BAB IV

HASIL DAN PEMBAHASAN

4.1 Tahapan Implementasi dengan RapidMiner

Data penjualan *smartphone* Erafone Suzuya Baru Rantauprapat akan diimplementasikan ke dalam aplikasi *RapidMiner* 9.10. Proses implementasi dilakukan dengan menjadikan variabel Minat sebagai atribut Label (target) untuk prediksi, dengan atribut-atribut lainnya sebagai atribut independen.

4.1.1 Import Data

Tahap import data adalah proses memasukkan data dari sumber eksternal ke dalam lingkungan perangkat lunak atau sistem yang akan digunakan untuk analisis atau pemrosesan lebih lanjut.



Gambar 4.1 Proses Membuka Aplikasi RapidMiner

Gambar 4.1 menunjukkan tampilan awal (*splash screen*) saat membuka aplikasi *RapidMiner* Studio versi 9.10. Berikut penjelasan detailnya:

- 1. Di bagian kiri atas terdapat logo *RapidMiner* yang merupakan identitas visual dari perangkat lunak ini.
- Terdapat tulisan besar "Studio" yang menandakan nama produk, yaitu RapidMiner Studio.
- 3. Tertera versi aplikasi yaitu Version 9.10, menunjukkan bahwa pengguna menggunakan versi terbaru atau spesifik dari *RapidMiner* Studio.
- 4. Tulisan "*Reading Configuration Files*" menandakan bahwa aplikasi sedang memuat file konfigurasi yang diperlukan untuk menjalankan program dengan benar.

🗋 늘 📙 🔹 🕨	· ·	Views:	Design	Results	Turbo Prep	Auto Model	Deployments			Find dat	a, operatorsetc	,0	All Studio	Ŧ
Repository ×	Process							,			Parameters	×		
	-							-			Process			
🕒 Import Data 🛛 🗉 🔻	Process							à 🖪	4 📽	: 🔎 🖸	Flocess			
Training Resources (connected)	Process										logverbosity	init	*	٦
Samples											locfile			ີຄ
Community Samples (connected)) inp									res (logine			J
Local Repository (Legacy)											resultfile			0
DB (Legacy)														
											random seed	2001		Ø
											send mail	never		- D
<														J."
											encoding	SYSTE	EM 🔻	٢
Operators ×						irst.								
Search for Operators														
Data áccoso (62)														
Blending (82)														
Cleansing (28)														
Modeling (167)											Tide advan	<u>sed param</u>	leters	
Scoring (14)											Change cor	npatibility ((9.10.014)	£ _
Validation (30)														
🕨 📴 Utility (85)											Help \times			
Get more operators from the											Proce			^
Marketplace				A Recommer	dations unavailabl	0					RapidMi	bar Studie	o Core	~

Gambar 4.2 Lembar Kerja Utama Aplikasi RapidMiner

Gambar 4.2 menampilkan tampilan utama (main *workspace*) dari aplikasi *RapidMiner* Studio setelah aplikasi berhasil dibuka dan siap digunakan untuk melakukan proses analisis data.

	Import Data - Select the d	ata location.				× Find data	operatorsetc	<u>р</u> А	II Studio 🔻
Repository ×	Proces Prc	Sele	ct the data	ocation.		• 10	Parameters	×	
Training Resources (connected) Pro	ocess 📑 File Tugas Akhir			•	← 등 ☆ ☆ 🐸	a -	logverbosity	init	• 0
Samples	Bookmarks	File Name	Size	Туре	Last Modified		logfile		- C
Community Samples (connected) in	🜟 Last Directory	data_penjualan.xlsx	10 KB	Microsoft Excel Work	she Apr 13, 2025	∧ res (
DB (Legacy)							resuluile		
							random seed	2001	
							send mail	never	•
<							encoding	SYSTE	M V G
Operators ×								C.C.L.	
Search for Operators									
Data Access (63)						~			
Blending (82)	data penjualan xisx	1							
Cleansing (28)							Inde advance	ed parame	eters
Modeling (167)	All Files					· ·	Change cor	nnatibility (S	9 10 014)
Scoring (14)				The selected file w	vill be imported as: Excel Ch	ange	• •		
Validation (30)									
Utility (85)				- Previou	us → <u>N</u> ext X <u>C</u> ai	ncel	Help ×		
Get more operators from the							Proce	~ ~	^
Marketplace		A Re	commendations	unavailable					

Gambar 4.3 Direktori File Import Data

Gambar 4.3 di atas menunjukkan jendela pemilihan file data saat proses import data ke dalam aplikasi *RapidMiner*. Gambar ini merupakan tahap awal dalam mengimpor dataset penjualan smartphone ke dalam *RapidMiner*, dimana pengguna memilih file *Excel* yang berisi data yang akan dianalisis lebih lanjut menggunakan berbagai operator dan algoritma *machine learning* di aplikasi tersebut. Gambar di atas menunjukkan tampilan saat proses *import* data dari file *Excel* sedang berlangsung. File data penjualan diunggah ke aplikasi *RapidMiner*. Gambar ini memperlihatkan tahap awal pengambilan data dari sumber eksternal (*file Excel*) ke dalam lingkungan kerja *RapidMiner*. Selanjutnya klik *Next*.

