

DATA MINING METODE K-NEAREST NEIGHBOR UNTUK MEMPREDIKSI KELULUSAN TEPAT WAKTU MAHASISWA FTI UNIVERSITAS BUDI LUHUR

Asprilla Ananda Wicaksana¹, Windarto²

¹Teknik Informatika, Fakultas Teknologi Informasi, Universitas Budi Luhur, DKI Jakarta, Indonesia

²Manajemen Informatika, Fakultas Teknologi Informatika, Universitas Budi Luhur, DKI Jakarta, Indonesia

Email: ¹asprillast@email.com, ²windarto@budiluhur.ac.id

(* : corresponding author)

Abstrak- Lulus tepat waktu merupakan permasalahan yang sulit bagi beberapa mahasiswa yang menjalani pendidikan untuk mendapatkan gelar strata-1 (S1). Penelitian ini difokuskan untuk memprediksi kelulusan mahasiswa Budi Luhur dengan akurat menerapkan metode algoritma K-Nearest Neighbors karena K-Nearest Neighbors merupakan algoritma yang bisa digunakan untuk mengolah data yang bersifat numerik dan tidak memerlukan skema estimasi parameter perulangan yang rumit, ini berarti bisa diaplikasikan untuk dataset berukuran besar. Dengan begitu, maka metode pengukuran jarak terdekat yang dapat diimplementasikan salah satunya yaitu Manhattan Distance dikarenakan pada metode ini terdapat kesesuaian untuk mengukur kedekatan data yang bersifat numerik. Hasil penelitian ini digunakan sebagai parameter mahasiswa untuk mengetahui hasil mahasiswa tersebut lulus tepat waktu atau tidak lulus tepat waktu. Hasil uji coba data set yaitu 2 angkatan yakni 2017/2018, dan 2018/2019 yang berjumlah 653 mahasiswa. Setelah melakukan proses prediksi dan klasifikasi pada sistem, didapatkan nilai akurasi terbesar yaitu 99.4% dari hasil 2 nilai K, yaitu 3 dan 5, karena dengan menggunakan nilai pengujian 3 dan 5 dapat mendapatkan akurasi yang mendekati nilai 100%. Berdasarkan hasil penelitian tersebut dapat disimpulkan bahwa dengan program yang dibuat menggunakan algoritma K-Nearest Neighbor dan pengukuran jarak menggunakan Manhattan Distance dapat mempercepat dan mempermudah perhitungan hasil ketepatan lulus mahasiswa secara akurat.

Kata Kunci: Java, Algoritma K-Nearest Neighbor, Manhattan Distance, Prediksi Kelulusan Tepat Waktu Mahasiswa.

DATA MINING K-NEAREST NEIGHBOR METHOD TO PREDICT ON TIME GRADUATION OF FTI STUDENTS BUDI LUHUR UNIVERSITY

Abstract- *Graduating on time is a difficult problem for some students who are studying to get a bachelor's degree (S1). This research is focused on predicting the graduation of Budi Luhur students by accurately applying the K-Nearest Neighbors algorithm method because K-Nearest Neighbors is an algorithm that can be used to process numerical data and does not require a complicated iteration parameter estimation scheme, this means it can be applied to datasets big size. That way, one of the methods for measuring the shortest distance that can be implemented is the Manhattan Distance because this method has the suitability for measuring the closeness of numerical data. The results of this study are used as student parameters to determine the results of these students graduating on time or not graduating on time. The results of the data set trials were 2 batches, namely 2017/2018 and 2018/2019 with a total of 653 students. After carrying out the prediction and classification process on the system, the greatest accuracy value is 99.4% from the results of 2 K values, namely 3 and 5, because using test values 3 and 5 can get an accuracy close to 100%. Based on the results of this study, it can be concluded that the program created using the K-Nearest Neighbor algorithm and measuring distances using Manhattan Distance can speed up and simplify the calculation of the results of student graduation accuracy accurately.*

Keywords: Java, K-Nearest Neighbor Algorithm, Manhattan Distance, Prediction of On Time Student Graduation.

