**Indonesian Journal of Artificial Intelligence and Data Mining (IJAIDM)**

Vol. 1, No. 1, March 2018, pp. xx-xx

p-ISSN: XX-XX | e-ISSN: XX-XX  1

**The Application Of Fuzzy K-Nearest Neighbour Methods for A Student Graduation Rate**

**1Imam Ahmad, 2Heni Sulistiyani, 3Hendrik Saputra**

1,2,3Sistem Informasi, Fakultas Teknik Dan Ilmu Komputer, Universitas Teknokrat Indonesia

Email: 1imamahmad@teknokrat.ac.id, 2henisulistiani@teknokrat.ac.id, 3hendriksa68@yahoo.com

****

**Article Info ABSTRACT**

**Article history:**

Received XX, 2018

Revised XX, 2018

Accepted XX, 2018

**Keywords:**

Prediction

Fuzzy

Student Graduation Rate

K-Nearest Neighbor

The absence of prediction system that can provide prediction analysis on the graduation rate of students becomes the reason for the research on the prediction of the level of graduation rate of students. Determining predictions of graduation rates of students in large numbers is not possible to do manually because it takes a long time. For that we need an algorithm that can categorize predictions of students' graduation rates in computing. The Fuzzy Method and KNN or K-Nearest Neighbor Methods are selected as the algorithm for the prediction process. In this study using 10 criteria as a material to predict students' graduation rate consisting of: NPM, Student Name, Semester 1 achievement index, Semester 2 achievement index, Semester 3 achievement index, Semester 4 achievement index, SPMB, origin SMA, Gender , and Study Period. Fuzzyfication process aims to change the value of the first semester achievement index until the fourth semester achievement index into three sets of fuzzy values ​​are satisfactory, very satisfying, and cum laude. Make predictions to improve the quality of students and implement KNN method into prediction, where there are some attributes that have preprocess data so that obtained a value, and the value is compared with training data, so as to produce predictions of graduating students will be on time and graduating students will be late. This study produces a prediction of student pass rate and accuracy.

 *Copyright © 2018 Puzzle Research of Data Technology*

**Corresponding Author:**

Imam Ahmad

Sistem Informasi, Fakultas Teknik dan Ilmu Komputer

Universitas Teknokrat Indonesia

Jl. Zaenal Abidin Pagar Alam No. 9-11, Labuhan Ratu, Kota Bandar Lampung 35132, Lampung

Email: imamahmad@teknokrat.ac.id

**I. PENDAHULUAN**

Data Mining didefinisikan sebagai proses penemuan pola dalam data. Berdasarkan tugasnya, *data mining* dikelompokkan menjadi deskripsi, estimasi, prediksi, klasifikasi, *clustering* dan asosiasi.[6] Selain itu data mining merupakan salah satu teknik yang dapat melakukan prediksi dengan teknik penggalian data. Penggalian data berdasarkan data pendidikan di universitas dapat meningkatkan kualitas pembelajaran mahasiswa di universitas.

Melakukan Prediksi Tentang prediksi tingkat kelulusan mahasiswa sangatlah penting dimana informasi yang dihasilkan oleh hasil prediksi ini dapat dijadikan sebagai bahan pertimbangan bagi pihak manajemen program studi untuk melakukan langkah secara persuasif dalam rangka meningkatkan persentasi kelulusan mahasiswa yang memenuhi standar mutu yang telah ditetapkan kampus. Prediksi dalam bentuk komputasi merupakan satu kegiatan matematis. Prediksi ini telah dilakukan bertahun-tahun sebelum penguasaan komputer, yaitu menggunakan kalkulator. Prediksi tingkat kelulusan mahasiswa dapat membantu pihak manajemen dalam mengambil keputusan. Bagaimana menganalisa kelulusan mahasiswa dan mendapatkan parameter terbaik dari metode yang digunakan untuk dapat memprediksi kelulusan mahasiswa.

