

## IMPLEMENTASI ALGORITMA FUZZY K NEAREST NEIGHBOR UNTUK MEMPREDIKSI MAHASISWA BERPOTENSI DROP OUT

Devi Sugianti, Arief Soma Darmawan

Program Manajemen Informatika, STMIK WidyaPratama

Jl. Patriot 25 Pekalongan

Telp (0285)427816

email : [devi.sugianti9807@gmail.com](mailto:devi.sugianti9807@gmail.com), [ariefsoma24@gmail.com](mailto:ariefsoma24@gmail.com)

### ABSTRAK

Keberhasilan pendidikan tercermin dari kualitas Perguruan Tinggi adalah tingginya tingkat keberhasilan mahasiswa dan rendahnya kegagalan mahasiswa. STMIK Widya Pratama mempunyai 4 program studi dengan jumlah mahasiswa yang drop out atau non aktif yang cukup banyak, untuk itu perlu di ketahui karakteristik dari mahasiswa yang drop out dengan menganalisa data yang lama. Algoritma Fuzzy K Nearest Neighbor definisikan pemberian label kelas data uji yang akan diprediksi. Algoritma Fuzzy Nearest Neighbor adalah pengembangan dari algoritma K-NN yang digabungkan dengan teori Fuzzy. Tahapan penelitian yang dilakukan adalah pengumpulan data, pengolahan data, metode yang diusulkan dan eksperimen. Dengan data yang digunakan sebagai data latih 382 untuk mahasiswa angkatan 2016. Dengan menguji nilai  $K=3$  dan  $K=5$  dari mahasiswa dengan IPS dari semester 1 sampai semester 5 (3.2 2.78 3.6 3.3 0.83) hasil eksperimen yang dilakukan menghasilkan  $K=3$  dengan kelas 1 = 0.32 kelas 2 = 0.67 lalu jika  $K=5$  menghasilkan kelas 1 = 0.28 kelas 2 = 0.71. Hasil menunjukkan bahwa masuk ke kelas 2 dengan nilai keanggotaan terbesar, kelas 2 merupakan prediksi mahasiswa registrasi ulang pada semester 6. Fuzzy K-Nearest Neighbor dapat digunakan untuk mencari nilai membership pada masing-masing kelas.

*Kata Kunci: Fuzzy K NN, Drop Out*

### 1. PENDAHULUAN

Keberhasilan pendidikan jika sudah memenuhi tujuan pendidikan nasional dan proses belajar dan mengajar yang efektif dan efisien. Sedangkan prestasi belajar mahasiswa dapat didasarkan atas kesuksesan dan keunggulan pihak perguruan tinggi. (nurhayati, kusrini, & taufiq, 2015). Cerminan kualitas Perguruan Tinggi adalah tingginya tingkat keberhasilan mahasiswa dan rendahnya tingkat kegagalan mahasiswa (husnah, 2018). Kemampuan mahasiswa dalam menyelesaikan waktu tepat studi merupakan faktor dalam menentukan kualitas perguruan tinggi (hastuti, 2012). STMIK Widya Pratama merupakan salah satu Perguruan Tinggi yang mempunyai 4 Program Studi, yaitu Manajemen Informatika, Komputerisasi akuntansi, Sistem informasi dan Teknik Informatika. Berikut ini jumlah mahasiswa yang bersatus aktif

TABEL 1.1

Tahun Ajaran	Semester	
	Gasal	Genap
2016/2017	1720	1509
2017/2018	1674	1400
2018/2019	1468	1225
2019/2020	1295	1090

Dari data pada tabel 1.1 Mutu pendidikan dan Status Akreditasi akan terpengaruh dengan banyaknya jumlah mahasiswa yang drop out atau non aktif.

