# Data Mining Classification to Predict Student Graduation Using the Naive Bayes Method

#### Jefri<sup>a1</sup>, Zaehol Fatah<sup>a2</sup>

<sup>a</sup> Information Systems Study Program, Faculty of Science and Technology, Ibrahimy University, East Java, Indonesia - 68374 <u><sup>1</sup>renorisma903@gmail.com</u>, <sup>2</sup>zaeholfatah@gmail.com

#### Abstrak

Data mining membantu memberikan keputusan yang tepat dan cermat. Kelulusan mahasiswa tepat waktu merupakan salah satu point penilaian dalam proses akreditasi perguruan tinggi. Namun kelulusan mahasiswa tidak selalu dapat dideteksi secara cepat sehingga dapat mengurangi penilaian suatu perguruan tinggi dalam proses akreditasi. Permasalah inilah yang muncul untuk mengetahui mahasiswa nantinya bisa lulus tepat waktu atau tidak. metode klasifikasi untuk prediksi lulusan mahasiswa menggunakan algoritma Naïve Bayes. kelulusan mahasiswa tepat waktu atau tidak, yang diharapkan hasilnya dapat memberikan informasi dan masukan bagi pihak perguruan tinggi dalam membuat kebijakan kedepannya. Dari hasil pengujian ini didapatkan hasil bahwa dengan menerapkan algoritma Naïve Bayes sistem dapat memprediksi kelulusan mahasiswa dengan tepat waktu, Setelah membandingkan beberapa literatur, dapat disimpulkan bahwa metode ini dapat digunakan untuk prediksi tersebut dengan tingkat keakuratan 90%. Kajian literatur ini penting sebagai faktor pendukung bagi penelitian.

Kata kunci: Data Mining, Naive Bayes, Prediction, Student Graduation

#### Abstract

Data mining helps provide precise and careful decisions. Student graduation on time is one of the assessment points in the higher education accreditation process. However, student graduation cannot always be detected quickly, which can reduce the assessment of a university in the accreditation process. This problem arises to find out whether students will be able to graduate on time or not. Classification method for predicting student graduates using the Naïve Bayes algorithm. Whether a student graduates on time or not, it is hoped that the results will provide information and input for the university in making future policies. From the results of this test, it was found that by applying the Naïve Bayes algorithm the system can predict student graduation in a timely manner. After comparing several literatures, it can be concluded that this method can be used for this prediction with an accuracy rate of 90%. This literature review is important as a supporting factor for research.

Keywords : Data Mining, Naive Bayes, Prediction, Student Graduation

#### 1. Introduction

Higher education is an educational institution that provides high-level learning services, which is an optional final stage in formal education. These higher education institutions are generally in the form of universities, academies, institutes or high schools. The types of higher education include vocational, academic and professional. Based on the level, higher education provides diploma, bachelor's, master's, specialist and doctoral programs[1].

Universities are expected to provide quality education for students so as to produce students who have competence in their fields. In Indonesia, the quality of a higher education institution is determined by the accreditation grade issued by the National Accreditation Board for Higher Education (BAN-PT). There are many aspects that can be used as a benchmark for the quality of higher education. One of them can be seen from the number of students who can complete their studies within the specified time or in other words, students graduate on time.

The more students who graduate on time, the better the performance of the higher education institution, so that the student graduation rate on time becomes one of the assessment criteria. accreditation for a university.

Many factors can support an increase in the number of students graduating on time, including:

1. Effective Academic Management

2. Relevant Curriculum

3. Academic Guidance Support

By implementing effective academic management, a relevant curriculum, and strong academic guidance support, universities can significantly increase the number of students graduating on time. This not only benefits students, but also increases the reputation and credibility of the educational institution[2]. Graduation data can be analyzed and processed in detail using data mining techniques.

Data mining is a process of extracting valuable knowledge or information from large and complex datasets. The main goal of data mining is to identify patterns, relationships, or information that may not be directly visible in the data, so that it can provide deeper and more valid information[3]. Data mining in another sense is a term used to describe the discovery of knowledge in databases, which uses statistical, mathematical, artificial intelligence and machine learning techniques to extract and identify useful information and knowledge assembled from various large databases[4].

Student graduation rates are predicted using a system. However, some universities do not yet have a system to predict delays in student graduation, so universities cannot prevent this.

How can the Naïve Bayes algorithm be applied to predict student graduation on time? The Nave Bayes algorithm is one of the most effective and efficient inductive learning algorithms for machine learning and data mining. Nave Bayes performance is competitive in the classification process even though it uses the assumption of attribute independence (no relationship between attributes). Naive Bayes is used for data classification techniques using probability and statistical methods that predict future opportunities based on previous experience, so it is known as Bayes' Theorem. This theorem is combined with Naive where it is assumed that the conditions between one attribute and another attribute are mutually independent. Naive Bayes classification assumes that the presence or absence of certain characteristics of one class has nothing to do with the characteristics of other classes[5].

From several explanations regarding graduate predictions, there are several differences, including using the Naive Bayes algorithm with a classification method using the attributes of student name, NPM, entry route, student GPA and gender. With these attributes and criteria data, the algorithm can predict student graduation rates. So, with this student graduation prediction system, it can help to find out whether students graduate on time or not, with the hope that the results can provide information and input for universities in making future policies.