Sebuah penelitian membahas tentang Analisis Prediksi Tingkat Pengunduran Diri Mahasiswa dengan Metode K-Nearest Neighbor. Dimana dalam penelitian yang dilakukan oleh penulis mengangkat masalah mengenai pentingnya analisis prediksi tentang pengunduran diri mahasiswa di STIKOM UYELINDO Kupang karena menentukan prediksi tingkat pengunduran diri mahasiswa dalam jumlah besar tidak mungkin dilakukan secara manual karena membutuhkan waktu yang cukup lama. Untuk itu dibutuhkan sebuah algoritma yang dapat mengkategorisasikan prediksi tingkat pengunduran diri mahasiswa secara otomatis menggunakan komputer. Hasil yang didapat dari penelitian ini yaitu bertujuan untuk menghasilkan sebuah data yang dapat diuji untuk mendapat prediksi dan keakuratan klasifikasi. [8]

Adapun penelitian lainnya membahas tentang Graduation Prediction of Gunadarma University Students Using Algorithm and Naïve Bayes C4.5 Algorithm, Permasalahan yang dihadapi adalah banyaknya mahasiswa yang lulus tidak tepat waktu. Untuk mengetahui tingkat kelulusan mahasiswa dalam satu tahun ajaran dapat dilakukan suatu prediksi berdasarkan data-data mahasiswa pada tahun ajaran pertama. Algoritma yang digunakan adalah C45 dan naïve bayes. Hasil dari penelitian ini adalah akurasi untuk metode naïve bayes adalah 80,85% dengan presentasi kesalahan 19,05% Akurasi ketepatan hasil prediksi C4.5 85,7% dan presentasi kesalahannya adalah 14,3% [11].

Penelitian yang dilakukan oleh Karamouiz, [12] dengan judul Sensitivy Analysis of Neural Network for Identifying the Factors For College Students Success, Masalah yang dikaji adalah tingkat kelulusan yang dianggap sebagai indikator efektivitas suatu lembaga institusi, Metode yang digunakan yaitu NN (Neural Network). Dari hasil data training yang dilakukan, diperoleh kategori yang lulus adalah 86,04 % dan data training yang tidak sukses adalah 68,21 % dan error yang diperoleh untuk kedua kategori tersebut adalah 0,18 %.

Penelitian yang dilakukan oleh Qudri, Kalyankar (2010) dengan judul Drop Out Feature of Student Data for Academic Performance Using Decision Tree techniques. Masalah dalam penelitiannya adalah prestasi akademik siswa sangat penting bagi lembaga pendidikan karena program-program strategis dapat direncanakan untuk meningkatkan atau mempertahankan prestasi siswa selama periode mereka studi di lembaga. Metode yang digunakan adalah Decision Tree, yakni algoritma J4.8. Hasil dari penelitian ini adalah sebuah pohon keputusan yang dapat dijadikan rule bagi prediksi siswa yang putus sekolah.[9]

Penelitian yang dilakukan oleh Mustakim, Giantika Oktaviani F (2016) Program Studi Sistem Informasi Fakultas Sains dan Teknologi, Universitas Islam Negeri Sultan Syarif Kasim, Riau dengan judul Algoritma K-Nearest Neighbor Classification Sebagai Sistem Prediksi Predikat Prestasi Mahasiswa. Dimana dalam penelitian yang dilakukan oleh penulis mengangkat masalah tentang Persoalan Indeks Prestasi Mahasiswa. Atribut yang digunakan dalam proses prediksi adalah Jenis Kelamin, Jenis Tinggal, Umur, Jumlah Satuan Kredit Semester (SKS), dan Jumlah Nilai Mutu (NM), sehingga dengan menerapkan algoritma K-Nearest Neighbor dapat dilakukan prediksi berdasarkan kedekatan dari histori data lama (training) dengan data baru (testing). Hasil yang didapat dari penelitian ini adalah sistem prediksi yang dapat dijadikan sebagai acuan bagi mahasiswa untuk meningkatkan prestasi dan predikat perkuliahan dimasa yang akan datang.[7]

***K-Nearest Neighbor***

K-nearest neighbor (KNN) termasuk kelompok *instance-based learning*. Algoritma ini juga merupakan salah satu teknik *lazy learning*. KNN dilakukan dengan mencari kelompok objek dalam data training yang paling dekat (mirip) dengan objek pada data baru atau data testing [12]. Contoh kasus, misal diinginkan untuk mencari solusi terhadap masalah seorang pasien baru dengan menggunakan solusi dari pasien lama. Untuk mencari solusi dari pasien baru tersebut digunakan kedekatan dengan kasus pasien lama,solusi dari kasus lama yang memiliki kedekatan dengan kasus baru digunakan sebagai solusinya. Terdapat pasien baru dan 4 pasien lama, yaitu P,Q,R, dan S (Gbr, 1). Ketika ada pasien baru maka yang diambil solusi adalah solusi dari kasus pasien lama yang memiliki kedekatan terbesar.