1. PENDAHULUAN

Lulus tepat waktu merupakan satu indikator keberhasilan mahasiswa dalam memperoleh gelar sarjana. Mahasiswa akan dinyatakan lulus tepat waktu jika menyelesaikan studinya di perguruan tinggi dengan mendapatkan 144 sks dalam waktu empat tahun atau delapan semester. Dalam kenyataan yang ada mahasiswa tidak selalu bisa menyelesaikan pendidikan sarjananya dalam waktu empat tahun atau delapan semester di perguruan tingginya. Mahasiswa yang telah menyelesaikan program studi sarjananya selanjutnya akan

mendaftarkan diri sebagai calon wisudawan yang nantinya akan melalui proses wisuda. Peneliti melihat dari data mahasiswa Teknik Informatika di Universitas Budi Luhur dalam kurun waktu 2 tahun, dari tahun 2017/2018 hingga tahun 2018/2019 terdapat data populasi mahasiswa sebanyak 956 mahasiswa Teknik Informatika yang berbeda pada setiap tahun ajarannya dengan rincian di tahun 2017 terdapat 455 data dan di tahun 2018 terdapat 501 data. Namun tidak semua mahasiswa pada angkatan 2017 dan 2018 lulus tepat pada waktunya. Hal ini berdampak pada bertambahnya jumlah data akademik dari seluruh mahasiswa yang masih terdaftar. Oleh karena itu diperlukan kebijakan khusus untuk meminimalisir jumlah mahasiswa yang tidak lulus tepat waktu. Agar mahasiswa yang masuk dan sudah lulus bisa seimbang. Sistem ini menggunakan metode Data Mining yaitu K-Nearest Neighbor. K-Nearest Neighbor merupakan metode yang bisa digunakan untuk mengolah data yang bersifat numerik dan tidak membutuhkan skema estimasi parameter perulangan yang rumit, ini berarti bisa diaplikasikan untuk dataset berukuran besar. Dengan begitu, maka metode pengukuran jarak terdekat yang dapat diimplementasikan salah satunya yaitu Manhattan Distance dikarenakan pada metode ini terdapat kesesuaian untuk mengukur kedekatan data yang bersifat numerik. Berdasarkan hal tersebut peneliti ingin melakukan penelitian untuk memprediksi ketepatan lulus mahasiswa dengan metode K-Nearest Neighbor.

2. METODE PENELITIAN

Sistem yang digunakan untuk penelitian ini menggunakan sistem berbasis desktop yang dikembangkan menggunakan bahasa pemrograman Java, agar sistem tersebut nantinya dapat bersifat multi platform. Untuk mengklasifikasi data IPS (indeks prestasi semester) yang ingin diketahui prediksinya, pertama-tama user harus meng-import file dataset ke dalam sebuah database melalui menu dataset. File yang akan di-import yakni berjenis file spreadsheet dengan sebuah ekstensi *.xls ataupun *.xlsx. File dataset tersebut yang akan menjadi sebagai model pembelajaran algoritma klasifikasi K-Nearest Neighbor.

2.1 Data Cleaning (Pembersihan Data)

Proses data cleaning ini dilakukan untuk membersihkan data yang dianggap sebagai outlier atau data yang mengandung noise. Kriteria yang dianggap sebagai outlier adalah data dengan nilai IPS (Indeks Prestasi Semester) mahasiswa yang bernilai 0.

2.2 Metode KNN Pada Proses Prediksi

Tahap selanjutnya, user dapat melakukan sebuah evaluasi atau pengujian terhadap model yang sebelumnya telah di-import ke dalam sebuah database melalui menu evaluasi. Pada tahapan ini, user akan dapat melakukan evaluasi atau pengujian terhadap model yang sudah tersedia dengan membagi dataset yang terdapat pada database menjadi 2 bagian yang dimana salah satunya akan menjadi bentuk dataset training dan dataset testing. Dan juga terdapat 2 buah metode pengujian yang dapat dilakukan user, yaitu metode pengujian single testing dan juga multi testing. Tentunya pada kedua proses pengujian tersebut memiliki sebuah peran yang berbeda-beda, namun tergantung dari kebutuhan user. Proses single testing dilakukan apabila user ingin mengetahui tingkat keberhasilan sistem dengan sebuah parameter nilai ketetanggaan terdekat (nilai K). Sedangkan pada proses multi testing dilakukan apabila user ingin membandingkan beberapa nilai-nilai parameter ketetanggaan terdekat (nilai K) yang berdasarkan hasil tingkat keberhasilan sebuah sistem.