Based on the existing problems, it is necessary to have a system to predict student graduation rates based on existing variables. Apart from that, a suitable algorithm is needed so that it can produce good accuracy values. With the system created, it is hoped that universities can make policies so that students can graduate on time.

# 2. Research Method / Proposed Method

This research began with problem formulation and literature study. A certain amount of data is required obtained through observation and documentation methods to be able to solve the problem. Next, the required student data was collected, with the acquisition of 146 data sets of students from the class of 2018 who had completed their studies in 15 attributes. Then in the second stage, data pre-processing is carried out to obtain good data before the data is processed using Naïve Bayes. If data pre-processing has been completed, 146 student data in 15 attributes will be used to carry out the mining process. Third, the mining process by applying the naïve Bayes algorithm. The fourth stage is experimentation and model testing. The fifth stage is evaluation and validation of test results. the results of the accuracy values obtained, test the accuracy method in the concept of data mining. Accuracy calculations can be assisted by applying the Naive Bayes calculation formula. The results of applying the Naive Bayes

formula create very good accuracy based on the calculations of applying the Naive Bayes model [6].

# 3. Literature Study

#### 3.1 Data collection

The techniques used to collect research data consist of:

1. Observation method

Efforts to obtain or collect data directly by observing in the field, which will then be useful in calculations when predicting the performance of systems running in the company[7]. 2. Library method

Collect data by studying books, scientific journals, and information from the internet that supports research[8].

#### 3.2 Data analysis

Student data in this research is data from Kaggle in the form of gender, student status, married status, age, Semester Achievement Index (IPS) from semester 1 to semester 8, and Cumulative Achievement Index (GPA). With the target class classification, namely graduation status which includes whether it is on time or late[9].

#### 3.3 Naive Bayes algorithm

Naive Bayes is the simplest form of Bayesian network classifier. In naive Bayes, each feature node has a class node as its parent, but does not have the parents of any other feature nodes. However, a large body of work in supervised learning has shown that such a simple naive Bayes classifier can compete with state-of-the-art classifiers such as C4.5 and is still one of the top 10 data docking algorithms in the world[10].

#### 3.4 Rapidminer

Rapidminer is a software for data science created by Rapidminer, Inc which provides integrated data, text mining, machine learning, and predictive analysis. As well as assisting with steps in a machine learning process such as visualization results, optimization, data preparation and model validation. Rapidminer brings artificial intelligence to the enterprise through an open and extensible data science platform. Built for analytics teams, Rapidminer brings together the entire data science cycle from data preparation to machine learning to predictive deployment (Hofmann & Klinkenberg, 2016)[11].

In simple terms, Rapidminer is an application that is used to process data using various techniques and methods in data mining, so that it can become useful information. Rapidminer is an open-source software that is useful for processing mining data. Rapidminer processes it by extracting patterns from datasets and combining them with statistical methods, artificial intelligence and databases. The aim of using Rapidminer is to obtain high quality information from the processed text. Rapidminer is used in various studies for examining researched data[12].

#### 4. Result and Discussion

This section contains the results and discussion of the research topic, which can be created first as a research methodology. This section also represents explanations in the form of explanations, pictures, tables and others. The number of words in this section ranges. Data collection

The data collection process was carried out by taking it from Kaggle. The data used is in the form of a Microsoft Excel file with .xlsx format. The data comes from the 2018 incoming class of students who have completed their studies, with a total of 146 data and 14 attributes. An example of the data used can be seen in the following image. This data was taken with the aim of further analysis in research projects, specifically for the purposes of Naive Bayes classification using RapidMiner. An example of the data used is in the picture.