Gambar. 1 Ilustrasi kasus algoritma KNN

Misal D1adalah jarak antara pasien baru dengan pasien P,D2 adalah jarak antara pasien baru dengan pasien Q,D3 adalah jarak antara pasien baru dengan pasien R,D4 adalah jarak antara pasien baru dengan pasien S. Dari ilustrasi gambar terlihat bahwa D2 yang paling terdekat dengan kasus baru. Dengan demikian maka solusi dari kasus pasien Q yang akan digunakan sebagai solusi dari pasien baru tersebut. Untuk mendefinisikan jarak antara dua titik pada data training (x) dan titik pada data testing (y) maka digunakkan rumus Euclidean, Seperti yang ditunjukan pada persamaan dibawah ini.

D (x,y) = $\sqrt{\sum\_{k-1}^{n}(x\_{k}-y\_{k})}$2

Keterangan :

D : jarak kedekatan.

X : data training.

Y : data testing.

n : jumlah atribut individu antara 1 s.d. n

f : fungsi similitary atribut i antara kasus X dan kasus Y.

i : Atribut individu antara 1 sampai dengan n

Langkah – langkah untuk menghitung metode Algoritma *K-Nearest Neighbor:*

* Menentukan Parameter K (Jumlah tetangga paling dekat)
* Menghitung kuadrat jarak *Euclid* (*query instance*) masing-masing objek terhadap data sampel yang diberikan.
* Kemudian mengurutkan objek-objek tersebut ke dalam kelompok yang mempunyai jarak *Euclid* terkecil.
* Mengumpulkan kategori Y (Klasifikasi Nearest Neighbor)
* Dengan menggunakkan kategori *Nearest Neighbor* yang paling mayoritas maka dapat diprediksi nilai *query instance* yang telah dihitung.

**2. METODOLOGI PENELITIAN**

Tahapan penelitian merupakan kegiatan penelitian yang dilakukan secara terencana, teratur, dan sistematis untuk mencapai tujuan tertentu. Tahapan penelitian ini juga merupakan pengembangan dari kerangka penelitian, dan terbagi lagi menjadi beberapa sub menu bagian. Berikut adalah tahapan penelitian :



Gambar.2 Tahapan Penelitian

***Fuzzyfication Kriteria Indeks Prestasi***

Indeks Prestasi Komulatif (I)

 Nilai Cumlaude

2,76 – 4,00 I[1]=3,50

 Sangat Memuaskan 2,00 - 3,50 I[1]=2,76

 Memuaskan 1,00 -2.76 I[1]=2,00



Gambar. 3 Grafik Kriteria IP [1]

Fungsi keanggotaan untuk himpunan MEMUASKAN :

$$Memuaskan [i]= \left\{\begin{matrix}0&i\leq 1,00 atau i\geq 2,76\\\frac{i-1}{2-1}&1\leq i\leq 2,00\\\frac{2,76-i}{2,76-2,00}&2,00\leq i\leq 2,76\end{matrix}\right.$$

Fungsi keanggotaan untuk himpunan SANGAT MEMUASKAN :

$$Sangat Memuaskan [i]= \left\{\begin{matrix}0&i\leq 2,0 atau i\geq 3,5\\\frac{i-2,0}{2,76-2,00}&2,0\leq i\leq 2,76\\\frac{3,5-i}{3,5-2,76}&2,76\leq i\leq 3,5\end{matrix}\right.$$

