

PREDIKSI KELULUSAN MAHASISWA STMIK DHARMAWACANA MENGUNAKAN ALGORITMA K-NEAREST NEIGHBORS

¹Febri Sugandi

[¹febri.sugandi@dharmawacana.ac.id](mailto:febri.sugandi@dharmawacana.ac.id)

¹STMIK DHARMAWACANA

Abstract: Prediction of graduation is a key factor for each school based on academic performance each semester. STMIK Dharmawacana has difficulties in predicting, so in predicting it requires machine learning. In this article the K-NN (K-Nearest Neighbor) algorithm is used in discussions at STMIK Dharmawacana using the Rapidminer application, with student graduation data for 2014-2018 and test data, namely class of 2019, for informatics technology graduates and information systems graduates. The results showed that the convergence of K-NN with K-5 and K-Fold Cross Validation with k=4 accuracy reached 97.15%. This shows that K-NN is an algorithm that can predict.

Keywords: Prediction, K-Nearest Neighbor, STMIK Dharmawacana

Abstrak: Prediksi kelulusan merupakan faktor kunci bagi setiap sekolah berdasarkan kinerja akademik setiap semester. STMIK Dharmawacana mengalami kesulitan didalam memprediksi, maka didalam memprediksi diperlukanya machine learning. Dalam artikel ini algoritma K-NN (K-Nearest Neighbour) digunakan dalam pembahasan pada STMIK Dharmawacana menggunakan aplikasi Rapidminer, dengan data kelulusan mahasiswa tahun 2014-2018 dan data uji yaitu angkatan tahun 2019, untuk sarjana teknologi informatika dan sarjana sistem informasi. Hasil penelitian menunjukkan bahwa konvergensi K-NN dengan K-5 dan K-Fold Cross Validation dengan k=4 akurasi mencapai 97,15%. Hal ini menunjukkan bahwa K-NN merupakan algoritma dapat dalam memprediksi.

Kata Kunci: Prediksi, K-Nearest Neighbor, STMIK Dharmawacana

I. PENDAHULUAN

Salah satu keunggulan aplikasi SIAKAD (Sistem Informasi Akademik) di perguruan tinggi adalah dapat menggambarkan indeks kelulusan hasil akademik setiap mahasiswa untuk menilai kinerjanya dalam mengikuti mata kuliah

yang dipelajari setiap semester. Perguruan tinggi, melalui lembaga pengolah data khusus seperti BAAK (Baka Akademik dan Kemahasiswaan), sering memberikan laporan hasil semester yang disampaikan kepada mereka yang disampaikan kepada Lembaga Layanan Pendidikan Tinggi (LLDikti) setiap semester, meminta

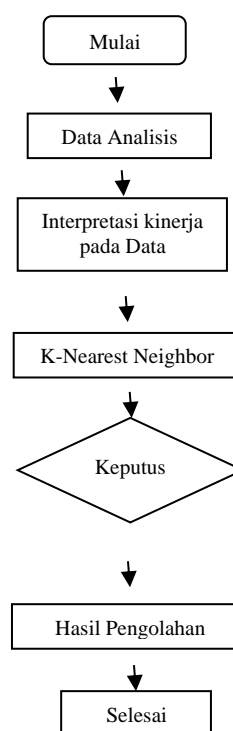
keakuratan kelengkapan data tentang prestasi akademik siswa. peringkat. Kebutuhan ini dapat dikembangkan melalui penelitian pada kampus.

Guna meningkatkan kualitas kelulusan dan meningkatkan rekognisi STMIK Dharma-wacana, perlu dilakukan prediksi mahasiswa mana yang akan lulus tepat waktu. Memprediksi keberhasilan siswa dapat memberikan wawasan yang bermanfaat tentang data besar siswa. Sehingga diperlukan suatu metode untuk mengatasi permasalahan tersebut dengan metode data mining.

