

Penerapan Teknik Klasifikasi Untuk Prediksi Kelulusan Mahasiswa Menggunakan Algoritma K-Nearest Neighbour

Implementation of Classification Method to Predict Student Graduation Using K-Nearest Neighbor Algorithm

Arie Yandi Saputra¹, Yogi Primadasa²

^{1,2}Sistem Informasi STMIK Bina Nusantara Jaya Lubuklinggau

E-mail : ¹arielahat@gmail.com, ²yogiak45@gmail.com

Abstrak

Tidak stabilnya kelulusan mahasiswa program studi Sistem Informasi pada STMIK Bina Nusantara Jaya Lubuklinggau menjadi tugas yang berat bagi Program Studi. Bertambahnya mahasiswa tiap tahunnya menyebabkan penumpukan data mahasiswa. Prediksi ketepatan kelulusan mahasiswa dirancang untuk mendukung prodi untuk membimbing mahasiswa agar kelulusan tepat waktu. Dengan mengetahui prediksi status kelulusan mahasiswa berjalannya perkuliahan, maka dari itu prodi di bantu pembimbing akademik dapat memberi perhatian khusus terhadap mahasiswa yang di prediksi tidak lulus tepat waktu (terlambat) sehingga mahasiswa tersebut dapat memperbaiki indeks prestasinya tiap semester agar dapat lulus tepat waktu. Dalam penelitian ini untuk memprediksi kelulusan mahasiswa menggunakan algoritma K-Nearest Neighbor. K-Nearest Neighbor merupakan metode klasifikasi, dimana kelas yang paling banyak muncul (mayoritas) yang akan menjadi kelas hasil klasifikasi. Penelitian ini mengambil sample data mahasiswa semester V STMIK Bina Nusantara Jaya Lubuklinggau, menggunakan 9 data training dan 1 data testing. Dengan menggunakan k=5 yang diterapkan menggunakan metode K-Nearest Neighbor untuk prediksi kelulusan mahasiswa.

Kata Kunci : Algoritma K-Nearest Neighbor, Prediksi, Klasifikasi, Mahasiswa

Abstract

The unstable graduation of the students of Information System program at STMIK Bina Nusantara Jaya Lubuklinggau becomes a difficult task for the study program. The growing number of students each year causes the accumulation of student data. The prediction of students' graduation accuracy is designed to support the study program to guide students to graduate on time. By knowing the prediction of the graduation status of college students, the academic advisors can give special attention to the students who are not predicted to graduate on time (late) so that the student can improve their achievement index every semester in order to pass on time. In this study to predict students' graduation using K-Nearest Neighbor algorithm. K-Nearest Neighbor is a classification method, whereby the most emerging class (the majority) will be the classified class. This research takes the sample data of students of semester V STMIK Bina Nusantara Jaya Lubuklinggau, using 9 training data and 1 test data. Using $k = 5$ applied using K-Nearest Neighbor method to predict students' graduation.

Keywords : K-Nearest Neighbor Algorithm, Prediction, Classification, student

1. PENDAHULUAN

Perguruan tinggi dituntut untuk menyelenggarakan pendidikan yang berkualitas bagi mahasiswa sehingga menghasilkan sumber daya manusia yang memiliki daya saing. Dalam perguruan tinggi mahasiswa merupakan asset yang sangat penting bagi institusi pendidikan oleh karena itu perlu diperhatikan tingkat kelulusan mahasiswa tepat pada waktunya[1]. Tidak stabilnya kelulusan mahasiswa program studi Sistem Informasi pada STMIK Bina Nusantara Jaya Lubuklinggau menjadi

tugas yang berat bagi prodi. Bertambahnya mahasiswa tiap tahunnya menyebabkan penumpukan data mahasiswa.

Prediksi ketepatan kelulusan mahasiswa dirancang untuk mendukung prodi dalam membimbing mahasiswa agar kelulusan tepat waktu. Dengan mengetahui prediksi status kelulusan mahasiswa berjalannya perkuliahan, maka dari itu prodi di bantu pembimbing akademik dapat memberi perhatian khusus terhadap mahasiswa yang di prediksi tidak lulus tepat waktu sehingga mahasiswa tersebut dapat memperbaiki indeks prestasinya tiap semester agar dapat lulus tepat waktu.

