

# PENERAPAN ALGORITMA *K-NEAREST NEIGHBORS* UNTUK PREDIKSI KELULUSAN MAHASISWA PADA STMIK SINAR NUSANTARA SURAKARTA

Agus Panoto ( [ragilpanoto@gmail.com](mailto:ragilpanoto@gmail.com) )  
Yustina Retno Wahyu Utami ( [yustina.retno@gmail.com](mailto:yustina.retno@gmail.com) )  
Wawan Laksito YS ( [wlaksito@yahoo.co.id](mailto:wlaksito@yahoo.co.id) )

## ABSTRAK

*Penelitian ini difokuskan untuk menerapkan perhitungan Algoritma K-Nearest Neighbors (K-NN) pada aplikasi prediksi kelulusan mahasiswa yang nantinya diharapkan bisa sebagai alat bantu untuk memprediksi kelulusan mahasiswa. Dalam hal ini penulis memilih algoritma K-Nearest Neighbors (K-NN) karena K-Nearest Neighbors (K-NN) merupakan algoritma 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. Input dari sistem ini adalah Data sampel mahasiswa yang sudah dinyatakan lulus yang akan digunakan sebagai data training dan data testing. Kriteria yang digunakan dalam penelitian ini adalah jenis kelamin, status nikah, status kerja, IP Semester 1-6, capaian SKS, Status Kelulusan. Output dari sistem ini berupa hasil prediksi kelulusan mahasiswa. Hasil pengujian menunjukkan bahwa prediksi dari Algoritma KNN dari 20 data testing hasil prediksinya adalah 18 benar dan hanya 2 yang salah sehingga dapat dikatakan bahwa akurasi prediksi dengan naive bayes menggunakan 40 data training untuk dan 20 data testing, Algoritma KNN mempunyai akurasi 90%.*

*Kata kunci : Prediksi Kelulusan, algoritma K-Nearest Neighbors, Sistem Pendukung Keputusan*

## I. PENDAHULUAN

STMIK Sinar Nusantara merupakan salah satu perguruan tinggi yang sukses menarik banyak mahasiswa di setiap periodenya. Namun ada beberapa hal yang tidak seimbang antara masuk dan keluarnya mahasiswa yang telah menyelesaikan studinya. Mahasiswa yang masuk dalam jumlah besar, tetapi mahasiswa yang lulus tepat waktu sesuai dengan ketentuan jauh sangat kecil dibandingkan masuknya. Sehingga terjadi penumpukan mahasiswa dalam jumlah tinggi disetiap periode kelulusan.

Oleh karena itu untuk meningkatkan kualitas dan akreditasi pada perguruan tinggi STMIK Sinar Nusantara, maka haruslah ada filter pada mahasiswa yang mendaftar untuk masuk dan perlu di prediksi mahasiswa yang lulus tepat waktu karena berpengaruh langsung terhadap kapasitas ruang, rasio dosen terhadap mahasiswa, dan kapasitas tempat parkir. Prediksi tingkat kelulusan mahasiswa merupakan salah satu metode yang tepat untuk membentuk pola-pola yang mungkin memberikan indikasi yang bermanfaat pada data mahasiswa yang dalam jumlah besar.

Dalam hal ini penulis memilih algoritma *K-Nearest Neighbors* karena merupakan

algoritma 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 [1].

Sistem Prediksi Tingkat Kelulusan Mahasiswa ini di mulai dengan Input Data Training yaitu data mahasiswa yang sudah lulus di STMIK Sinar Nusantara, kemudian input data testing, setelah itu dilanjutkan dengan perhitungan KNN kemudian langkah selanjutnya diperoleh hasil prediksi dari data testing yaitu status kelulusan cepat, tepat, terlambat atau sangat terlambat.

## II. METODE PENELITIAN

### A. Data

Pada penelitian ini digunakan data sebagai berikut:

#### 1. Data *training*

Data ini berupa data sampel mahasiswa yang sudah dinyatakan lulus. Data ini memiliki atribut jenis kelamin, status kerja, status nikah, umur, IP smester 1-6 dan capaian SKS dengan status kelulusan cepat, tepat, terlambat, sangat terlambat.

