**JURTI**, Vol.1 No.1, Juni 2017, ISSN: 2579-8790 ◼

Model Algoritma K-Nearest Neighbor

Dalam Menganalisa Kinerja Akademik Mahasiswa

Prodi Teknik Informatika Universitas Mulawarman

**Adelowys Sinaga\*1, Masna Wati2, Novianti Puspitasari3**

1,2,3Jurusan Teknologi Informasi dan Komunikasi, Universitas Mulawarman, Samarinda

Barong Tongkok Kampus Gn.Kelua Universitas Mulawarman, Telp: 0541753133

e-mail: \*1**adeloywsinaga11@gmail.com**, 2masnawati.ssi@gmail.com, 3miechan.novianti.com

# ***Abstrak***

*Analisa kinerja akademik pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer dan Teknologi Informasi (FKTI), Universitas Mulawarman sangat diperlukan. Analisa kinerja ini bertujuan untuk mengantisipasi drop out (DO) mahasiswa. Dalam penelitian ini, algoritma K-Nearest Neighbor (K-NN) telah digunakan untuk menganalisa data akademik mahasiswa tiga angkatan (2012-2014) sebanyak 158 Data. Adapun alat bantu analisa menggunakan software Microsoft Excel dan Rapid Miner. Sebanyak 8 variabel terdiri dari Jenis Kelamin, Umur, IP Semester 1, IP Semester 2, IP Semester 3, IP Semester 4, IPK Hingga Semester 4 dan SKS, serta memiliki label predikat kelulusan Cepat, Tepat Waktu dan Terlambat. Hasil dari penelitian ini menunjukan bahwa Ratio terbaik yang ditandai dengan pola dasar 70:30 dengan 112 data training dan 46 data testing mendapatkan nilai akurasi 76,09%, precision 45,57%, recall 43,30%, specificity 78,87%, dan error rate 23,91%. Sehingga dapat diketahui bahwa kinerja akademik mahasiswa dengan IP Semester 2 menjadi faktor yang paling berpengaruh.*

***Kata kunci****— Universitas, Data Mining, Kinerja, Analisa, K-Nearest Neighbor.*

# 1. PENDAHULUAN

Pendidikan tinggi sebagai bagian dari sistem pendidikan nasional memiliki peran strategis dalam mencerdaskan kehidupan bangsa dan memajukan ilmu pengetahuan dan teknologi dengan memperhatikan dan menerapkan nilai humaniora serta pembudayaan dan pemberdayaan bangsa Indonesia yang berkelanjutan melalui perguruan tinggi yang merupakan sebuah institusi yang memiliki peran makro dalam pengembangan kualitas sumber daya manusia sebagai penyalur minat individu terhadap bidang tertentu sesuai minat dan bakat serta mengasah individu untuk menerapkan ilmu yang telah didapat demi stabilitas nasional. Program Studi Teknik Informatika di Fakultas Ilmu Komputer dan Teknologi Informasi, Universitas Mulawarman, Samarinda, Kalimantan Timur merupakan salah satu penyelenggara pendidikan tinggi yang ada di Indonesia, juga turut serta dalam pengembangan pendidikan, penelitian dan sumber daya manusia yang bermartabat, inovatif di bidang informatika yang berwawasan wirausaha dan mampu bersaing di tingkat Nasional dan Internasional di pulau Kalimantan. Berhubungan dengan kualitas akademik, program studi berkewajiban untuk memantau perkembangan studi dari mahasiswa. Pada program studi untuk penentuan kelulusan selama ini belum menggunakan sebuah metode untuk mengantisipasi terjadinya *Drop Out* pada mahasiswa, program studi hanya memberikan peringatan kepada mahasiswa agar lulus tepat waktu dan tidak terjadi *Drop Out,* oleh karena itu dibutuhkan sebuah analisis untuk mengetahui kinerja akademik mahasiswa dengan metode *K-Nearest Neighbor.*

Tujuan dari penelitian ini adalah untuk mengetahui analisa metode *K-Nearest Neighbor* dalam memprediksi kelulusan mahasiswa agar dapat mengetahui cepat, tepat waktu dan terlambatnya mahasiswa menjalankan perkuliahan pada Program Studi Teknik Informatika Fakultas Ilmu Komputer dan Teknologi Informasi Universitas Mulawarman.

