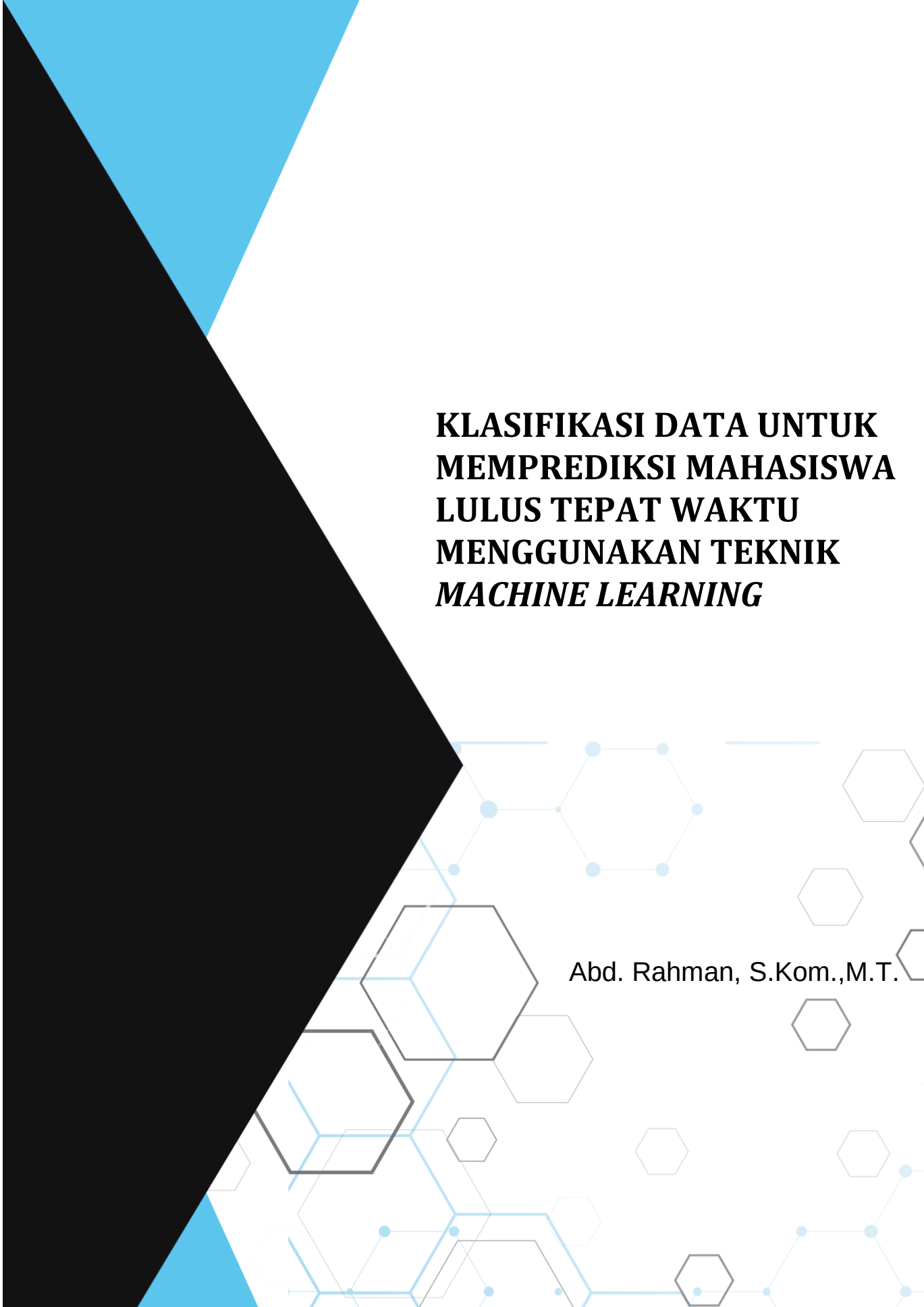




KLASIFIKASI DATA UNTUK MEMPREDIKSI MAHASISWA LULUS TEPAT WAKTU MENGUNAKAN TEKNIK *MACHINE LEARNING*

Abd. Rahman, S.Kom.,M.T.

**KLASIFIKASI DATA UNTUK
MEMPREDIKSI MAHASISWA LULUS
TEPAT WAKTU MENGGUNAKAN
TEKNIK *MACHINE LEARNING***

ABD. RAHMAN, S.KOM.,M.T.

KATA PENGANTAR

Perkembangan sistem komputer beserta produk pendukungnya dirasakan sangat pesat saat ini. Ada banyak hal yang bisa dilakukan dengan sistem komputer, termasuk menirukan perilaku manusia dalam menangani dan memecahkan berbagai permasalahan.

Diantara hal yang bisa dilakukan adalah klasifikasi dan prediksi. Dengan bantuan sistem komputer dalam hal ini teknik *machine learning*, diharapkan dapat membantu manusia dalam menyelesaikan persoalan klasifikasi dan prediksi dengan cepat dan akurat.

Buku ini memberikan wawasan bagi pembaca tentang bagaimana menggunakan teknik *machine learning* dalam melakukan klasifikasi data mahasiswa untuk keperluan prediksi kelulusan mahasiswa yang tepat waktu.

Demi penyempurnaan buku ini, penulis sangat mengharapkan kritik, saran, dan gagasan dari para pembaca, dan untuk itu diucapkan terima kasih.

Penulis,

DAFTAR ISI

Kata Pengantar	2
Daftar Isi	3
Daftar Tabel	4
Daftar Gambar	5
A. PENDAHULUAN	7
B. IMPLEMENTASI TEKNIK <i>MACHINE LEARNING</i> UNTUK KLASIFIKASI DAN PREDIKSI	18
C. PENUTUP	28
REFERENSI	29

DAFTAR TABEL

Tabel 1. Confusion Matrix	10
Tabel 2. Data Asli untuk Keperluan Pelatihan	20
Tabel 3. Data Asli untuk Keperluan Pengujian	21

DAFTAR GAMBAR

Gambar 1 Antarmuka aplikasi RapidMiner.	12
Gambar 2. Diagram Alir Algoritma Naïve Bayes Classifier	18
Gambar 3. Import Data Pelatihan dan Pengujian	22
Gambar 4. Naïve Bayes pada Kanvas Process	23
Gambar 5. Operator Apply Model pada Kanvas Process	23
Gambar 6. Operator Performance pada Kanvas Process	24
Gambar 7. Keterhubungan Antara Data dan Operator	24
Gambar 8. Tampilan Tab Example Set	25
Gambar 9. Tampilan Tab Simple Distribution dalam Bentuk Tabel ..	25
Gambar 10. Tampilan Tab PerformanceVector	26

A. PENDAHULUAN

1. Latar Belakang

Kelulusan tepat waktu bagi mahasiswa merupakan suatu keniscayaan dalam sebuah proses bisnis perguruan tinggi. Persentase jumlah mahasiswa yang lulus tepat waktu sangat mempengaruhi nilai akreditasi dari sebuah program studi. Sesuai dengan Pedoman Penilaian Pemantauan dan Evaluasi Peringkat Akreditasi dari BAN-PT, jika persentase antara jumlah lulusan tepat waktu dan jumlah mahasiswa baru pada tahun yang sama lebih besar atau sama dengan 50% maka diberi skor 4, sedangkan jika persentasenya dibawah 50% maka skornya adalah 1.