Teknik klasifikasi merupakan cara mengklasifikasi data ke dalam satu atau beberapa kelas yang sudah didefinisikan sebelumnya[2]. Algoritma *k-nearest neighbor* salah satu teknik klarifikasi data yang kuat, dengan cara mencari kasus dengan menghitung kedekatan antara kasus baru dengan kasus lama berdasarkan pencocokan bobot[3]. KNN adalah suatu metode algoritma *supervised learning*, dimana kelas yang paling banyak muncul (mayoritas) yang akan menjadi kelas hasil klasifikasi[4][3]. Data yang berdimensi q , pada algoritma K-NN, dapat menghitung jarak data satu ke data yang lain[5]

Penelitian sebelumnya oleh Muhammad Sukri[6] memprediksi kelulusan tepat waktu bagi mahasiswa baru. Dari hasil pengujian dengan menerapkan algoritma K-Nears Neighbour dan menggunakan data sampel alumni tahun wisuda 2004 s.d 2010 untuk kasus lama dan data alumni tahun wisuda 2011. Pada metode K-Nears Neighbour ini memiliki tingkat akurasi yang cukup besar yaitu sebesar 83,36%. Tingkat akurasi ini dapat diperbaiki dengan menambah jumlah data training.

Penelitian sebelumnya oleh Mustakim[7] Prediksi dilakukan terhadap mahasiswa proram studi sistem informasi angkatan 2014/2015 sebagai data testing dengan jumlah 50 data, serta berdasarkan dari data angkatan 2012/2013 sebagai data training dengan jumlah 165 data yang menghasilkan pengujian akurasi sebesar 82%.

Pada penelitian sebelumnya oleh Chavid[8] penelitian mengenai difteri menggunakan metode KNN dilakukan perhitungan similaritas pada kasus lama dengan kasus baru. Hasil diagnose penyakit difteri berdasarkan gejala yang dialami mendapatkan hasil tingkat akurasi sebesar 95,17%.

Oleh karena itu dengan adanya sebuah prediksi kelulusan mahasiswa dengan teknik klasifikasi menggunakan algoritma *k-nearest neighbor* diharapkan agar mahasiswa dapat lulus pada tepat waktu.

2. METODE PENELITIAN

2.1 Prediksi

Prediksi merupakan kegiatan memprediksi nilai-nilai sebuah variabel berdasarkan nilai yang diketahui dari variabel tersebut atau variabel yang berhubungan. Menurut sifatnya, teknik prediksi terbagi menjadi 2 jenis yaitu teknik kualitatif dan teknik kuantitatif. Teknik kuantitatif dikelompokkan menjadi 2 jenis[9] :

1. Model Time Series (Runtun Waktu)

Pada model time series, peramalan masa mendatang dilakukan berdasarkan nilai data masa lalu atau disebut data historis. Tujuan metode ini adalah menemukan pola dalam deret data historis dan memanfaatkan pola deret tersebut untuk peramalan masa mendatang. Data time series juga menggambarkan suatu objek dari waktu ke waktu atau periode secara historis dan terjadi berurutan. Keuntungan dalam menggunakan model ini adalah peramalan dapat dilakukan secara lebih sederhana dibandingkan dengan model kausal.