# 2. METODE PENELITIAN

## Landasan Teori

### Evaluasi Proses Pembelajaran

Penentuan prestasi untuk seorang mahasiswa terhadap penguasaan suatu mata kuliah ditentukan berdasarkan hasil kuis, ujian, praktikum, dan/atau tugas lain yang mencangkup ranah kognitif, afektif, dan psikomotorik. Penilaian keberhasilan akademik mahasiswa didasarkan pada nilai bobot rata-rata atau Indeks Prestasi (IP). Indeks Prestasi dibedakan atas Indeks Prestasi Semester (IPS) dan Indeks Prestasi Komulatif (IPK). IPS dan IPK dihitung dari mata kuliah yang tertulis dalam kartu rencana studi, kecuali untuk mata kuliah yang termasuk dalamkelompok kegiatan akhir akademik yang belum memiliki nilai. Jumlah sks yang telah lulus dan IPK dipakai sebagai dasar evaluasi keberhasilan studi seorang mahasiswa Program Vokasi (D3) dan Sarjana (S1). Evaluasi keberhasilan studi mahasiswa pada akhir tahun pertama, akhir tahun kedua, dan akhir tahun ketiga dilakukan oleh fakultas untuk menentukan seorang mahasiswa diperbolehkan melanjutkan studi bila memenuhi syarat. [1]

1. *Data Mining*

*Data Mining* adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine* *learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar. Istilah Data Mining memiliki hakikat sebagai disiplin ilmu yang tujuan utamanya adalah untuk menemukan, menggali, atau menambang pengetahuan dari data atau informasi yang kita miliki. Data Mining, sering juga disebut sebagai *Knowledge Discovery in Database (KDD).* *KDD* adalah kegiatan yang meliputi pengumpulan, pemakaian data, historis untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar. [2]

### K-Nearest Neighbor

### Algoritma K-Nearest Neighbor merupakan salah metode klasifikasi dalam Data mining yang termasuk dalam algoritma supervised learning. Algoritma supervised learning merupakan algoritma yang bertujuan untuk menemukan pola pada data baru yang didasarkan pada analisa pola dalam Data Training. Adapun komponen fundamental dalam proses klasifikasi yaitu, Class untuk merupakan variabel dependen yang menunjukkan label dari suatu objek yang telah terklasifikasi, Predictors merupakan variabel independen yang menunjukkan karakter (atribut) dari data yang telah terklasifikasi, Training Dataset merupakan set data yang mengandung kedua komponen di atas dan digunakan sebagai model Training untuk mengenali class yang cocok berdasarkan predictors yang ada, Testing Dataset mengandung data baru yang akan di klasifikasikan berdasarkan Data Training dan nantinya dari akurasi dari klasifikasi tersebut akan di evaluasi. Algoritma K-Nearest Neighbor (KNN) juga merupakan salah satu contoh dari instance-based learning, yang mana Data Training yang telah terklasifikasi digunakan sebagai pembanding dalam menemukan persamaan pada Data Testing dengan Data Training yang ada.

### Langkah pertama dalam metode KNN yaitu menentukan nilai k pada Data Training bila k = 5 maka 5 record dari Data Training tersebut diambil berdasarkan jarak yang paling dekat dari record baru. Kemudian record baru tersebut dibandingkan berdasarkan dari kelima record dari Data Training yang memiliki jarak yang paling mendekati dengan record baru tersebut. [3]

### Model CRISP-DM

### Proses data mining berdasarkan CRISP-DM terdiri dari 6 fase yaitu, Business Understanding merupakan pemahaman tentang substansi dari kegiatan data mining yang akan dilakukan, kebutuhan dari perspektif bisnis. Kegiatannya antara lain menentukan sasaran atau tujuan bisnis, memahami situasi bisnis, menentukan tujuan data mining dan membuat perencanaan strategi serta jadwal penelitian. Data Understanding merupakan fase pengumpulan data awal, mempelajari data untuk bisa mengenal data yang akan dipakai, mengidentifikasikan masalah yang berkaitan dengan kualitas data, mendeteksi subset yang menarik dari data untuk membuat hipotesa awal. Data Preparation, fase yang padat karya. Aktivitas yang dilakukan antara lain memilih table dan field yang akan ditransformasikan ke dalam database baru untuk bahan data mining (set data mentah). Modeling, fase menentukan tehnik data mining yang digunakan, menentukan tools data mining, teknik data mining, algoritma data mining, menentukan parameter dengan nilai yang optimal. Evaluation, fase interpretasi terhadap hasil data mining yang ditunjukkan dalam proses pemodelan pada fase sebelumnya. Evaluasi dilakukan secara mendalam dengan tujuan menyesuaikan model yang didapat agar sesuai dengan sasaran yang ingin dicapai dalam fase pertama. Deployment, fase penyusunan laporan atau presentasi dari pengetahuan yang didapat dari evaluasi pada proses data mining [4].