Untuk mengetahui karakteristik mahasiswa yang drop out atau tidak aktif maka perlu dianalisa dari data yang lampau. Sehingga dapat diketahui sejak dini pola mahasiswa yang drop out atau tidak aktif tersebut (yanti, 2020). Untuk menemukan pola atau pengetahuan dari kumpulan data disebut data mining (nurhayati, kusrini, & taufiq, 2015). Data yang ada didalam database dapat dianalisis dengan algoritma kalsifikasi, diantaranya *decision tree*, *naive bayes*, *nearest neighbor*. Pada penelitian yang dilakukan (nurhayati, kusrini, & taufiq, 2015) dengan judul prediksi mahasiswa drop out menggunakan metode support vector machine. Karena SVM diharapkan mendapatkan parameter yang digunakan secara tepat. Data individu dan evaluasi IPK yang dijadikan sebagai variabel input, sedangkan untuk variabel output nya adalah mahasiswa drop out dan non drop out. Pada penelitian (husnah, 2018) dengan judul Penerapan metode C4.5 untuk klasifikasi mahasiswa berpotensi drop out. Bahawa terdapat kurang lebih 1000 mahasiswa dengan mahasiswa drop out sekitar 200 mahasiswa pada tahun 2017. Dengan metode C4.5 ditemukan 17 rule untuk menentukan mahasiswa berpotensi Drop out. Atribut yang digunakan untuk penelitian jenis kelamin, umur, agama, asal daerah, kelas, nilai IPS smt1, nilai IPS smt2, nilai IPS smt 3, nilai IPS smt IV. Pada penelitian (yanti, 2020) dengan judul Prediksi mahasiswa berpotensi non aktif menggunakan data mining dalam decision tree dan algoritma C4.5

dengan hasil penelitian didapatkan akurasi sangat baik. Penelitian ini dapat digunakan untuk menghindari mahasiswa drop out secara sepihak kriteria yang digunakan jadwal kuliah, nilai absensi, nilai gagal, pembayaran uang kuliah. Adapun yang membedakan penelitian ini dengan penelitian sebelumnya adalah penggunaan metode. Pada penelitian ini menggunakan metode *Fuzzy K Nearest Neighbor*. Algoritma *Fuzzy K Nearest Neighbor* menyampikan definisi pemberian label kelas data uji yang akan diprediksi. Algoritma *Fuzzy Nearest Neighbor* adalah pengembangan dari algoritma *K-NN* yang digabungkan dengan teori *Fuzzy* (Prasetyo, 2012). Aturan klasifikasi *fuzzy* dihasilkan dari keanggotaan kelas *fuzzy set* sampel. Hasil dari kedua *crisp* dan aturan *Fuzzy KNN* dibandingkan pada dua dataset, algoritma *fuzzy* terbukti mendominasi *crisp* dengan memiliki tingkat kesalahan lebih rendah dan dengan memproduksi nilai-nilai keanggotaan yang berfungsi sebagai alat ukur dalam klasifikasi. *Fuzzy KNN* digunakan karena tingkat akurasi sebesar 77,35 % dengan 1138 *true positive* dan 163 *true negative* (Ahmad, Sulistiyani, & saputra, 2018). Akurasi *fuzzy K-Nearest Neighbor* memiliki akurasi yang stabil dan mendapat akurasi yang tinggi dibandingkan dengan *K-Nearest Neighbor* (Anugrah, indriati, & dewi, 2018). Berdasarkan penelitian (nugraha, putri, & wihandika, 2017) penentuan gizi balita dengan menggunakan metode *Fuzzy K-NN* mampu melakukan akurasi sebesar 84,37 % dari data latih sebanyak 160 dan nilai  $k=4$

## 2 Landasan Teori

### 2.1 K NN

Algoritma *Nearest Neighbor* merupakan algoritma yang melakukan klasifikasi berdasarkan kedekatan lokasi (jarak) suatu data dengan data yang lain. Jumlah tetangga terdekat atau dinyatakan  $k$  ditentukan oleh user. Berikut langkah-langkah metode *K NN*. Sebagai berikut.

1. Menghitung jarak *Eucliden* dengan rumus

$$d(x_i, x_j) = \sqrt{\sum_{r=1}^n (a_r(x_i) - a_r(x_j))^2}$$

Di mana :

$d(x_i, x_j)$  = jarak *Euclidean*

$x_i$  = record ke- $i$

$x_j$  = record ke- $j$

$a_r$  = data ke- $r$

2. Mengurutkan berdasarkan nilai *Euclidean distance*
3. Menentukan  $k$  record klasifikasi terdekat
4. Target output merupakan kelas yang mayoritas.