Fungsi keanggotaan untuk himpunan CUMLAUDE :

$$Cumlaude [i]= \left\{\begin{matrix}0&i\leq 2,76 atau i\geq 4,0\\\frac{i-2,76}{3,5-2,76}&2,76\leq i\leq 3,5\\1&i\geq 3,5\end{matrix}\right.$$

**3. HASIL**

Tahap penjelasan hasil penelitian ini dilakukan setelah praprocess data telah selesai dirubah, kemudian dari data yang telah diubah kedalam bentuk fuzzy dilakukan proses pengelompokan terhadap atribut Indeks Prestasi Semester Satu, Dua, Tiga, Empat, dan Asal SMA. Pengelempokan ini dimaksudkan agar dapat mengurangi kesalahan-kesalahan yang timbul pada saat mengoperasikan sistem.Sistem yang dibuat ini digunakan dapat lebih baik karena belum adanya aplikasi prediksi tentang tingkat kelulusan mahasiswa. Berikut akan diberikan penjelasan hasil penelitian yang ditulis secara rinci.

***Pengolahan Data Awal***

Jumlah data awal yang diperoleh dari pengumpulan data yaitu sebanyak 1.685 data, namun tidak semua data dapat digunakan dan tidak semua atribut digunakan karena harus melalui beberapa tahap pengolahan awal data (*preparation data*). Untuk mendapatkan data yang berkualitas, beberapa teknik yang dilakukan adalah sebagai berikut:

1. *Data validation,* untuk mengidentifikasi dan menghapus data yang ganjil (*outlier/noise*), data yang tidak konsisten, dan data yang tidak lengkap (*missing value*). *Missing data* terlihat

2. *Data integration and Transformation,* untuk meningkatkan akurasi dan efisiensi algoritma. Data yang digunakan dalam penulisan ini bernilai kategorikal. Data ditransformasikan ke dalam *software RapidMiner*.

3. *Data size reduction and dicrtization,* untuk memperoleh data set dengan jumlah atribut dan record yang lebih sedikit tetapi bersifat informatif. Dalam penelitian ini atribut yang tidak relevan seperti indeks prestasi semester satu, dua, tiga dan empat diubah menjadi bentuk *fuzzyfication* dan SMA dikelompokan menjadi angka 0-3.

****

Gambar. 4 Data Sampel

***Prediksi K-Nearest Neighbor***

Pada Tahap Ini Penulis Menjelaskan tentang prediksi K-Nearest Neighbor yang dilakukan secara manual/ menggunakan excel, terdapat 50 data sampel angkatan 2008 yang merupakan data training sedangkan 1 data sampel mahasiswa angkatan 2013 yang merupakan data testing, Berikut Penjelasan mengenai prediksi k-nearest neighbor :

***Menentukan Jumlah K/ Tetangga Terdekat***

Langkah Selanjutnya yang akan dilakukan yaitu Menentukan terlebih dahulu K/ Jumlah tetangga terdekat dari prediksi kelulusan mahasiswa, k yang digunakan dalam penelitian ini adalah 10. Menurut Hastie et al (2008), dengan k=5 atau 10 dapat digunakan untuk memperkirakan tingkat kesalahan yang terjadi, sebab data training pada setiap fold cukup berbeda dengan data training yang asli. Secara keseluruhan, 5 atau 10- fold cross validation sama-sama direkomendasikan dan disepakati bersama.[4]

Tabel. 1 Data Testing Sebelum Preprocess

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| NPM | Nama | Ip Sem 1 | Ip Sem 2 | Ip Sem 3 | Ip Sem 4 |
| 13312055 | Budi Andriawan | 3,6 | 3,38 | 3,2 | 3,05 |

|  |  |  |  |
| --- | --- | --- | --- |
| SPMB | Asal SMA | JK | Masa Studi |
| 330 | SMK N 4 BDL | Laki-Laki | 4 thn 3 bln 20 hr |

Tabel. 2 Data Testing Sesudah Preprocess

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| NPM | Nama | Ip Sem 1 | Ip Sem 2 | Ip Sem 3 | Ip Sem 4 |
| 13312055 | Budi Andriawan | 1 | 0,837838 | 0,594594595 | 0,391892 |

|  |  |  |  |
| --- | --- | --- | --- |
| SPMB | Asal SMA | JK | Masa Studi |
| 330 | 1 | Laki-Laki | 4 thn 3 bln 20 hr |

***Menghitung Jarak/ Distance***

Setelah Melakukan Proses *Preprocess data* pada *data training* dan *data testing*, langkah selanjutnya yang akan dilakukan yaitu menghitung jarak/ *distance* pada data sampel. Dalam penelitian ini penulis menggunakan perhitungan jarak menggunakan *euclidean distance*.

