step. In this step, the file location of the training data must be determined.

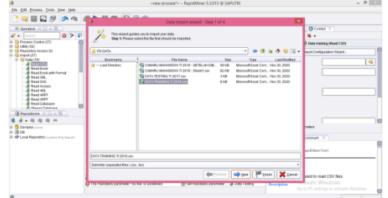


Figure 16 Data Import Form of Training Data

After selecting the training data, Next was clicked, and the Data Import Wizard form from the second step appeared.

	-		of willow three 2 of 4		23	
C (people)	10 M.F. 1944	the cash front out the			Const :	
Ø . 1	Ja Sep 2 Peace	specty for the line should be par-	ndard too citeren an argented	É.		
it: D Prozess Collect (27)					6 Outo training (Read CSU)	
IF TO Reputitory Access (8)	Fialinating		Column Separation		real-Configuration Woard.	
(0 the impart (27) (0 the Date (19)	FileEncodeg	and the first of the local division of the l	S Comme (C)	Otere		-
-4 Diversion	Tretan	,	Citerator ?	014		191
of Real Days	C) restress		C Regular Danesson	George I	61 E	
- A Prior 100,	i Big Cannets		Except Character			
Read SHQ					-	
d fired rife.			Citike Darme	1		
- Real INTY - Real 1977	Jana, Kalanan	Provine	Janta Dattar	Hangahasi	0	
Anat Distance	4		PRUM	Yes.	C1	
Pagesticas of 21 or 22	1	6	PMON	741		
84.000		0	PMDH- CHEF	Tes		_
tores			PRESA	745	New York Company	
+ 15 CH	1 C	4	PEDI	Tes		
······································	à		CAP	540	and served in	
					0	
	Rin Gilote	Dist	Criptwinetut	Thursday	aprilliant Core)	
			Care Contra			
			des Greating 1	biest Migran X La	ned to read CBV fies.	

Figure 17 Data Import Wizard of Step 2 of Training Data

Then, the second step form appeared, as illustrated in Figure 3.16. Then, Column Separation Comma was clicked. The comma separated one attribute from another since the data would merge if the file were in CSV format. By clicking Next, the third step form for the Data Import Wizard emerged.

Image: Control in the	<u>) 🛶 🖩 🖬 🖓 🧆 n</u>	1.1				Da	ta import wiz	rd - Step 3 of 4				ж.		
Bit Markov Andráz Adia	Process Cantrol (07)	4) ()	Ž	Step 3: In Flat	and defines t	ach all thus can be sames of the	be annotated. T	data dows will contain all	SDJA STAL O	e is its name not set this	- arter with property. If	1	Outs training (Read CSV)	
Image: Constraint of the	in inspectatory recess (1)		Annotation	- 41	10	and a	184					1	rport Configuration Weard	
Image: Control of the sector of the	(D) 100 Date (197)		Name	Jania Kalar	Proveni	Jania Daffar	Planaglaithaan					a [
Image: Section of Example 1 Image: Section of Example 1 <t< td=""><td></td><td></td><td></td><td>L</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></t<>				L										
- -				L	C	PMCK.	Yes.					- 1		
Image: Second	Paul IM.			P	C	PMDK	Ves.							
Image: Second				L	D	081	Yes-					-		
A main starting L A ACC No A main starting L L A ACC No A main starting L L A ACC No No A main starting L L A ACC No No No A main starting L L A No No No No A main starting				P	8	PNUM	Yes							
A max start - L B OT M A max start - L A				L	E	PMDK	Yes							
A metanoma B L A Off Mail A mail A ma	Read 1077			L.	8	CRT	Yes							
Image manual state L A OP Mail Image manual state -				L.	<i>4</i> ,		Yes							
d d d d m		· · ·		L	6		100							
Impacts Impacts <t< td=""><td></td><td></td><td></td><td>P</td><td>t</td><td>CBT</td><td>100</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></t<>				P	t	CBT	100							
B L A M244 Ma P (act Ampetity survey top base) L C C C Ma Ma P E CH Ma Ma Ma Ma Ma L L C CH Ma Ma <td>4-9994</td> <td></td>	4-9994													
B · L A PROM Mail P (actil Repetitivy Lower Veg Name) ·	Samples			L	8	87	Yes						laters.	
	8 00			L			Ves							
- L B CPF No - L A PNDK No - L B ZF No	P Local Republicity (Locara Very Deputy)			Ł	C								owners ×	
- L A MOK Vo - L B R7 Vo				P	£							1		
- L A MOK Yes				L	8							1	and Different Council	
				L	A							- 1	provident control	
				L	8									_
				P	6	Phylippe	Yes.					. I.		