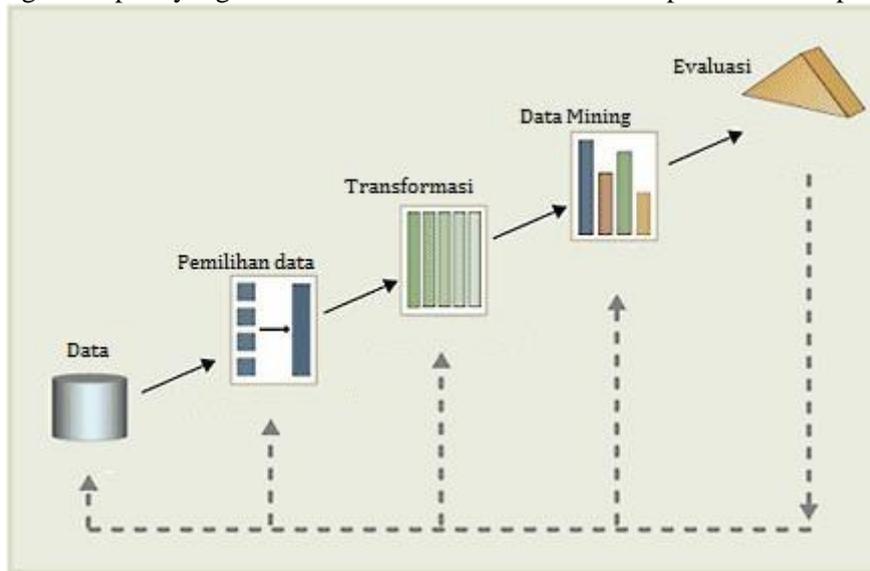
2. Model Kausal

Model kausal merupakan suatu model yang mengasumsikan faktor yang diramalkan menunjukkan suatu hubungan sebab akibat dalam suatu atau lebih variabel bebas dan menggunakannya untuk meramalkan nilai mendatang dari suatu variabel tak bebas. Keuntungan dalam menggunakan model ini adalah dapat menghasilkan tingkat keberhasilan yang lebih besar dalam pengambilan keputusan yang bijaksana.

2.2 Data Mining

Data mining merupakan suatu istilah yang biasa digunakan untuk memperjelaskan penemuan pengetahuan di dalam *database*[10].

Knowledge discovery data dapat memproses seluruh data non-trivial untuk mengetahui pola dalam data, yang mana pola yang ditemukan memiliki sifat sah dan dapat / mudah dipahami[11].



Gambar 1 Proses KDD

Adapun tahapan KDD adalah :

1. Data
Membuat himpunan data target, penetapan himpunan data dan memfokuskan pada subset variabel atau sampel data, dimana penelitian akan dilakukan.
2. Pemilihan Data
Langkah pertama pemrosesan data dan pembersihan data adalah tindakan dasar seperti penghapusan noise. Sebelum melakukan proses data mining, maka diperlukan proses cleaning pada data yang menjadi fokus dalam KDD
3. Transformasi
Pada tahap ini merupakan tahapan proses kreatif dan sangat tergantung pada pola informasi yang akan dicari dalam basis data.
4. Data Mining
Dalam pemilihan algoritma data mining untuk melakukan pencarian proses data mining yaitu antara lain teknik, metode atau algoritma dalam data mining sangat bervariasi. Penetapan metode atau algoritma yang tepat tergantung pada tujuan dan proses KDD secara keseluruhan.
5. Evaluasi
Tahap ini merupakan tahapan pemeriksaan, apakah pola yang ditemukan bertentangan dengan fakta atau hipotesa yang ada sebelumnya.

2.3 Teknik Data Mining

Fungsi data mining merupakan untuk mengklasifikasikan pola yang harus ditemukan dalam data mining. Berikut operasi-operasi dan teknik yang berhubungan dengan data mining[12] :

1. Operasi Predictive Modeling : (*classification, value prediction*)
2. Database segmentation : (*demographic clustering, neural clustering*)
3. Link Analysis : (*association discovery, sequential pattern discovery, similar timesequence discovery*)
4. Deviation detection : (*statistics, visualization*)

2.4 Klasifikasi

Klasifikasi merupakan proses penemuan model yang menggambarkan dan membedakan kelas data, atau cara mengklasifikasi data kedalam satu atau beberapa kelas yang sudah didefinisikan sebelumnya[2]. Teknik klasifikasi yang banyak digunakan secara luas, diantaranya adalah *Neural, Rough sets, K-nearest neighbor, Bayesian classifiers, network*, dan lain-lain.