NAMA	JENIS KELAMIN	STATUS MAHASISWA	UMUR	STATUS NIKAH	IPS 1	IPS 2	IPS 3	IPS 4	IPS	5 IF	\$6	IPS 7	IPS 8	IPK	STATUS KELULUSAN
UNAMA	LAKI - LAKI	MAHASISWA	24	BELUM MENIKAH	3.17	2.	.7 3.	23	2.41	3	2.47	1.75		0	2.75 TEPAT
EYLA TRIYANA PRATIWI	PEREMPUAN	MAHASISWA	26	BELUM MENIKAH	3.6	3.	5 3.	42	2.85	3.31	2.95	2.18			3.39 TEPAT
ERIS SOFIYAN PRAYOGA	LAKI - LAKI	MAHASISWA	29	BELUM MENIKAH	2.67	2.6	i6 2.	93	3.14	2.92	2.64	2.88		0.5	2.81 TEPAT
ADITYA AKBAR NUGRAHA	LAKI - LAKI	MAHASISWA	27	BELUM MENIKAH	2.48	2.8	6 2.	09	2.55	2.55	2.43	2.55	2.	.17	2.82 TEPAT
ERNA EKA RIYANTI	PEREMPUAN	MAHASISWA	25	BELUM MENIKAH	3.19	3.0	18 3.	31	2.83	3.36	2.73	3.06		0	3.09 TEPAT
FARID DWI NORYANTO	LAKI - LAKI	MAHASISWA	24	BELUM MENIKAH	3.1	2.9	8 3.	17	3.25	3.41	3.08	3.43		3	3.23 TEPAT
DAFIK HADI WINOTO	LAKI - LAKI	MAHASISWA	24	BELUM MENIKAH	2.98	2.6	8 2.	23	2.86	2.25	2.64	1.52		2.1	2.54 TEPAT
WAHYU FITRIYANTO	LAKI - LAKI	MAHASISWA	24	BELUM MENIKAH	3.45	3.1	5 3.	54	3.78	3.42	3.88	2.5		4	3.56 TEPAT
MAM SURYO SUSANTO	LAKI - LAKI	MAHASISWA	26	BELUM MENIKAH	3.31	3.0	12 3.	48	3.7	3.19	3.21	2.58		4	3.4 TEPAT
AHMAD SUTOPO	LAKI - LAKI	MAHASISWA	24	BELUM MENIKAH	2.62	2.7	3 2.	11	3.36	2.79	3.17	2.45		0	2.97 TEPAT
IERI SUFA'AT	LAKI - LAKI	MAHASISWA	24	BELUM MENIKAH	3.24	3.0	6 2.	79	2.73	3.02	2.59	3.08		0	3.08 TEPAT
MURYA AMIEN NUR PRABOWO	LAKI - LAKI	MAHASISWA	26	BELUM MENIKAH	2.83	3.0	12 3.	02	3.63	3.21	3.09	2.57		3	3.33 TEPAT
EDY PURNOMO	LAKI - LAKI	MAHASISWA	24	BELUM MENIKAH	2.83		3 2	54	3.36	2.79	3.28	3.28		3	3.15 TEPAT
DIDIK KURNIAWAN	LAKI - LAKI	MAHASISWA	24	BELUM MENIKAH	3	3.0	18 3.	02	3.07	2.88	2.76	3		2.5	2.92 TEPAT
YDIA EVITA SANDRA DEWI	PEREMPUAN	MAHASISWA	25	BELUM MENIKAH	3.52	3.5	2 3.	31	3.33	3.3	2.94	3.46		0	3.26 TEPAT
DAVID KURNIAWAN	LAKI - LAKI	MAHASISWA	36	BELUM MENIKAH	2.25	0.	.5 0.	16	0.32	1.13	0.37	0.28	0.	.11	1.02 TEPAT
DEVI KISTIANI	PEREMPUAN	MAHASISWA	25	BELUM MENIKAH	3.52	3.4	6 3.	69	3.41	3.4	0	0			3.43 TEPAT
ADI PURWADI	LAKI - LAKI	MAHASISWA	27	BELUM MENIKAH	2.6	2.9	1 2	66	2.91	3.25	2.88	2.83		0	2.86 TEPAT
ARIF WIBOWO	LAKI - LAKI	MAHASISWA	30	BELUM MENIKAH	2.52	2.5	9 2.	11	3	2.73	2.5	2.48		0	2.68 TEPAT
NOVAZAEN DWI SAPUTRO	LAKI - LAKI	MAHASISWA	28	BELUM MENIKAH	2.75	3.0	15 3.	21	2.92	3.41	2.86	2.5		0	3.04 TEPAT
MOHAMAD ADE FADHORI	LAKI - LAKI	MAHASISWA	30	BELUM MENIKAH	2.93	2.5	2 1.	95	2.37	3.07	2.91	1.73		1.6	2.81 TEPAT
SUBAEDAH	PEREMPUAN	MAHASISWA	23	BELUM MENIKAH	3.52	3.4	8 3.	71	3.83	3.25	3.65	2.33		3.5	3.56 TEPAT
DANI PRIMAYANTI	PEREMPUAN	MAHASISWA	24	BELUM MENIKAH	3.48	3.7	3	3.6	3.52	2.9	3.32	3.21		0	3.45 TEPAT
DERA BAHTIAR WIDIYANTO	LAKI - LAKI	MAHASISWA	23	BELUM MENIKAH	2.52	3.	.3 3.	79	3.71	3.65	2.1	3.59		4	3.52 TEPAT
MUKHAMAD YAHYA WICAKSON	LAKI - LAKI	MAHASISWA	24	BELUM MENIKAH	2.45	2.6	6 2	93	3.08	3.29	3.18	2.1		3	3.1 TEPAT
WAHYU DWI UTOMO	LAKI - LAKI	MAHASISWA	28	BELUM MENIKAH	2.62	1.7	5 1.	95	2.17	1.95	2.18	1.45	1.	.08	2.34 TEPAT
AHMAD FIRDAUS MABRURI	LAKI - LAKI	MAHASISWA	25	BELUM MENIKAH	2.33		3 2.	19	0.44	2.17	0.59	0.26			1.72 TEPAT
HARYOKO ABDUL HAMID	LAKI - LAKI	MAHASISWA	26	BELUM MENIKAH	3.24	3.4	6 3.	56	3.83	3.73	3.74	2.33		3.5	3.58 TEPAT
RSA GILANG ARDHANI	LAKI - LAKI	MAHASISWA	25	BELUM MENIKAH	2.95	2.9	1 3.	44	3.38	3.33	3.36	2.08		0	3.1 TEPAT
NEORITA NUR FATIMAH	PEREMPUAN	MAHASISWA	25	BELUM MENIKAH	3.48	3.	5 2.	77	3.59	3.5	3.35	3.27		0	3.31 TEPAT
JLIL ALBAB	LAKI - LAKI	MAHASISWA	23	BELUM MENIKAH	2.93	3.1	.8 3.	31	3.25	3.44	3.22	3.07		3.5	3.33 TEPAT
DEWI KHOIRUN NISA'	PEREMPUAN	MAHASISWA	27	BELUM MENIKAH	3	3.3	3 3.	27	2.96	3.16	3.12	2.61		0	3.08 TEPAT
AGUS NUGROHO	LAKI - LAKI	MAHASISWA	23	BELUM MENIKAH	2.83	3.1	1	3.1	2.96	3.2	3.53	2.48		3	3.27 TEPAT
MUKHAMAD LUQMAN HABIBI	LAKI - LAKI	MAHASISWA	25	BELUM MENIKAH	2.88	3.2	5 3.	42	3.29	3.33	2.75	3		0	3.13 TEPAT
AMBAR SETIYANI	PEREMPUAN	MAHASISWA	25	BELUM MENIKAH	3.71	3.7	9 3.		3.91	3.75	3.94	2.58		4	3.85 TEPAT
AVINA SEPTIANA	PEREMPUAN	MAHASISWA	27	BELUM MENIKAH	3.07			3.2	3	3.21	2.47	2.55	2	.75	3.1 TEPAT
MARIA ULFA	PEREMPUAN	MAHASISWA	24	BELUM MENIKAH	3.19				2.77	3.09	2.83	3.11		2.9	3.08 TEPAT
VANDHY SETYARACHMAN	LAKI - LAKI	MAHASISWA	31	BELUM MENIKAH	2.86				3.27	2.98	2.85	2.12			3.21 TEPAT
CHASAN MUBAROK	LAKI - LAKI	MAHASISWA	25	BELUM MENIKAH	2.98				3.42	3.5	3.47	2.58		3.5	3.52 TEPAT