Pada artikel ini menggunakan metode K-NN untuk mengukur tingkat kelulusan mahasiswa pada STMIK Dharmawacana. STMIK Dharmawacana sendiri belum pernah mengukur tingkat kelulusan mahasiswa. Oleh karena itu akan dilakukan penelitian untuk mengukur kelulusan mahasiswa STMIK Dharmawacana dengan menggunakan metode K-Nearest Neighbor dalam pengklasifikasian kelulusan siswa menggunakan variabel-variabel yang dapat dijadikan kriteria untuk menentukan kelulusan mahasiswa. Salah satu metode klasifikasi yang mungkin berguna adalah algoritma K-NN untuk mengukur tingkat akurasi yang dapat mendeskripsikan kelulusan mahasiswa di STMIK Dharmawacana.

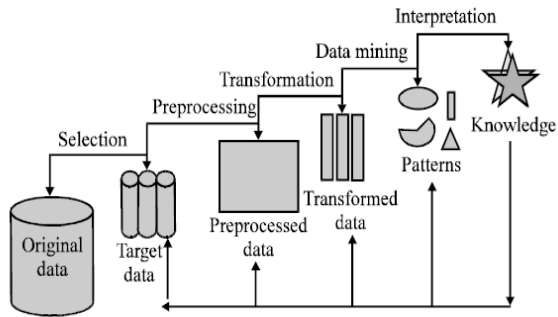
II. METODE

Pada gambar menunjukkan tahapan-tahapan penelitian menggunakan algoritma K-NN.



Gambar 1. Tahapan Metode Penelitian

Langkah penting dalam data mining adalah preprocessing. Pada data yang akan di proses untuk penambangan tidak selalu dalam kondisi yang baik dalam melakukan pemrosesan. Terkadang, pada data masih terdapat berbagai masalah sehingga dapat mempengaruhi hasil proses data itu sendiri, seperti data yang rusak (nilai hilang, outlier, atau format yang tidak sesuai).



Gambar 2. Posisi data preprocessing pada data mining

Jumlah data yang akan diolah adalah 1017, dimana 917 merupakan data latihan dan 100 data uji. Selama pengumpulan data langsung dalam format excel. Data kemudian akan dibagi menjadi dua bagian menjadi data latihan dan data uji menggunakan cross validation.

III. HASIL DAN PEMBAHASAN

Analisis Kebutuhan

Pada proses pengumpulan data menggunakan data kelulusan mahasiswa S1 Teknik Informatika dan Sistem Informasi pada STMIK Dharmawacana angkatan 2014 sampai 2018 sebagai data latihan dan angkatan 2019 merupakan data uji. Pada data raw atribut yang digunakan adalah NM (Nama_Mahasiswa), JK (Jenis_Kelamin), TTL (Tempat_Tanggal_Lahir), Alamat, IPK (Indeks_Prestasi_Kumulatif), UKM (Unit_Kegiatan_Mahasiswa), Penghasilan

Orang Tua dan TPA (Test_Potensi_Akademik, dan kelulusan (Tepat_Waktu dan Tidak_Tepat_Waktu). Pada tabel dibawah ini perolehan data penelitian :

Tabel 1. Data Latihan (Raw Data Latihan)

Nama	N	Tempat Tanggal Lahir	Alamat	Progres Dulu	IPK	UKM	Penghasilan Peng. Tua	Penghasilan Peng. Mahasiswa	KELOMPOK
BAGAS PRATIWI	1	BOJONEGARA 25-08-1994	Des. Sidareja- Kota Banjar Lampung- Perse Lampung	02 Mei 2019	2,8	1	1.000.000	10.000	Tepat Waktu

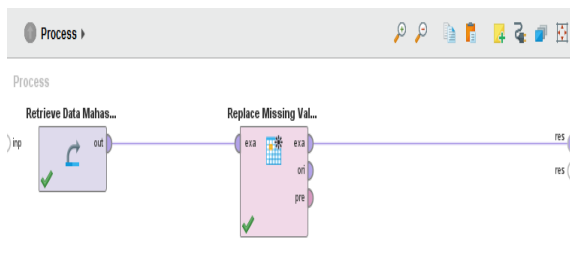
Tabel 2. Data Uji (Raw Data Uji)

Nama	N	Tempat Tanggal Lahir	Alamat	Progres Dulu	IPK	UKM	Penghasilan Peng. Tua	Penghasilan Peng. Mahasiswa	Substansi
ABDUL RAZAK	1	Banda Aceh 15 Desember 1994	Des. Pening- Kota Bandar Lampung- Perse Lampung	02 Mei 2019	2,8	0	1.200.000	10.000	?