🗋 📒 🔚 🔹		*		B	View	Design		Results 1	urbo Prep	Auto Model	Deployment	5		Find data	operatorsetc	🔎 All S	itudio 🖛
Repository ×			Impo	rt Data - Sele	ect the cells to i	mport.							×	rameters	×		
C Import Data	= •													Process			
• Training Resources toom	ected)			_			Select	the cells to	import.					verbosity	init		• 0
Samples					142											16	In
Community Samples (con	nected)		She	et Sheet1	▼ Cell	range: All		Select	All 🗸 D	efine header ro	w: 1 .						
Local Repository (Lepery)														uttile			0
Connections				A	В	c	D	E	F	G	н	1					
data			1	Apple	Oppo	Samsung	Vivo	Xiaomi	Omset	Jumlah P	Kategori	Minat	^	dom seed	2001		(D)
processes			2	30.000	10.000	25.000	1.000	7.000	47556486	73.000	Kecil	Sedang		dmail	neuer		• m
• B DB (redeck)			3	12.000	2.000	5.000	1.000	2.000	16475495	22.000	Kecil	Rendah		O TINDE	ivevei		
			4	24.000	16.000	23.000	3.000	4,000	42302923	70.000	Kecil	Rendah		oding	SYSTEM		- 0 -
Operators ×			5	21.000	27.000	31.000	2.000	13.000	49383423	94.000	Kecil	Sedang		Hide advanc	ed parameters		
			6	30.000	14.000	40.000	1.000	10.000	64000450	95.000	Menengah	Sedang		Change con	000000000000000000000000000000000000000		
Search for Operators			7	37.000	20.000	41.000	6.000	7.000	63242639	111.000	Menengah	Sedang		MINDER FOR	A STRAIGHT LA. LA. W. T.	*	
Data Access (63)		^	8	34.000	16.000	35.000	3.000	3.000	61180990	91.000	Menengah	Sedang					
Blending (82)			9	61.000	21.000	36,000	3.000	10.000	85848557	131.000	Besar	Tinggi		lp ×			
Cleansing (28)			10	42.000	23.000	32.000	4.000	7.000	59412178	108.000	Kecil	Sedang		Proces			1
Modeling (167)		п.	11	42.000	18.000	34.000	4.000	6.000	60462089	104.000	Menengah	Sedang		BapidMir	ner Studio Core		- 1
Scoring (14)			12	35.000	18.000	53.000	2.000	6.000	65722838	114.000	Menengah	Sedang		and the second s			
Validation (30)			13	70.000	23.000	59.000	7.000	9.000	11653779_	168.000	Besar	Tinggi		ropsis			
Utility (85)		~	14	66.000	9.000	53.000	4.000	5.000	93490600	137.000	Besar	Tinggi	v	eroot operal every proces	tor which is the oi is.	uter most ope	rator
Get more operators from the Marketplace	2								[- Previous	\longrightarrow Next	X Cano	el	arrintion			

Gambar 4.4 Select the Cells to Import

Gambar 4.4 memperlihatkan jendela yang terbuka berjudul "*Import Data -Select the cells to import*", dan sedang berada pada tahap pemilihan data yang akan digunakan dalam proses analisis. Gambar ini menunjukkan bahwa *RapidMiner* sudah berhasil membaca dan menampilkan dataset dengan benar, dan pengguna tinggal menekan "*Next*" untuk melanjutkan proses berikutnya, seperti pengaturan tipe data, definisi atribut label.

Gambar 4.4 menunjukkan preview data yang akan diimpor ke dalam *RapidMiner*, dengan beberapa variabel (kolom) yang merepresentasikan informasi penjualan *smartphone*. Berikut penjelasan variabel-variabel tersebut.

- Kolom Apple, Oppo, Samsung, Vivo, Xiaomi merupakan kolom yang berisi angka penjualan unit untuk masing-masing merek *smartphone*. Tipe data **Numerik** (kemungkinan integer).
- 2. Kolom Omset menunjukkan total pendapatan atau omzet dari penjualan *smartphone*.
- 3. Kolom Jumlah Penjualan adalah total jumlah unit *smartphone* terjual secara keseluruhan.

- 4. Kolom Kategori Omset adalah kategori klasifikasi omset berdasarkan nilai tertentu seperti "Kecil," "Menengah," dan "Besar.
- 5. Kolom Minat merupakan variabel target yang menunjukkan tingkat minat pembeli terhadap produk. Kategori seperti "Sedang," "Rendah," dan "Tinggi."

<new process=""> - RapidMiner Studio Economic Studio Economica View Connection</new>	ducational 9.1	0.014	@ DESKTOP-I	R1KAUK7										-	D	\times
		impor	rt Data - Form	at your column	ns.		1		···	.)		× ind data,	operatorsetc	P	All Studio	•
Repository ×	Proces					Format yo	ur columns.						Parameters	×		
Training Resources (connected) Samples	Process		Replace	errors with m	issing values 🛈								logverbosity	init	•	
Community Samples (connected Local Repository (Legacy) DD -) inp		Vivo Integer	♦ ▼ Xia Inte	aomi 🛛 🗢 👻 araer nge role	Omset ¢	Jumlah Pe.	0 -	Kategori O	♥ ▼ Minat oolync ×	ominal 🕈 🔻	res (resultfile			
DB (Legacy)		1 2	1		Please ent	er the new role:				Sedar Rend	ng ah	Â	random seed	2001		Ð
<		3 4 5	2	label						v Sedar	ng ng	-	send mail	never	-	
Operators ×		6 7	6 3					2 🍾	<u>)</u> К 🔀 <u>C</u> ar	Sedar Sedar	ng		encoding	5151	EM	
Data Access (63)		8 9	3 4	10		858485574 594121785	131 108		Besar Kecil	Tingg Sedar	i ng					
Blending (82) Cleansing (28) Modeling (467)		10 11	4	6		604620894 657228384	104 114		Menengah Menengah	Sedar	ng		🐰 Hide advanc	ed paran	neters.	
 Scoring (14) Validation (30) 		12	< 7	٥		1165377030	169	11	Recor	Tinga (> S no problem	s.	Change com	patibility	<u>(9.10.014</u>	a
Get more operators from the								+	- Previous	→ <u>N</u> ext	X Cancel		Help ×			^
Marketplace						A Recommend	lations unavailable						RapidMir	iS her Studi	io Core	~

Gambar 4.5 Set Role Variabel Target

Gambar 4.5 menunjukkan proses pengaturan variabel Minat sebagai Label (target). Variabel lain seperti Apple, Oppo, Samsung, Vivo, Xiaomi, Omset, dan Jumlah Penjualan digunakan sebagai atribut.