### Confussion Matrix

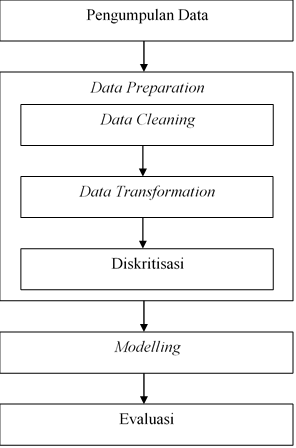
Dalam algoritma klasifikasi penting untuk mengetahui apakah objek yang terklasifikasi tersebut bernilai benar atau salah. Metode *Confussion Matrix* adalah metode yang berbentuk Tabel matriks digunakan untuk mengevaluasi adanya kesalahan dalam klasifikasi yang terjadi pada setiap objek. *Confussion Matrix* mengevaluasi seberapa akurat metode klasifikasi tersebut dalam memprediksikan *record* pada kelasnya. [5]

## Penelitian Relevan

Penelitian relevan yaitu membandingkan penelitian dengan judul penelitian sebelumnya yang sudah ada dengan judul atau metode yang sama, bertujuan untuk membuat perbedaan serta pengembangan dari penulisan sebelumnya. Dalam penelitian ini penulis memaparkan beberapa penelitian terdahulu yang relevan dengan permasalahan yang akan penulis teliti. Penelitian terdahulu dengan judul Model Algoritma *K-Nearest Neighbor (K-NN)* Untuk Prediksi Kelulusan Mahasiswa oleh Abdul Rohman (2014), hasil dari penelitian tersebut didapatkan tingkat akurasi pengujian model kelulusan mahasiswa dengan metode Algoritma *K-NN* dipengaruhi oleh jumlah klastering data, Akurasi dan nilai AUC paling tinggi adalah dengan mengklaster data k=5, didapatkan akurasi 85,15% dan AUC 0,888 [6]. Penelitian terdahulu dengan judul Komparasi Algoritma *Naïve Bayes* *Classifier* Dan *Tree C4.5* Untuk Evaluasi Kinerja Akademik Mahasiswa Program Studi Teknik Informatika Universitas Mulawarman oleh Wahyu Indrawan (2017) didapatkan bahwa bahwa faktor yang paling berpengaruh yaitu keaktifan mahasiswa, tempat kelahiran dan umur. Algoritma *Naïve Bayes Classifier* memiliki akurasi sebesar 76,79% dan presisi 80% serta *recall* 81,92% sedangkan C4.5 memiliki akurasi sebesar 78,57% dan presisi 90% serta *recall* 76,72% [7]. Penelitian dengan judul Penerapan Data mining Untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma *Naïve Bayes Classifier* oleh Mujib Ridwan, Hadi Suryono, M. Sarosa (2013) mendapatkan hasil *precision, recall,* dan *accuracy* menggunakan metode Algoritma *Naïve Bayes Classifier* masing-masing 83%, 50%, dan 70% [8]. Penelitian dengan judul Implementasi *Data mining* untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma *Naïve Bayes Classifier* oleh M. Syukri Mustafa, Muh Rizky Ramadhan, Angelina P. Thenata (2017) mendapatkan hasil akurasi 92,3% [9]. Penelitian dengan judul Kajian Perbandingan Teknik Klasifikasi Algoritma *C4.5*, *Naïve Bayes* Dan *CART* *(Classifcation and Regression Tree)* Untuk Prediksi Kelulusan Mahasiswa (Studi Kasus : Stmik Rosma Karawang) oleh Priati (2016), hasil pengujian menggunakan Algoritma *C4.5* mendapatkan Variabel paling berpengaruh adalah Jumlah SKS dan IP Semester 4 serta mendapatkan akurasi Algoritma *C4.5* sebesar 95,6%, *Naïve Bayes* 89,5 dan *CART* 95,6% [10].

## Tahapan Penelitian

Penelitian ini akan memiliki tahapan penelitian yakni identifikasi masalah, menetapkan tujuan penelitian, pengumpulan data awal, pengolahan data awal, proses *CRISP-DM*, hingga didapatkannya hasil dari model *K-Nearest Neighbor* melalui tahap *Modelling* dan Evaluasi. Tahapan penelitian secara lengkap dapat dilihat pada Gambar 1.