Dalam proses klasifikasi data terdiri dari 2 langkah, yaitu *learning (fase training)* dan klasifikasi. Proses *learning* dibuat untuk mengalisa data *training* lalu direpresentasikan berupa *rule*

klasifikasi. Sedangkan proses klasifikasi, dimana data tes digunakan untuk memperkirakan akurasi dari *rule* klasifikasi. Model tersebut dibangun dengan menganalisa *database tuple*. Setiap *tuple* diasumsikan menjadi *predefined class* yang ditentukan oleh suatu atribut yang disebut *class label attribute*[3]. Dapat di ilustrasikan pada gambar 2 di bawah ini :



Gambar 2 Model Klasifikasi

Proses klasifikasi didasarkan pada empat komponen[12] :

1. *Kelas*
Variabel dependen berupa kategori yang mempresentasikan “label” yang terdapat pada objek.
2. *Predictor*
Variabel independen yang direpresentasikan oleh karakteristik data.
3. *Training dataset*
Satu set data yang mempunyai nilai dari kedua komponen yang digunakan untuk menentukan kelas yang cocok berdasarkan predictor.
4. *Testing dataset*
Berupa data baru yang diklasifikasikan oleh model data yang telah di buat dan akurasi klasifikasi di evaluasi.

2.5 *K-Nearest Neighbor* (KNN)

Algoritma *k-nearest neighbor* salah satu teknik klarifikasi data yang kuat, dengan cara mencari kasus dengan menghitung kedekatan antara kasus baru dengan kasus lama berdasarkan pencocokan bobot[3]. *K-Nearest Neighbor* adalah suatu metode algoritma *supervised learning*, dimana kelas yang paling banyak muncul (mayoritas) yang akan menjadi kelas hasil klasifikasi[7]. *K-Nearest Neighbor* termasuk dalam kelompok *instance-based learning*. *K-nearest neighbor* merupakan contoh algoritma berbasis pembelajaran, dimana data set pelatihan (*training*) disimpan, sehingga klasifikasi untuk record baru yang tidak diklasifikasi didapatkan dengan membandingkan *record* yang paling mirip dengan *training set*. Langkah-langkah *K-Nearest Neighbor*[6] :

1. Menentukan parameter K (jumlah tetangga paling dekat), Parameter K pada testing ditentukan berdasarkan nilai K optimum pada saat training.
2. Menghitung kuadrat jarak euclid (*euclidean distance*) masing-masing objek terhadap data sampel yang diberikan.
3. Mengurutkan objek-objek tersebut kedalam kelompok yang mempunyai jarak *Euclidian* terkecil
4. Mengumpulkan kategori Y (klasifikasi *nearest neighbor*)
5. Dengan menggunakan kategori mayoritas, maka dapat hasil klasifikasi

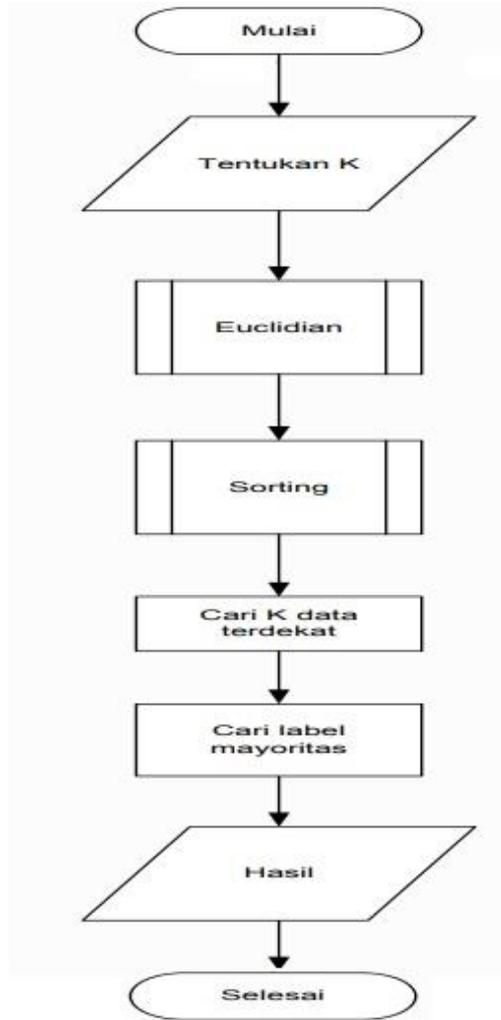
Secara umum untuk mendefinisikan jarak antara dua objek *x* dan *y*, digunakan rumus jarak *Euclidian* pada persamaan (1)[7]

$$D(x,y) = \sqrt{\sum_{k=1}^n (X_{\text{training}} - Y_{\text{testing}})^2} \quad (1)$$

dengan

- | | |
|-----------------------|---|
| X_{training} | : data training ke- <i>i</i> , |
| Y_{testing} | : data testing, |
| <i>i</i> | : record (baris) ke- <i>i</i> dari tabel, |
| <i>n</i> | : jumlah data training. |