## 2. Data testing

Data ini digunakan sebagai data untuk pengujian sistem yang setatus kelulusannya belum diketahui dan mempunyai atribut yang sama dengan data *training*.

## B. Tahapan Penelitian

Pada penelitian ini, tahapan penelitian yang dilakukan secara garis besar meliputi beberapa kegiatan yaitu pembuatan proposal, implementasi Algoritma KNN, pengujian.

## C. Tahapan Implementasi

Impelementasi Sistem Prediksi Tingkat Kelulusan ini dilakukan dengan menggunakan paket perangkat lunak *xampp* dan *browser mozilla firefox*. Penulis menggunakan komputer dengan spesifikasi Intel dualcore 2,30 GHZ, RAM 2 GB dan Harddisk 120 GB dengan *operating system* Microsoft Windows 7.

## D. Tahapan Pengujian

Dari pengujian Sistem ingin mencari tingkat akurasi Algoritma KNN untuk Prediksi Masa Studi Mahasiswa.

## III. TINJAUAN PUSTAKA

### A. Prediksi

Prediksi mirip seperti klasifikasi dan estimasi, tapi hasilnya untuk memprediksi masa depan. Contoh, memprediksi harga barang tiga bulan mendatang, memprediksi prosentase kenaikan angka kematian karena kecelakaan tahun mendatang jika kecepatan berkendara dinaikkan. Metode dan teknik untuk klasifikasi dan estimasi, jika cocok, bisa juga digunakan untuk prediksi, termasuk metode statistik [2].

### B. Kelulusan

Kelulusan adalah tanda keberhasilan seseorang yang telah menempuh ujian. Bagi seorang siswa atau mahasiswa merupakan batu loncatan untuk melangkah ke jenjang selanjutnya [3].

### C. *K-Nearest Neighbors*

Algoritma KNN merupakan sebuah metode untuk melakukan klasifikasi terhadap obyek baru berdasarkan (K) tetangga terdekatnya [4][5]. KNN termasuk algoritma supervised learning, dimana hasil dari query instance yang baru, diklasifikasikan berdasarkan mayoritas dari kategori pada KNN. Kelas yang paling banyak muncul yang

akan menjadi kelas hasil klasifikasi. Adapun Urutan Algoritma KNN sebagai berikut

- Menentukan parameter K (jumlah tetangga paling dekat)
- Menghitung kuadrat jarak euclidian (*euclidean distance*) masing-masing obyek terhadap data sampel yang diberikan

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

- Mengurutkan objek-objek tersebut ke dalam kelompok yang mempunyai jarak euclid terkecil
- Mengumpulkan kategori Y (klasifikasi nearest neighbors)

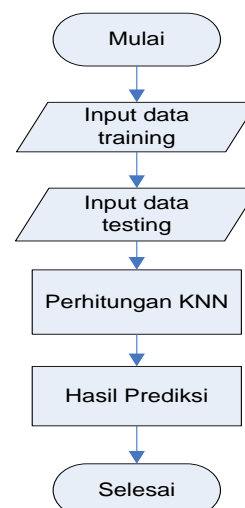
Dengan menggunakan kategori mayoritas, maka dapat diprediksikan nilai query instance yang telah dihitung.

## IV. HASIL DAN PEMBAHASAN

### A. Persiapan Data

Pada tahap pengujian ini menggunakan data mahasiswa yang sudah dinyatakan lulus dengan kategori cepat, tepat, terlambat, sangat terlambat.

### B. Diagram Alir Prediksi Kelulusan



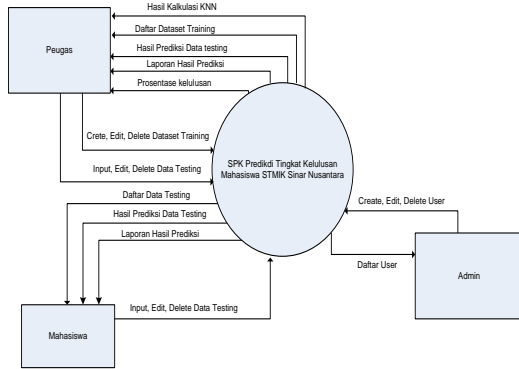
Gambar 1. Diagram Alir Prediksi Tingkat Kelulusan

Diagram alir data diatas menggambarkan algoritma secara umum semua proses yang ada dalam Sistem Prediksi Tingkat Kelulusan. Proses diawali dengan melakukan input data training yaitu data yang didapat dari mahasiswa yang sudah lulus, kemudian input data testing yaitu data yang ingin di prediksi status kelulusannya, setelah

itu dilakukan perhitungan dengan algoritma KNN, kemudian diperoleh hasil prediksi.