<new process=""> – RapidMiner Studio Ec</new>	ducational 9.1	0.014	@ DESKTOP	-R1KAUK7											-	o ×	
Eile Edit Process View Connection	ons <u>S</u> etting:	s Ex	tensions I	Help													
	·	Impor	rt Data - Forr	nat your co	olumns.		-		_			×	Find data,	operatorsetc	<i>P</i>	VI Studio 🔻	<u> </u>
Repository ×	Proces													Parameters	×		
🔂 Import Data 🛛 = 🔻	Prc						Formaty	your	columns.				• •	Process			
	Process													logverbosity	init	•	٢
Samples Gommunity Samples (conner			Replac	e errors w	iui missing v	aiues 🛈											
Local Repository (Legecy)) inp	,	Vivo	0 v	Xiaomi	0 v	Omset	۰.	Jumlah Pe 🔷 🤘	Kategori O 🚸 🧃	• Minat 🛛 🗢	Ŧ	res	logfile			0
Connections			Integer		Integer		integer		Integer	polynominal	polynominal label			resultfile			Ð
tata 📒		1	1		7		475564869		73	Kecil	Sedang	^					
processes		2	1		2		164754955		22	Kecil	Rendah			random seed	2001		Ð
alfia (4/14/25 12:06 PM - 7		3	3		4		423029238		70	Kecil	Rendah					-	
DB (Legacy)		4	2		13		493834237		94	Kecil	Sedang	=		send mail	never	•	e
		5	1		10		640004503		95	Menengah	Sedang			encodina	SYSTE	м т	Ð
Operators ×		6	6		7		632426390		111	Menengah	Sedang						
		7	3		3		611809900		91	Menengah	Sedang						
Search for Operators		8	3		10		858485574		131	Besar	Tinggi						
Data Access (63)		9	4		7		594121785		108	Kecil	Sedang						
Blending (82)		10	4		6		604620894		104	Menengah	Sedang						
Cleansing (28)		11	2		6		657228384		114	Menengah	Sedang			R Hide advanc	ed paramr	eters	
 Modeling (167) 		12	7		0	_	1165377030		168	Recor	Tinggi	× ×		Change com	an at billity (9 10 014)	
Scoring (14)											🕗 no probl	ems.			pononity (2.10.0141	
Validation (30)																	
Utility (85)										← Previous –	→ <u>N</u> ext X <u>C</u> a	ncel		Help ×			
Get more operators from the Marketplace							A Recomm	endatio	ns unavailable					Proces	SS her Studio	o Core	^ ~

Gambar 4.6 Format dan Tipe Data

Gambar 4.6 menunjukkan tahap pemberian format dan tipe data pada kolom-kolom dataset di *RapidMiner* setelah proses import data. Berikut penjelasan detailnya:

a. Format your columns

mengindikasikan bahwa langkah ini bertujuan untuk menentukan tipe data yang tepat untuk setiap kolom dalam dataset.

b. Preview Data

tabel di tengah menampilkan sebagian dari dataset yang telah diimpor, memperlihatkan nilai-nilai pada setiap kolom.

c. Vivo, Xiaomi, Omset, Jumlah Penjualan

keempat kolom ini memiliki tipe data *Integer*, ditunjukkan oleh ikon angka bulat kecil di samping nama kolomnya. Ini menandakan bahwa kolom-kolom ini berisi angka bulat (bilangan cacah).

d. Kategori Omset, Minat

Kedua kolom ini memiliki tipe data Polynomial.

Bit End Process Vew Connections Settings Edensions Help	<new process=""> – RapidMiner Studio Educati</new>	tional 9.1	0.014 @ DESKTOP-R1KAUK7			- 0 ×
Import Data - Where to store the data? Repository Import Data - Where to store the data? Import Data - Where to store the data - Tab Import	Elle Edit Process View Connections	Settings	s Extensions Help			
Repealed y Import Data Import Data <th></th> <th>•</th> <th>Import Data - Where to store the data?</th> <th>Find data, e</th> <th>operatorsetc</th> <th>🔎 All Studio 👻</th>		•	Import Data - Where to store the data?	Find data, e	operatorsetc	🔎 All Studio 👻
Samples Samples Samples Samples Samples Samples Samples Somaling (Samples) Samples	Repository X	Proces	Where to store the data?	• IN	Parameters	×
Commands Samples Comm				- 11		
Community Samples (come Local Repository (Lago) Community Samples (come Local Repository (Lago) Community Samples (come Come data Percesses Percess	Samples	ocess	▼ 🜉 Local Repository (Legacy)		logverbosity	Init 🔍 🛈
Connections Socient (14) Socient (15) Socient (16) So	Community Samples (connect		📇 data		loofile	0
Connections data for connections for conneconnections for conneconnections for connections for co	▼ 🜉 Local Repository (Legacy)		processes	res	-	
Search for Operators <	Connections		🚀 alfia (4/14/25 12:06 PM – 7 kB)		resultfile	📒 🗇
Control Section (26) Control Cost (20) Cost (tata 📒					
Control (Control (Contro) (Control (Control (Contro) (Control (Contro) (Control (Contro)	processes				random seed	2001
Cleaning (2) Cleaning (2) Cleaning (2) Cleaning (2) Cleaning (2) Location //Local Repository/data/data_penjualan Location //Local Repository/data/data_penjualan Location //Local Repository/data/data_penjualan Cleaning (2)	alfia (4/14/25 12:06 PM - 7				send mail	never T
Operators Search for Operators > Data Access (63) > Data Access (63) > Defining (22) > Cleansing (28) > Modeling (167) > Scoring (14) > Validation (30) > Utility (65)	C DB (Legacy)					
Operators × Search for Operators > > Data Access (3) > > Scoring (14) > > Validation (30)					encoding	SYSTEM 🔻 🛈
Search for Coperators Data Access (63)	Operators ×					
Data Access (63) Deta	Search for Operators					
Stand Access (0.5) Scone (14) Cation //Local Repository/data/data_penjualan Cation //Local Repository/Local Repository/Local Repo	b Data (cross (62)					
Sections (28) Sect	Blending (82)					
Modeling (167) Modeling (167) Scoring (14) Location //Local Repository/data/data_penjualan Change compatibility (9 10.014) Location //Local Repository/data/data_penjualan Change compatibility (9 10.014) Kernetations unavailable	Cleansing (28)				-	
Scoring (14) Location //Local Repository/data/data_penjualan Location //Local Repository/data/data_penjualan Change compatibility (9:10.014) Help × Change compatibility (9:10.014) Help × Process Process Process Process Process Process Process	Modeling (167)	1	Name data_penjualan		Hide advance	d parameters
Validation (30) V	Scoring (14)				Change comp	atibility (9.10.014)
Utility (85) Cancel Help × Cancel Marketilaze Accommendations unavailable Process Process Process Process	Validation (30)	L	Location //Local Repository/data/data_penjualan			
Get more operators from the Marketulace A Recommendations unavailable	Utility (85)		- Previous 🛛 Einish 🗶 Cancel		Help ×	
Markeblace. A Recommendations unavailable	Get more operators from the				Ducces	^
) RADIONIDE SUBIOLOPE 1)	Marketplace		🛕 Recommendations unavailable		RapidMine	s er Studio Core 🗸 🗸