Figure 18 Data Import Wizard of Step 3 of Training Data

In the third step, the contents of the training data appeared. After clicking Next, the Data Import Wizard of the fourth step emerged. Then, the attribute column serving as the label was selected and modified. The re-registration attribute was designated as the label attribute. Afterward, Finish was clicked, and the process of importing training data was complete. Subsequently, testing data were inputted in the same manner as training data.

	<new 5.1015="" life="" p="" process*x="RapidMiller" saputre<=""></new>	- U
Die Edit Docess Jools New Help		
👌 🖬 🖬 🖉 🔊 🛝 🖉	Data import wizard - Step 4 of 4	
Convertien (1) (C) Convertien (1) (C) Convertient (2) Convertient	The solid galaxy is to important this. Set 3 Registration was an object the advective to the solid set of the solid	Context II Sector Table Source training (Read CRV) evel Contgoration Woods.
Breet Exert Food Exert Food Exert Food Exert Food Exert Food Artific Food Artific Food APP Food AP		
Dependence 21 22 0 10 Dependence 21 20 Dependence 21 Dependence 21 20 Dependence 21 D	L 8 (87 Ta)	nas
2 🥐 Local Reporting (Local Try Tryon)	© Darres © Darres ettes © Darres ettes New Course Darre Organization Øld Darles ette Organization Missiope	see to read CBV files.
		to PC settings to activate Windows.

Figure 19 Data Import Wizard of Step 4 of Training Data

	<new -="" process*x="" ra<="" th=""><th>idMiner 5.3015 @ SAPUTRI</th><th></th><th>- 0 -</th></new>	idMiner 5.3015 @ SAPUTRI		- 0 -
Die Die Docese Inoje Diem Daie				
👌 🗑 🖩 🔛 🥬 🔊 🛝	🥔 🕨 🖩 🛐 🗟 🚳			
Coexitors (X 32 do 62	2 ¹ Process N Z XM, N		Parameters II O Control IC	
a Devent and Articlet	ψι + H0 - 2 ₩.Process >	0-0100000		
8 📮 UNIN- (52)	Then Process		d Data Tenting (Read C SV)	
E Supporting Access (5) (1) Support (27)			A Impot Carifiganilian Woard.	
C 😘 Data (19)	es) [Intertuning		cavitte	
- A Feat Excel	(- <u>.</u> -)		solumn separaters	
Read XM.			ware queter	
Read Access	Inte Institute			
Real AUT			pates charadar	
Field Extense			skip commerits	
Papentinies 32 52 6 10			🖓 parse numbers	
8 d - 1 1 1 1 1 1			decinal character	
D 😏 Samples (norm) In 18 CA			A Ehitten eigenparameters	
T @ Local Papersitory (Locard Yorg Deputy)			O map 31 12 Comment 31	
			interest (Republican Carry)	
	A Problems II D Log X			
	Se One polanital problem		Synopsis	
	Except One mandator parameter "coville" is undefined.	Fixed Location	This operator is used to read-C5V flies.	
			Beaription Activate Windows Go to PC settings to activate	

Figure 20 Importing Testing File

The Import Configuration Wizard was selected to bring up the Data Import Wizard form of the first step. During this step, the file location of the testing data for the research process was selected.

0		<new process*=""> - RapidMiner 5.3.01</new>	5 @ SAPUTRI		-	5
Die Eqs Loosus Tonis New Helb						
📑 📦 🖩 🕁 🖉 🔊 🛝		Data import wizard - Step 1	of 4			
County 1, 12 0 0 P 81 County 0 P 81 County 0 P 81 County (2)	2	ruis import your data. the file that should be imported.			Context Totals Tending (Read C SV)	
18 Repository Access (8) 19 Repository Access (8)	a PR.DKTA			। 🖗 🤨 🖼 🗉	reari Configuration Woard	
Australiant Australia	5	Restance Constructions of the Construction Construction of the Construction Construction of the Construction Construction of Construction	12 KD Historit Excel D 3 KB Microsoft Excel D	Lad ModRed inn. Nev 20, 2023 inn. Mev 20, 2029 inn. Mev 20, 2029 inn. Mev 20, 2029	refere	
	DATA TESTING TI 2017 DV					_
	Contember segmented files (core, for)		*	•) 2000 - 1		
	·	(4-		Description	used to read CBV files. Wrate Winckows to PC settings to activate Window	

Figure 21 Import Data Form of Testing Data

After selecting data testing, Next was clicked, and the Data Import Wizard form of the second step appeared.