Figure 2. 1 Sample Data

RapidMiner Studio 10.3 provides an efficient and effective environment for performing data testing and predictive analysis. The structured testing process and advanced features offered ensure that users can obtain accurate and reliable results. Here is the data testing process and display of RapidMiner Studio 10.3



Figure 2. 2 Display of Rapidminer Studio 10.3

Open the Rapidminer application, after it appears as shown, the start with display will appear then select blank process to open a new worksheet/open a file that has been saved as in Figure 2.3

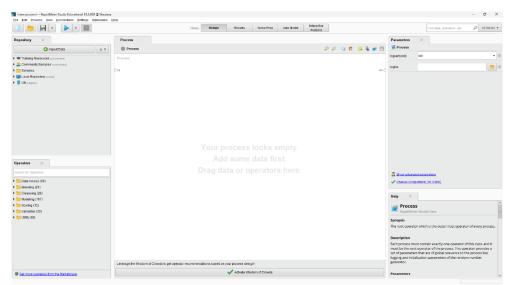


Figure 2. 3 Worksheet views

Next is the data input process and requires the read excel operator by dragging and dropping into the worksheet like the data retrieval process using double click on the read excel operator

Fie Edit Process New Connections Settings Estension	нар											
		Views:	Design R	tesults Turbo Pre	Auto Model	Interactive Analysis				Find data, operatorselc	🔎 Al Studi	• •
Repository ×	Process								Parameters	×		
🕒 Import Data 🗉 💌	Process						o 🔎 🐚 🖪	📮 ≩ 💣 🗄	📩 Read Excel			
Training Resources (connected)	Process									Impert Configuration Waard.		Ð
<ul> <li>Sommunity Samples (connected)</li> </ul>									excel file			Ð
Samples	Dre								excerne			<u> </u>
Local Repository (Local)     DB (Locat)									sheetnumber	1		Φ
P De Lugacy									date format	Entervalue	• 8	0
		Read Excel										
Operators ×												
r 3												
T Data Access (59)	~								Show advance	albiin (10.3.000)		
* 5 Files (19)									V Shine Carl	<u>annin (10.3.100)</u>		
* 🦰 Read (13)												
📩 Read COV									Help ×			
📩 Read Excel									Read E	ĸcel		^
📩 Read URL									Read E RapidMine	r Studio Core		10
📩 Read SPSS										t Read Data Eles, Xis, Xisk Micros	ofs Spreadsheets	
📩 Read Stata									Deterets			
📩 Read ARFF									Synopsis			
📩 Read XRFF										ids an ExampleSet from the speci	fied Excel file.	
📩 Read DBase									Jump to Tutorial	Process		
📩 Read C4.5												
📩 Read ElbTeX									Description			
📩 Read DasyLab	V I manage the illighter of C	Drowcis to get operator recommenda	loor barad an usur pro	unana designal						be used to load data from Micro		
We found "Recommender Extension", "RapidMiner Finance and Econom" and 62 more results in the Marketplace. Show	Cereage the Wiscom or C	anneas to get operator recommendat								is operator is able to read data fr 3. The user has to define which o		
and Econom" and 62 more results in the Marketplace. Show met			× *	chuste Wisdom of Crowd					In the workbook	should be used as data table. The	table must have	a 👃

Figure 2. 4 Excel Read Operators

Click on the Read Excel module to open the file selection window. In this window, you will be given the option to search for and select the data file you want to import. Use the 'Browse' or 'Select File' button to open the file selection dialog. Navigate to the directory where the Excel file you want to use is stored, select the data file that suits your analysis needs. Make sure that the selected file has the correct format and that the required data is complete.