Dimana matriks *distance* adalah jarak skala dari kedua vektor x dan y dari matriks dengan ukuran n dimensi. Pada fase training, algoritma ini hanya melakukan penyimpanan vektor-vektor fitur dan klasifikasi data training sample. Pada fase klasifikasi, fitur-fitur yang sama dihitung untuk testing data (yang klasifikasinya tidak diketahui). Jarak dari vektor baru yang ini terhadap seluruh vektor training sample dihitung dan sejumlah k buah yang paling dekat diambil. Alur dalam K-NN dapat dilihat pada gambar di bawah ini :



Gambar 3 Alur K-Nearest Neighbour

3. HASIL DAN PEMBAHASAN

3.1 Analisa Masalah

Pada proses ini penelitian ini menggunakan algoritma K-NN. Data yang digunakan adalah data mahasiswa STMIK Bina Nusantara Jaya Lubuklinggau semester V. Berikut data sample anak semester V :

Tabel 1 Data Mahasiswa Semester V

Nama	Nim	Index Prestasi				
		S1	S2	S3	S4	S5
Shinta Marzania	2015010002	3.28	2.89	3.00	2.95	3.24

Rani Okta Felani	2015010003	3.00	3.12	3.42	2.79	2.88
Riki Mardiansyah	2015010007	2.68	2.79	3.23	2.63	1.84
Rio Saputra	2015010019	1.68	1.84	1.53	2.32	2.88
Lindah	2015010029	3.21	3.58	3.35	3.00	3.11
Juliyanti	2015010033	2.68	2.79	2.95	3.68	3.23
Duwita Hartini	2015010043	2.95	4.00	3.23	3.42	3.58
Gusti Randa	2015010048	1.53	2.95	2.68	2.95	3.12
Evy Suryani	2015010054	3.00	3.58	3.23	2.68	2.32
Yuansi Oktaria	2015010061	2.95	3.12	3.58	2.68	3.42

Setelah data sample di dapatkan peneliti membagi data tersebut menjadi 2 data, yaitu data training dan data testing. Data training dan data testing dapat dilihat dibawah ini :

1. Data Training

Data training terdiri dari NIM, IP (Index Prestasi) mahasiswa dari semester satu sampai lima dan Keterangan (kategori kelulusan). Data training dapat dilihat pada tabel 2 di bawah ini

Tabel 2 Data Training

Nama	Nim	Index Prestasi					Keterangan
		S1	S2	S3	S4	S5	
Shinta Marzania	2015010002	3.28	2.89	3.00	2.95	3.24	Tepat Waktu
Rani Okta Felani	2015010003	3.00	3.12	3.42	2.79	2.88	Tepat Waktu
Riki Mardiansyah	2015010007	2.68	2.79	3.23	2.63	1.84	Terlambat
Rio Saputra	2015010019	1.68	1.84	1.53	2.32	2.88	Terlambat
Lindah	2015010029	3.21	3.58	3.35	3.00	3.11	Tepat Waktu
Juliyanti	2015010033	2.68	2.79	2.95	3.68	3.23	Tepat Waktu
Duwita Hartini	2015010043	2.95	4.00	3.23	3.42	3.58	Tepat Waktu
Gusti Randa	2015010048	1.53	2.95	2.68	2.95	3.12	Terlambat
Evy Suryani	2015010054	3.00	3.58	3.23	2.68	2.32	Terlambat

2. Data Testing

Data testing terdiri dari NIM dan IP (Index Prestasi) mahasiswa dari semester satu sampai lima. Data testing dapat dilihat pada tabel 3

Tabel 3 Data Testing

Nama	Nim	Index Prestasi					Keterangan
		S1	S2	S3	S4	S5	
Yuansi Oktaria	2015010061	2.95	3.12	3.58	2.68	3.42	?

3.2 Penerapan Algoritma K-NN

Setelah data training dan data testing sudah didapatkan langkah selanjutnya melakukan langkah-langkah dalam Algoritma K-NN.