Oleh karena itu dibutuhkan berbagai langkah antisipatif agar persentase tersebut jangan sampai dibawah 50%. Salah satu upaya yang dapat dilakukan adalah mempersiapkan melakukan pemantauan terhadap potensi yang dapat menyebabkan mahasiswa lulus tidak tepat waktu. Bisa pula dengan memantau perkembangan capaian indeks prestasi setiap semester masing-masing mahasiswa. Dari sini bisa diambil gambaran apakah nantinya mahasiswa tersebut bisa lulus tepat waktu atau tidak.

Untuk yang berpotensi lulus tidak tepat waktu, selanjutnya bisa dilakukan pembinaan dan pendampingan dengan melibatkan dosen penasehat akademik, prodi dan pihak-pihak lain yang terlibat.

Kegiatan pemantauan terhadap potensi lulus tepat waktu atau tidak, dapat dilakukan dengan bantuan sistem

komputer. Dengan bantuan sistem komputer kegiatan tersebut dapat dilakukan dengan lebih efektif dan efisien. Bantuan sistem komputer dalam hal ini dapat dilakukan dengan menggunakan teknik *machine learning*.

2. *Machine Learning*

Secara harfiah, *machine learning* diambil dari dua kata dalam Bahasa Inggris, yakni *machine* yang berarti mesin, dan *learning* yang berarti sedang belajar atau dapat juga dikatakan dengan pembelajaran. Jika kedua kata tersebut itu disatukan, maka berarti mesin yang sedang belajar, atau pembelajaran mesin. Berarti bahwa, mesin melakukan pembelajaran (Hurwitz & Kirsch, 2018). Mesin disini, bisa diwakili oleh komputer melalui program aplikasi yang melakukan proses pembelajaran terhadap data untuk tujuan tertentu.

Mesin (program komputer) melakukan proses pembelajaran data untuk berbagai tujuan, antara lain pengelompokan (*clustering*), klasifikasi (*classification*), pengenalan pola (*pattern recognition*), memprediksi (*prediction*), meramal (*forecasting*), dan tugas-tugas lainnya sebagaimana manusia biasa melakukannya (Agarwal, 2013).

Mesin tidak serta merta bisa mempelajari data, seperti juga dengan manusia yang memerlukan beberapa teknik dan strategi dalam melakukan pembelajaran. Mesin juga serupa dengan itu. Dalam teknik *machine learning* ada beberapa teknik pembelajaran yang digunakan, yakni *supervised*

learning, unsupervised learning dan reinforcement learning.(Abhishek Vijayvargia, 2018)

Supervised learning atau pembelajaran terbimbing atau terawasi, yang dapat dianalogikan seperti guru mengarahkan muridnya dalam melakukan pembelajaran. Dimana murid melakukan tugas yang diberikan oleh guru, kemudian guru melakukan validasi kebenaran atas tugas tersebut. Jika tugas tersebut dianggap benar, maka ini menandakan murid tersebut mampu melaksanakan tugasnya. Sebaliknya, jika tugas tersebut dianggap salah, maka murid tersebut diarahkan untuk mengulang lagi tugasnya sampai benar. Dalam *machine learning* membutuhkan seperangkat data untuk dilakukan pengolahan dan data target sebagai pembimbing.

Unsupervised learning adalah teknik yang berkebalikan dengan *supervised learning*, dimana teknik ini tidak diawasi atau tidak dibimbing. Teknik ini cocok digunakan untuk data dengan volume yang sangat besar, beragam dan tidak berlabel. Sebagai contoh ialah data dari aplikasi media sosial seperti Twitter, Facebook, Instagram dan lain-lain. Karena besarnya volume data dari aplikasi tersebut, maka data tersebut tidak dapat diklasifikasikan secara langsung.

Reinforcement learning merupakan teknik pembelajaran dengan analogi mirip tingkah laku. Dalam prosesnya teknik ini melakukan observasi kemudian menjalankan mekanisme *try and error*. Untuk hasil yang dianggap benar, akan disimpan dan dijadikan penguat untuk pelatihan pembelajaran berikutnya.

Dalam penelitian ini, salah satu fungsi *machine learning* yang akan digunakan adalah klasifikasi. Klasifikasi adalah proses menemukan model (atau fungsi) yang menggambarkan dan membedakan kelas atau konsep data. Model diturunkan berdasarkan analisis sekumpulan data pelatihan (yaitu, objek data yang label kelasnya diketahui). Model yang digunakan untuk memprediksi label kelas objek yang label kelasnya tidak diketahui (Han et al., 2011). Tahapan klasifikasi data dibagi dalam 2 langkah proses. Langkah pertama adalah tahapan learning yang disebut sebagai fase training. Fase ini merupakan tahapan algoritma klasifikasi untuk membentuk aturan klasifikasi data yang diambil melalui analisis data training. Tahap kedua adalah klasifikasi, dimana data tes digunakan untuk memperkirakan akurasi dari aturan klasifikasi.

3. Naïve Bayes Classification

Naïve Bayes Classification merupakan metode probabilistik untuk membangun pengklasifikasi. Asumsi karakteristik dari pengklasifikasi Naïve Bayes Classification adalah untuk mempertimbangkan bahwa nilai fitur tertentu tidak tergantung pada nilai fitur lainnya, mengingat variabel kelas. (Viet et al., 2021)

Dalam klasifikasi dengan metode Naïve Bayes Classification, hal yang paling penting adalah menemukan probabilitas posterior yaitu probabilitas label yang diberikan beberapa fitur yang diamati. Dengan bantuan teorema Bayes,

kita dapat menyatakannya dalam bentuk persamaan berikut ini:

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

Dimana:

X : data dengan kelas yang belum diketahui

H : hipotesis data dengan kelas spesifik

$P(H|X)$: probabilitas H berdasarkan kondisi X (posterior probabilitas)

$P(H)$: probabilitas hipotesis H (prior probabilitas)

$P(X|H)$: probabilitas X berdasarkan kondisi hipotesis H

$P(X)$: probabilitas X

Atau dapat juga dilihat sebagai :

$$\text{Posterior} = \frac{\text{prior} \times \text{likelihood}}{\text{evidence}}$$