pository ×					Analysis			
	Proces	•					Parameters	×
🕒 Import Data 📃 👻	Pro	cess			g q	🐚 🖪 🔒 🖬 🕄	📩 Read Excel	
Training Resources scorrected		Import Data - Select the dat	a location.			×		Import Configuration Wizard
Community Samples (connected)							excel file	
Samples	Dre		Select the data locat	ion.			ayrei Ing	
Local Repository (Local)							sheet number	1
DB (Legery)		Downloads			v	+ = + ☆ ≝ =+		
		Bookmarks	File Name	Size		Last Modified	date format	Entervalue
		+ Last Directory	- geocia	JAC	Type File Folger	JUN 29, 2023		
			gradient-elegant-certificate (4)		File Folder File Folder	Dec 28, 2022 Aug 15, 2024		
			gradient-golden-kusury-background gradient-golden-kusury-horizontal-business-cand-template (1)		File Folder	Aug 23, 2024		
			gradient-golden www.insteam.com/englane(1) gradient-orcanizational-chart-infographic-with-photo		File Folder	Aug 15, 2024		
			aretaros		File Folder	Oct 5, 2024		
			hidayahulah		File Folder	Sep 25, 2024		
			hidayatallah (1)		File Folder	Oct 4, 2024		
			LAPORAN		File Folder	Jan 24, 2024		
			Inear-fat-organizational-charl-with-photo		File Folder	Sep 15, 2023		
			🚞 marela		File Folder	Oct 10, 2024		
			moon_get		File Folder	Sep 25, 2024		
			Pid Unio 2023 putra Pid Unio 2023 putra_files		File Folder	Dec 29, 2023		
erators 🖂			real-astate-house-property-facebook-cover-barrner-template		File Folder	Jan 3, 2023		
			realistic-trames-template		File Folder	May 30, 2024		
×			simple-blue gradient background vector-business		File Folder File Folder	Aug 14, 2024 Oct 10, 2024	3 Show advance	ed parameters
Data Access (59)			BougStore-Template		File Folder	Jun 15, 2024	Change comp	
Files (19)			urzigper-master		File Folder	Jun 15, 2024	V Children Come	0000000000
* Read (13)			vMix Pro 24.0.0.72 x84		File Folder	Nov 28, 2022		
			while-banner-with-triangle-halftone-shapes		File Folder	Sep 10, 2024	Help ×	
a Read COV			06 -DAFTAR-DOSEN-PEMBINGBING-NO-TP xisx	15 KB	Microsoft Excel Worksheet	May 31, 2023		
📥 Read Excel			😣 Kelulusan Testxis	51 KB	Microsoft Excel 97-2003 Wo	Oct 19, 2024	🔸 Read E	kcel
📩 Read URL						~	RapidMine	r Studio Core
📩 Read SPSS		Kelulusan Testxis						t Read Data Elles, Nis, Nisk Microsoft Spreadsheets
📩 Read Stata		Excel (also, als)					Detects	
📩 Read ARFF		c.man (.max, .10)					Synopsis	
Read XNFF						→ Next X Cancel	This operator rea	ids an ExampleSet from the specified Excel file.
📩 Road DBase							Jump to Tutorial	Process
all Read C45								
📩 Read BibTeX							Description	
t Read DasyLab								be used to load data from Microsoft Excel
We found "Recommender Extension", "RapidMiner Finance	Leverage t	the Wisdom of Crowds to get	toperator recommendations based on your process design!					is operator is able to read data from Excel 95, 97,
and Econom" and 62 more results in the Marketplace. Show			Activate Western of Crowds					3. The user has to define which of the spreadshee

Figure 2. 5 Student data import

After selecting the file, the next step is to verify that the file has been properly loaded into RapidMiner. The Read Excel module will preview the data of the selected file, check the column headers, data format and ensure there are no errors in the imported data. If necessary, make additional settings such as setting the delimiter, sheet to be read, and cell range.