Gambar 4.7 Direktori Penyimpanan File

Gambar 4.7 menunjukkan jendela penentuan lokasi penyimpanan file di *RapidMiner* setelah proses import data selesai. Ini adalah repositori lokal pada komputer pengguna, tempat dataset akan disimpan secara *default*. Pengguna dapat memberikan nama untuk dataset yang telah diimpor pada kolom "*Name*". Dalam gambar ini, nama filenya adalah "data penjualan". Klik tombol "*Finish*" untuk menyelesaikan proses import dan menyimpan data ke lokasi yang ditentukan. Gambar ini menggambarkan tahap akhir dari proses import data ke *RapidMiner* dimana pengguna menentukan lokasi dan nama file untuk menyimpan dataset yang telah berhasil diimpor ke dalam sistem agar dapat digunakan untuk analisis lebih lanjut.

				Desig	Dae	ulta Turba D	ture ture thed	al Deploym	ente	The distance of		All Ohu	lia =
	1 1 			views: Design	i nes		Auto Mod	er Deploym	ents	Find data, o	peratorsetc	All Stud	10 +
pository ×	Result History		ExampleSet (//L	ocal Repository/d	ata/data_pen	jualan) ×	ExampleSet (//Local Reposito	ory/data/data_penju	ıalan) ×		arameters	×
			_		1							Process	
		Open in	🗧 Turbo Prep	Auto Model				F	Filter (24 / 24 examp	les): all	•	logverbosity	Œ
 Training Res Samples 	Data	Row No.	Minat	Apple	Орро	Samsung	Vivo	Xiaomi	Omset	Jumlah Penj	Katego		0
🕨 🚨 Community S		1	Sedang	30	10	25	1	7	475564869	73	Kecil ^	lognie	a
Local Repos	Σ	2	Rendah	12	2	5	1	2	164754955	22	Kecil	resultfile	Œ
🔻 🧮 data	Statistics	3	Rendah	24	16	23	3	4	423029238	70	Kecil	random seed	Œ
📕 data_p		4	Sedang	21	27	31	2	13	493834237	94	Kecil		
processe:	(5	Sedang	30	14	40	1	10	640004503	95	Meneng	send mail	Œ
alfia (4/14)	Visualizations	6	Sedang	37	20	41	6	7	632426390	111	Meneng	encoding	Œ
		7	Sedang	34	16	35	3	3	611809900	91	Meneng		
		8	Tinggi	61	21	36	3	10	858485574	131	Besar		
		9	Sedang	42	23	32	4	7	594121785	108	Kecil		
	Annotations	10	Sedang	42	18	34	4	6	604620894	104	Meneng		
		11	Sedang	35	18	53	2	6	657228384	114	Meneng		
		12	Tinggi	70	23	59	7	9	1165377930	168	Besar		
		13	Tinggi	66	9	53	4	5	934906000	137	Besar	Hide advartight adv	ced
		14	Sedang	46	7	46	6	12	843497000	117	Besar	Change	
												🖌 compatibili	hv .

Gambar 4.8 Preview Dataset

Gambar 4.8 menampilkan *preview* dataset yang telah berhasil diimpor dan disimpan dalam *repository* lokal *RapidMiner*. Tabel ini menunjukkan data mentah yang siap untuk dianalisis atau diproses lebih lanjut. Setiap baris mewakili satu contoh (*record*) data penjualan *smartphone*. Di bagian bawah tabel tertulis bahwa dataset berisi 24 baris dengan 1 atribut khusus dan 8 atribut reguler. Dataset ini

merupakan dasar untuk analisis lebih lanjut menggunakan algoritma *machine learning Naïve Bayes* atau *Support Vector Machine* (SVM) dalam *RapidMiner*.

4.1.2 Pra-pemrosesan Data

Setelah data berhasil diimpor ke dalam *RapidMiner*, langkah selanjutnya adalah melakukan *pra-pemrosesan* data. Tahap ini sangat penting untuk memastikan kualitas dan konsistensi data sebelum digunakan dalam proses analisis atau pemodelan *machine learning*.



Gambar 4.9 Retrieve Data

Gambar 4.9 menunjukkan tampilan proses awal dalam *RapidMiner* Studio, yaitu langkah *Retrieve* "data penjualan" yang berfungsi untuk mengambil dataset yang sudah disimpan di *repository* lokal agar dapat digunakan dalam *prapemrosesan* data.

Result History		ExampleSet (Retrieve data_pe	njualan) ×				
		Name	- Type	Missing	Statist Filter (9 / 9 attributes	:): Search for Attributes	▼ •
Data	~	Label Minat	Polynominal	0	Least Rendah (2)	Most Sedang (15)	Valu A
Statistics	~	Apple	Integer	0	Min 12	Max 70	Aver 36
	~	Орро	Integer	0	Min 2	Max 39	Aver 19.
Visualizations	~	Samsung	Integer	0	Min 5	Max 59	Aver 42.
	~	Vivo	Integer	0	Min O	Max 7	Aver 2.5
Annotations	~	Xiaomi	Integer	0	Min 2	Max 13	Aver 7.7
	~	Omset	Integer	0	Min 164754955	Max 1165377930	Aver 662
	<	1	Integer	0	Min 22	Max 168	Aver 109
	Showi	ng attributes 1 - 9			Examples: 24 Spec	cial Attributes: 1 Regular Attril	butes: 8

Gambar 4.10 Handle Missing Value

Gambar 4.10 di atas menunjukkan bahwa tidak ada *missing values* dalam dataset yang dianalisis. Hal ini sangat penting karena keberadaan nilai yang hilang dapat mempengaruhi kualitas model prediksi serta akurasi analisis data.

