

BAB III

HASIL DAN PEMBAHASAN

3.1 Pengumpulan Data

Proses pengumpulan data yang diambil langsung dari Program Studi Teknik Informatika UMKT berupa data jumlah kelulusan mahasiswa baru angkatan 2017 – 2020 dan data akademik dari semester 1 sampai semester 7 angkatan 2017 – 2020. Data tersebut dapat divisualisasikan dalam bentuk Tabel 3.1 dan 3.2.

Tabel 3.1 Data Akademik

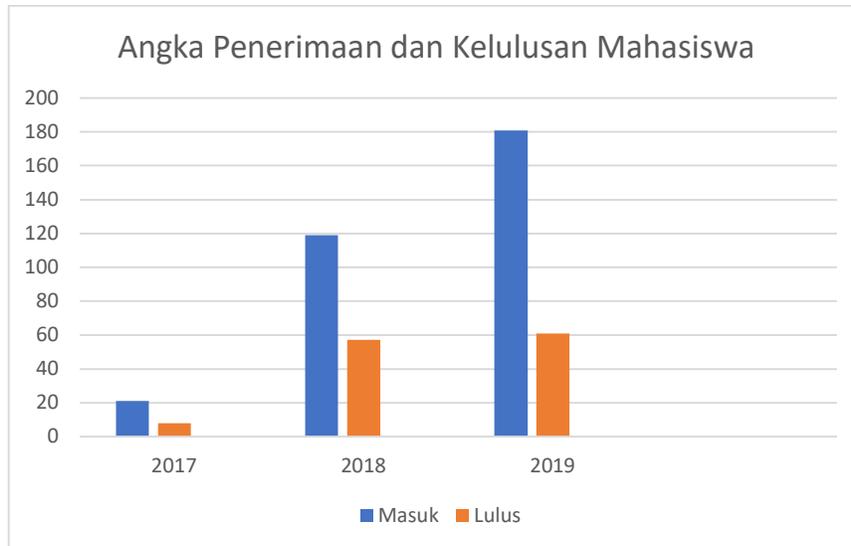
NIM	NAMA	Prodi	SKS 1	IPS 1	...	IPS 7	Status Mahasiswa
17111024410001	ABDUL BASIR SALSABILA	TI	20.0	3.325	...	3.7941	TEPAT
17111024410002	AGUS SASWONO	TI	20.0	3.45	...	3.5882	TEPAT
1811102441004	ANDI FIRANTI	TI	20.0	3.02	...	2.11	DIKELUARKAN
...
1911102441186	ARIA HADI PRADANA	TI	20.0	3.3		3.4545	AKTIF

Pada proses pengumpulan data akademik (Tabel 3.1) yang diperoleh dari Prodi Teknik Informatika yang memiliki atribut data yaitu NIM, Nama, SKS (Sistem Kredit Semester), IPS (Indeks Prestasi Semester), dan Status Mahasiswa.

Tabel 3.2 Data Yudisium

NIM	NAMA	Tempat Tanggal Lahir	Tahun Masuk	IPK	Status Kelulusan
17111024420002	Agus Saswono	Samarinda, 19 Agustus 1997	2017	3.48	Lulus
17111025510003	Andi Devina Safira	Muara Jawa, 7 Agustus 1999	2017	4	Lulus
1811102441001	Anton Saputra	Samarinda, 16 September 1999	2017	4	Lulus
...
1911102441167	Ade Salwa	Tarakan, 25 Mei 2001	2019	3.58	Lulus

Pada proses pengumpulan data yudisium (Tabel 3.2) yang diperoleh dari Prodi Teknik Informatika melalui SK Yudisium pada angkatan 2017 – 2019. Yang dimana, pada data ini mencakup mahasiswa yang sudah lulus.



Gambar 3.1 Data Penerimaan dan Kelulusan Mahasiswa

Pada gambar 3.1 merupakan perbandingan antara penerimaan dan kelulusan mahasiswa pada angkatan 2017 – 2019. Yang dimana, pada angkatan 2017 mahasiswa masuk berjumlah 21 orang dan lulus berjumlah 8 orang, pada angkatan 2018 mahasiswa masuk 119 orang dan lulus berjumlah 57 orang, pada angkatan 2019 mahasiswa masuk 181 orang dan lulus 61 orang.