2 😑 🔚 🔹 🕨 🔳						Views	: Desi	on R	esults	Turbo Pre	p Auto	Model	Interactive Analysis					Find data, operat	orsetc 🔑 All	50.d
spository ×		Process	•														Parameters	×		
🕒 Import Data	= *	Pro	CRSS										ρ,	ρ 🐚 🖪	<b>4</b> 4		📩 Read Excel			
Training Rescurbes (connected)	F	Process	Import	Data - Selec	t the cells to i	mport.										×		Impert Configuration	on Weard	
Community Samples (connected)																	excel file			
Samples	Dr							Sele	ct the	cells to in	nport.						ayres ino			
Local Repository (Local)			1.1					_									sheet number	1		
DB (Legery)			Sheet	Sheet1	Cell	ange: A.O		Sele	d Al	✓ Define	reader row.	10					date format	Entervalue	*	
				A	в	с	D	1		6	н	1	J	к	L					
			1	NAMA	JENS K	STATUS	UMUR	STATUS	IPS1	IPS 2	IPS 3	IPS 4	IPS 5	IPS 6	IPS7	IP.A				
			2	UNADA	LAKI-LA.	MAHASI.	24	BELUM	3.17	27	3.23	2.41	3	2.47	1.75	0				
			3	LEYLAT	PEREMP	MAHASL.	28	BELUM	3.6	3.5	3.42	2.85	3.31	2.95	2.18					
			4	VERIS S	LAKI+LA	MAHASL.	29	BELUM	2.67	2.66	2.93	3.14	2.92	2.64	2.00	9.0				
			5	ADITYAA	LAKI-LA.	M4HA3L.	27	BELUM	2.48	2.85	2.09	2.55	2.55	2.43	2.55	2.				
			6	ERNAE	PEREMP	MAHASL.	25	BELUM	3.19	3.08	3.31	2.83	3.36	2.73	3.06	0				
			7	FARID D	LAKI - LA	MAHASI	24	BELUM	3.1	2.98	3.17	3.25	3.41	3.08	3.43	3				
			8	DAFIK H	LAKI - LA	MAHASI	24	BELUM	2.98	2.68	2.23	2.85	2.25	2.64	1.52	2.				
			9	WARMO	LAKI - LA	MAHASL.	24	BELUM	3.45	3.15	3.54	3.78	3.42	3.88	2.5	4				
rators ×			10	IMAM SU	LAKI + LA		28	BELUM	3.31	3.02	3.48	3.7	3.19	3.21	2.50	4				
	×		11	AHMAD	LAKI-LA	MAHASL.	24	BELUM	2.62	2.73	2.11	3.36	2.79	3.17	2.45	0				
			12	HERISU	LAKI - LA.,		24	BELUM	3.24	3.08	2.79	2.73	3.02	2.59	3.08	0	Show advance	ed parameters		
Data Access (59)	^		13	MURYA	LAKI - LA		28	BELUM	2.83	3.02	3.02	3.63	3.21	3.09	2.57	3	Change comp	sability (10.3.000)		
Files (19)			14	EDY PU		M4HASL.	24	BELUM	2.83	3	2.54	3.35	2.79	3.28	3.28	3				
Read (13) Read CSV			15	DIDIKK	LAKI - LA		24	BELUM	3	3.08	3.02	3.07	2.88	2.76	3	2!	Help ×			
A Read Excel			16	LYDIA E	PEREMP		25	BELUM	3.52	3.52	3.31	3.33	3.3	2.94	3.48	0				
A Read URL			17	DAVID K DEVLKIS	LAKI-LA	MAHASI.	35	BELUM	2.29	0.5	0.15	0.32	1.13	0.37	0.28	0.1	🛓 Read E	xcel In Studio Core		
A Read SPSS			18	ADI PUR.	PEREMP	MAHASL.	25	BELUM	3.52	3.45	2.68	3.41 2.91	3.4	0	2.83	-			lsx Microsoft Spreadshe	
Aread Stata			20	ARIE VI	LAKI-LA.	MAHASL.	30	BELUM	2.52	2.91	2.00	2.01	2.73	2.50	2.03	0	Deterets	a case was case or o		
Read ARFF			20	<	CAR + DA.	MATAOL.	30	BELOW	1	2.09	211	3	215	2.0	2.90	>	Synopsis			
📥 Read XRFF													- Breaker	s → N	est 🗙	Cancel	This operator rea	ads an ExampleSet from	the specified Excel file.	i.
📩 Read DBase					_	_	_	_	_	_	_	_					Jump to Tutorial	Process		
📩 Read C4.5																				
📩 Read BibTeX																	Description			
📩 Read DasyLab	~	everage t	he Wise	inm of Cone	tis to get one	stor recomme	edations ha	sed on your pro	ness desi	cm)								n be used to load data f		
Vo Isund "Recommender Extension", "RapidMiner Fina and Econom" and 62 more results in the Marketplace. ted.	ance					recording	and the Dis			stom of Crawd							2000, XP, and 200	03. The user has to defin	ad data from Excel 95, 9 we which of the spreads table. The table must h	she

Figure 2. 6 file verification

Next, click next on the selected data. After that, determine the data type, then the data class is given a label attribute with the change role option and click finish, then it will look like Figures 7 and 8.

1 🔤 🔣 - 🕨 🕨 🔳		there. Design Provide Table 9	na Adv. Boder mail.yok	Salaan weeden oo 🖉 Altica
CHEV >	FROM .			fatoria 1
O input D in	E etxee		<mark>ه ۳</mark> ۲۸ ۸	
( NAMES OF A DESCRIPTION OF A DESCRIPTIO	140000 C			O the parameters in directly
Contracting Company (Including) Sumption	De			214
inst Paper (1937-2015) 28 magnetic				
	ŕ			
ien x				
Notice (2)				A conservation provides
Kasalan (2) J. Hata Bares				1.46
Have being the state				P Notice Dayes: Periodice Technics
				Fair Separate Linearization, Model, Lincology Lowerse, Sciences of Landson's Separate Reportation Section
				typepala.
				The Operation generative Science Constitution and a second second second
				(0, 10, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0
				Second patients
	Long and in Marco of Country is and one sta-	the sum there is any server a constrained		<ul> <li>Solve Regist for Egit Maximum characterizations and the articular grant market even with a small data web, the simple to use and</li> </ul>
A state "North Cold", Neuralize Easter Free State Procession and in Section Sciences, "Cold and		A 10.00 (10.00 (10.00))		An approximately provident of synchronic methods and a single structure of the single structure