1. Menentukan Paramater K

Disini paramater yang dipakai adalah K=5

2. Menghitung kuadrat jarak terkecil (*euclidean distance*) masing-masing objek terhadap data sample yang diberikan, dengan rumus persamaan :

$$D(x,y) = \sqrt{\sum_{k=1}^n (X_{training} - Y_{testing})^2} \quad (1)$$

Tabel 4 Data training dan testing

Nama	Nim	Index Prestasi					Keterangan
		S1	S2	S3	S4	S5	
Shinta Marzania	2015010002	3.28	2.89	3.00	2.95	3.24	Tepat Waktu
Rani Okta Felani	2015010003	3.00	3.12	3.42	2.79	2.88	Tepat Waktu
Riki Mardiansyah	2015010007	2.68	2.79	3.23	2.63	1.84	Terlambat
Rio Saputra	2015010019	1.68	1.84	1.53	2.32	2.88	Terlambat
Lindah	2015010029	3.21	3.58	3.35	3.00	3.11	Tepat Waktu
Juliyanti	2015010033	2.68	2.79	2.95	3.68	3.23	Tepat Waktu
Duwita Hartini	2015010043	2.95	4.00	3.23	3.42	3.58	Tepat Waktu
Gusti Randa	2015010048	1.53	2.95	2.68	2.95	3.12	Terlambat
Evy Suryani	2015010054	3.00	3.58	3.23	2.68	2.32	Terlambat
Yuansi Oktaria	2015010061	2.95	3.12	3.58	2.68	3.42	?

Contoh perhitungan data mahasiswa ke 1 terhadap ke 10 (data testing)

$$D(x,y) = \sqrt{\sum_{k=1}^n (X_{training} - Y_{testing})^2} \quad (1)$$

$$d_1, d_{10} = ((d_{11}-d_{101})^2 + (d_{12}-d_{102})^2 + (d_{13}-d_{103})^2 + (d_{14}-d_{104})^2 + (d_{15}-d_{105})^2)^{1/2}$$

$$d_1, d_{10} = ((3.28 - 2.95)^2 + (2.89 - 3.12)^2 + (3 - 3.58)^2 + (2.95 - 2.68)^2 + (3.24-3.42)^2)^{1/2}$$

$$d_1, d_{10} = ((0.33)^2 + (-0,23)^2 + (-0,58)^2 + (0,27)^2 + (-0,18)^2)^{1/2}$$

$$d_1, d_{10} = (0,1089 + 0,0529 + 0,3364 + 0,0729+0,0324)^{1/2}$$

$$d_1, d_{10} = (0,6035)^{1/2}$$

$$d_1, d_{10} = 0,776853$$

Berikut hasil dari *euclidean distance* masing-masing objek terhadap data testing yang diberikan

Tabel 5 Euclidean distance data testing

Nama	Index Prestasi					Keterangan	Euclidean distance data testing
	S1	S2	S3	S4	S5		
Shinta Marzania	3.28	2.89	3.00	2.95	3.24	Tepat Waktu	0,7769
Rani Okta Felani	3.00	3.12	3.42	2.79	2.88	Tepat Waktu	0,5760
Riki Mardiansyah	2.68	2.79	3.23	2.63	1.84	Terlambat	1,6743
Rio Saputra	1.68	1.84	1.53	2.32	2.88	Terlambat	2,8062
Lindah	3.21	3.58	3.35	3.00	3.11	Tepat Waktu	0,7284
Juliyanti	2.68	2.79	2.95	3.68	3.23	Tepat Waktu	1,2707
Duwita Hartini	2.95	4.00	3.23	3.42	3.58	Tepat Waktu	1,2125
Gusti Randa	1.53	2.95	2.68	2.95	3.12	Terlambat	1,7373
Evy Suryani	3.00	3.58	3.23	2.68	2.32	Terlambat	1,2436

- Mengurutkan objek-objek tersebut ke dalam kelompok yang mempunyai jarak *euclidian* terkecil. Setelah objek-objek tersebut di urutkan maka menentukan peringkat dari data yang mempunyai jarak terkecil ke terbesar, hasil pengurutan peringkat jarak dapat di lihat pada tabel 6 di bawah ini :