Elle Edit Process View Connections Settings E	Extensions Help							
	Views: Design	Results Turbo Prep	Auto Model	Deployments		Find data, opera	torsetc 🔎 All Studio	•
Repository × Proc	ess				Р	arameters ×		
🕒 Import Data 🛛 = 💌 🔘 P	rocess		🔎 🔎 🗈	🖬 😼 🗸		Normalize		
Samples Process	5					create view		0
Samples (connected)	eve data_penju Normalize				res att	ribute filter type	all	0
Local Repository (Legacy) Connections					res (invert selection		٢
data data_penjualan (4/14/25 9/27 /						include special attri	butes	٢
DB (Legacy)		Close			m	ethod	range transformation	0
					mi	in	0.0	1
Operators ×					m	ax	1.0	0
Cleansing (3) Normalize					2	Hide advanced para	ameters itv (9.10.014)	
Scale by Weights					н	elp ×		
Modeling (1)			and the first		-	Normalize	e	^
Transformation (1)	e the wisdom of Growds to get operator recom	nmendations based on your proce	ss aesigní			RapidMiner Stu	idio Core	
No results were found.	×	Activate Wisdom of Crowds				aling, Features, Attrib	anuaruze, <u>z-rransformation,</u> outes, <u>Variables, Columns</u>	~

Gambar 4.11 Normalisasi Data

Gambar 4.11 menunjukkan proses normalisasi data yang dilakukan dalam aplikasi *RapidMiner* Studio sebagai bagian dari tahapan pra-pemrosesan data.

Operator pertama, *Retrieve* data penjualan, berfungsi untuk mengambil dataset penjualan yang sudah disimpan di *repository* lokal agar dapat diproses lebih lanjut. Operator kedua adalah *Normalize* yang digunakan untuk melakukan normalisasi pada dataset. Normalisasi bertujuan mengubah skala nilai atribut agar berada dalam rentang tertentu, biasanya antara 0 dan 1. *Method range transformation* mengindikasikan metode yang digunakan untuk menormalkan data, yaitu transformasi rentang (*range transformation*). Nilai minimum setelah normalisasi diatur ke 0, dan Nilai maksimum setelah normalisasi diatur ke 1.

<	dMiner Studio Educati	ional 9.10.014 @ DE Settings Extensio	SKTOP-R1KAUK7								- 0 ×
	•		Views:	Design	Results	Turbo Prep	Auto Model	Deployments		Find data, operal	torsetc 🔎 All Studio 🔻
Repository ×	Result History	Ex.	ampleSet (Norm	alize) ×							Parameters ×
Impo Training Resource		Open in 🔢	Turbo Prep	🐴 Auto Model			F	Filter (24 / 24 exam)	oles): all	•	Normalize create view
Samples	Data	Minat	Apple	Орро	Samsung	Vivo	Xiaomi	Omset	Jumlah Penj	Kategori	attribute filter type
Community Samp		Sedang	0.310	0.216	0.370	0.143	0.455	0.311	0.349	Kecil ^	
 Local Repository (Connections 	Σ	Rendah	0	0	0	0.143	0	0	0	Kecil	invert selection
T 📩 data	Statistics	Rendah	0.207	0.378	0.333	0.429	0.182	0.258	0.329	Kecil	include special attributes ①
data_penju:		Sedang	0.155	0.676	0.481	0.286	1	0.329	0.493	Kecil	
processes	(Sedang	0.310	0.324	0.648	0.143	0.727	0.475	0.500	Menenga	method • ①
DB (Legacy)	Visualizations	Sedang	0.431	0.486	0.667	0.857	0.455	0.467	0.610	Menenga	min 0.0 ①
		Sedang	0.379	0.378	0.556	0.429	0.091	0.447	0.473	Menenga	
		Tinggi	0.845	0.514	0.574	0.429	0.727	0.693	0.747	Besar	max 1.0 0
		Sedang	0.517	0.568	0.500	0.571	0.455	0.429	0.589	Kecil	
	Annotations	Sedang	0.517	0.432	0.537	0.571	0.364	0.440	0.562	Menenga	
		Sedang	0.397	0.432	0.889	0.286	0.364	0.492	0.630	Menenga	
		Tinggi	1	0.568	1	1	0.636	1	1	Besar	
		Tinggi	0.931	0.189	0.889	0.571	0.273	0.770	0.788	Besar	_
		Sedang	0.586	0.135	0.759	0.857	0.909	0.678	0.651	Besar 🗸	Hide advanced parameters
<		< ExampleSet (24 e	xamples, 1 specia	il attribute, 8 regula	ır attributes)					>	Change compatibility (9.10.014)

Gambar 4.12 Hasil Normalisasi Data

Gambar 4.12 menunjukkan hasil normalisasi dataset pada aplikasi *RapidMiner*. Kolom-kolom seperti Apple, Oppo, Samsung, Vivo, Xiaomi, Omset, dan Jumlah Penjualan sudah dikonversi ke rentang nilai antara 0 dan 1 (sesuai parameter normalisasi *Min-Max*). Kolom Minat tetap dipertahankan sebagai atribut target (label) dan tidak dinormalisasi. Kolom Kategori juga tetap sebagai atribut kategorikal (nominal) dan tidak mengalami normalisasi. Data asli dari

kolom "Apple" seperti 30.000, 12.000, dst., kini menjadi 0,310, 0,207, dst. Begitu juga nilai "Omset", misalnya 475564865, menjadi 0,311.