### 2.2. Fuzzy K- Nearest Neighbor

Algoritma *Fuzzy K Nearest Neighbor* memberikan nilai keanggotaan kelas pada data uji bukan menempatkan data uji pada kelas

tertentu. *Fuzzy K Nearest Neighbor* merupakan metode klasifikasi yang digunakan untuk memprediksi data uji menggunakan nilai derajat keanggotaan data uji pada setiap kelas. Adapun langkah-langkah dari perhitungan *fuzzy K-Nearest Neighbor* adalah sebagai berikut:

1. Inisialisasi *fuzzy*

$$u_{ij} = \begin{cases} 0,51 + \left(\frac{n_j}{K}\right) * 0,49, & \text{jika } j=i \\ \left(\frac{n_j}{K}\right) * 0,49 & , \text{jika } j \neq i \end{cases}$$

Di mana :

$u_{ij}$  = nilai keanggotaan kelas  $i$  pada vektor  $j$

$n_j$  = jumlah anggota kelas  $j$  pada suatu dataset  $K$

$K$  = banyaknya tetangga terdekat

$J$  = kelas target

2. Menghitung jarak *eucliden* data uji terhadap data latih
3. Mengurutkan berdasarkan nilai *eucliden* terkecil
4. Menentukan  $k$  record terdekat
5. Menghitung derajat keanggotaan data baru terhadap masing-masing kelas menggunakan persamaan

$$u_i(x) = \frac{\sum_{j=1}^K u_{ij} \left(1 / \|x - x_j\|^{m-1}\right)}{\sum_{j=1}^K \left(1 / \|x - x_j\|^{m-1}\right)}$$

Di mana :

$U_i(x)$  = nilai keanggotaan data  $x$  ke kelas  $i$

$K$  = jumlah tetangga terdekat yang digunakan

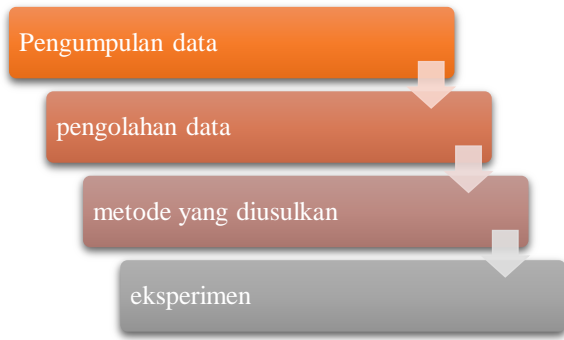
$x-x_j$  = selisih jarak data  $x$  ke data  $x_j$  dalam  $K$  tetangga terdekat

$m$  = bobot pangkat (weight exponent) yang besarnya  $m > 1$

6. Memilih kelas yang memiliki nilai keanggotaan terbesar sebagai hasil

## 3. METODE PENELITIAN

Penelitian *fuzzy K-NN* yang dilakukan agar dilakukan secara terencana, teratur dan sistematis agar mencapai tujuan penelitian ini maka tahapan yang dilakukan adalah sebagai berikut



Gambar 2. Tahapan Penelitian

**3.1 Pengumpulan data**

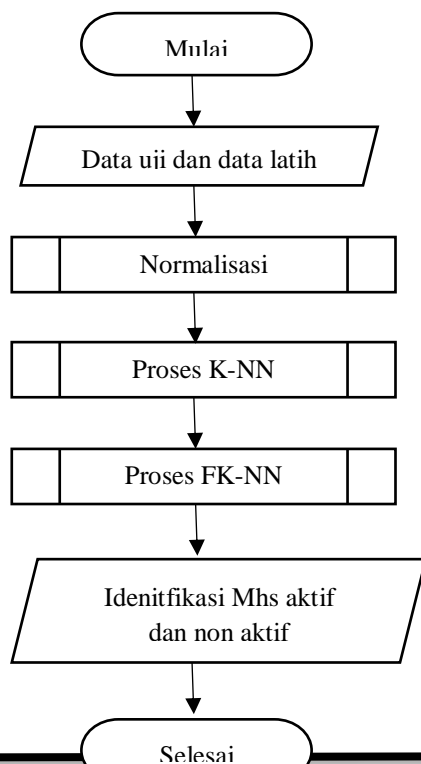
Pada tahapan pengumpulan data yang dilakukan adalah mengumpulkan data *training* dari database akademik STMIK Widya Pratama