Nilai *evidence* selalu tetap untuk setiap kelas pada satu sampel. Nilai dari posterior tersebut nantinya akan dibandingkan dengan nilai-nilai posterior kelas lainnya untuk menentukan ke kelas apa suatu sampel akan diklasifikasikan. Sementara klasifikasi data yang bersifat kontinu dapat menggunakan formula densitas gaussian dengan formula dibawah ini :

$$P(X_i = x_i | Y = Y_j) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i - \mu)^2}{2(\sigma)^2}}$$

Dimana:

P : peluang

X_i : atribut ke-i

X_i : nilai atribut ke-i

Y : kelas yang dicari

μ : rerata seluruh atribut

σ : standar deviasi

Setelah proses klasifikasi selesai, dan telah menghasilkan hasil pengelompokan masing-masing kelas sesuai atributnya, selanjutnya adalah melakukan proses prediksi. Proses prediksi membutuhkan suatu uji validitas untuk menentukan hasil prediksi bersifat benar atau setidaknya dapat dipercaya. Uji validitas ini dinamakan juga dengan uji performa. Uji ini umumnya mengukur tingkat akurasi, presisi, dan sensitifitas dari kinerja prediksi. Hasilnya adalah sebuah matrik yang disebut dengan *confusion matrix*. Matrik ini terdiri dari tabel dua dimensi yakni “Aktual” dan “Diprediksi”, selanjutnya kedua dimensi ini akan menyisir pada kondisi *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), *False Negative* (FN). Lebih jelasnya seperti pada tabel dibawah.

		Nilai Kebenaran	
		1	0
Nilai Prediksi	1	TP	FP
	0	TN	FN

Tabel 1. *Confusion Matrix*

Dari tabel diatas, kebenaran dan prediksi diwakili dengan bilangan 1 dan 0. Bilangan 1 mewakili kebenaran, dan bilangan 0 mewakili kesalahan antara keadaan sebenarnya dengan prediksinya. Secara rinci penjelasan tabel 1 dijelaskan dibawah ini :

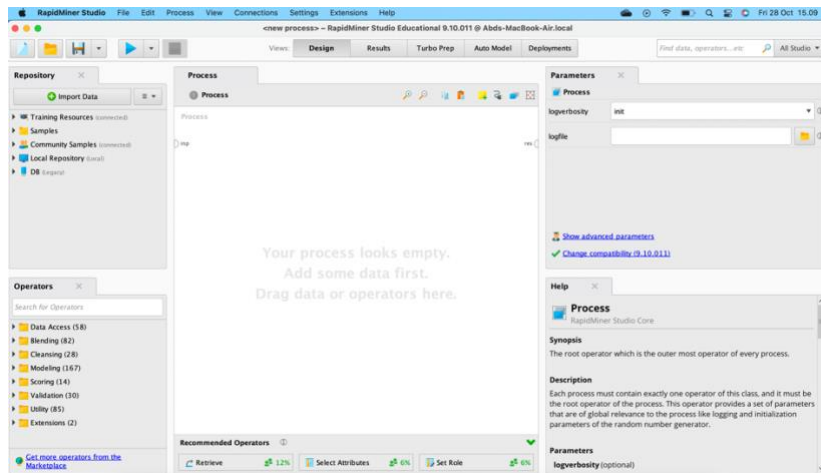
- *True Positives* (TP): Keadaan dimana kelas sebenarnya dan kelas prediksi dari titik data adalah 1.
- *True Negatives* (TN): Keadaan dimana ketika kelas sebenarnya dan kelas prediksi dari titik data adalah 0.
- *False Positive* (FP): Keadaan dimana kelas sebenarnya dari titik data adalah 0 dan kelas titik data yang diprediksi adalah 1.
- *False Negative* (FN): Keadaan dimana kelas sebenarnya dari titik data adalah 1 dan kelas prediksi titik data adalah 0.

Dari tabel 1 diatas dapat dihitung tingkat akurasi dari kinerja metodenya dengan menggunakan persamaan sebagai berikut

$$\text{Akurasi} = \frac{TP + TN}{TP + TN + FP + FN} * 100\%$$

4. RapidMiner Studio

Untuk keperluan klasifikasi dalam rangka memprediksi mahasiswa lulus tepat waktu dalam penelitian ini, digunakan alat bantu berupa program aplikasi RapidMiner. Merupakan perangkat lunak yang berfungsi untuk membantu melakukan analisis terhadap *data mining* dan *text mining* dalam kerangka *machine learning*. RapidMiner menyediakan berbagai fitur untuk analisis data, digunakan untuk tujuan klasterisasi, klasifikasi, prediksi, dan tujuan lainnya (Aprilla Donny et al, 2013). Adapun versi yang digunakan dalam penelitian ini adalah RapidMiner Studio versi 9.10.011 dengan lisensi resmi tipe *Educational Edition License*.



Gambar 1. Antarmuka aplikasi RapidMiner

5. Kelulusan Tepat Waktu

Jodi Letkiewicz, dkk. (2015) dalam laporan riset mereka menyampaikan bahwa mahasiswa yang banyak terlibat dengan lingkungan kampus, dalam hal ini adalah tinggal di lingkungan kampus merupakan faktor penentu kelulusan mahasiswa tepat waktu. Mereka beranggapan bahwa bilamana mahasiswa berada dekat dengan kampusnya, memungkinkan mereka akan lebih fokus belajar, serta mengakses berbagai fasilitas dan prasarana belajar (Letkiewicz et al, 2014).

Mereka merujuk kepada pernyataan Rektor Universitas Gordon Gee dari laporan riset Bush dan Pyle tahun 2011 dengan menyimpulkan bahwa memperluas area tempat tinggal mahasiswa di lingkungan kampus akan mengarah kepada “tingkat retensi dan kelulusan yang lebih tinggi, nilai yang lebih baik, peningkatan keterlibatan dalam kegiatan kampus, dan lebih banyak pendampingan di antara siswa” .

Pendapat Jodi bisa saja benar, namun itu hanya sedikit dari banyak faktor. Dan pada umumnya faktor-faktor penentu bisa bergantung kepada pola pikir dan keadaan masyarakat dari setiap negara. Sementara penulis sepakat dengan pendapat R.A. Pertiwi yang mengatakan bahwa faktor kelulusan dapat ditentukan dengan faktor belajar.(Anggun Pertiwi & Mohamad Afendi, 2013). Maksudnya disini adalah, dengan semangat belajar yang tinggi serta mengeksplorasi pengetahuan yang sesuai dengan bidang yang sedang diambil, apakah belajar mandiri maupun belajar bersama dengan sesama teman, baik diskusi maupun individu, serta belajar melalui sumber-sumber referensi maupun belajar melalui bimbingan di lembaga kursus dan sejenisnya, akan meningkatkan kompetensi berpikir dan kompetensi memecahkan masalah. Dengan meningkatnya kemampuan berfikir dan memecahkan masalah, akan mendorong mahasiswa cepat lulus karena kemampuannya dalam menghadapi berbagai penugasan dalam kegiatan perkuliahan.