4.1.3 Pembagian Data

Setelah data berhasil diimpor dan dilakukan *pra-pemrosesan* seperti normalisasi, langkah selanjutnya adalah pembagian data (data *splitting*). Proses ini sangat penting dalam *machine learning* untuk memastikan model yang dibangun dapat diuji secara objektif.



Gambar 4.13 Pembagian Data

Gambar 4.13 di atas menunjukkan proses pembagian data (*data splitting*) dalam *RapidMiner*. Operator "*Split* Data" digunakan untuk membagi dataset menjadi dua bagian berdasarkan rasio tertentu. Rasio Pembagian 0.7 dan 0.3, yang berarti bahwa data akan dibagi dengan proporsi 70% untuk pelatihan dan 30% untuk pengujian. Ini merupakan praktik umum dalam *machine learning* agar model dapat dilatih dengan cukup data sambil tetap memiliki set data terpisah untuk evaluasi kinerja model.

Open in Image: Turbo Prep Auto Model Statistics Filter (17 / 17 examples): all Row No. Minat Apple Oppo Samsung Vivo Xiaomi Omset 1 Rendah 0 0 0 0.143 0 0 2 Sedang 0.155 0.676 0.481 0.286 1 0.329	
Row No. Minat Apple Oppo Samsung Vivo Xiaomi Omset 1 Rendah 0 0 0 0.143 0 0 2 Sedang 0.155 0.676 0.481 0.286 1 0.329	Ŧ
Instantiation Rendah 0 0 0.143 0 0 2 Sedang 0.155 0.676 0.481 0.286 1 0.329	J
2 Sedang 0.155 0.676 0.481 0.286 1 0.329 Statistics 2 0.400 0.204 0.504 0.510 0.412 0.707 0.475	0 ^
Statistics 0.04422 0.0442 0.0442 0.0707 0.475	0
3 Secang 0.310 0.324 0.548 0.143 0.727 0.475	0
4 Sedang 0.431 0.486 0.667 0.857 0.455 0.467	0
5 Sedang 0.379 0.378 0.556 0.429 0.091 0.447	0
Visualizations 6 Tinggi 0.845 0.514 0.574 0.429 0.727 0.693	0
7 Sedang 0.397 0.432 0.889 0.286 0.364 0.492	0
8 Sedang 0.586 0.135 0.759 0.857 0.909 0.678	0
9 Tinggi 0.672 0.541 0.907 0 0.818 0.677	0
Annotations 10 Tinggi 0.466 0.649 0.630 0.429 1 0.576	0
11 Sedang 0.431 0.405 0.722 0 0.818 0.485	0
12 Tinggi 0.224 0.784 0.981 0 0.818 0.465	0
13 Sedang 0.224 0.730 0.944 0.143 0 0.540	0
14 Sedang 0.086 0.568 0.870 0.429 0.273 0.489	0 🗸
ExampleSet (17 examples: 1 special attribute: 8 regular attributes)	>

Gambar 4.14 Hasil Pembagian Data

Gambar 4.14 di atas menunjukkan hasil dari proses pembagian (*splitting*) data menggunakan *RapidMiner*, dengan 17 baris data pelatihan hasil pembagian. Sisanya, sebanyak 7 baris data digunakan untuk data pengujian.

4.1.4 Implementasi Metode Naïve Bayes

Bagian ini menjelaskan proses Implementasi Metode *Naïve Bayes*. Ini merupakan tahap di mana algoritma *Naïve Bayes* diterapkan pada data pelatihan yang telah disiapkan sebelumnya.

<new process*=""> - RapidMiner Studio</new>	Educational 9.10.014 @ DESI	CTOP-R1KAUK7								_	o ×
Eile Edit Process View Connect	ions <u>S</u> ettings Extensions	elp <u>H</u> elp									
	• •	Views:	Design	Results	Turbo Prep	Auto Model	Deployments]	Find data,	operatorsetc 🔑	All Studio 🔻
Repository ×	Process									Parameters ×	
🕒 Import Data 🛛 = 💌	Process						,o ,o 👔	n 🗖 😽	द 🧉 🖾	Naive Bayes	
Community Samples (conner	Process									✓ Iaplace correction	٢
Local Repository (Legacy)											
Connections) inp						C		105		
T data	Retrieve data_penju	Normalize		plit Data	Naive Bayes	L			ies (
data_penjualan (4/16/	C out)-	exa 🔛	exa exa	T par	tra mod						
processes			pre	put p	v						
DB (Lenery)		1									
< >											
Operators ×											
naive bayes 🗙											
🕶 📇 Modeling (2)											
Predictive (2)											
🔻 🛅 Bayesian (2)											
Valve Bayes										R Hide advanced par	ameters
V Naive Bayes (Kernel)											
										Heln X	
											^
Get more operators from the Marketplace				🛕 Recomme	ndations unavailabl	e				Naive Bay RapidMiner Stu	res Idio Core

Gambar 4.15 Implementasi Operator Naïve Bayes

Gambar 4.15 menampilkan *workflow* proses di *RapidMiner* yang mengilustrasikan tahapan implementasi algoritma *Naïve Bayes* untuk analisis data penjualan *smartphone*. Operator *Naive Bayes* adalah inti dari proses ini, dimana algoritma *Naïve Bayes* diterapkan pada data pelatihan untuk membangun model klasifikasi.