2.3 Metode K-NN Pada Proses Klasifikasi

Proses input atau memasukkan nilai K sebagai proses parameter utama algoritma K-Nearest Neighbor, sistem tersebut akan melakukan sebuah proses klasifikasi terhadap dataset yang sudah tersedia. Setelah proses tersebut telah selesai dilakukan, maka sistem akan memberikan sebuah evaluasi kinerja dari algoritma K-Nearest Neighbor dalam melakukan klasifikasi model dengan metode evaluasi Confusion Matrix. Proses klasifikasi ini untuk mendapatkan data IPS (indeks prestasi semester) yang ingin diketahui prediksinya, dan juga user dapat menggunakan menu klasifikasi yang pada dasarnya merupakan sebuah fungsionalitas utama pada sistem. Pada tahapan ini, pertama-tama user harus memilih jenis file spreadsheet yang berekstensi *.xls ataupun *.xlsx yang didalamnya berisi data-data mahasiswa yang diprediksi sebuah peluang kelulusannya. Selanjutnya seperti proses sebelumnya, dengan memasukkan nilai K yang sebagai parameter utama pada algoritma K Nearest Neighbor, proses klasifikasi terhadap model yang sudah disimpan dalam database dapat dilakukan dan kemudian sistem tersebut menghasilkan sebuah output yang berupa hasil dari prediksi kelulusan mahasiswa. Hasil dari proses tersebut nantinya dapat disimpan oleh user dalam jenis format file spreadsheet yang digunakan sebagai laporan atau sebuah bahan pertimbangan untuk selanjutnya.

2.4 Data Penelitian

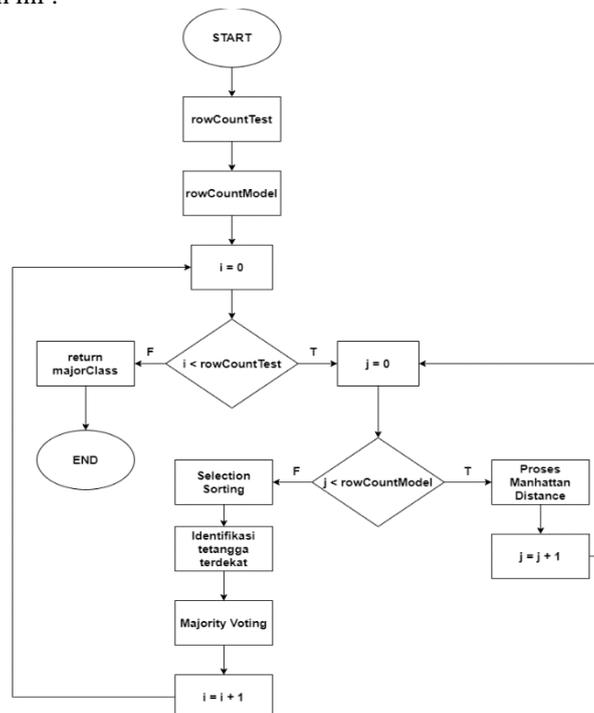
Data penelitian yang digunakan adalah data yang didapatkan dari data mahasiswa FTI Universitas Budi Luhur tahun 2017 sampai 2020 dengan keterangan data populasi sebanyak 1894 data dan data sampel sebanyak 1232 data.

Tabel 1. Data Penelitian

Data Populasi		Data Sampel	
Tahun Angkatan	Jumlah Data	Tahun Angkatan	Jumlah Data
2017	455	2017	317
2018	501	2018	336
2019	535	2019	335
2020	403	2020	244
Total Data	1894	Total Data	1232

2.5 Flowchart Algoritma KNN

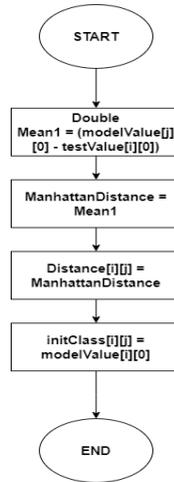
Flowchart proses Algoritma *K-Nearest Neighbor* merupakan alur proses utama dari metode yang diterapkan. Proses tersebut memiliki beberapa proses-proses yang akan dijelaskan secara rinci pada flowchart selanjutnya. Berikut flowchart proses Algoritma *K-Nearest Neighbor* yang digambarkan pada Gambar 1 di bawah ini :