0	-	new process*> - RapidMir	er \$3.015-@ SAPUTRI			- 0 0
En En Louis Jon Hab	• • • • • • •	3	Data imp	ort wizerd - Step 2 of 4		
Convertions Convertions Convertions Convertion Convertion		File Reading File Reading File December The December	dee you'ts import your dats. specify thee the file attrautilitie particular to particu	Column Separation Column Separation Column Separation Column 1' Column 1'	C Space C Tale Science E	
Paposterias 2 3 4 6	4 ^{1.1.1.1.1} 0 ¹	Janta Johanna B L L L L L L L L L L L L L L L L L L	Profesi A A A A A A A Exer	Jera "Deter Patok Patok RT RT RT RT RT RT Object value	Maragishasi 194 195 195 195 195 195 195 195 195 195 195	
	Probleme Diag in Probleme Diag in Probleme Diag in Probleme Diag in Diag			Get Dresser Inter opposite of the Description	The Age Same	X Gancer

Figure 22 Data Import Wizard Form of Step 2 of Testing Data

Then, the second step form appeared, as depicted in Figure 3.21. Then, Column Separation Comma was selected. The comma separated one attribute from another since the data would merge if they were in CSV format. Subsequently, Next was clicked, and the Data Import Wizard form of the third step appeared.

6 Operation () () () () () () () () () () () () ()	2 trees - 2 Marson -	1 %	that dravations	e daferan ika	els additione in a statute celle	to be and stand to	to manimum restant annotation of an attribute on its name - a row with data down and cardian difficult instance, do call of this property 2 file, concern an experiment from	
CBMby (52)	Party Processo	ANNER						
Contraction of the second seco	10.0	199149	.mmit. Matter	Provinci	WH.D#	a melepolitare		a 1.
Cont (10)	-		*	*	PERK	188		18
Cala Transformation (114)	RAIN TRAMES	Index Brown	P	A	PREK	185		18
The dation of CTINE	40 3 40 - 4	N 1/24 -	A	A	147	188		
and provide on which		100 M	S	A	1417	100		- 8
	A COLORADO DE LA COLO		h	A	PNDK	199		
		· · ·	Ł		187	100		
			L		107	184		
	And Andrewson (h		41	189		
	Anis Income				10	789		
B Preventeries (0. 12 to 10	4. 3		č	e	PREK	194		
		Page 10	S		PREM I	198		
4-000-		-	-	-	PARK			
D Tarques		-			PREN	199		
8 ce			-		PIEK	100		
@ Const Reporting Concerning Security			-		and a			
			-		101	100		
	-				207	182		
	A.Painares of the states				at.	100		
	S in protectured	-	-	-	100		And the second s	2
							Carol Marine Marine See	
	bistriage			_			THE OPTIMUM OF AN ADDRESS OF ADDRESS	1

Figure 23 Data Import Wizard Form of Step 3 of Testing Data

In the third step, the contents of the testing data appeared. After clicking Next, the Data Import Wizard of the fourth step emerged. The attribute column serving as the label was determined and modified. The re-registration was selected as the label attribute. Then, Finish was clicked, and the procedure of importing testing data was complete.

N Tarrier - I D		0 \$ G \$	12		f pailine year to	ingentrying		(permit) - Stag. 4 of 4	
Concernation (S7)	4 - 10 -	2 Breast +	10	Flagidding	e autorigena cielana	ACT AND	taking inducing	what they can be used for to the testimote of all they can be used for to the testimote of the real (There erities).	
Approximation (ii)	1		8 1-	#10#s	The games of	he lowo	[]*****	an on half lines. Debuteral	
Gapart (16) Capa Trautomatics (176)	1	alla (Research) (Annue R	9	-	-	9			
Calified Mitep (1116)	1 8	4-1-4-	and, for	Avenu	[iwie_lists	-weight	INFC.		
of some sets that			here.	1010	· patera ·	inner4.			
			100.00	ann	· [stend ·	[hered	•		
				4	PREX	1986			6
		CONTRACTOR OF	0	*	2903	140			
	1 2	A sta	1	-	40	100			
B Regosterate of 11 - 15		3	L	4	PADR	100			
4-848			6	4	81	140			
S target			100		.41				10
Son Pure free free free free free free free			Giana					(i)e	un anun [] iften anit grant
ALCON AND ALCONDUCT.				ine Course			Les.	Capital Value	Manage.
		Multine lifety =							
	28 100 0100	enutural						delignation and in-	M gant X Conte
		Recept.						TAT INCOME VALUE	(1