Dalam kbbi.web.id kata lulus diartikan dengan, dapat masuk, dapat lepas, terperosok, dan berhasil. Dalam konteks pendidikan, “berhasil” merupakan kata yang tepat untuk mendefinisikan seorang mahasiswa yang keluar dari perguruan tinggi karena telah melalui proses pendidikan. Dengan kata lain seorang mahasiswa dibiarkan bebas untuk keluar dari perguruan tinggi dengan gelar sarjana karena telah menyelesaikan semua tugas-tugas perkuliahan, baik

ketika belajar pada mata kuliah umum, dasar kejuruan mata kuliah konsentrasi hingga tugas akhir.

Kelulusan setiap mahasiswa, untuk setiap perguruan tinggi memiliki aturan tersendiri, secara umum adalah telah menyelesaikan beban mata kuliah. Syarat kelulusan di IAIN Palopo menetapkan beban mata kuliah sebesar 144 Satuan Kredit Semester (SKS), termasuk diantaranya Praktek Pengalaman Lapangan, Kuliah Kerja Nyata (KKN), seminar proposal dan seminar hasil (skripsi).

Secara normal, kelulusan sarjana adalah 8 semester atau 4 tahun pendidikan. Namun, tidak menutup kemungkinan dapat ditempuh sampai 3,5 tahun masa studi. Kelulusan mahasiswa yang dicapai paling lama 4 tahun, dipandang sebagai mahasiswa lulus tepat waktu. Untuk mahasiswa yang lulus lebih dari 4 tahun dipandang sebagai kelulusan yang terlambat atau tidak tepat waktu. Semua mahasiswa memiliki peluang untuk lulus tepat waktu dan tidak tepat waktu bergantung kepada kemampuan sikap, pengetahuan dan keterampilan yang dimilikinya.

6. Riset Terkait

Penelitian tentang *machine learning* telah banyak dilakukan oleh para peneliti. Beberapa diantaranya yang cukup relevan dengan topik penelitian ini diuraikan seperti berikut.

Azis Zed Ali Muhammad Bajabir melakukan penelitian tentang penerapan metode Naïve Bayes untuk melakukan

prediksi penentuan karyawan tetap. Dimana hasil penelitiannya mengemukakan bahwa metode tersebut memiliki akurasi sebesar 94% dalam melakukan prediksi berdasarkan atribut absensi, kepribadian dan prestasi (A. Zed & A.L.I. Muhammad, 2018).

Metode yang sama juga diteliti oleh Devi Hedryana dengan akurasi sebesar 74,6% untuk data *training* dan 77,2% untuk data *testing* .(Devi Hedryana, n.d.)

Sementara Erene Fajrila melakukan perbandingan penggunaan metode regresi logistik biner dan Naïve Bayes Classifier dalam mengklasifikasi ketepatan waktu kelulusan (Erene Fajrila, 2018).

Penelitian tentang prediksi kelulusan tepat waktu juga dilakukan oleh Susi Mashlahah, metode yang digunakan adalah *Decision Tree*, dimana akurasi prediksi yang dihasilkan adalah 82,7% (Susi Mashlahah, 2013).

Metode jaringan saraf tiruan *backpropagation* diteliti penggunaannya oleh Fahmi Izhari dengan akurasi mencapai 85,37%. Objek penelitiannya sendiri adalah prediksi kelulusan mahasiswa. (Rudy Ansari, 2016)

Rindy Anggun Pertiwi, Indahwati, Farit Mohammad Efendi melakukan analisis CHAID untuk identifikasi kelulusan tepat waktu mahasiswa dalam penelitiannya, kesimpulannya adalah setiap mahasiswa yang memiliki IPK $\geq 3,39$ dapat lulus dari fakultas ekonomi manajemen, sedangkan IPK 2,77 dari MIPA juga dapat lulus (Anggun Pertiwi & Mohamad Afendi, 2013).

Dari beberapa referensi penelitian diatas, maka peneliti mencoba untuk mengimplementasikan salah satu teknik dalam *machine learning* yaitu algoritma Naïve Bayes Classifier untuk melakukan klasifikasi untuk memprediksi kelulusan mahasiswa tepat waktu. Diharapkan dari penelitian ini dapat diperoleh hasil prediksi secara efektif dan akurat.

B. IMPLEMENTASI TEKNIK *MACHINE LEARNING* UNTUK KLASIFIKASI DAN PREDIKSI

Kegiatan penelitian ini dilakukan dengan dasar pengalaman dan pengamatan di lapangan, bahwa selama ini rasio antara lulusan dan penerimaan mahasiswa baru memiliki selisih yang cukup besar, yakni masih di bawah 50%. Sementara salah satu indikator penilaian dalam akreditasi program studi dan institusi adalah kelulusan mahasiswa tepat waktu. Jika kurang dari 50% skornya adalah 1, sedangkan jika lebih atau sama dengan 50% skornya adalah 4. Satu indikator dengan skor rendah sudah pasti akan mempengaruhi penilaian akreditasi secara keseluruhan.

Berangkat dari hal tersebut, maka peneliti mencoba untuk melakukan penelitian tentang bagaimana melakukan prediksi kelulusan mahasiswa tepat waktu dengan bantuan teknik *machine learning*. Diharapkan dengan tersedianya hasil prediksi tersebut, dapat diambil langkah-langkah antisipatif terhadap mahasiswa yang masuk ke kategori lulus tidak tepat waktu, supaya dapat pula lulus tepat waktu, atau paling tidak bisa lulus tidak terlalu jauh dari kategori lulus tepat waktu.

Diharapkan dengan selesainya penelitian ini nanti, dapat memberikan sumbangsih terhadap khazanah ilmu pengetahuan dan implementasinya dalam kehidupan bermasyarakat.

Dalam menyusun penelitian ini tentunya didukung oleh berbagai sumber referensi seperti buku, artikel dari berbagai jurnal, karya ilmiah populer dan lain sebagainya.