Gambar 1. Flowchart Proses Algoritma *K-Nearest Neighbor*

2.6 Flowchart Proses Manhattan Distance

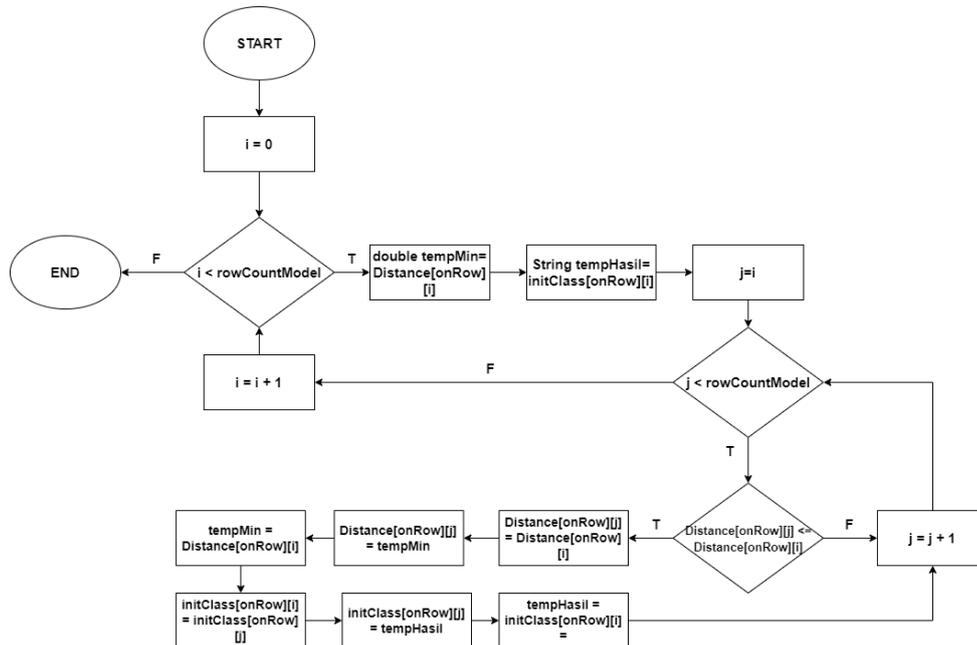
Flowchart proses *Manhattan Distance* merupakan proses yang terjadi pada proses algoritma *K-Nearest Neighbor*. Proses ini terdapat proses perhitungan jarak antara data model (*data training*) dengan data uji (*data testing*) yang dilakukan pada data uji dan baris demi baris dengan keseluruhan data model yang ada. Nantinya hasil dari pengukuran jarak yang sudah dilakukan tersebut disimpan pada *variable* Array 2 dimensi untuk dilakukan proses lebih lanjut. Berikut flowchart proses *Manhattan Distance* yang digambarkan pada Gambar 2 di bawah ini :



Gambar 2. Flowchart Proses Manhattan Distance

2.7 Flowchart Proses Selection Sorting

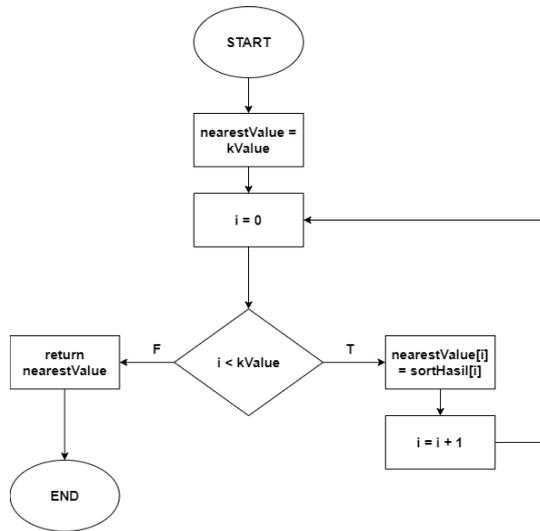
Flowchart proses *Selection Sorting* merupakan bentuk alur proses yang terjadi pada proses algoritma *K-Nearest Neighbor*. Pada proses ini, hasil (*target class*) akan diurutkan sesuai kedekatan jarak yang mengidentifikasi suatu *class*. Berikut *flowchart* proses *Selection Sorting* yang digambarkan pada Gambar 3 di bawah ini :



Gambar 3. Flowchart Proses Selection Sorting

2.8 Flowchart Proses Identifikasi Tetangga Terdekat

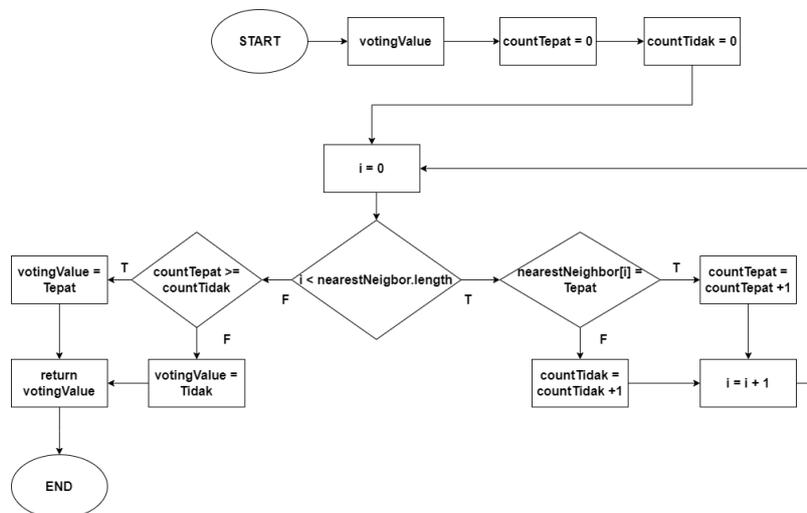
Flowchart proses Identifikasi Tetangga Terdekat merupakan bentuk alur proses yang terjadi pada algoritma *K-Nearest Neighbor*. Pada proses ini, terdapat proses mengidentifikasi ataupun memisahkan data-data yang terdekat sebanyak nilai *K* (sebuah parameter yang telah diinput pada proses sebelumnya). Berikut *flowchart* proses Identifikasi Tetangga Terdekat yang digambarkan pada Gambar 4 di bawah ini :