Figure 2. 7 Data import

	-	•					Views: Design	Results	Turbo	Prep Auto Mod	el Intera Anal	active lysis				Find data, operatorsetc	🔎 All Studio 🔹			
sult History	ExampleSet (Read Excel) × 🔰 SimpleOutlobution (Naive Bayes) ×														Repository ×					
	Open in	Turbo Prep	👫 Auto Model	🚠 Interactive	Analysis	🔁 Import Data	= •													
Data	Row No.	STATUS KEL	NAMA J	JENIS KELA	STATUS MA	UMUR	STATUS NIK	IPS 1	IPS 2	IPS 3	IPS 4	IPS 5	IP\$6	IPS 7		Training Resources (connected)     A Community Samples (connected)				
	39	TEPAT	CHASAN MU	LAKI - LAKI	MAHASISWA	25	BELUM MENL.	2,980	3.180	3.730	3.420	3.500	3.470	2.580	^	Samples				
Σ	112	TEPAT	BAMBANG ZA	LAKI - LAKI	MAHASISWA	28	BELUM MENL.	3.240	3.270	3.500	3.220	2.770	2.940	2.250		Local Repository (Local)     BB (Legacy)				
Italistics	49	TEPAT	AWIT RIKA JA	LAKI - LAKI	MAHASISWA	23	BELUM MENL.	2.740	2.860	2.610	2.950	2.860	2.680	3.160		P DB (Legec))				
	36	TEPAT	AVINA SEPTI	PEREMPUAN	MAHASISWA	27	BELUM MENL.	3.070	2.850	3.200	3	3.210	2.470	2.550						
	85	TEPAT	AVIF ROHAN	LAKI - LAKI	MAHASISWA	40	BELUM MENL.	2.830	2.640	2.890	3.140	2.850	3.420	3.030						
ualizations	81	TEPAT	AUFAL WIDAD	LAKI - LAKI	MAHASISWA	23	BELUM MENL	2.500	2.080	2.250	2.270	2.210	2.160	2.860						
	136	TEPAT	ASROFAH	PEREMPUAN	MAHASISWA	23	BELUM MENL.	2.640	2.590	3.090	3.140	2.540	3.310	1.630						
	141	TEPAT	ARY JULI SE	LAKI - LAKI	MAHASISWA	23	BELUM MENI.	1.980	2.500	2.140	2.770	2.610	2.930	2.820						
notations	93	TEPAT	ARIS HIDAYAT	LARI-LARI	MAHASISWA	24	BELUM MENL.	2.710	3.050	3.040	3.140	2.930	3	3.630						
instations	19	TEPAT	ARIF WIBOWO	LAKI - LAKI	MAHASISWA	30	BELUM MENL.	2.520	2.590	2.110	3	2.730	2.500	2.480						
	83	TEPAT	ARGI CAHYA	PEREMPUAN	MAHASISWA	24	BELUM MENL.	2.170	2.660	2.850	2.860	2.640	2.830	1.570						
	135	TEPAT	ANIS WATUL	PEREMPUAN	MAHASISWA	24	BELUM MENU.	2.880	2.770	2.610	3.230	2.710	3.450	3.110						
	126	TEPAT	ANIK FARYA	PEREMPUAN	MAHASISWA	25	BELUM MENL.	2.900	3.070	2.440	2.410	2.400	2.750	2.080						
	75	TEPAT	ANIF FARIZI	LAR - LAR	MAHASISWA	25	BELUM MENU.	2.740	2.640	2.390	2.820	2.700	3.060	2.700						
	60	TEPAT	ANDI PRASE	LAKI - LAKI	MAHASISWA	23	BELUM MENL.	2.710	2.410	2.800	2.950	3.180	3.100	2.500						
	132	TERLAMBAT	AMIR ALFIAN	LAKI - LAKI	BEKERJA	23	BELUM MENU.	2.740	2.680	2.350	2.550	2.110	2.500	1.350						
	35	TEPAT	AMBAR SETL.	PEREMPUAN	MAHASISWA	25	BELUM MENL.	3.710	3.790	3.950	3.910	3.750	3.940	2.580						
	76	TEPAT	ALI NURSID	LAKI - LAKI	MAHASISWA	22	BELUM MENI.	2.600	2.200	2.270	2.550	2.450	2.390	2.550						
	122	TEPAT	ALI AKBAR D	LAKI - LAKI	MAHASISWA	28	BELUM MENL.	2.980	2.660	2.700	2.730	2.800	3.330	2.430						
	139	TEPAT	AJI PRASETYO	LAKI - LAKI	MAHASISWA	23	BELUM MENL.	3.050	3.100	3.330	3.170	3.170	3.260	2.330						
	10	TEPAT	AHMAD SUT	LAKI - LAKI	MAHASISWA	24	BELUM MENL.	2.620	2.730	2.110	3.360	2.790	3.170	2.450						
	27	TEPAT	AHMAD FIRD	LARI-LARI	MAHASISWA	25	BELUM MENU.	2.330	3	2.190	0.440	2.170	0.590	0.260						
	88	TEPAT	AHMAD FAIZIN	LAKI - LAKI	MAHASISWA	23	BELUM MENL.		2.890	3.250	2.830	2.860	2.970	1.310						
	73	TEPAT	AGUSTINUS	LAKI - LAKI	MAHASISWA	23	BELUM MENU.		2.200	2.160	1.800	2.110	1	1.190						
	129	TEPAT	AGUS SUPRI	LAKI - LAKI	MAHASISWA	24	BELUM MENL.		3.230	3.420	3.430	3.250	3.090	2.420						
	134	TEPAT	AGUS SETRI	LAKI - LAKI	MAHASISWA	24	BELUM MENL.	3.070	3.210	2.940	3.360	3.150	2.770	3.470	~					

Figure 2. 8 Student Data Test Prediction Results

The number of testing data used was 146 students using the Naïve Bayes method. The prediction results were that 144 students graduated on time.