Proses Pembersihan Data (Data Cleaning)

Dalam penelitian ini tahapan pembersihan data dilakukan dahulu untuk

membuat kualitas data karena berawal dari data yang rusak (noise) sehingga mempengaruhi kualitas data, untuk mendapatkan data yang berkualitas yaitu dilakukan modifikasi, mengisi data yang kosong dan dapat melakukan penghapusan pada data yang rusak.



Gambar 3. Tahap Pembersihan Data (Preprocessing Data)

Selection Data

Pemilihan data dilakukan dengan menyeleksi data yang dibutuhkan dan menghapus atau membuang data yang tidak diperlukan. Data kelulusan di STMIK Dharmawacana S1 Teknik Informatika dan Sistem Informasi angkatan 2014 sampai 2018 digunakan sebagai data latih, sedangkan angkatan 2019 digunakan sebagai data uji.

Tabel 3. Seleksi Data

No	Atribut	Ket
1	NM (Nama Mahasiswa)	√
2	JK (Jenis Kelamin)	X
3	TTL	X
4	Alamat	X

5	PS (Program Studi)	X
6	IPK	√
7	UKM	√
8	Penghasilan Orang Tua	√
9	TPA	√
10	Kelulusan	√

Transformasi Data

Pada data batch tahun 2014 hingga 2018 dan 2019 di STMIK Dharmawacana, data tersebut akan dikonversi menjadi Nama Mahasiswa, IPK (Indeks Prestasi Mahasiswa), UKM (Usaha Kegiatan Mahasiswa), P.O.T (Penghasilan Orang Tua), TPA (Test Potensi Akademik) serta Kelulusan Mahasiswa, seperti berikut:

Tabel 4. Atribut Input

Atribut	Keterangan
Indeks Prestasi Kumulatif	> 3.00 = Tinggi 2.70 – 3.00 = Sedang < 2.70 = Rendah
Unit Kegiatan Mahasiswa	< 1 = Tinggi 1 – 2 = Sedang > 2 = Rendah
Penghasilan Orang Tua	> 3500000 = Tinggi 2000000 – 3500000 = Sedang < 2000000 = Rendah
Test Potensi Akademik	> 80 = Tinggi 60 - 80 = Sedang < 60 = Rendah

Dimana atribut yang dijadikan sebagai target adalah sebagai berikut :

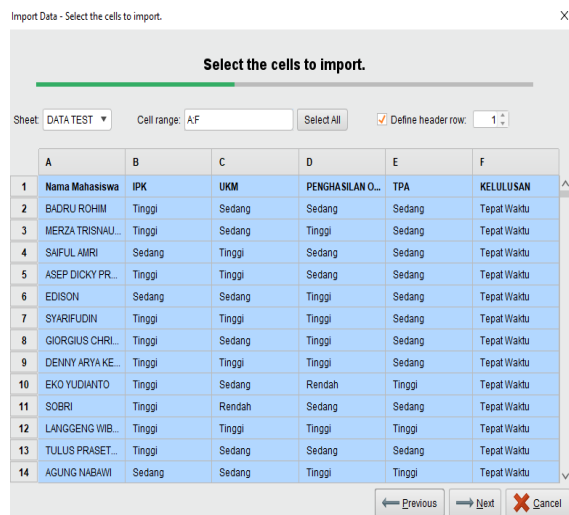
Tabel 5. Atribut Output

Atribut	Keterangan
Kelulusan	Tepat Waktu Tidak Tepat Waktu

Pembahasan

Import Data

Pengimporan dirapidminer menggunakan data 2014 sampai 2019 sebanyak 1017 data yang mempunyai 6 atribut, sebagai berikut :