Gambar 4. Flowchart Proses Identifikasi Tetangga Terdekat

2.9 Flowchart Proses Majority Voting

Flowchart proses *Majority Voting* merupakan bentuk alur proses yang terjadi pada proses algoritma *K-Nearest Neighbor*. Pada proses ini, melakukan pengambilan hasil (*target class*) terbanyak dari tetangga terdekat yang sudah dilakukan sebelumnya. Berikut flowchart proses *Majority Voting* yang digambarkan pada Gambar 5. di bawah ini :

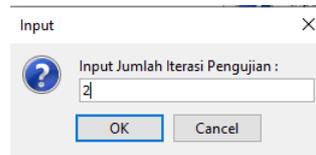


Gambar 5. Flowchart Proses Majority Voting

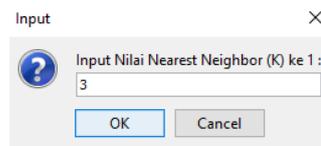
3. HASIL DAN PEMBAHASAN

3.1 Hasil Uji Coba Program

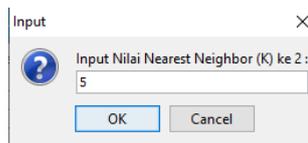
Pada bagian hasil uji coba program, dilakukan dengan 2 tipe pengujian yakni *multi testing* dan *single testing* yang bertujuan untuk mengetahui kinerja sistem yang berdasarkan nilai akurasi dari nilai ketetanggaan terdekat (*number of nearest neighbor*) yang terbaik pada proses klasifikasi. Pada proses pengujian *multi testing* dengan menggunakan nilai (K) yang sama pada masing-masing pengujian, jumlah iterasi pada pengujian *multi testing* yakni sebanyak 2 kali untuk menentukan nilai (K) terbaik seperti Gambar 6 dan meng-*input* iterasi secara berturut-turut sesuai dengan jumlah iterasi yang sudah ditentukan sebelumnya seperti pada Gambar 7 dan Gambar 8 berikut ini :



Gambar 6. Jumlah Iterasi Pengujian *Multi Testing*



Gambar 7. Iterasi *Multi Testing* Input Nilai (K) Ke-1



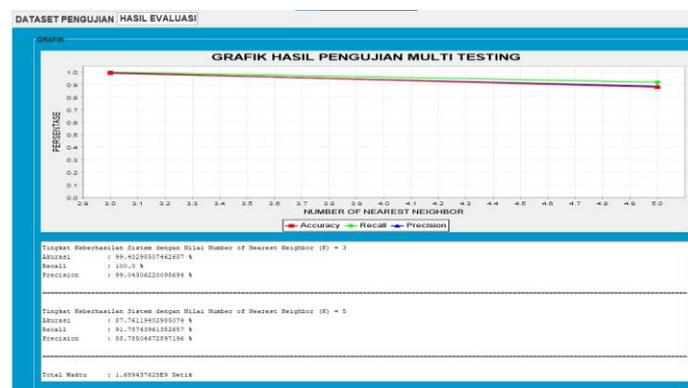
Gambar 8. Iterasi *Multi Testing* Input Nilai (K) Ke-2

3.2 Pengujian *Multi Testing*

Proses pengujian *multi testing* memiliki tujuan untuk mengevaluasi model yang telah ada menggunakan beberapa parameter nilai (K) dan proses tersebut dilakukan sebanyak 2 kali pada tahun angkatan dengan perbandingan yaitu 1 : 2.

3.2.2 Pengujian *Multi Testing* Pertama

Pada proses pengujian *multi testing* ketiga, akan menggunakan parameter nilai (K) yaitu 3 dan 5. *Dataset* yang akan digunakan sebagai *dataset* uji yakni 2019 dengan jumlah data sebanyak 335 data dan yang sebagai *dataset* model yakni 2017 dan 2018 dengan total jumlah data sebanyak 653 data. Adapun hasil pengujian *multi testing* ketiga berupa grafik dan tabel seperti pada Gambar 9 dan Tabel 2 berikut :



Gambar 9. Grafik Hasil Pengujian *Multi Testing* Pertama

Tabel 2. Tabel Hasil Pengujian *Multi Testing* Pertama

K	Accuracy	Recall	Precision
3	99.4 %	100.0 %	99.0 %
5	87.7 %	91.7 %	88.7 %

Dengan Keterangan sebagai berikut :