Metode penelitian adalah rencana dan struktur penyelidikan yang digunakan untuk memperoleh bukti-bukti empiris dalam menjawab pertanyaan penelitian. Tahapan-tahapan yang akan dilakukan dalam penelitian ini adalah sebagai berikut :

1. Tahapan penelitian

- a. Identifikasi masalah
- b. Perumusan masalah
- c. Rancangan penelitian
- d. Pengumpulan data

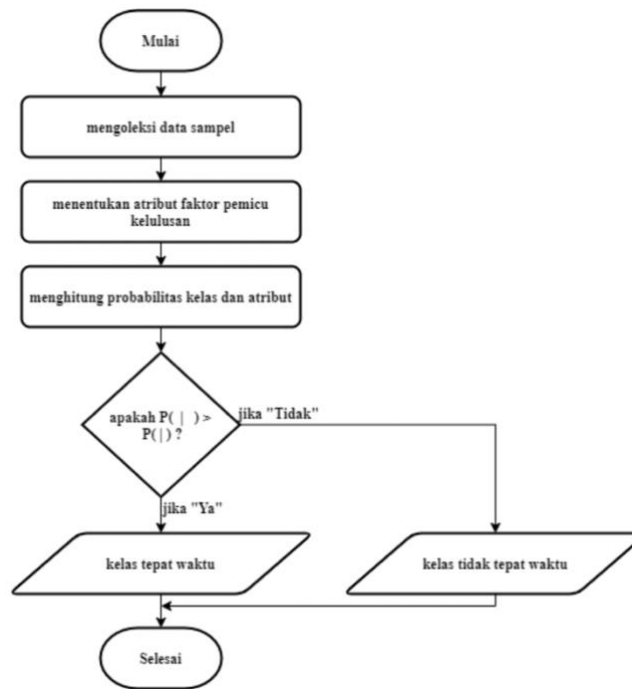
2. Analisis data

Analisis data dilakukan setelah tahapan penelitian dilaksanakan. Di dalam analisis data ini yang dilakukan dalam penelitian adalah menganalisis data yang telah terkumpul dari tahapan penelitian sebelumnya.

3. Pengolahan data

Setelah data dianalisis, selanjutnya adalah pengolahan data tersebut. Data diolah dalam bentuk pengklasifikasian menggunakan algoritma Naïve Bayes Classifier. Klasifikasi data dibagi menjadi dua, yakni mahasiswa lulus tepat waktu dan mahasiswa lulus tidak tepat waktu.

Cara kerja algoritma Naïve Bayes Classifier dalam mengklasifikasikan data mahasiswa lulus tepat waktu dapat digambarkan dalam diagram alir seperti berikut.



Gambar 2. Diagram Alir Naïve Bayes Classifier

Setelah proses klasifikasi data selesai dilakukan, selanjutnya adalah proses prediksi. Algoritma akan melakukan prediksi berdasarkan data pengujian dan data pelatihan.

4. Uji validitas

Pada tahap ini, dilakukan proses lanjutan dari hasil klasifikasi dan prediksi pada tahap selanjutnya. Tahap ini akan menguji seberapa besar tingkat akurasi prediksi yang berhasil dicapai.

Penelitian ini menggunakan data yang bersumber dari data lulusan mahasiswa angkatan 2016, 2017 dan 2018. Lulusan tersebut berasal dari program studi Perbankan Syariah dan Ekonomi Syariah Fakultas Ekonomi dan Bisnis Islam IAIN Palopo.

Untuk data pelatihan dipilih sebanyak 40 mahasiswa, sementara untuk data pengujian dipilih sebanyak 10 mahasiswa.

Keseluruhan data tersebut tersaji seperti pada tabel 2 dan 3 dibawah ini.

NO	IPS1	IPS2	IPS3	IPS4	SKS4	MASA STUDI	STATUS
1	3,72	3,58	3,48	3,66	91,00	45	TW
2	3,65	3,40	3,55	3,40	91,00	45	TW
3	3,91	3,92	3,77	3,67	91,00	45	TW
4	3,75	3,27	3,36	3,59	91,00	45	TW
5	3,75	3,63	3,67	3,77	91,00	45	TW
6	3,81	3,61	3,65	3,77	91,00	45	TW
7	3,63	3,58	3,68	3,72	91,00	45	TW
8	3,91	4,00	4,00	3,91	91,00	43	TW
9	3,52	3,38	3,34	3,26	92,00	47	TW
10	3,63	3,64	3,59	3,44	92,00	48	TW
11	3,58	3,37	3,43	3,32	92,00	47	TW
12	3,45	3,32	3,37	3,26	92,00	48	TW
13	3,45	3,37	3,48	3,35	92,00	45	TW
14	3,38	3,55	3,60	3,49	92,00	45	TW
15	3,67	3,59	3,56	3,45	92,00	46	TW
16	3,64	3,34	3,32	2,97	92,00	67	TTW
17	3,57	3,45	3,28	2,80	92,00	67	TTW
18	3,07	3,17	3,05	2,78	90,00	66	TTW
19	3,45	2,98	3,18	2,95	90,00	67	TTW
20	3,36	2,56	3,07	2,99	85,00	68	TTW
21	3,64	3,34	3,32	2,97	92,00	46	TW
22	3,11	3,21	3,05	2,78	90,00	69	TTW
23	3,75	3,27	3,36	3,59	91,00	45	TW
24	3,81	3,51	3,65	3,67	91,00	45	TW
25	3,58	3,37	3,43	3,32	92,00	47	TW
26	3,75	3,27	3,36	3,59	91,00	45	TW
27	3,64	3,34	3,32	2,97	92,00	67	TTW
28	3,54	3,29	3,71	3,02	92,00	66	TTW
29	2,68	3,08	2,46	2,66	84,00	68	TTW

30	3,45	3,32	3,37	3,26	92,00	48	TW
31	3,61	3,59	3,51	3,38	92,00	48	TW
32	3,69	3,73	3,74	3,75	92,00	47	TW
33	3,65	3,61	3,54	3,65	94,00	45	TW
34	3,75	3,68	3,55	3,76	94,00	45	TW
35	3,75	3,72	3,71	3,75	94,00	43	TW
36	3,54	3,74	3,48	3,73	87,00	47	TW
37	3,58	3,73	3,46	3,66	87,00	47	TW
38	3,46	3,72	3,38	3,57	87,00	47	TW
39	3,58	3,65	3,36	3,67	89,00	43	TW
40	3,54	3,93	3,75	3,80	89,00	43	TW

Tabel 2. Data asli untuk keperluan pelatihan

NO	IPS1	IPS2	IPS3	IPS4	SKS4	MASA STUDI	STATUS
1	0,58	3,04	3,40	2,99	74	67	TTW
2	3,54	3,29	3,71	3,02	92	66	TTW
3	2,68	3,08	2,46	2,66	84	68	TTW
4	3,14	3,15	3,45	2,95	92	68	TTW
5	2,00	1,60	3,02	2,49	70	68	TTW
6	3,70	3,73	3,75	3,69	92	42	TW
7	2,82	2,73	2,86	2,88	90	48	TW
8	3,61	3,42	3,43	3,34	92	48	TW
9	3,54	3,50	3,58	3,49	92	48	TW
10	3,61	3,59	3,51	3,38	92	48	TW

Tabel 3. Data asli untuk keperluan pengujian

Untuk mengetahui lebih jelas perihal data pada tabel diatas, berikut ini diuraikan penjelasannya.