<new process*=""> – RapidMiner Studio E File Edit Process View Connection</new>	ducational 9.10.014 @ DESK	TOP-R1KAUK7 Help							-	0 X
	•	Views: De	sign Results	Turbo Prep	Auto Model	Deployments]	Find data,	operatorsetc 🔎	All Studio 🔻
Repository ×	Process								Parameters ×	
🕒 Import Data 🛛 = 💌	Process					P P 🔋	। 🖪 🖪 व	🖌 🥶 🖻	Apply Model	
Community Samples (conner	Process								application parameters	 (1)
Local Repository (Legacy)									create view	D
Connections) inp							res		
T data	Retrieve data_penju	Normalize	Split Data	Naive Bayes				res		
data_penjualan (4/18/	C out	exa exa	exa par	tra mod			(1		
processes	 ✓ <u>1</u> 	pre	par)	1						
The second		v	1	Ap	ply Model					
Operators ×				uni V	mod					
apply model 🗙										
The Modeling (1)										
Time Series (1)										
Forecasting (1)									Hide advanced para	motore
Apply Forecast									a ride advanced para	inecers.
The second secon									 Change compatibilit 	<u>y (9.10.014)</u>
Apply Model										
									Help ×	
Get more operators from the									Apply Mod	el ^
Marketplace			🛕 Recomm	nendations unavailable	•				RapidMiner Stud	dio Core 🗸 🗸

Gambar 4.16 Apply Model

Gambar 4.16 menunjukkan proses *Apply Model* dalam *RapidMiner* setelah model *Naïve Bayes* selesai dilatih. Data pelatihan dihubungkan ke operator *naïve bayes*, sedangkan data pengujian dihubungkan ke *operator apply model*.



Gambar 4.17 Pengukuran Performance

Gambar ini menampilkan *workflow* lengkap di *RapidMiner* yang digunakan untuk mengukur kinerja (*performance*) model klasifikasi, dalam hal ini menggunakan algoritma *Naïve Bayes*. Operator performance mengukur kinerja hasil klasifikasi menggunakan akurasi. Hasil evaluasi performa akan ditampilkan setelah proses dijalankan sehingga dapat dianalisis seberapa baik model memprediksi kelas target pada data uji.

4.1.5 Implementasi Metode Support Vector Machine (SVM)

Bagian ini menjelaskan proses Implementasi Metode *Support Vector Machine* (SVM). Setelah implementasi dan evaluasi *Naïve Bayes*, penelitian ini melanjutkan dengan menerapkan algoritma SVM pada dataset yang sama untuk perbandingan kinerja.



Gambar 4.18 Implementasi SVM pada RapidMiner

Gambar 4.18 menampilkan proses yang serupa dengan implementasi Naïve Bayes, tetapi menerapkan metode SVM di dalam perangkat lunak machine learning yang digunakan (seperti RapidMiner). Pada proses penerapan SVM, sebelum melakukan normalisasi, dilakukan perubahan tipe data terlebih dahulu dengan menggunakan operator Nominal to Numerical. Algoritma SVM bekerja dengan konsep hyperplane di ruang fitur berdimensi tinggi. Untuk menghitung jarak, margin, dan memisahkan kelas, SVM membutuhkan input berupa data numerik (angka). Data kategorikal atau nominal tidak bisa langsung diproses karena tidak memiliki nilai numerik yang dapat dihitung secara matematis. Operator "Nominal to Numerical" mengubah atribut kategori (misalnya label kelas seperti "Rendah", "Sedang", "Tinggi") menjadi representasi angka. Dengan mengubah data nominal menjadi numerik, model SVM dapat melakukan perhitungan matematis yang diperlukan untuk pelatihan dan prediksi. Konversi ini

membantu model memahami pola dalam data secara lebih efektif dibandingkan jika data tetap dalam bentuk nominal. Operator Nominal to *Numerical* sangat penting saat menggunakan algoritma seperti SVM yang hanya menerima input numerik agar proses pelatihan dan prediksi berjalan lancar serta menghasilkan performa model yang optimal.

4.2 Hasil

Hasil dari masing-masing algoritma akan menunjukkan seberapa baik mereka dapat memprediksi minat pembeli *smartphone* berdasarkan fitur-fitur yang telah ditentukan.

🐒 PerformanceVector (Performance) 🛛 🗙			🚦 ExampleSet (//Local Repository/data/data_penjualan) 🛛 🛛 🛛					
Result History			📕 ExampleSet (Apply Model) 🛛 🛛					
%	Criterion accuracy	Table View Plot View	W				^	
Performance		accuracy: 85.71%					. 1	
			true Sedang	true Rendah	true Tinggi	class precision		
		pred. Sedang	4	1	0	80.00%		
Description		pred. Rendah	0	0	0	0.00%		
		pred. Tinggi	0	0	2	100.00%		
		class recall	100.00%	0.00%	100.00%			
Annotations								
		<					~	

Gambar 4.19 Hasil Performance Naïve Bayes

Gambar 4.19 di atas merupakan hasil evaluasi performa model *Naïve Bayes* yang ditampilkan dalam bentuk *confusion matrix* pada aplikasi *RapidMiner*. Akurasi 85.71% adalah metrik utama yang menunjukkan persentase prediksi yang benar dari model *Naïve Bayes* terhadap seluruh data uji. Artinya, model berhasil memprediksi minat pembeli dengan benar dalam 85.71% kasus. Bagian utama dari gambar ini adalah *confusion matrix*. *Confusion matrix* adalah tabel yang menunjukkan perbandingan antara prediksi model (kolom) dan nilai aktual (baris). Mari kita uraikan setiap sel:

a. True Sedang

Ini adalah baris yang menunjukkan kasus-kasus di mana minat pembeli sebenarnya adalah "Sedang". Model memprediksi 4 kasus dengan benar sebagai "Sedang". Model tidak salah memprediksi kasus "Sedang" sebagai "Rendah". Model tidak salah memprediksi kasus "Sedang" sebagai "Tinggi".

b. True Rendah

Ini adalah baris yang menunjukkan kasus-kasus di mana minat pembeli sebenarnya adalah "Rendah". Model salah memprediksi 1 kasus "Rendah" sebagai "Sedang". Model tidak memprediksi kasus "Rendah" dengan benar. Model tidak salah memprediksi kasus "Rendah" sebagai "Tinggi".

c. True Tinggi

Ini adalah baris yang menunjukkan kasus-kasus di mana minat pembeli sebenarnya adalah "Tinggi". Model tidak salah memprediksi kasus "Tinggi" sebagai "Sedang". Model tidak salah memprediksi kasus "Tinggi" sebagai "Rendah". Model memprediksi 2 kasus dengan benar sebagai "Tinggi".