Figure 24 Data Import Wizard Form of Step 4 of Testing Data

The last step was to transfer the Naïve Bayes model into the operator, serving as the research method. Drag and drop was employed to navigate to the processes view. Afterward, Modeling – Classification and Regression – Bayesian Modeling – Naïve Bayes were clicked. As demonstrated in Figure 3.20, the read CSV of Training Data was linked to the Naïve Bayes operator. The Naïve Bayes algorithm was applied to process the data.

ð	//Local Repository/processes/TA DATA MINING* - RapidMiner 5.3.015 @	SAPUTRI	- 5 *
Die Bit Docuse Ince San Red			
े 📦 🖬 🖬 🖉 🔊 🛝	# 🕨 🖩 🛐 🗟 🚯		
		2 • 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	Cone Cone
	A Proteiners II. 3 Log II.		Spanpels .
	10 Increases and Increase Free	Lacation Active	The root operator which is the outer most operator of ter Velvorgiprocess.

Figure 25 Read CSV File of Training Data Connected to the Naïve Bayes Operator

The subsequent step was to incorporate the Apply model operator into the process view, performed by selecting Application Model – Apply Model. Furthermore, the Naïve Bayes operator mod port was connected to the Apply model port, and the read CSV output port of testing data was to the UNL Apply model port.

A Designation of the second se	//Local Repairing/processes/1	A DATA MINING? - RepidMines 53.015	@ SAPUTIN		10. X
the DM Dougon lines New					
C 👒 🖬 🖓 🕬 👂					
D Oparation 14 52 m 10				Context	
e - 0 > 1	4 • 10 - 2 EPocas •	a • 🗆 🛊 🗟 🕸 •		•	
Abtude Weighting (21) Chattering and Tegmer	Then Discourse			Process	
🖶 🤤 Association and flam B-		3	(iconequi	int	
Consisten and Depen Sinitiato Computation		q-	ingline		14
Book Application (1) Book Application (1) Book Application (1) Book Application (1) Contraction (1) Contraction (1) Contraction (1) Book Application Contraction (1) Book Application Application Application		•	 4 holden egent garamiter 		
0 Scol			O weip 11 17 Cares	an IC	
10 @ Local Repository parameters			Process		1
	APastern a Glag II		27 march		_
0	Unecage	Fors Looden	The root operator who process.	this the autor most operator of every Activation Windows Go to PC antingging activent Week	

Figure 26 The Naïve Bayes Model and Read CSV of Testing Data Connected to the Apply Model

The Apply model served as a repository for the outcomes of data computations processed using the Naïve Bayes algorithm, subsequently confirmed by data testing. Therefore, the objective of predictive data testing was to calculate the training data results.

It was performed by selecting Evaluation - Performance Measuring - Performance as the next step to add a performance operator. This operator was applied to determine the amount of precision and inaccuracy associated with the computation results in the Apply model, as illustrated in Figure 3.26.

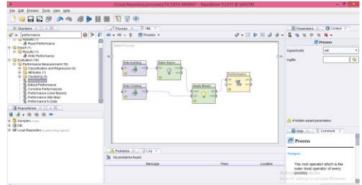


Figure 27 Operator Performance

The lab output of the Apply model was connected to the lab input for operator performance. After everything had been hooked, the per and exa ports on the operator performance were connected to the res port on the process view's right side.