Contoh Perhitungan Dengan Data Sampel D4 (Ana Francisca) Dibandingkan Dengan Data Testing D51 (Budi Andriawan) Adalah:

D (x,y) = $\sqrt{\sum\_{k-1}^{n}(x\_{k}-y\_{k})}$2

Perhitungan ini melibatkan atribut-atribut yang terdapat dalam data training dan data testing dimana penulis memasukkan atribut indeks prestasi semester satu sampai dengan indeks prestasi semester empat.

=D(D4,D51)$ $

$=\sqrt{\begin{array}{c}\left(0,77027-1\right)^{2}+\left(1-0,837838\right)^{2}+\\\left(0,716216216-0,594594595\right)^{2}\\+( 0,202703- 0,391892)\^2)\end{array}}$

= 0,36.

Hasil Perhitungan Jarak *Eulcidean Distance* Sampel untuk Data D4 dan Data D51 yang merupakan Data Testing.

Tabel. 3 Data Sampel Dengan Jarak *Euclidean*

|  |  |  |
| --- | --- | --- |
| No. | Nama | Distance |
| 1 | D4,D51 | 0,36 |
| 2 | D5,D51 | 0,05 |
| 3 | D13,D51 | 0,7 |
| 4 | D14,D51 | 0,45 |
| 5 | D33,D51 | 0,65 |
| 6 | D40,D51 | 0,23 |
| 7 | D32,D51 | 0,75 |
| 8 | D16,D51 | 1,14 |
| 9 | D46,D51 | 0,95 |
| 10 | D49,D51 | 0,45 |

**Menghitung Data Berdasarkan Jarak**

Setelah mendapatkan hasil jarak selanjutnya diurutkan dari nilai terbesar ke terkecil seperti pada tabel 4.5.

Tabel. 4 Pengurutan Nilai Jarak

|  |  |  |  |
| --- | --- | --- | --- |
| No. | Nama | Distance | Ranking |
| 1 | D16,D51 | 1,14 | 1 |
| 2 | D46,D51 | 0,95 | 2 |
| 3 | D32,D51 | 0,75 | 3 |
| 4 | D13,D51 | 0,7 | 4 |
| 5 | D33,D51 | 0,65 | 5 |
| 6 | D14,D51 | 0,45 | 6 |
| 7 | D49,D51 | 0,45 | 7 |
| 8 | D4,D51 | 0,36 | 8 |
| 9 | D40,D51 | 0,23 | 9 |
| 10 | D5,D51 | 0,05 | 10 |

***Menentukan Prediksi Menggunakan Kategori Mayoritas***

Dari 10 Jarak yang telah didapatkan dari perhitungan menggunakan *Euclidean distance*, selanjutnya adalah menentukan prediksi menggunakan kategori mayoritas yang didapat dari klasifikasi data training.

Tabel. 5 Penentuan Prediksi

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| No. | Nama | Distance | Ranking | Klasifikasi |
| 1 | D16,D51 | 1,14 | 1 | Terlambat |
| 2 | D46,D51 | 0,95 | 2 | Tepat Waktu |
| 3 | D32,D51 | 0,75 | 3 | Tepat Waktu |
| 4 | D13,D51 | 0,7 | 4 | Tepat Waktu |
| 5 | D33,D51 | 0,65 | 5 | Tepat Waktu |
| 6 | D14,D51 | 0,45 | 6 | Tepat Waktu |
| 7 | D49,D51 | 0,45 | 7 | Terlambat |
| 8 | D4,D51 | 0,36 | 8 | Tepat Waktu |
| 9 | D40,D51 | 0,23 | 9 | Terlambat |
| 10 | D5,D51 | 0,05 | 10 | Tepat Waktu |

Dalam Data Training yang sudah dilakukan perhitungan K-NN dengan Data Testing didapatkan kategori Lulus akan Tepat Waktu berjumlah 7, Sedangkan kategori Lulus akan Terlambat berjumlah 3. Karena konsep K-NN menggunakan kategori mayoritas sebagai hasil prediksi, Maka untuk Data baru yaitu mahasiswa bernama Budi Andriawan akan menghasilkan Lulus Tepat Waktu.