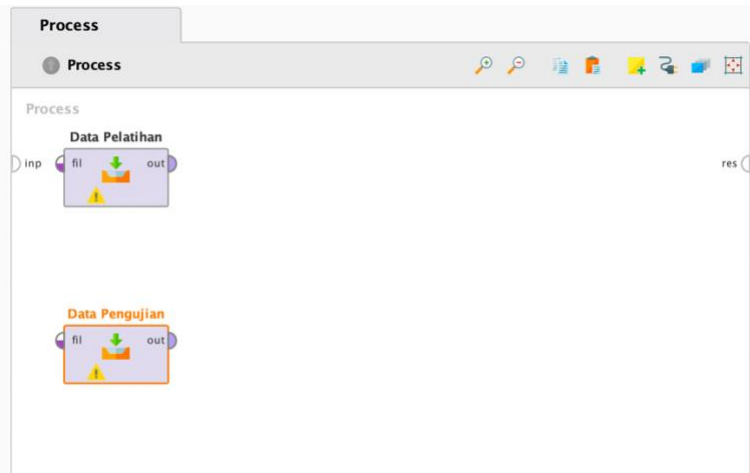
- 1) IPS, merupakan singkatan dari Indeks Prestasi Semester, dalam data diatas merupakan data dari semester 1 sampai 4, diwakili dengan singkatan IPS1, IPS2, IPS3, dan IPS4.
- 2) SKS4, merupakan jumlah SKS yang telah dikumpulkan oleh mahasiswa pada semester 4.

- 3) Masa Studi, yakni masa studi yang ditempuh oleh mahasiswa sampai memperoleh gelar sarjana, dihitung dalam satuan bulan.
- 4) Status, merupakan status kelulusan mahasiswa. TW untuk Tepat Waktu, dan TTW untuk Tidak Tepat Waktu.

Selain data tersebut diatas, terdapat pula data NIM, Nama, dan Angkatan, tetapi data tersebut bukan merupakan pemicu kelulusan dan tidak akan dihitung dalam proses klasifikasi, oleh karena itu data tersebut tidak dimasukkan dalam tabel data sampel.

Data tersebut diatas, kemudian dilakukan proses klasifikasi dan prediksi menggunakan algoritma Naïve Bayes Classifier yang dijalankan dengan bantuan aplikasi RapidMiner. Kelebihan RapidMiner dalam melakukan analisis data diskrit dan numerik, membuat aplikasi ini populer dikalangan peneliti. Kelebihannya juga didukung oleh fungsi-fungsi dari metode-metode analisis data serta kemudahannya dalam menggunakannya. Model prediksi kelulusan tepat waktu dan tidak tepat waktu mahasiswa dalam penelitian ini menggunakan RapidMiner Studio versi 9.10.011 yang dioperasikan menggunakan komputer laptop dengan spesifikasi, CPU Apple M1, RAM 8 GB, SSD 256GB.

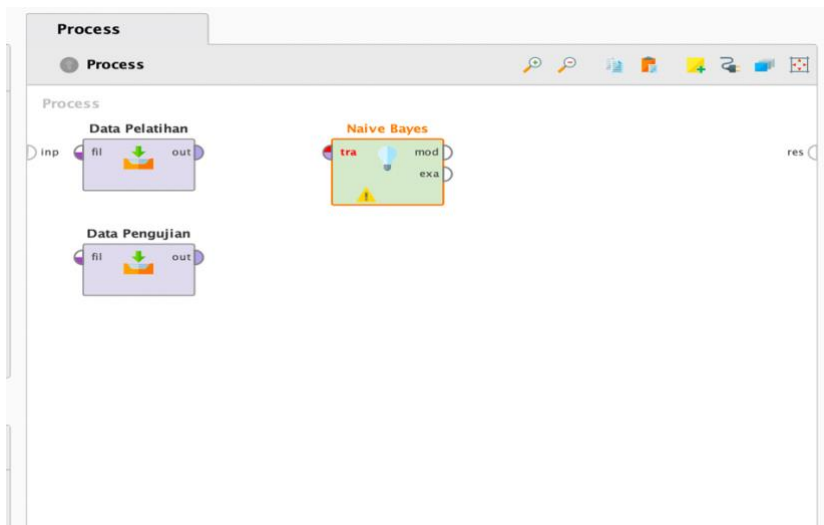
Tahap pertama data pelatihan dan pengujian diimport terlebih dahulu ke dalam kanvas process di lembar kerja RapidMiner.



Gambar 3. Import Data Pelatihan dan Pengujian

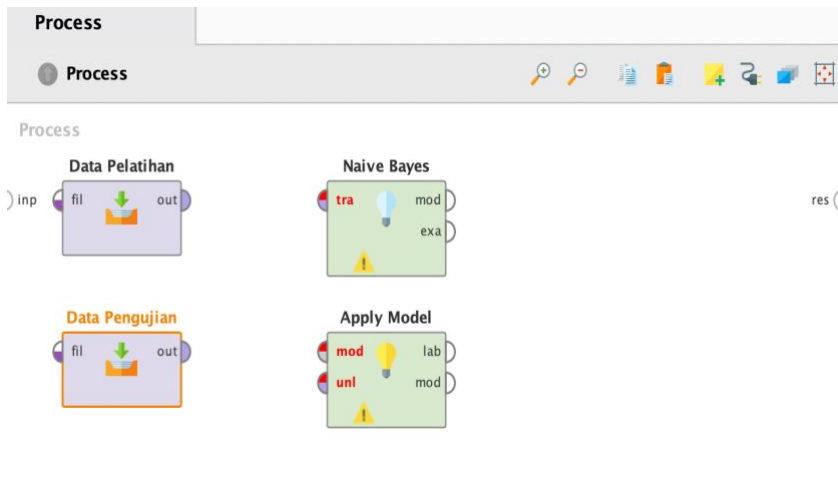
Saat import data, ada tampilan *Import Configuration Wizard*, pastikan kolom Status di ubah role nya menjadi Label. Ini maksudnya adalah kolom status dinyatakan sebagai patokan untuk klasifikasi.

Selanjutnya masukkan operator Naïve Bayes Classifier ke dalam kanvas process.