3.2 Data Preprocessing

Pada tahap ini *preprocessing* ini ialah tahap memproses data menjadi lebih berkualitas sehingga data yang dihasilkan menjadi lebih akurat yaitu dengan cara memisahkan data yang tidak perlu dan menggabungkan kedua data yang diperoleh dari Prodi Teknik Informatika, serta mengeliminasi beberapa atribut yang tidak diperlukan. *Data preprocessing* dapat dibagi menjadi 3 tahap yaitu *Data Merging*, *Data Cleaning*, *Data Normalization*.

3.2.1 Data Merging

Pada proses ini dilakukan penggabungan data antara dua data yang berbeda dari Prodi Teknik Informatika. Pada proses penggabungan data, mahasiswa yang tidak masuk atau tidak memiliki data pada salah satu IPS dan SKS akan di eliminasi secara otomatis. Setelah

menggabungkan data, selanjutnya adalah menghapus beberapa atribut yang tidak dipakai yaitu Nama Mahasiswa, Tempat Tanggal Lahir, Program Studi, Status Mahasiswa, dan IPK.

Tabel 3.3 Penggabungan kedua data dan mengeliminasi atribut tidak diperlukan.

NIM	Kelulusan	SKS 1	IPS 1	SKS 7	IPS 7	Status
17111024410001	Lambat	20	33250	17	37941	Cuti
17111024410002	Tepat	20	34500	17	35882	Lulus
1811102441010	Tepat	20	35500	17	38824	Lulus
...
1911102441186	Lambat	20	3.300	22	3.455	Lulus

3.2.2 Data Cleaning

Pada tahap ini, yaitu dilakukan penghapusan data dengan cara memfilter nilai pada atribut “Status”. Yang dimana, filter yang dilakukan yaitu pada nilai yang bertuliskan “non-aktif” dan “dikeluarkan”. Setelah melalui proses ini, maka data yang dihasilkan pada angkatan 2017 berjumlah 12 data, angkatan 2018 berjumlah 90 data, dan angkatan 2019 berjumlah 111 data. Yang dimana jika ditotalkan berjumlah 213 data.

NIM	SKS 1	IPS 1	SKS 7	IPS 7	Status	Kelulusan
17111024410001	20	33250	17	37941	Non-Aktif	-
17111024410002	20	34500	17	35882	Lulus	-
1811102441010	20	35500	17	38824	Lulus	-
...
1911102441186	20	3.300	22	3.455	Lulus	-

NIM	SKS 1	IPS 1	SKS 7	IPS 7	Status	Kelulusan
17111024410002	20	34500	17	35882	Lulus	Tepat
1811102441010	20	35500	17	38824	Lulus	Tepat
...
1911102441186	20	3.300	22	3.455	Lulus	Terlambat

3.2.3 Data Normalization

Pada tahap ini adalah mengisi nilai pada data yang hilang. Yang dimana jika menggabungkan data IPS dan data yudisium maka ada beberapa data mahasiswa yang tidak

memiliki keterangan lulus. Hal ini dikarenakan data yudisium hanya berisi data mahasiswa yang lulus tepat waktu. Sehingga nilai pada mahasiswa yang tidak masuk dalam data yudisium akan hilang. Karena penelitian ini bertujuan untuk memprediksi lulus tepat atau terlambat, maka mahasiswa yang belum memiliki status kelulusan akan di beri label “Lambat”. Hal ini sesuai dengan panduan akademik berdasarkan ketepatan waktu lulus, sehingga kelulusan setiap angkatan (2017 – 2020) dapat dilihat pada Tabel 3.4. *Data normalization* akan di tunjukkan pada Tabel 3.5 dan Tabel 3.6.

Tabel 3.4 Tahun Kelulusan Angkatan

Angkatan	Lulus Tepat Waktu
2017	2021
2018	2022
2019	2023
2020	2024

Tabel 3.5 *Missing value data* pada atribut kelulusan

NIM	Kelulusan	SKS 1	IPS 1	SKS 7	IPS 7
17111024410001	?	20	33250	17	37941
1811102441002	Tepat	20	34500	17	35882
...
1911102441003	Tepat	20	35500	17	38824

Pada Tabel 3.5 dapat diketahui bahwa ada data yang tidak memiliki nilai pada atribut “Kelulusan” yang menyebabkan terjadinya *missing value data*. Selanjutnya pada Tabel 3.6 yaitu mengganti nilai yang hilang pada atribut kelulusan tersebut menjadi “Lambat” sesuai angkataannya seperti pada Tabel 3.6.