***Pengelompokan Data***

Data yang digunakan untuk melakukan perhitungan Dalam Penerapan Fuzzy dan KNN kedalam Rapidminer merupakan Data yang telah diproses secara *fuzzyfication*. Berikut ini adalah data indeks prestasi yang digunakan untuk dalam proses prediksi kelulusan mahasiswa.

Tabel. 6 Kelompok Data IP Semester 1

|  |  |  |  |
| --- | --- | --- | --- |
| **No** | **Nama** | **IP Sem 1** | **Fungsi Keanggotaan** |
| **Memuaskan** | **Sangat Memuaskan** | **Cumlaude** |
| 1 | Adik Abdul Rohim | 2,81 | 0,000 | 0,932 | 0,067 |
| 2 | Agus Ade Putra | 3,24 | 0,000 | 0,351 | 0,648 |
| 3 | Ana Francisca | 3,33 | 0,000 | 0,229 | 0,770 |
| 4 | Apriyani | 3,71 | 0,000 | 0,000 | 1,000 |
| 5 | Bayu Adi Kurnia | 2,33 | 0,565 | 0,434 | 0,000 |
| 6 | Desi Jumiarti | 2,38 | 0,5 | 0,5 | 0,000 |
| 7 | Dwi Anita | 2,86 | 0,000 | 0,135 | 0,864 |
| 8 | Edwin Novanda | 2,14 | 0,815 | 0,184 | 0,000 |
| 9 | Erick Setiawan | 2,86 | 0,000 | 0,135 | 0,864 |
| 10 | Febi Herumanika | 2,05 | 0,934 | 0,065 | 0,000 |

***Pengujian Model***

Pengujian model dalam penelitian ini menggunakan *Cross Validation* adalah teknik validasi dengan membagi data secara acak kedalam *k* bagian dan masing-masing bagian akan dilakukan proses klasifikasi. Dengan menggunakan *cross validation* akan dilakukan percobaan sebanyak k. Data yang digunakan dalam percobaan ini adalah data *training* untuk mencari nilai *error rate* secara keseluruhan. Secara umum pengujian nilai k dilakukan sebanyak 10 kali untuk memperkirakan akurasi estimasi. Dalam penelitian ini nilai k yang digunakan berjumlah 10 atau 10-*fold Cross Validation*.



 Gambar. 6 Ilustrasi 10 *Fold Cross Validation*

Algoritma *K-Nearest Neighbor*

***Pengujian***

Hasil Pengujian menggunakan tabel *confusion matrix* dapat dilihat pada gambar dibawah ini:



Gambar. 7 *Confusion Matrix*

Dari Tabel Confusion matrix diatas dapat dilihat tingkat akurasi prediksi kelulusan mahasiswa menggunakan K-NN adalah sebesar 77,35 % dengan class precision untuk prediksi tepat waktu yaitu 79,86 % dan prediksi terlambat 63,42 % , sedangkan untuk class recall yaitu 92,37 % untuk prediksi yang bernilai tepat waktu. Dan class recall yaitu 36,22 % untuk prediksi yang bernilai terlambat.