N Reality	0 > 4		0.0			O Const 1
A Start Parkwas Const III A Start Parkwas A Start		4. 4			nganak taja di shtila ese di nge 1 fil Proces	Recent
		A Pedarte di Ling II			trager .	
		Relage	Plant	Location	outer root	water which is the specifier of every

Figure 28 The Operator Performance was Connected to the Res Port in the Process View

After everything was connected, the run button on the toolbar was selected. The calculation results appeared in a few moments.

h .								//Loc	al Rep	ository/p
<u>Eile</u>	Edit	Process	Tools	⊻iew	<u>H</u> elp	V				
	0		3 Ø	1	-		88	1	F	۲

Figure 29 Running the RapidMiner

				A TEXTING		rited - Pail	CARDING AND			Branches M.H. + D.
E Date Vi	re () ma:Da	Active () P	tel tiles: () /	Altaniai Ohats 🖸	Analatina 1				0.4 -	84-999 **
Dairwas	1(110X every)	to. Converted	density 1	(orlettile stuge				Valuetilias childer nimite al		a giargina yoon
ALC: NO	Herminsteinen		contamine)	paintering.		France	(weil, Dather			# # Coal Nexadan 1
28	180	0.000	0.442	180	L		PEB:		6	a a cramena com a com
67	80.	0.683	0.007	1960	Ł	0	PSB			
18.	44	0.000	0.449	180	h	A	P58			
10	4.5	0.084	0.441	191	h.		FELP			
	80.	0.535	0.404	180	16.	2	PS0			
142	80	0.886	3.68	1980	1.		PEOP			
	1945	0.060	4.812	80	10	4	FMDA			
	185	0.088	0.912	10	*	4	PMDK			
	yes	0.228	4712	84	. L.	A	RT			
	180	0.000	8.172	10	6	A	87			
		0.187	5.64)	44	6	A	FMDK			
		0.238	4.02	040	L.	A	RI			
	185	0.238	0.112	44	£.	A	#1			
	181	6.288	4.712	**	6	Α.	407			
	181	0.194	0.001	194	1	A	RT.			
80 - E	295	0.088	0.612	184	10	Α.	PMON			
	780	0.001	8.778		4	4	C017			
	284	0.068	8.402	.83	*	A.	FMOX			
0	181	11,059	18.901	184	27	8	FMDH		10	

Figure 30 Calculation Results

RapidMiner software calculation with the prediction method from testing and training data resulted in a prediction column (Re-registration). This column provided statistics on prospective new students who re-registered (Yes) and did not (No). The performance vector was selected to determine the accuracy of the Naïve Bayes algorithm.

			onew process*> - RapidMiner 5.3.0	HS & SARJTRI	- 0 ×
(in Lat Ducata	Iosis: yew yes				
े 👒 🔒 🖉	9 04 8	🕨 II 🖩 🛐 🖉 🕕			
Resalt Deniew	X Beargielist (DATE	TENTING IN THE PARTICIPANT	a (Pedomana)		(III Republices, M. St. et B. 1)
Table Plot Veel 🔾	Test lifew O Amelations				B d - B d - 0 0 0 -
Reion Selector	· Wulticlass Classification	n Performance () Aventations		6	a 👍 🕳 🗈 🔅 Samples runnt
10.420	@ Table View O Plot	8 (\$ C0 8 - 49 Local Repair/Op Luciana Ving Sepuri			
call XC (optimiztic)	accuracy 45.625				
C C pessenetico		the yes	true-ma	dises precision	
C Ballin alar	two: iwa	1		18.67%	
	pred no class receil	684 0.17%	679 98.54%	48 79%	
BAR H 32 41					Entern Harmiter, 34, 52, 40 Hb
# A					
4) 5, 2820 5/39/34 AB at its wrong results of	lating model application. WWWNING Sergectrotribute			doe for attribute 'Jewa, Kalamin'. This will probably to doe for attribute 'Provinsi'. This will probably to	

Figure 31 Accuracy

3.7 Naïve Bayes Algorithm

The following data were generated using the Naïve Bayes algorithm.

No.	Gender	Province	Registration	Re-registration
			Pathway	
1	М	В	PNUAN	Yes
2	Μ	С	PMDK	Yes
3	F	С	PMDK	Yes
4	Μ	D	CBT	Yes
5	F	В	PNUAN	Yes
	Μ	В	CBT	Yes
	Μ	А	PNUAN	No
	Μ	E	CBT	No
706	Μ	В	RT	No

Table 3 Training Data of Prospective Students of Information Technology Class of 2016

Table 4 Testing Data of Pros	pective Students of Information	Technology Class of 2017
Table 4 Testing Data of Tios	pective students of information	Technology Class of 2017

No.	Gender	Province	Registration Pathway	Re-registration
1	F	А	PMDK	Yes
2	F	А	PMDK	Yes
3	Μ	А	RT	Yes
4	Μ	А	RT	Yes
5	Μ	А	PMDK	Yes
	Μ	А	PNUAN	No
	М	А	CBT	No
	Μ	А	RT	No
1169	Μ	А	CBT	No

This study utilized 706 records from Table 3.3 as training data and 1,169 from Table 3.4 as testing data, with 50% Yes data included in testing data. The following formula was

applied to calculate the probability value or prediction of prospective new students who would re-register.