- 1) *Accuracy* : Persentase dari jumlah data yang sudah diklasifikasikan dengan benar
- 2) *Recall* : Persentase dari kemampuan sistem untuk memanggil dokumen yang relevan
- 3) *Precision* : persentase dari kemampuan sistem untuk tidak memanggil dokumen yang tidak relevan

Dari ketiga hasil proses pengujian *multi testing* yang sudah dilakukan, didapatkan nilai parameter nilai K yang terbaik yakni 3. Adapun hasil dari keseluruhan pengujian *multi testing* dapat dilihat pada Tabel 3 :

Tabel 3. Tabel Hasil Pengujian *Multi testing*

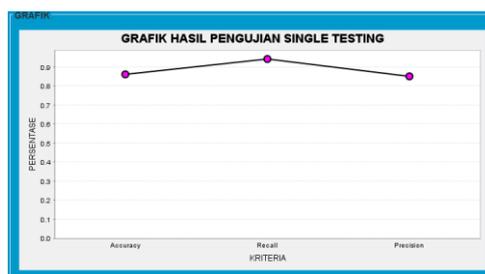
No	Dataset Training		Dataset Testing		Nilai K	Hasil Evaluasi	
	Tahun Angkatan	Jumlah Data	Tahun Angkatan	Jumlah Data		Akurasi	Rata-Rata
	2017	653	2019	335	3	99.4 %	93.5 %
	2018				5	87.7 %	

3.3 Pengujian Single Testing

Proses pengujian *single testing* bertujuan untuk mengevaluasi model yang ada menggunakan 1 buah parameter nilai (K) yang ditentukan oleh penulis yakni 7. Pada proses pengujian *single testing* ini akan menghasilkan *table Confusion Matrix* yang nantinya akan menampilkan kinerja dari algoritma *K-Nearest Neighbor* dalam proses pengklasifikasian *dataset* yang berupa nilai *accuracy*, *recall*, dan *Precision*. Pengujian ini dilakukan pada masing-masing tahun angkatan dengan perbandingan 1 : 2.

3.3.1 Pengujian Single Testing Pertama

Pada proses pengujian *single testing* pertama, akan menggunakan parameter nilai (K) yaitu 7. *Dataset* yang akan digunakan sebagai *dataset* uji yakni 2019 dengan total jumlah data sebanyak 335 data dan yang sebagai *dataset* model yakni 2017 dan 2018 dengan jumlah data sebanyak 653 data. Adapun hasil pengujian *single testing* pertama berupa grafik dan tabel seperti pada Gambar 10 dan Gambar 11 sebagai berikut :



Gambar 10. Grafik Hasil Pengujian *Single Testing* Pertama

CONFUSION MATRIX			
	PREDIKSI TEPAT	PREDIKSI TIDAK	CLASS RECALL
ACTUAL TEPAT	195	12	94.2 %
ACTUAL TIDAK	34	94	73.44 %
CLASS PRECISION	85.15 %	88.88 %	

***ACCURACY**
Overall Accuracy is 88.27 %

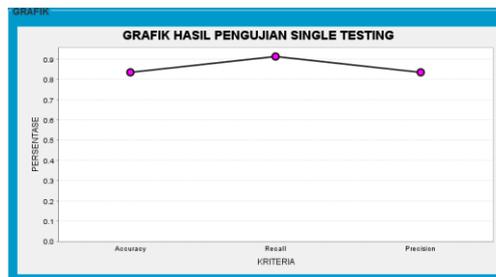
Gambar 11. *Confusion Matrix* Hasil Pengujian *Single Testing* Pertama

Dengan Keterangan sebagai berikut :

- 1) 335 data yang digunakan untuk *dataset* uji
- 2) 653 data yang digunakan untuk *dataset* model
- 3) 195 siswa yang diprediksi lulus dan diklasifikasikan lulus
- 4) 34 siswa yang diprediksi lulus dan diklasifikasikan tidak lulus
- 5) 12 siswa yang diprediksi tidak lulus dan diklasifikasikan lulus
- 6) 94 siswa yang diprediksi tidak lulus dan diklasifikasikan tidak lulus
- 7) 94.2 % tingkat *recall* yang diklasifikasikan lulus
- 8) 73.4 % tingkat *recall* yang diklasifikasikan tidak lulus
- 9) 85.1 % tingkat *precision* yang diprediksikan lulus
- 10) 88.6% tingkat *precision* yang diprediksikan tidak lulus
- 11) 86.2 % tingkat *accuracy*