**4. KESIMPULAN**

Berdasarkan rumusan masalah dan hasil penelitian terhadap Penerapan Metode Fuzzy dan K-Nearest Neighbour Untuk Prediksi Kelulusan Mahasiswa maka dapat disimpulkan bahwa: Telah berhasil dibangun Aplikasi untuk prediksi kelulusan mahasiswa Menggunakan metode fuzzy dan k-nearest neighbor dan tools Rapidminer yang dibuat sebagai acuan mahasiswa dalam meningkatkan mutu mahasiswa. Dan Telah berhasil Menampilkan Nilai Akurasi dari prediksi kelulusan mahasiswa Menggunakan tools Rapidminer Sebesar 77,35% dengan *true positive* berjumlah 1138 Data, dan *true negative* berjumlah 163Data. Dan *Class Precision* untuk prediksi Tepat Waktu Sebesar 79,86% dan *Class Precision* untuk prediksi Terlambat Sebesar 63,42%, memiliki *Class Recall* untuk *true* Tepat sebesar 92,37% dan *Class Recall* untuk *true Terlambat sebesar 36,22%.*

**UCAPAN TERIMA KASIH**

Kami Ucapkan terimakasih kepada sumber data yang telah menyediakan datanya dengan gratis sehingga dapat digunakan dalam penelitian ini.

**REFERENSI**

[1] Agus Wantoro, Adhie Thyo Priandika., 2017. *Komparasi Perhitungan Pemilihan Mahasiswa Terbaik Menggunakan Metode Statistik Klasik Dengan Logika Fuzzy (Tsukamoto Dan Mamdani) Studi Kasus STMIK Teknokrat,* Sistem Informasi STMIK Teknokrat.Vol 15 No. 1 2018

[2] Ernawati Iin, 2008, *Perbandingan algoritma C4.5 dan K-Nearest Neighbor untuk Prediksi Status Keaktifan Studi Mahasiswa*.

[3] Gorunescu, F. 2011, *Data Mining Concepts Models and Techniques. Craiova: Springer.*

[4] Hastie Trevor, Tibshirani Robert, Jerome Friedman, 2008, *The Elements of Statistical Learning Data Mining, Inference, and Prediction.* California: Springer*.*

[5] Ahmad I, Hermadi I, Arkeman Y, 2015. *Financial Feasibility Study Of Waste Cooking Oil Utilization for Biodiesel Production Using ANFIS.* TELKOMNIKA: Vol.13 No 3 Maret 2015*.*

[6] Karamouiz, Vrettos, 2009, *Sensitivy Analysis of Neural Network for Identifying the Factors For College Students Success.*

[7] Larose, Daniel.T. 2005. *Discovering Knowledge in Data.* New Jersey : John Willey & Sons, Inc.

[8] Mustakim, Giantika Oktaviani F,2016, Algoritma *K-Nearest Neighbor Classification* *Sebagai Sistem Prediksi Predikat Prestasi Mahasiswa, Sistem Informasi Fakultas Sains dan Teknologi*, Universitas Islam Negeri Sultan Syarif Kasim, Riau.

[9] Ndaumanu Imanuel Ricky, Kusrini, M. Rudyanto Arief **,** 2014, *Analisis Prediksi Tingkat Pengunduran Diri Mahasiswa STIKOM UYELINDO dengan Metode K-Nearest Neighbor,* Kupang: Magister Teknik Informatika.

[10] Qudri, Kalyankar, 2010, *Drop Out Feature of Student Data for Academic Performance Using Decision Tree techniques.*

[11] Pandie, Emerensye S.Y. 2012, *Implementasi Algoritma Data Mining K-Nearest Neighbour (KNN) Dalam Pengambilan Keputusan Pengajuan Kredit*, Jurusan Ilmu Komputer, Fakultas Sains dan Teknik, Universitas Nusa Cendana:Kupang.

[12] Suhartina, Ernastuti, 2010, *Graduation Prediction of Gunadarma University Students Using Algorithm and Naïve Bayes C4.5 Algorithm.*

[13] Wu X. 2009,*The Top Ten Algorithms in Data Mining*. New York:CRC Press.

**BIBLIOGRAFI PENULIS**

  **Imam Ahmad.S. Kom., M.Kom** lahir di Punggur 2 Agustus Lulus S1 dari program studi sistem informasi dan melanjutkan studi di Institut Pertanian Bogor Jurusan Ilmu Komputer

 **Heni Sulistiyani S.Kom., M.Kom** lahir di

**Hendrik Saputra,** lahir di Bandar Lampung pada 1 Oktober 1995, Lulus S1 dari program studi Informatika, Universitas Teknokrat Indonesia tahun 2018.