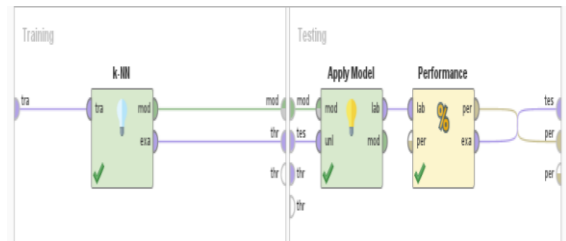


Gambar 4. Data Import

Pengujian data di Rapidminer

Rapidminer digunakan pada pengolahan data dipenelitian ini. Data yang akan diproses adalah angkatan 2014 sampai dengan 2019 yang berjumlah 1017 yang

kemudian akan dilakukan pengujian menggunakan metode K-NN. Pengolahan data menggunakan K-NN dapat dilihat seperti berikut:



Gambar 5. Pengolahan data pada rapidminer menggunakan metode K-NN

Confusion Matrix

Confusion matrix adalah ukuran dari kinerja klasifikasi menggunakan metode K-NN yang berupa akurasi, presisi dan recall (performa):

Tabel 6. Performa Kinerja K-NN

accuracy: 97.15% +/- 1.42% (micro average: 97.15%)

	true Tepat Waktu	true Tidak Tepat Waktu	class precision
pred. Tepat Waktu	911	8	99.13%
pred. Tidak Tepat Waktu	21	77	78.57%
class recall	97.75%	90.50%	

Berikut perhitungan akurasi, presisi dan recall:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$$

$$= \frac{911+77}{911+77}$$

$$= \frac{911 + 77 + 8 + 21}{911 + 77 + 8 + 21}$$

$$= 97,15\%$$

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$= \frac{911}{911 + 8}$$

$$= 99,13\%$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

$$= \frac{911}{911 + 21}$$

$$= 97,75\%$$

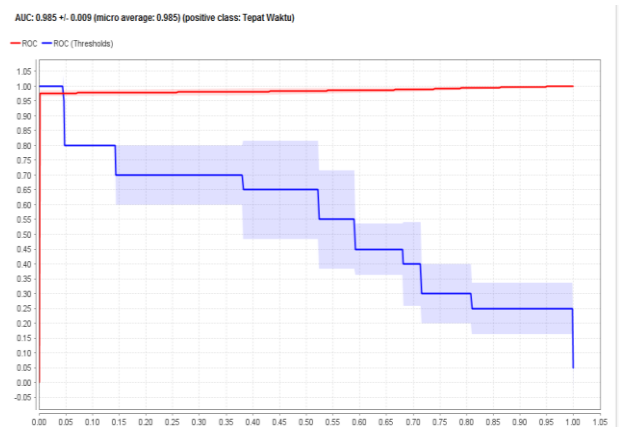
Perhitungan diatas didapati hasil accuracy 97.15%, Precission 99.13% dan Recall 97.75%, kelulusan angkatan 2014 sampai 2019 di STMIK Dharmawacana, berikut merupakan tabel AUC hasil daripada prediksi dengan metode K-NN.