Tabel 6 Pengurutan Peringkat Jarak

Nama	Keterangan	Euclidean distance data testing	Peringkat Jarak Terkecil
Shinta Marzania	Tepat Waktu	0,7769	3

Rani Okta Felani	Tepat Waktu	0,5760	1
Riki Mardiansyah	Terlambat	1,6743	7
Rio Saputra	Terlambat	2,8062	9
Lindah	Tepat Waktu	0,7284	2
Juliyanti	Tepat Waktu	1,2707	6
Duwita Hartini	Tepat Waktu	1,2125	4
Gusti Randa	Terlambat	1,7373	8
Evy Suryani	Terlambat	1,2436	5

4. Mengumpulkan kategori Y (klasifikasi *nearest neighbor*)
 Setelah mendapatkan peringkat pada jarak, maka pada langkah selanjutnya mengumpulkan kategori Y (klasifikasi *nearest neighbor*). Pada langkah 1, parameter (k) yang kita tentukan adalah 5. Hasil klasifikasi K-NN dapat dilihat pada tabel 7 di bawah ini :

Tabel 7 Hasil Klasifikasi K-NN

Nama	Keterangan	Euclidean distance data testing	Peringkat Jarak	Klasifikasi
Shinta Marzania	Tepat Waktu	0,7769	3	Ya
Rani Okta Felani	Tepat Waktu	0,5760	1	Ya
Riki Mardiansyah	Terlambat	1,6743	7	Tidak
Rio Saputra	Terlambat	2,8062	9	Tidak
Lindah	Tepat Waktu	0,7284	2	Ya
Juliyanti	Tepat Waktu	1,2707	6	Tidak
Duwita Hartini	Tepat Waktu	1,2125	4	Ya
Gusti Randa	Terlambat	1,7373	8	Tidak
Evy Suryani	Terlambat	1,2436	5	Ya

5. Dengan menggunakan kategori mayoritas, maka dapat hasil klasifikasi
 Setelah mengumpulkan kategori Y dengan K =5 maka langkah selanjutnya adalah mengumpulkan kategori mayoritas. Hasil dari kategori mayoritas dapat di lihat pada tabel 8 di bawah ini :

Tabel 8 Hasil Kategori Mayoritas

Nama	Keterangan	Euclidean distance data testing	Peringkat Jarak	Klasifikasi
Shinta Marzania	Tepat Waktu	0,7769	3	Ya
Rani Okta Felani	Tepat Waktu	0,5760	1	Ya
Lindah	Tepat Waktu	0,7284	2	Ya
Duwita Hartini	Tepat Waktu	1,2125	4	Ya
Evy Suryani	Terlambat	1,2436	5	Ya

Dari hasil klasifikasi diatas jadi, data testing termasuk kategori Tepat Waktu dikarenakan jumlah mayoritas Tepat Waktu berjumlah 4 dan Terlambat berjumlah 1. Hasil data testing dapat dilihat pada tabel 9 dibawah ini :

Tabel 9 Hasil Data Testing

Nama	Nim	Index Prestasi					Keterangan
		S1	S2	S3	S4	S5	
Shinta Marzania	2015010002	3.28	2.89	3.00	2.95	3.24	Tepat Waktu
Rani Okta Felani	2015010003	3.00	3.12	3.42	2.79	2.88	Tepat Waktu
Riki Mardiansyah	2015010007	2.68	2.79	3.23	2.63	1.84	Terlambat

Rio Saputra	2015010019	1.68	1.84	1.53	2.32	2.88	Terlambat
Lindah	2015010029	3.21	3.58	3.35	3.00	3.11	Tepat Waktu
Juliyanti	2015010033	2.68	2.79	2.95	3.68	3.23	Tepat Waktu
Duwita Hartini	2015010043	2.95	4.00	3.23	3.42	3.58	Tepat Waktu
Gusti Randa	2015010048	1.53	2.95	2.68	2.95	3.12	Terlambat
Evy Suryani	2015010054	3.00	3.58	3.23	2.68	2.32	Terlambat
Yuansi Oktaria	2015010061	2.95	3.12	3.58	2.68	3.42	Tepat Waktu

4. KESIMPULAN

Berdasarkan penerapan teknik klasifikasi menggunakan metode K-NN di atas maka peneliti mengambil kesimpulan sebagai berikut :