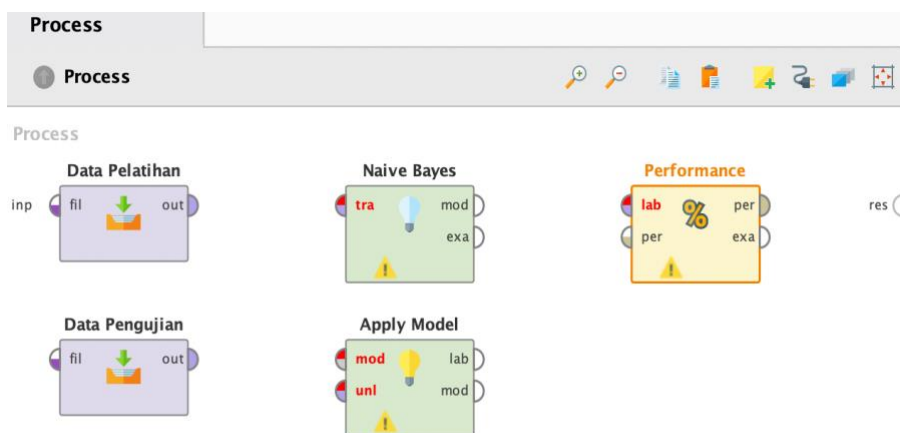


Gambar 4. Naïve Bayes pada Kanvas Process

Setelah itu masukkan pula operator Apply Model

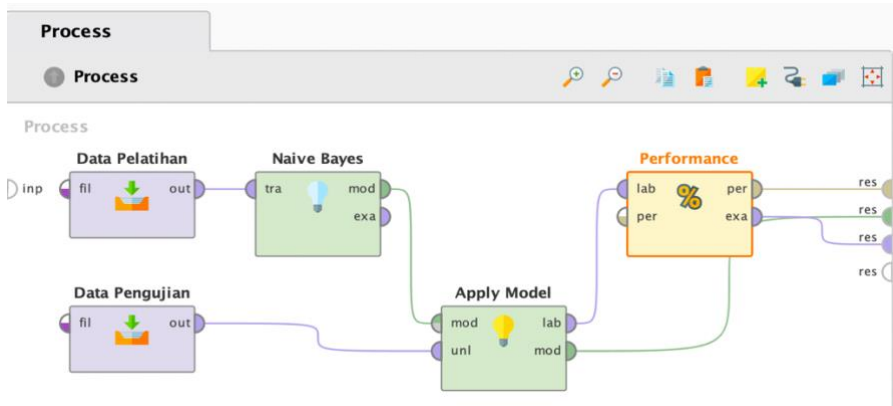


Gambar 5. Operator Apply Model pada Kanvas Process
Kemudian masukkan pula operator Performance untuk keperluan pengujian.



Gambar 6. Operator Performance pada Kanvas Process
Setelah semua data diimport dan operator dimasukkan kedalam kanvas process, selanjutnya menghubungkan kesemuanya. Dimulai dari menghubungkan Data Pelatihan dengan operator Naïve Bayes, selanjutnya dari Naïve Bayes dihubungkan ke operator Apply Model, kemudian dari Apply Model menuju ke

operator Performance, dan dari Performance menuju ke result. Untuk data Pengujian, dihubungkan ke operator Apply Model, dan dari Apply Model langsung menuju ke result.



Gambar 7. Keterhubungan Antara Data dan Operator

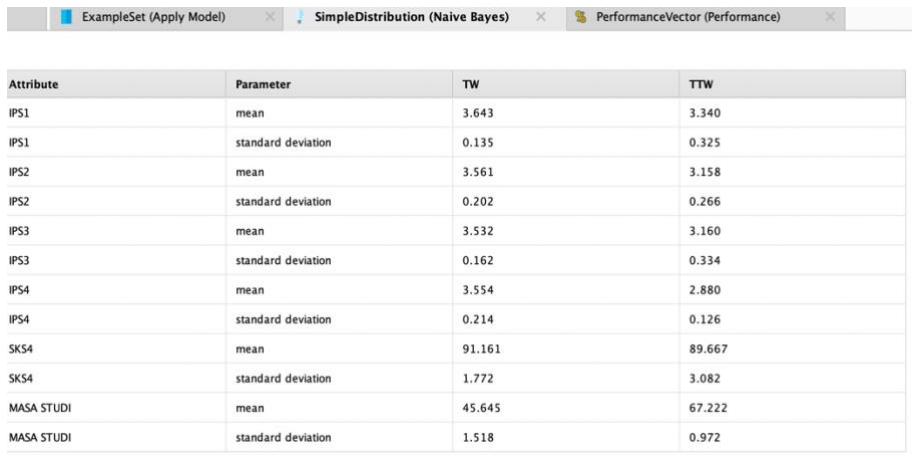
Setelah semuanya terhubung, selanjutnya adalah menjalankan keseluruhan proses dengan cara menekan tombol play. Jika tidak ada masalah, seharusnya akan tampil tiga tab, yakni Example Set (Apply Model), Simple Distribution (Naïve Bayes) dan Performance Vector (Performance).

Pada tab Example set akan menampilkan tabel dengan salah satu kolomnya ada hasil prediksi dari kolom Status yang diset rolenya menjadi label.

ExampleSet (Apply Model) × SimpleDistribution (Naive Bayes) × PerformanceVector (Performance) ×										
Open in Turbo Prep Auto Model		Filter (10 / 10 examples): all								
Row No.	STATUS	prediction...	confidence...	confidence...	IPS1	IPS2	IPS3	IPS4	SKS4	M/
1	TTW	TTW	0	1	0.580	3.040	3.400	2.990	74	67
2	TTW	TTW	0	1	3.540	3.290	3.710	3.020	92	66
3	TTW	TTW	0	1	2.680	3.080	2.460	2.660	84	68
4	TTW	TTW	0	1	3.140	3.150	3.450	2.950	92	68
5	TTW	TTW	0	1	2	1.600	3.020	2.490	70	68
6	TW	TW	1	0	3.700	3.730	3.750	3.690	92	42
7	TW	TW	1	0	2.820	2.730	2.860	2.880	90	48
8	TW	TW	1	0	3.610	3.420	3.430	3.340	92	48
9	TW	TW	1	0	3.540	3.500	3.580	3.490	92	48
10	TW	TW	1	0	3.610	3.590	3.510	3.380	92	48

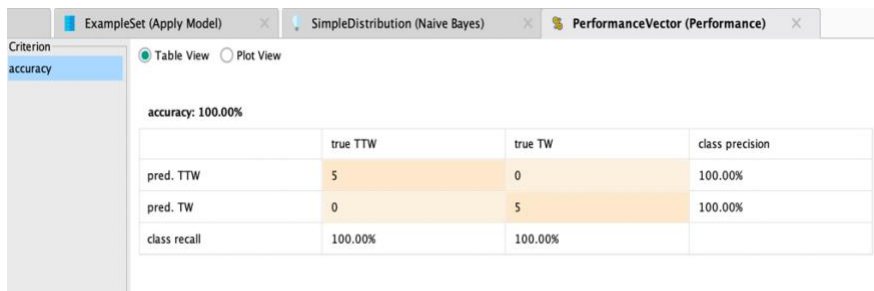
Gambar 8. Tampilan Tab Example Set

Pada tab Simple Distribution akan menampilkan hasil klasifikasi data. Bisa divisualisasikan dalam bentuk grafik atau tabel.