3.3.2 Pengujian *Single Testing* Kedua

Pada proses pengujian *single testing* kedua, akan menggunakan parameter nilai (K) yaitu 9. *Dataset* yang akan digunakan sebagai *dataset* uji yakni 2019 dengan total jumlah data sebanyak 335 data dan yang sebagai *dataset* model yakni 2017 dan 2018 dengan jumlah data sebanyak 653 data. Adapun hasil pengujian *single testing* kedua berupa grafik dan tabel seperti pada Gambar 12 dan Gambar 13 sebagai berikut :



Gambar 12. Grafik Hasil Pengujian *Single Testing* Kedua

CONFUSION MATRIX			
	PREDIKSI TEPAT	PREDIKSI TIDAK	CLASS RECALL
ACTUAL TEPAT	189	18	91.3 %
ACTUAL TIDAK	37	91	71.09 %
CLASS PRECISION	83.63 %	83.49 %	

*ACCURACY
Overall Accuracy is 83.58 %

Gambar 13. *Confusion Matrix* Hasil Pengujian *Single Testing* Kedua

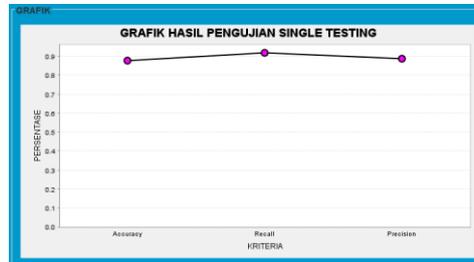
Dengan Keterangan sebagai berikut :

- 1) 335 data yang digunakan untuk *dataset* uji
- 2) 653 data yang digunakan untuk *dataset* model
- 3) 189 siswa yang diprediksi lulus dan diklasifikasikan lulus
- 4) 37 siswa yang diprediksi lulus dan diklasifikasikan tidak lulus
- 5) 18 siswa yang diprediksi tidak lulus dan diklasifikasikan lulus
- 6) 91 siswa yang diprediksi tidak lulus dan diklasifikasikan tidak lulus
- 7) 91.3 % tingkat *recall* yang diklasifikasikan lulus
- 8) 71.0 % tingkat *recall* yang diklasifikasikan tidak lulus
- 9) 83.6 % tingkat *precision* yang diprediksikan lulus
- 10) 83.4 % tingkat *precision* yang diprediksikan tidak lulus
- 11) 83.5 % tingkat *accuracy*

3.3.3 Pengujian *Single Testing* Ketiga

Pada proses pengujian *single testing* ketiga, akan menggunakan parameter nilai (K) yaitu 5. *Dataset* yang akan digunakan sebagai *dataset* uji yakni 2019 dengan total jumlah data sebanyak 335 data dan yang sebagai

dataset model yakni 2017 dan 2018 dengan jumlah data sebanyak 653 data. Adapun hasil pengujian *single testing* ketiga berupa grafik dan tabel seperti pada Gambar 14 dan Gambar 15 sebagai berikut :



Gambar 14. Grafik Hasil Pengujian *Single Testing* Ketiga

CONFUSION MATRIX			
	PREDIKSI TEPAT	PREDIKSI TIDAK	CLASS RECALL
ACTUAL TEPAT	190	17	91.79 %
ACTUAL TIDAK	24	104	81.25 %
CLASS PRECISION	88.79 %	85.95 %	

*ACCURACY
Overall Accuracy is 87.76 %

Gambar 15. *Confusion Matrix* Hasil Pengujian *Single Testing* Ketiga

Dengan Keterangan sebagai berikut :

- 1) 335 data yang digunakan untuk *dataset* uji
- 2) 653 data yang digunakan untuk *dataset* model
- 3) 190 siswa yang diprediksi lulus dan diklasifikasikan lulus
- 4) 24 siswa yang diprediksi lulus dan diklasifikasikan tidak lulus
- 5) 17 siswa yang diprediksi tidak lulus dan diklasifikasikan lulus
- 6) 104 siswa yang diprediksi tidak lulus dan diklasifikasikan tidak lulus
- 7) 91.7 % tingkat *recall* yang diklasifikasikan lulus
- 8) 81.2 % tingkat *recall* yang diklasifikasikan tidak lulus
- 9) 88.7 % tingkat *precision* yang diprediksikan lulus
- 10) 85.9 % tingkat *precision* yang diprediksikan tidak lulus
- 11) 87.7 % tingkat *accuracy*

4. KESIMPULAN

Dengan melalui proses pengerjaan dan pengujian yang dilakukan pada penelitian ini, maka dapat disimpulkan beberapa hal sebagai berikut :

- a. Algoritma *K-Nearest Neighbor* dengan metode *Manhattan Distance* sebagai pengukur jaraknya, dapat diimplementasikan ke dalam bentuk proses prediksi dan klasifikasi ketepatan lulus.
- b. Lama waktu proses tergantung dengan jumlah data serta spesifikasi *software* dan *hardware*.
- c. Aplikasi mampu memprediksi kelulusan tepat waktu mahasiswa dengan nilai indeks prestasi semester rata-rata dari akurasi sistem sebesar 97.24 % dari *dataset* uji sebanyak 653 data dan *dataset* model sebanyak 579 data.