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

The first step was determining the number of Yes and No training data, generating the following results.

- Yes: 103
- Male: 81
- Female: 22
- Province B: 58
- Registration Pathway of PMDK: 17
- No: 603
- Male: 396
- Female: 207
- Province A: 155
- Registration Pathway of RT: 84
- Total: 706 records

After determining the number of prospective new students who re-registered and did not register, the probability of each attribute from testing data was calculated as follows.

- Gender = Male, Province = B, Registration Pathway = PMDK
- F (Re-registration = Yes / Number of Students of Training Data) = 103 / 706 = 0.146
- F (Gender = Male = Yes / Re-registration) = 81 / 103 = 0.787
- F (Province = B = Yes / Re-registration) = 58 / 103 = 0.563
- F (Registration Pathway = PMDK = Yes / Re-registration) = 17 / 103 = 0.165 Subsequently, the probability of No was calculated using the same testing data as the calculation of Yes.
- Gender = Male, Province = A, Registration Pathway = RT
- P (Re-registration = No / Number of Students of Training Data) = 603 / 706 = 0.854
- P (Gender = Male = No / Not Re-registered) = 396 / 603 = 0.657
- P (Province = A = No / Not Re-registered) = 155 / 603 = 0.257
- P (Registration Pathway = RT = No / Not Re-registered) = 84 / 603 = 0.140

4. Conclusion

The conclusion obtained from the research is that the Naïve Bayes Algorithm can be used to predict the enrollment of prospective new students at Yogyakarta Muhammadiyah University. The Naïve Bayes algorithm in predicting the registration of prospective new students has an accuracy of 49.62%, and the information obtained in this study is that prospective new students who are predicted to register at the time are students who are male, students from provinces outside Java, and students with a non-scholarship registration path.

References

- [1] Kusrini, & Emha T. Luthfi., 2009, Data Mining Algorithm, Andi Offset, Yogyakarta. (in Indonesian)
- [2] Larose D.T, Discovering Knowledge in Data. New Jersey : John Willey & Sons, Inc, 2005.
- [3] Prasetyo, E. (2012). Data Mining Concepts and Applications using Matlab. Yogyakarta: Andi. *(in Indonesian)*
- [4] HERI, S. (2017). Bayesiyan Classification Algorithm for Predicting New Student Registration at Stikes Hang Tuah Pekanbaru (Doctoral dissertation, Universitas Islam Negeri Sultan Syarif Kasim Riau). (*in Indonesian*)

- [5] Sugianti, D. (2012). Bayesian Classification Algorithm for Predicting New Student Heregration at STMIK Widya Pratama. Jurnal Ilmiah ICTech, 10(2), 1-5. (*in Indonesian*)
- [6] Febrian, F. (2011). Comparative Analysis of Data Mining Classification Algorithms on Life Reinsurance Facultative Data Acceptance. STMIK Eresha, Jakarta, MasterThesis. (*in Indonesian*)
- [7] Bramer, M. (2007). Principles of data mining (Vol. 180). London: Springer.
- [8] Bassil, Y. (2012). A simulation model for the waterfall software development life cycle. arXiv preprint arXiv:1205.6904.
- [9] Ruparelia, N. B. (2010). Software development lifecycle models. *ACM SIGSOFT Software Engineering Notes*, *35*(3), 8-13.
- [10] Makhtar, M., Nawang, H., & WAN SHAMSUDDIN, S. N. (2017). ANALYSIS ON STUDENTS PERFORMANCE USING NAÏVE BAYES CLASSIFIER. Journal of Theoretical & Applied Information Technology, 95(16).
- [11] Maitra, S., Madan, S., Kandwal, R., & Mahajan, P. (2018). Mining authentic student feedback for faculty using Naïve Bayes classifier. Procedia computer science, 132, 1171-1183.