### C. Context Diagram

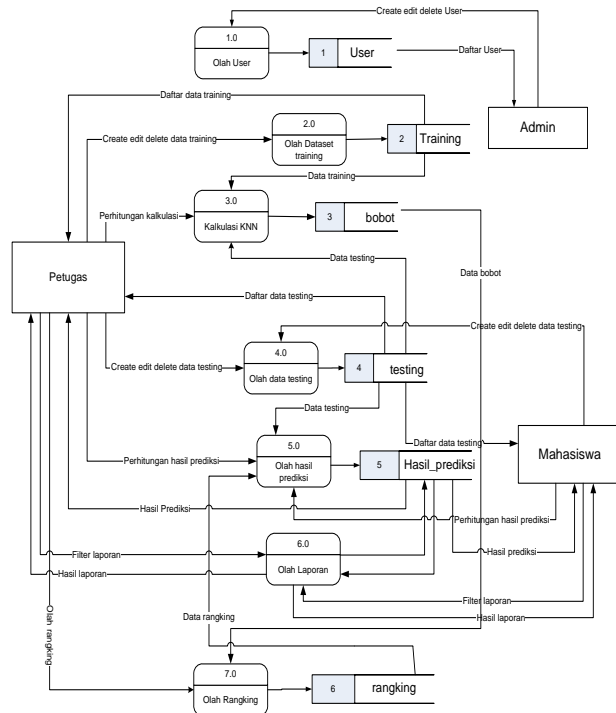
Context Diagram Aplikasi Prediksi Masa Studi Mahasiswa STMIK Sinar Nusantara menggambarkan system aliran program secara umum, yang digambarkan sebagai berikut :



Gambar 2. Context Diagram

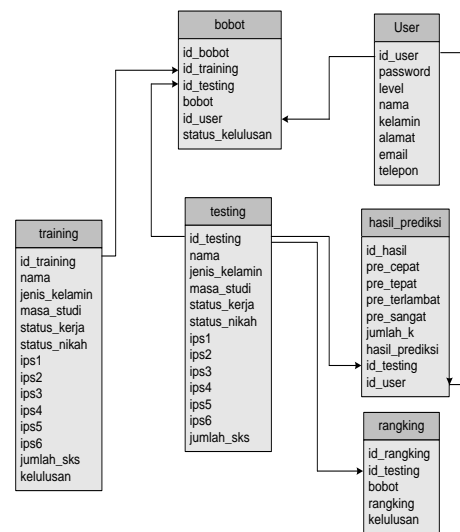
### D. Data Flow Diagram

Berikut ini Data Flow Diagram Aplikasi Prediksi Masa Studi Mahasiswa STMIK Sinar Nusantara



Gambar 3. Data Flow Diagram

### E. Perancangan Database



Gambar 4. Desain Database

Dari relasi tabel basis data diatas dapat dijelaskan bahwa hubungan antara testing dengan bobot adalah id\_testing yang hubungannya adalah one to many yaitu satu data testing dapat memiliki banyak bobot. Sedangkan hubungan antara user dan hasil prediksi adalah one to many satu user dapat melakukan banyak prediksi dan hubungan antara testing dan hasil prediksi adalah one to one yaitu satu data testing hanya memiliki satu hasil prediksi.

### F. Implementasi Antar Muka

#### 1. Halaman input data training

Halaman ini digunakan untuk memasukkan data training.



Gambar 3. Halaman input data training

## 2. Halaman input data *testing*

Halaman ini merupakan halaman yang digunakan untuk memasukkan data *testing*.

Gambar 4. Halaman Input Data testing

## 3. Halaman Kalkulasi KNN

Halaman ini merupakan tampilan dari data *training* dan data *testing* serta hasil prediksi setelah dihitung dengan Algoritma KNN.