UCAPAN TERIMA KASIH

Dalam kesempatan ini, penulis ingin memberikan ucapan terima kasih kepada semua pihak yang telah memberikan bimbingan dan masukan berupa kritik serta saran kepada penulis dalam penyusunan tugas akhir ini.

DAFTAR PUSTAKA

- [1] Baharuddin, M. M., Azis, H. and Hasanuddin, T. (2019) ‘Analisis Performa Metode K-Nearest Neighbor Untuk Identifikasi Jenis Kaca’, *ILKOM Jurnal Ilmiah*, 11(3), pp. 269–274. doi: 10.33096/ilkom.v11i3.489.269-274..
- [2] Jaya, T. S. (2018) ‘Pengujian Aplikasi dengan Metode Blackbox Testing Boundary Value Analysis (Studi Kasus: Kantor Digital Politeknik Negeri Lampung)’, *Jurnal Informatika Pengembangan IT (JPIT)*, 3(2), pp. 45–48. doi: 10.30591/jpit.v3i1.647.
- [3] Linda Tanti (2021) Metode Data Mining Clustering, Binus University school of information systems. Available at: <https://sis.binus.ac.id/2021/10/29/metode-data-mining-clustering/> (Accessed: 14 June 2023).
- [4] Malik, A. M. and Sibarani, A. J. P. (2018) ‘Aplikasi Prediksi Kelulusan Ujian Nasional Menggunakan Algoritma K-Nearest Neighbor dengan Pengukuran Jarak Manhattan Distance’, *Skanika*, 1(2), pp. 829–835.
- [5] Putra, P., M. H. Pardede, A. and Syahputra, S. (2022) ‘Analisis Metode K-Nearest Neighbour (Knn) Dalam Klasifikasi Data Iris Bunga’, *Jurnal Teknik Informatika Kaputama (JTik)*, 6(1), pp. 297–305
- [6] Rerung, R. R. (2018) ‘Penerapan Data Mining dengan Memanfaatkan Metode Association Rule untuk Promosi Produk’, 3(1), pp. 89–98. doi: 10.31544/jtera.v3.i1.2018.89-98.
- [7] Setio, P. B. N., Saputro, D. R. S. and Bowo Winarno (2020) ‘Klasifikasi Dengan Pohon Keputusan Berbasis Algoritme C4.5’, *PRISMA, Prosiding Seminar Nasional Matematika*, 3, pp. 64–71.
- [8] Sistem, R. and Elisa, E. (2018) ‘JURNAL RESTI Dengan Algoritma Apriori’, 2(2), pp. 472–478.
- [9] Sulistiawan, A., Diahsari, E. Y., & Situmorang, N. Z. (2020). (2020) ‘Implementasi Metode naive Bayes Untuk Prediksi Kelulusan Mahasiswa Tepat Waktu Prodi Informatika (Studi Kasus : Universitas Teknologi Yogyakarta)’, pp. 1–9. Available at: <http://eprints.uty.ac.id/4863/>.
- [10] Syahputra, A. (2022) “‘Algoritma Selection Sort’ “Implementasi Algoritma Selection Sort Untuk Pengurutan Nilai Ipk Mahasiswa Universitas Potensi Utama””, *Jurnal Teknik Informatika Kaputama (JTik)*, 6(2), pp. 390–398.
- [11] Umaidah, Y. et al. (2019) ‘JISICOM (Journal of Information System , Informatics and Computing) JISICOM (Journal of Information System , Informatics and Computing)’, 3(2), pp. 1–8.
- [12] Wijaya, A. C., Hasibuan, N. A. and Ramadhani, P. (2018) ‘IMPLEMENTASI ALGORITMA C5 . 0 DALAM KLASIFIKASI PENDAPATAN MASYARAKAT (STUDI KASUS : KELURAHAN MESJID KECAMATAN MEDAN KOTA)’, 5.
- [13] Zulfa, I., Rayuwati, R. and Koko, K. (2020) ‘Implementasi data mining untuk menentukan strategi penjualan buku bekas dengan pola pembelian konsumen menggunakan metode apriori’, *Teknika: Jurnal Sains dan Teknologi*, 16(1), p. 69. doi: 10.36055/tjst.v16i1.7601.