1. Data training yang dinilai adalah hasil dari Index Prestasi semester, sehingga kurang efektif dikarenakan hanya hasil dari Index Prestasi tidak dari nilai masing-masing mata kuliah.
2. Teknik klasifikasi menggunakan metode K-NN dapat memprediksi kelulusan.
3. *Output* yang dihasilkan adalah keterangan dari Data testing.
4. Berdasarkan penerapan dari teknik klasifikasi menggunakan metode K-NN didapatkan hasil bahwa mahasiswa Yuansi Oktaria mendapatkan keterangan Tepat Waktu
5. Banyaknya data training mempengaruhi hasil dari metode K-NN

5. SARAN

Penelitian ini menggunakan data beberapa mahasiswa STMIK Bina Nusantara Jaya Lubuklinggau untuk menerapkan teknik klasifikasi menggunakan metode K-NN, untuk penelitian lebih lanjut peneliti akan menggunakan semua data mahasiswa untuk memprediksi semua mahasiswa tingkat akhir agar penelitian ini menjadi lebih objektif.

DAFTAR PUSTAKA

- [1] A. Rohman, 2015 “*Model Algoritma K-Nearest Neighbor (K-Nn) Untuk Prediksi Kelulusan Mahasiswa*, Fakultas Teknik Universitas Pandanaran Semarang.”
- [2] Sumarlin, 2015 “*Implementasi Algoritma K-Nearest Neighbor Sebagai Pendukung Keputusan Klasifikasi Penerima Beasiswa PPA dan BBM*,” *J. Sist. Inf. Bisnis*, vol. 01, pp. 52–62,.
- [3] R. I. Ndaumanu and M. R. Arief, Kusri, 2014 “*Analisis Prediksi Tingkat Pengunduran Diri Mahasiswa dengan Metode K-Nearest Neighbor*,” *Jatiji*, vol. 1, no. 1, pp. 1–15,.
- [4] N. Krisandi, B. Prihandono, and Helmi, 2013 “*Algoritma K - Nearest Neighbor Dalam Klasifikasi Data Hasil Produksi Kelapa Sawit Pada PT. MINAMAS Kecamatan Parindu*,” *Bul. Ilm. Math.Stat.dan Ter.*, vol. 02, no. 1, pp. 33–38,.
- [5] L.Mei, 2014 “*Penerapan Algoritma Klasifikasi Nearest Neighbor (K-NN) Untuk Mendeteksi Penyakit Jantung*,” *Faktor Exacta*, Vol.7, no.4, Universitas Indraprasta,.
- [6] M. S. Mustafa and I. W. Simpen, 2014 “*Perancangan Aplikasi Prediksi Kelulusan Tepat Waktu Bagi Mahasiswa Baru Dengan Teknik Data Mining (Studi Kasus : Data Akademik Mahasiswa STMIK Dipanegara Makassar)*,” *Citec J.*, vol. Vol. 1, No, pp. 270–281,.
- [7] Mustakim, Giantika O, 2016 “*Algoritma K-Nearest Neighbor Classification Sebagai Sistem Prediksi Predikat Prestasi Mahasiswa*,” *J. Sains dan Teknol. Ind.*, vol. 13, no. 2, pp. 195–202,.
- [8] C. S. Fatoni and F. D. Noviantha, 2017 “*Case Based Reasoning Diagnosis Penyakit Difteri dengan Algoritma K-Nearest Neighbor*,” *Citec J.*, vol. 4, pp. 220–232,.
- [9] Tanjung I.H, 2012, “*Peramalan Jumlah Penjualan Distributor Telur Terhadap Permintaan Pasar Menggunakan Metode Average-Based Fuzzy Time Series (ABFTS)*”, Program Sarjana, Universitas Sumatra Utara, Medan
- [10] Kusri, Emha T. Luthfi, 2009, “*Algoritma Data Mining*”. Andi, Yogyakarta
- [11] Fayyad, Usama, 1996. *Advances in Knowledge Discovery and Data Mining*. MIT Press.
- [12] Ian H. Witten, Frank Eibe, Mark A. Hall, *Data Mining: Practical Machine Learning Tools and Technique 3rd Edition*, Elsevier, 2011.