Gambar 5. Halaman Kalkulasi KNN

## 4. Halaman Hasil Prediksi

Halaman ini merupakan halaman hasil dari prediksi data *testing*.

Gambar 6. Halaman hasil prediksi

## G. Pengujian Sistem

Pada pengujian sistem ini bertujuan mencari akurasi algoritma sehingga diketahui tingkat akurasi dari algoritma KNN dengan menggunakan data mahasiswa. Pada uji coba sistem ini digunakan 60 data mahasiswa dimana dibagi menjadi 2 yaitu 40 data training dan 20 data testing. Berikut ini data testing yang digunakan untuk mengukur akurasi algoritma KNN

Tabel 4. Tabel Perbandingan Hasil Pengujian

NAMA	STATUS KELULUSAN	HASIL PREDIKSI KNN	PREDIKSI
Agus	Tepat	Tepat	BENAR
Ani	Terlambat	Terlambat	BENAR
Adinda	Cepat	Cepat	BENAR
Anto	Sangat Terlambat	Sangat Terlambat	BENAR
Surati	Tepat	Tepat	BENAR
Andreas	Sangat Terlambat	Sangat Terlambat	BENAR
Ajeng	Cepat	Cepat	BENAR
Patro	Tepat	Tepat	BENAR
Satria	Cepat	Cepat	BENAR
Surti	Cepat	Cepat	BENAR
Jauri	Tepat	Tepat	BENAR
Lusi	Terlambat	Terlambat	BENAR
Hengki	Cepat	Cepat	BENAR
Handoko	Cepat	Cepat	BENAR
Danang	Terlambat	Terlambat	BENAR
Nurul Safarah	Sangat Terlambat	Sangat Terlambat	BENAR
Sandi	Sangat Terlambat	Sangat Terlambat	BENAR

NAMA	STATUS KELULUSAN	HASIL PREDIKSI KNN	PREDIKSI
Tiyas	Terlambat	Sangat Terlambat	SALAH
Ndari	Tepat	Cepat	SALAH
Ratih	Terlambat	Terlambat	BENAR



Gambar 7. Perbandingan Hasil Prediksi

Dari data perbandingan diatas dapat terlihat bahwa hasil prediksi dari KNN dari 20 data testing hasil prediksinya adalah 18 benar dan hanya 2 yang salah sehingga dapat dikatakan bahwa akurasi prediksi dengan KNN menggunakan 40 data training dan 20 data testing, Algoritma KNN mempunyai akurasi 90%.

## V. PENUTUP

### 5.1 Kesimpulan

Dari Hasil penelitian dapat disimpulkan bahwa:

1. Algoritma KNN dapat memecahkan masalah prediksi tingkat kelulusan mahasiswa dimana atributnya berupa data numerik.
2. Berdasarkan pengujian menggunakan 20 data uji terdapat 2 kesalahan prediksi kelulusan sehingga akurasi sistem 90%, dibuktikan dengan pengujian sistem.
3. Algoritma KNN hanya bisa mengolah data yang bersifat numerik.

### 5.2 Saran

1. Sebaiknya proses prediksi dilakukan menggunakan data training dalam jumlah banyak karena data training berpengaruh pada akurasi prediksi.
2. Hasil dari prediksi sebaiknya hanya dijadikan sebagai alat bantu dalam mengambil keputusan sedangkan

pengambilan keputusan tetap ditangan pengambil keputusan.

3. Aplikasi Prediksi Tingkat Kelulusan dapat ditingkatkan akurasi dengan mendeteksi anomali data training menggunakan algoritma deteksi anomali data.

## DAFTAR PUSTAKA

- [1] Gorunescu, Florin (2011). *Data Mining: Concept, Models, and Techniques*. San Fransisco: Morgan Kauffman.
- [2] Han, J., & Kamber, M. (2009). *Data Mining: Concept, Models, and Techniques*. Verlag Berlin Heidelberg: Springer
- [3] James H. McMillan. 2008. *Assessment Essentials for Standard-based Education*
- [4] Wu, Xindong & Kumar, Vipin. (2009). *The Top TEN Algorithms in Data Mining*. Boca Raton: CRC Press
- [5] Kusriani, & Luthfi, E. T. (2009). *Algoritma Data Mining*. Yogyakarta: Andi Publishing