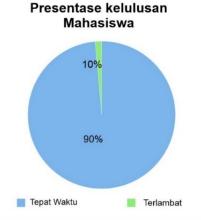


Figure 2. 9 Graphics of Student Graduation Percentage

Prediction results of data testing. The number of testing data used was 146 students using the Naïve Bayes method. The prediction results showed that 144 students graduated on time. cleaned then used. The attributes used are gender, student status, age, cumulative achievement index (GPA). The data taken is divided into two groups, which have two data classes, namely on time and late. Next, testing with RapidMiner Studio 10.3 as a tool to obtain the level of accuracy of the Naïve Bayes method. The following is a calculation to get the accuracy value with the following equation:

 $Akurasi = \frac{TP+TN}{TP+FN+FP+TN} = X100\%$ 

 $= \frac{50+40}{50+5+5+40} \ge 100\%$ 

= 90%

Predictions were made using training data using Naïve Bayes, resulting in 144 students predicted being on time, 2 students predicted being late.

## 5. Conclusion

This research shows that the Naïve Bayes method is effective in predicting student graduation on time with an accuracy rate of up to 90%. By utilizing attributes such as cumulative grade point average (GPA) and other academic data, this algorithm is able to identify relevant patterns to support policy decisions that increase on-time graduation. This implementation provides a practical solution for universities in managing student data to improve accreditation and institutional reputation.

## References

- [1] M. P. Dr. Achmad Fathoni Rodli and S. E. M. S. M. Nikma Yucha, *Manajemen Kinerja Institusi Perguruan Tinggi*. CV Rey Media Grafika, 2022. [Online]. Available: https://books.google.co.id/books?id=S6p-EAAAQBAJ
- [2] L. Setiyani, M. Wahidin, D. Awaludin, and S. Purwani, "Analisis Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode Data Mining Naïve Bayes: Systematic Review," *Fakt. Exacta*, vol. 13, no. 1, p. 35, 2020, doi: 10.30998/faktorexacta.v13i1.5548.
- [3] A. Wasik *et al.*, "Implementasi data mining untuk memprediksi penjualan accessoris handphone dan handphone terlaris menggunakan metode k-nearest neighbor (k-nn) 1," vol. 1, no. 2, pp. 469–479, 2024.
- [4] E. luthfi and U. Amikom, *Algoritma Data Mining*. Penerbit Andi, 2009. [Online]. Available: https://books.google.co.id/books?id=-Ojclag73O8C
- [5] A. S. Huda, R. M. Awangga, and R. N. S. Fathonah, *Prediksi Penerimaan Pegawai Baru Dengan Metode Naive Bayes*. in Data Science. Kreatif, 2020. [Online]. Available: https://books.google.co.id/books?id=aawGEAAAQBAJ
- [6] P. Studi, T. Informatika, A. A. Muin, P. Studi, and S. Informasi, "Metode Naive Bayes Untuk Prediksi Kelulusan (Studi Kasus: Data Mahasiswa Baru Perguruan Tinggi)," vol. 2, no. 1, 2016.
- [7] T. P. Hamakonda, *Pengantar klasifikasi persepuluhan Dewey*. Gunung Mulia, 1978. [Online]. Available: https://books.google.co.id/books?id=l4yxjMA8rQcC
- [8] W. I. Rahayu, M. H. K. Saputra, R. M. Awangga, and R. Habibi, *Penerapan Metode Naive Bayes dan Skala Likert Pada Aplikasi Prediksi Kelulusan Mahasiswa*. Kreatif, 2020. [Online]. Available: https://books.google.co.id/books?id=CZv9DwAAQBAJ
- [9] S. P. M. S. Ni'matuzahroh and M. P. Susanti Prasetyaningrum, OBSERVASI: TEORI DAN APLIKASI DALAM PSIKOLOGI. in 1. UMMPress, 2018. [Online]. Available: https://books.google.co.id/books?id=CMh9DwAAQBAJ
- [10] M. Zed, *Metode Penelitian Kepustakaan*. Yayasan Pustaka Obor Indonesia, 2008. [Online]. Available: https://books.google.co.id/books?id=zG9sDAAAQBAJ
- [11] L. J. Moleong and T. Surjaman, *Metodologi penelitian kualitatif*. Remadja Karya, 1989. [Online]. Available: https://books.google.co.id/books?id=YXsknQEACAAJ

[12] M. Hasanah, N. H. Harani, N. Riza, R. M. Awangga, and R. Habibi, *Implementasi Barcode Dan Algoritma Regresi Linear Untuk Memprediksi Data Persediaan Barang*. Kreatif. [Online]. Available: https://books.google.co.id/books?id=db38DwAAQBAJ