Class Precision menunjukkan akurasi prediksi untuk setiap kelas. *Precision* untuk kelas "Sedang" adalah 80%, yang berarti dari semua prediksi yang diklasifikasikan sebagai "Sedang", 80% benar-benar "Sedang". *Precision* untuk kelas "Rendah" adalah 0% karena tidak ada prediksi benar untuk kelas ini. *Precision* untuk kelas "Tinggi" adalah 100%. *Class Recall* menunjukkan kemampuan model untuk mengidentifikasi semua contoh dari kelas tertentu. *Recall* untuk kelas "Sedang" adalah 100%, yang berarti model berhasil mengidentifikasi semua contoh "Sedang". *Recall* untuk kelas "Rendah" adalah 0%, dan recall untuk kelas "Tinggi" adalah 100%.

Class F1-Score sebesar 88,89% pada *naïve bayes* menunjukkan bahwa model memiliki keseimbangan yang baik antara *precision* dan *recall*. Nilai ini mengindikasikan bahwa model cukup efektif dalam mengklasifikasikan minat pembeli dengan akurasi tinggi, meskipun ada kemungkinan adanya beberapa kesalahan klasifikasi.

Result History	ExampleS	et (Apply Model) $ imes$	🐒 PerformanceVector (Performance) ×			
%	Criterion accuracy	Table View Plot View accuracy: 57.14%					
Performance							
			true Sedang	true Rendah	true Tinggi	class precision	
		pred. Sedang	4	1	2	57.14%	
Description		pred. Rendah	0	0	0	0.00%	
		pred. Tinggi	0	0	0	0.00%	
		class recall	100.00%	0.00%	0.00%		
Annotations							
							~

Gambar 4.20 Hasil Performance Support Vector Machine (SVM)

Gambar 4.20 menampilkan hasil evaluasi performa model *Support Vector Machine* (SVM) pada aplikasi *RapidMiner*. Akurasi 57.14% ini adalah metrik utama yang menunjukkan persentase prediksi yang benar dari model SVM terhadap seluruh data uji. Artinya, model hanya berhasil memprediksi minat pembeli dengan benar dalam 57.14% kasus. Bagian utama dari gambar ini adalah *confusion matrix. Confusion matrix* adalah tabel yang menunjukkan perbandingan antara prediksi model dengan nilai aktual. Mari kita analisis:

a. True Sedang

Model memprediksi benar 4 kasus minat pembeli "Sedang".

b. True Rendah

Model gagal memprediksi semua kasus minat pembeli "Rendah". Ini menunjukkan masalah dalam kemampuan model untuk mengidentifikasi minat rendah.

c. True Tinggi

Model gagal memprediksi semua kasus minat pembeli "Tinggi". Ini juga menunjukkan masalah serius dalam kemampuan model untuk mengidentifikasi minat tinggi.

Class precision menunjukkan akurasi prediksi untuk setiap kelas. Hanya kelas "Sedang" yang memiliki *precision* 57.14%, sedangkan kelas "Rendah" dan "Tinggi" memiliki *precision* 0%. Ini berarti ketika model memprediksi "Sedang", hanya 57.14% prediksinya yang benar. Prediksi "Rendah" dan "Tinggi" semuanya salah.

Class recall menunjukkan kemampuan model untuk mengidentifikasi semua contoh dari setiap kelas. *Recall* untuk kelas "Sedang" adalah 100%, artinya model berhasil mengidentifikasi semua contoh minat "Sedang" yang ada dalam data uji. Namun, *recall* untuk kelas "Rendah" dan "Tinggi" adalah 0%, yang berarti model sama sekali gagal mengidentifikasi contoh-contoh dari kelas tersebut.

Class F1-Score sebesar 73% pada SVM menunjukkan bahwa meskipun model SVM dapat memberikan beberapa prediksi yang akurat, ada ruang untuk perbaikan.

4.3 Pembahasan

Matriks evaluasi adalah alat penting untuk menilai kinerja model klasifikasi. Dalam konteks ini, kita akan membandingkan hasil dari dua algoritma *Naïve Bayes* dan *Support Vector Machine* (SVM).



Gambar 4.21 Perbandingan Akurasi Naïve Bayes dan SVM

Gambar 4.21 menampilkan perbandingan akurasi antara dua algoritma klasifikasi *Naïve Bayes* dan *Support Vector Machine* (SVM). Gambar di atas didapatkan dengan cara membandingkan hasil akurasi dari algoritma *Naïve Bayes* dan SVM. Gambar ini memberikan gambaran jelas tentang performa kedua algoritma klasifikasi pada dataset tertentu, di mana *Naïve Bayes* terbukti lebih unggul dengan tingkat akurasi sebesar 85,71% dibandingkan SVM sebesar 57,14%. *Precision* dan *recall* untuk *Naïve Bayes* lebih tinggi di semua kelas

dibandingkan SVM, maka dapat disimpulkan bahwa *Naïve Bayes* memberikan prediksi yang lebih andal. *F1-score Naïve Bayes* juga lebih tinggi, menunjukkan keseimbangan yang baik antara *precision* dan *recall*. Model ini tidak hanya mampu memprediksi dengan benar tetapi juga memiliki tingkat kesalahan yang rendah.

Berdasarkan hasil pengujian dan analisis yang telah dilakukan, dapat disimpulkan bahwa algoritma *Naïve Bayes* menunjukkan kinerja yang lebih baik daripada SVM dalam menentukan minat pembeli *smartphone* di Erafone Suzuya Baru Rantauprapat. Hal ini terlihat jelas dari nilai *akurasi, precision, recall,* dan *F1-score* yang secara konsisten lebih tinggi pada *Naïve Bayes*. Meskipun SVM memiliki keunggulan dalam menangani data kompleks, dalam konteks dataset ini, kesederhanaan dan kecepatan *Naïve Bayes* menghasilkan model yang lebih akurat dan efisien. Oleh karena itu, *Naïve Bayes* direksomendasikan sebagai algoritma yang lebih tepat untuk diterapkan dalam analisis minat pembeli smartphone di kasus studi ini.