Row No.	KELULUSAN	prediction(K...	confidence...	confidence...	MM	Nama Maha...	IPK	UKM	PENGHASIL...
1	Tepat Waktu	Tepat Waktu	1	0	13111001	BADRU ROHM	Tinggi	Sedang	Sedang
2	Tepat Waktu	Tepat Waktu	1	0	13111014	EKO WAHYU...	Tinggi	Tinggi	Tinggi
3	Tepat Waktu	Tepat Waktu	1	0	13111017	IMAM SAPTA...	Sedang	Tinggi	Sedang
4	Tepat Waktu	Tepat Waktu	1	0	13111020	AHMAD YUSUF	Tinggi	Tinggi	Sedang
5	Tepat Waktu	Tepat Waktu	1	0	13111036	RISTA RISTA...	Tinggi	Tinggi	Tinggi
6	Tepat Waktu	Tepat Waktu	1	0	13111040	MUHAMMAD...	Sedang	Sedang	Tinggi
7	Tepat Waktu	Tepat Waktu	1	0	13111063	REZA RAMAD...	Sedang	Tinggi	Tinggi
8	Tepat Waktu	Tepat Waktu	1	0	13111066	MHROB ALF...	Sedang	Tinggi	Tinggi
9	Tepat Waktu	Tepat Waktu	1	0	13111069	M ALDI	Sedang	Tinggi	Tinggi
10	Tepat Waktu	Tepat Waktu	1	0	13111074	ADI SUBARK...	Tinggi	Sedang	Tinggi
11	Tepat Waktu	Tepat Waktu	1	0	13111076	JIMMI MANALU	Sedang	Sedang	Tinggi
12	Tepat Waktu	Tepat Waktu	1	0	13111085	ADITHA RAMA...	Sedang	Sedang	Rendah
13	Tidak Tepat...	Tidak Tepat...	0.400	0.600	13111086	RISKY RAMA...	Rendah	Sedang	Tinggi

Gambar 6. Hasil Prediksi Algoritma K-Nearest Neighbor

IV. SIMPULAN

Hasil dari prediksi kelulusan siswa STMIK Dharmawacana dengan metode K-NN didapat akurasi sebesar 97,15%, presisi sebesar 99,13% dan Recall sebesar 97,75%. Hal ini menunjukkan bahwa keunggulan dari metode K-Nearest Neighbour dapat digunakan dalam memprediksi.



Gambar 7. Diagram ROC K-NN

DAFTAR RUJUKAN

- Abdullah Baz, Fatima Alshareef, Ebtihal Alshareef, Hosam Alhakami, Tahani Alsubait. 2020. "Predicting Students' Academic Performance Using Naïve Bayes". *IJCSNS International Journal of Computer Science and Network Security*.
- Alpaydin, Ethem. 2014. *Introduction to Machine Learning*. Second Edition. The MIT Press: Cambridge, Massachusetts.
- Artay, Ketut. 2015. "Implementation of Naive Bayes Classifier Method to Predict Graduation Time of IBI Darmajaya Scholar". *International Conference on Information Technology and Business*:1-2.
- Bonaccorso, Giuseppe. 2017. *Machine Learning Algorithm*. Packt: Birmingham-Mumbai.
- Budiharto, Widodo. 2016. *Machine Learning & Computational Intelligence*. Penerbit Andi: Yogyakarta.
- Cormen, Thomas H, dkk. 2009. *Introduction to Algorithm Third Edition*. The MIT Press: United States of America.
- Effendy, Faried & Purbandini. 2018. "Klasifikasi Rumah Tangga Miskin Menggunakan Ordinal Class Classifier". *Jurnal Teknologi dan Sistem Informasi* 4 (1): 30–36.
- G.Vadivu, K.Sornalakshmi. 2017. "Applying Machine Learning Algorithms for Student Employability Prediction Using R". *Int. J. Pharm. Sci.Rev. Res.*, Vol.43(1),No.11, pp: 38-41.
- Hastuti, K. 2012. "Analisis Komparasi Algoritma Klasifikasi Data Mining untuk Prediksi Mahasiswa Non Aktif". *Seminar Nasional Teknologi Informasi & Komunikasi Terapan* 14(1): 241–249.
- Mustakim & Giantika Oktaviani F. 2016. "Algoritma K-Nearest Neighbor Classification Sebagai Sistem Prediksi Predikat Kelulusan Mahasiswa". *Jurnal Sains, Teknologi dan Industri*: 1-7.
- Mohammad Imron Satia Angga Kusumah. 2018. "Penerapan Metode Klasifikasi Data Mining untuk Memprediksi Kelulusan Siswa menggunakan Algoritma K-Nearest Neighbor (K-NN)". *CITISEE*, ISBN: 1-3.