Attribute	Parameter	TW	TTW
IPS1	mean	3.643	3.340
IPS1	standard deviation	0.135	0.325
IPS2	mean	3.561	3.158
IPS2	standard deviation	0.202	0.266
IPS3	mean	3.532	3.160
IPS3	standard deviation	0.162	0.334
IPS4	mean	3.554	2.880
IPS4	standard deviation	0.214	0.126
SKS4	mean	91.161	89.667
SKS4	standard deviation	1.772	3.082
MASA STUDI	mean	45.645	67.222
MASA STUDI	standard deviation	1.518	0.972

Gambar 9. Tampilan Tab Simple Distribution dalam Bentuk Tabel
 Pada tab PerformanceVector, akan tersaji hasil akurasi atas prediksi yang telah dilakukan



Criterion: accuracy

Table View Plot View

accuracy: 100.00%

	true TTW	true TW	class precision
pred. TTW	5	0	100.00%
pred. TW	0	5	100.00%
class recall	100.00%	100.00%	

Gambar 10. Tampilan Tab PerformanceVector

Bisa dilihat dari hasil proses diatas, terlihat akurasi prediksi mencapai 100%, dengan kata lain algoritma Naïve Bayes Classifier dapat memprediksi status kelulusan pada data pengujian dengan benar semuanya. Terdapat 5 (lima) dengan status lulus tepat waktu, dan 5 (lima) pula dengan status lulus tidak tepat waktu.

C. PENUTUP

1. Kesimpulan

Berdasarkan penelitian yang telah dilakukan, maka dapat disimpulkan bahwa salah satu teknik dalam *Machine learning* yakni algoritma Naïve Bayes Classifier mampu melakukan klasifikasi data untuk memprediksi kelulusan mahasiswa tepat waktu dengan akurat.

2. *Future Work*

- a. Mencoba menggunakan metode jenis *unsupervised learning* untuk kasus prediksi kelulusan mahasiswa tepat waktu. Karena pada penelitian ini masih menggunakan metode jenis *supervised learning* yang mana status kelulusan sudah ditentukan terlebih dahulu, sehingga bisa digunakan untuk melakukan prediksi terhadap mahasiswa yang masih aktif.
- b. Mencoba mengkombinasikan algoritma Naïve Bayes Classifier dengan algoritma lainnya dalam menyelesaikan kasus sejenis.

REFERENSI

- A. Zed, & A.L.I. Muhammad. (2018). *Penerapan metode naive bayes untuk prediksi menentukan karyawan tetap pada pt. ysp industries Indonesia*.
- Abhishek Vijayvargia. (2018). *Machine Learning with Python: An Approach to Applied Machine Learning*. BPB Publication.
- Agarwal, S. (2013). Data Mining: Data Mining Concepts and Techniques. 2013 *International Conference on Machine Intelligence and Research Advancement*, 203–207. <https://doi.org/10.1109/ICMIRA.2013.45>
- Anggun Pertiwi, R., & Mohamad Afendi, F. (2013). *ANALISIS CHAID UNTUK IDENTIFIKASI KETEPATAN WAKTU LULUS BERDASARKAN KARAKTERISTIK MAHASISWA* (Vol. 2, Issue 1).
- Aprilla Donny Aji Baskoro Lia Ambarwati I Wayan Simri Wicaksana Editor, D. C., & Sanjaya, R. (n.d.). *Belajar Data Mining dengan RapidMiner*.
- Devi Hedryana. (n.d.). *DATA MINING UNTUK MEMPREDIKSI KELULUSAN MAHASISWA PENDIDIKAN MATEMATIKA UIN RADEN INTAN LAMPUNG MENGGUNAKAN NAIVE BAYES*.
- Erene Fajrila. (2018). *Perbandingan Klasifikasi Ketepatan Waktu Kelulusan Mahasiswa Menggunakan Regresi Logistik Biner dan Naive Bayes Classifier*. Ull.
- Han, J., Kamber, M., & Pei, J. (2011). *Data Mining. Concepts and Techniques, 3rd Edition (The Morgan Kaufmann Series in Data Management Systems)*.
- Hurwitz, J., & Kirsch, D. (2018). *Machine Learning IBM Limited Edition*. <http://www.wiley.com/go/permissions>.
- Letkiewicz, J., Lim, H., Heckman, S., Bartholomae, S., Fox, J. J., & Montalto, C. P. (2014). The Path to Graduation: Factors Predicting On-Time Graduation Rates. *Journal of College Student Retention: Research, Theory & Practice*, 16(3), 351–371. <https://doi.org/10.2190/CS.16.3.c>
- Rudy Ansari. (2016). *PREDIKSI KELULUSAN MAHASISWA DENGAN JARINGAN SYARAF TIRUAN. Jurnal Teknologi Informasi Universitas Lambung Mangkurat (JTIULM)*, 1(1), 18–23. <https://doi.org/10.20527/jtiulm.v1i1.4>

Susi Mashlahah. (2013). *PREDIKSI KELULUSAN MAHASISWA MENGGUNAKAN METODE DECISION TREE DENGAN PENERAPAN ALGORITMA C4.5.*

Viet, T. N., Minh, H. le, Hieu, L. C., & Anh, T. H. (2021). THE NAÏVE BAYES ALGORITHM FOR LEARNING DATA ANALYTICS. *Indian Journal of Computer Science and Engineering*, 12(4), 1038–1043.
<https://doi.org/10.21817/indjcse/2021/v12i4/211204191>

KLASIFIKASI DATA UNTUK MEMPREDIKSI MAHASISWA LULUS TEPAT WAKTU MENGUNAKAN TEKNIK *MACHINE LEARNING*

Perkembangan sistem komputer beserta produk pendukungnya dirasakan sangat pesat saat ini. Ada banyak hal yang bisa dilakukan dengan sistem komputer, termasuk menirukan perilaku manusia dalam menangani dan memecahkan berbagai permasalahan.

Diantara hal yang bisa dilakukan adalah klasifikasi dan prediksi. Dengan bantuan sistem komputer dalam hal ini teknik *machine learning*, diharapkan dapat membantu manusia dalam menyelesaikan persoalan klasifikasi dan prediksi dengan cepat dan akurat.

Buku ini memberikan wawasan bagi pembaca tentang bagaimana menggunakan teknik *machine learning* dalam melakukan klasifikasi data mahasiswa untuk keperluan prediksi kelulusan mahasiswa yang tepat waktu.