**3.2 Pengolahan data**

Dari tahap pengumpulan data dilakukan pengolahan didapatkan data mahasiswa angkatan 2016 untuk 4 program studi yaitu: progdi MI, KA,SI,TI. Berjumlah 382. Data latih yang diperoleh berupa data latih registrasi untuk smt 6. Dimana dari data latih tersebut ada dua kemungkinan yaitu aktif atau non aktif pada semester berikutnya. *Fuzzy K-NN* digunakan untuk melakukan klasifikasi mahasiswa aktif dan non aktif dari data *training* tersebut

**3.3 Metode yang diusulkan**

Metode yang diusulkan adalah *fuzzy nearest neighbour*. Adapun *flowchart* klasifikasi secara umum.



Gambar 3. *Flowchart* klasifikasi *fK-NN*

**3.4 Eksperimen**

Dari data 382 mahasiswa dilakukan eksperimen dan pengujian terhadap nilai *K*,

**5. HASIL dan PEMBAHASAN**

Dengan data yang di jadikan pusat adalah:

IPS 1	IPS 2	IPS 3	IPS 4	IPS 5	status
3.2	2.78	3.6	3.3	0.83	1

Dari data tersebut di hitung berdasarkan *K=3* dan *K=5*

Jika *K=3* maka dihasilkan sebagai berikut:

No	Kelas	jarak
1	1	0.49
2	0	0.9
3	0	1.18

$$\text{Kelas 1} = \frac{(0 * 0.49^{\frac{-2}{-2-1}}) + (1 * 0.92^{\frac{-2}{-2-1}}) + (1 * 1.18^{\frac{-2}{-2-1}})}{0.49^{\frac{-2}{-2-1}} + 0.92^{\frac{-2}{-2-1}} + 1.18^{\frac{-2}{-2-1}}}$$

$$= \frac{0 + 1.23 + 0.79}{4.16 + 1.23 + 0.79} = \frac{2.02}{6.18} = 0.32$$

$$\text{Kelas 2} = \frac{(1 * 0.49^{\frac{-2}{-2-1}}) + (0 * 0.92^{\frac{-2}{-2-1}}) + (0 * 1.82^{\frac{-2}{-2-1}})}{0.49^{\frac{-2}{-2-1}} + 0.92^{\frac{-2}{-2-1}} + 1.18^{\frac{-2}{-2-1}}}$$

$$= \frac{4.16 + 0 + 0}{4.16 + 1.23 + 0.79} = \frac{4.16}{6.18} = 0.67$$

Karena nilai keanggotaan terbesar ada di kelas 2 (0.67), data uji (3.2 2.78 3.6 3.3 0.833) diprediksi masuk ke kelas 2 ( kelas 1 dalam set data). Dengan menggunakan *K=3* maka dapat dilihat class nya adalah 2 yaitu bahwa mahasiswa tersebut melakukan registrasi.

Jika *K=5* maka dihasilkan sebagai berikut:

No	Kelas	Jarak
1	1	0.49
2	0	0.9
3	0	1.18
4	1	1.40
5	1	1.42

$$\begin{aligned} & \text{Kelas 1} \\ & = \frac{(0 * 0.49^{\frac{-2}{2-1}}) + (1 * 0.9^{\frac{-2}{2-1}}) + (1 * 1.18^{\frac{-2}{2-1}}) * (0 * 1.4^{\frac{-2}{2-1}}) + (0 * 1.42^{\frac{-2}{2-1}})}{0.49^{\frac{-2}{2-1}} + 0.9^{\frac{-2}{2-1}} + 1.18^{\frac{-2}{2-1}} + 1.4^{\frac{-2}{2-1}} + 1.42^{\frac{-2}{2-1}}} \\ & = \frac{0 + 1.23 + 0.79 + 0 + 0}{4.16 + 1.23 + 0.79 + 0.51 + 0.50} \\ & = \frac{2.02}{7.19} = 0.28 \end{aligned}$$