Tabel 3.6 Setelah melakukan *replace missing value* pada data

NIM	Kelulusan	SKS 1	IPS 1	SKS 7	IPS 7
17111024410001	Lambat	20	33250	17	37941
1811102441002	Tepat	20	34500	17	35882
...
1911102441003	Tepat	20	35500	17	38824

3.3 Pembagian

Setelah melalui tahap diatas, selanjutnya yaitu pembagian data. Pembagian data ini adalah memisahkan data yang akan digunakan untuk membuat model dan validasi (2017 – 2019) dan data mahasiswa yang akan diprediksi (2020).

3.3.1 Data Mahasiswa 2017-2019

Pada pembagian data di atas, tahap selanjutnya adalah membagi data validasi menggunakan *split data* yaitu *data training* 75% dan *data testing* 25%. Yang dimana, dari banyaknya data yaitu 213 mahasiswa digunakan menjadi *data training* sebanyak 75% yaitu 160 data dan 25% menjadi *data testing* yaitu sebanyak 53 data. *Data training* menjadi data pertama untuk mengajarkan model pola pada algoritma K-NN dan keterkaitan dalam data dengan tujuan membuat prediksi. Sementara itu, *data testing* merupakan bagian lain dari dataset yang tidak digunakan selama pelatihan model dengan tujuan untuk melatih kinerja model pola pada perhitungan algoritma K-NN.

3.3.2 Data Prediksi Mahasiswa 2020

Penjelasan dari pembagian diatas yaitu data mahasiswa angkatan 2017 – 2019 digunakan untuk membuat model dan validasi model yakni seberapa akurat algoritma yang dipakai untuk memprediksi data. Selanjutnya percobaan pada data prediksi yaitu data mahasiswa angkatan 2020 yang berjumlah 165 data yang tidak memiliki atribut kelulusan, sehingga kelulusannya akan diprediksi menggunakan algoritma KNN.

3.4 Klasifikasi K-NN

Setelah persiapan data, langkah selanjutnya ialah melakukan klasifikasi mahasiswa angkatan 2020 dengan menggunakan K-NN yang memanfaatkan data mahasiswa 2017-2019. Untuk melakukan penghitungan data menggunakan algoritma K-NN dapat dijelaskan sebagai berikut.

Tabel 3.7 Data Percobaan

NO	NIM	IPS 1	IPS 2	IPS 3	IPS 4	KELULUSAN
1	1811102441089	2,575	3,079	3,000	3,450	TEPAT
2	1811102441091	3,500	3,132	3,088	3,325	TERLAMBAT
3	1811102441092	3,525	3,158	3,318	3,350	TERLAMBAT
4	1811102441093	3,625	3,395	3,237	3,775	TEPAT
5	1811102441094	2,875	3,368	3,316	3,400	TEPAT
6	1811102441095	2,850	2,211	3,029	3,250	TERLAMBAT
7	1811102441096	2,875	3,316	3,250	3,275	TEPAT
8	1811102441097	3,500	3,395	3,273	3,025	TERLAMBAT
9	1811102441098	3,075	3,263	3,158	3,650	TERLAMBAT
10	1811102441100	3,675	3,553	3,591	3,925	TEPAT
11	1234567890	3,700	2,700	3,000	2,500	

Adapun data pada Tabel 3.7 sebagai contoh terdapat 10 mahasiswa yang memiliki kelulusan dan 1 mahasiswa yang belum diketahui. Untuk memprediksi kelulusan mahasiswa tersebut menggunakan algoritma K-NN, maka perlu dilakukan perhitungan jarak antara data yang akan dievaluasi dan data yang digunakan untuk pelatihan menggunakan persamaan perhitungan jarak K-NN (2.1) di bawah ini:

$$d = \sqrt{(2,575 - 3,700)^2 \times (3,079 - 2,700)^2 \times (3,000 - 3,000)^2 \times (3,450 - 2,500)^2} = 1,52042435$$

$$d = \sqrt{(\dots)^2 \times (\dots)^2} = \dots$$

$$d = \sqrt{(3,675 - 3,700)^2 \times (3,553 - 2,700)^2 \times (3,591 - 3,000)^2 \times (3,925 - 2,500)^2} = 1,76276475$$

Dengan menghitung jarak antara data yang akan dievaluasi di atas (data ke-11) ke semua data latih maka di peroleh hasil sebagai berikut:

Tabel 3.8 Perangkingan Jarak

NO	KELULUSAN	JARAK	RANKING
1	TEPAT	1,52042435	9
2	TERLAMBAT	0,95639051	2
3	TERLAMBAT	1,03152734	3
4	TEPAT	1,47306902	8
5	TEPAT	1,42727475	6
6	TERLAMBAT	1,23510105	4
7	TEPAT	1,31261557	5
8	TERLAMBAT	0,93412974	1
9	TERLAMBAT	1,43361489	7

10	TEPAT	1,76276475	10
11

Adapun jarak antara data yang telah diketahui pada Tabel 3.8. Langkah selanjutnya ialah mengurutkan data dan meranking mulai dari jarak terdekat. Karena pada penelitian ini nilai $K = 5$ maka, data yang diperlukan ialah data yang masuk dalam ranking 5 besar. 1 data diantaranya memiliki Label Kelulusan “Tepat”, sedangkan 4 data diantaranya memiliki Label Kelulusan “Terlambat”. Dikarenakan label kelulusan “Terlambat” lebih banyak (4 Data), maka data ke – 11 diprediksi memiliki label “Terlambat”. Pada penelitian ini, untuk melakukan pembuatan model yang menggunakan semua data, mengevaluasi, dan memprediksi mahasiswa 2020 menggunakan alat bantu yaitu *RapidMiner Studio* versi 9.10.

3.5 Evaluation

Langkah selanjutnya ialah mengevaluasi algoritma K-NN menggunakan *data testing*. Hasil prediksi terhadap *data testing* ditunjukkan pada Tabel 3.9.

Tabel 3.9 Hasil Percobaan pada *Data Testing*

NO	NIM	Kelulusan,	Prediksi
1	17111024410002	TEPAT	TERLAMBAT
2	17111024410005	TERLAMBAT	TERLAMBAT
3	17111024410011	TERLAMBAT	TEPAT
4	1811102441008	TEPAT	TERLAMBAT
5	1811102441017	TEPAT	TEPAT
6	1811102441019	TERLAMBAT	TEPAT
7	1811102441024	TEPAT	TEPAT
8	1811102441030	TEPAT	TEPAT
9	1811102441032	TEPAT	TERLAMBAT
10	1811102441036	TERLAMBAT	TERLAMBAT
...
53	1911102441182	TERLAMBAT	TEPAT

Perlu diketahui bahwa dari data di atas label “Tepat” bernilai “*Positive*” dan label “Terlambat” bernilai “*Negative*”. Dari hasil data diperoleh *True Negative* adalah 18 data, *False*

Negative adalah 6 data, *True Positive* adalah 20 data dan *False Positive* adalah 9 data. Selanjutnya ialah menghitung hasil penghitungan algoritma dengan menggunakan persamaan (2.2), (2.3), (2.4). sehingga menghasilkan nilai dibawah ini

<i>Accuracy</i> 71.70%	<i>True Positive</i>	<i>True Negative</i>	<i>Class Precision</i>
<i>Pred. Positive</i>	20	9	68.97%
<i>Pred. Negative</i>	6	18	75%
<i>Class Recall</i>	76.92	66.67%	

Adapun hasil tersebut didapat dengan menghitung hasil menggunakan persamaan dibawah :

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN} = \frac{38}{53} \times 100 = 71.70\%$$

$$Presisi = \frac{TP}{TP + FP} = \frac{20}{29} = 0.68965517 = 68.97\%$$

$$Recall = \frac{TP}{TP + FN} = \frac{20}{26} = 0.76923077 = 76.92\%$$

3.6 Prediksi 2020

Setelah melalui evaluasi terhadap model K-NN dan mendapatkan hasil akurasi sebesar 71,70%, maka langkah selanjutnya adalah menerapkan model ke data angkatan 2020 dan memperoleh hasil bahwa dari 165 mahasiswa terdapat 74 mahasiswa yang diprediksi “Terlambat” dan 91 mahasiswa yang diprediksi “Tepat”. Adapun hasil prediksi setiap mahasiswa dapat dilihat pada Tabel 3.10 dan grafik kelulusan mahasiswa 2020 disajikan pada Gambar 3.2.

Tabel 3.10 Hasil Prediksi Angkatan 2020

NO	NIM	Prediksi
1	2011102441008	TERLAMBAT
2	2011102441009	TERLAMBAT
3	2011102441012	TERLAMBAT

4	2011102441013	TEPAT
5	2011102441014	TEPAT
6	2011102441015	TEPAT
7	2011102441016	TEPAT
8	2011102441017	TERLAMBAT
...
165	2011102441020	TEPAT



Gambar 3.2 Grafik Perbandingan pada Angkatan 2020