$$\begin{aligned} & \text{Kelas 2} \\ & = \frac{(1 * 0.49^{\frac{-2}{2-1}}) + (0 * 0.9^{\frac{-2}{2-1}}) + (0 * 1.18^{\frac{-2}{2-1}}) + (1 * 1.4^{\frac{-2}{2-1}}) + (1 * 1.42^{\frac{-2}{2-1}})}{4.16 + 1.23 + 0.79 + 0.51 + 0.5} \\ & = \frac{4.16 + 0.51 + 0.5}{7.19} \\ & = \frac{5.17}{7.19} = 0.71 \end{aligned}$$

Karena nilai keanggotaan terbesar ada di kelas 2 (0.71) data uji (3.2 2.78 3.6 3.3 0.833) diprediksi masuk ke kelas 2 ( kelas 1 dalam set data). Dengan menggunakan  $K=5$  maka dapat dilihat class nya adalah 2 yaitu bahwa mahasiswa tersebut melakukan registrasi

Dari pengujian dengan menggunakan  $K=3$  dan  $K=5$  di simpulkan bahwa mahasiswa tersebut melakukan registrasi atau aktif pada semester 6.

## 6. KESIMPULAN

Berdasarkan penelitian yang dilakukan dapat disimpulkan bahwa:

1. Metode *fuzzy K Nearest Neighbor* digunakan dalam menentukan status keaktifan mahasiswa pada semester 6 dengan melalui normalisasi dari data latih dan data uji sebanyak 382 untuk angkatan 2016 untuk program studi TI, SI, MI, KA.
2. Dari data latih dan data uji digunakan untuk mencari nilai *euclidean distance* nya dengan menentukan nilai  $k$ .  $K$  yang digunakan adalah  $K=3$  dan  $K=5$ . *Fuzzy K-Nearest Neighbor* digunakan untuk mencari nilai *membership* pada masing-masing kelas.
3. *Fuzzy K Nearest Neighbor* lebih memiliki akurasi yang stabil dan mendapat akurasi yang tinggi. Karena nilai inisialisasi ini berperan mendapatkan nilai *membership* penentu kelas keluaran

## Daftar Pustaka

Ahmad, I., Sulistiyani, h., & saputra, h. (2018). Using Fuzzy K-Nearest Neighbor for Predicting University Students

Graduation in Teknokrat. *Indonesia Journal of artificial Intelligence and Data Mining*, 47-51.

Anugrah, A. s., indriati, & dewi, c. (2018). Implementasi Algoritma fuzzy k nearest neighbor untuk penentuan lulus tepat waktu (studi kasus: fakultas ilmu komputer universitas brawijaya). *jurnal pengembangan teknologi informasi dan ilmu komputer*, 1726-1732.

hastuti, k. (2012). analisis komparasi algoritma klasifikasi data mining untuk prediksi mahasiswa non aktif. *seminar nasional teknologi informasi & komunikasi terapan* (pp. 241-249). semarang: semantik.

husnah, a. (2018). penerapan metode c4.5 untuk klasifikasi mahasiswa berpotensi drop out. *ilkom*, 244-250.

nugraha, s. d., putri, r. r., & wihandika, r. c. (2017). penerapan Fuzzy K nearest Neighbor (FK-NN) dalam menentukan status gizi balita. *jurnal pengembangan teknologi informasi dan ilmu komputer*, 925-932.

nurhayati, s., kusrini, & taufiq, e. (2015). prediksi mahasiswa drop out menggunakan metode support vector machine. *sisfotenika*, 82-93.

Prasetyo, E. (2012). *Data Mining konsep dan aplikasi menggunakan Matlab*. Yogyakarta: andi.

yanti, n. (2020). prediksi mahasiswa berpotensi non aktif menggunakan data mining dalam decision tree dan algoritma c4.5. *Jurnal informasi & teknologi*, 23-29.