Gambar 1 Alur Tahapan Penelitian

### 2.2.1 Cross Industry Standart Process Data Mining (CRISP-DM)

Tahapan penelitian akan dilakukan penulis dengan menggunakan metode *CRISP-DM* yang terdiri dari 6 fase.

## 1. Bussines Understanding

## Penelitian ini telah mengetahui kinerja Algoritma K-Nearest Neighbor (KNN) dalam menganalisa kinerja akademik mahasiswa Program Studi Teknik Informatika, FKTI, Universitas Mulawarman. Dalam penelitian ini terdapat pula batasan masalah yang digunakan, diantara lainnya adalah Data Training dan Data Testing menggunakan set data dan data record dari data diri serta nilai mahasiswa lulusan program studi teknik informatika dari angkatan 2012 sampai dengan tahun 2014.

## 2. Data Understanding

## Data dalam penelitian ini diperoleh dari bagian akademik FKTI Universitas Mulawarman. Adapun data waktu kelulusan menggunakan data dari angkatan 2012 hingga angkatan 2014, yang mana didapatkan dari buku yudisium FKTI, Universitas Mulawarman mulai dari bulan Maret 2016 hingga bulan Maret 2018 dengan jumlah data yang diteliti sebanyak 158 data.

## 3. Data Preparation

Pada proses pengolahan data awal atau *preparation* peneliti akan melakukan berbagai proses agar mendapatkan dataset yang dapat diolah sesuai dengan tujuan penelitian.

## 4. Modelling

Algoritma yang akan digunakan dalam penelitian ini adalah algoritma *K-Nearest Neighbor* untuk evaluasi kinerja akademik. Berikut model K-Nearest Neighbor jika digambarkan dalam bentuk diagram alur yang akan dijelaskan pada Gambar 2.

## 

## Gambar 2 Flowchart algoritma K-Nearest Neighbor

### 5. Evaluation

Peneliti melakukan tahapan evaluasi dengan menghitung akurasi pemodelan *K-Nearest Neighbor* dengan metode *Confussion Matrix*

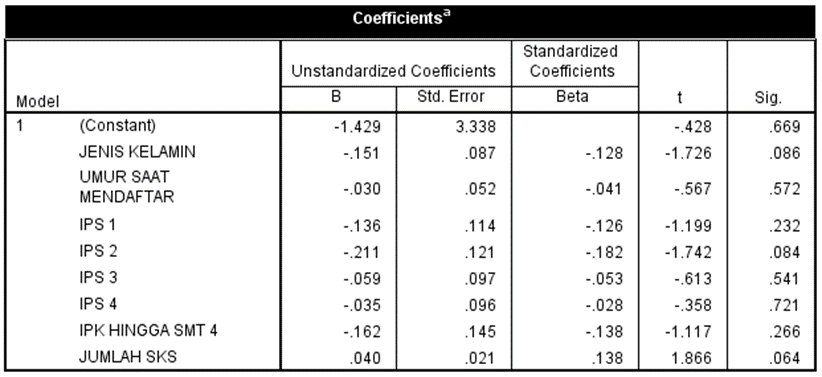
### 6. Deployment

Setelah didapatkan hasil dari pemodelan algoritma *K-Nearest Neighbor,* hasil dapat dijadikan strategi baru yang dapat mengetahui bagaimana data mahasiswa yang diprediksi berpotensi lulus tidak sesuai target dan dapat mengantisipasi terjadinya *drop out.*

# 3 HASIL DAN PEMBAHASAN

## 3.1 Hasil

Analisa variabel bertujuan untuk mengetahui variabel mana yang paling berpengaruh terhadap sebuah model data. Hal ini dapat dilihat pada nilai koefisien dari masing-masing variabel yang telah dicari menggunakan Aplikasi *IBM SPSS Statistic 23*. Hasil dari nilai koefisien variabel yang telah didapatkan bisa dilihat pada gambar 3.



Gambar 3. Nilai Koefisien Variabel

Dari gambar 3. diketahui bahwa nilai koefisien masing-masing variabel yaitu, Jenis Kelamin = -0,151, umur =-0030, ips1= -0,136, ips2 = -0,211, ips3 = -0,059, ips4 = -0,035, ipk4 = -0,162, dan jumlah sks = 0,40. Menentukan variabel yang paling berpengaruh adalah dengan mencari nilai koefisien yang paling menjauhi nilai Nol. Maka dapat disimpulkan bahwa variabel yang paling berpengaruh dalam penelitian ini adalah Indeks Prestasi Semester 2 dengan nilai koefisien = -0,211

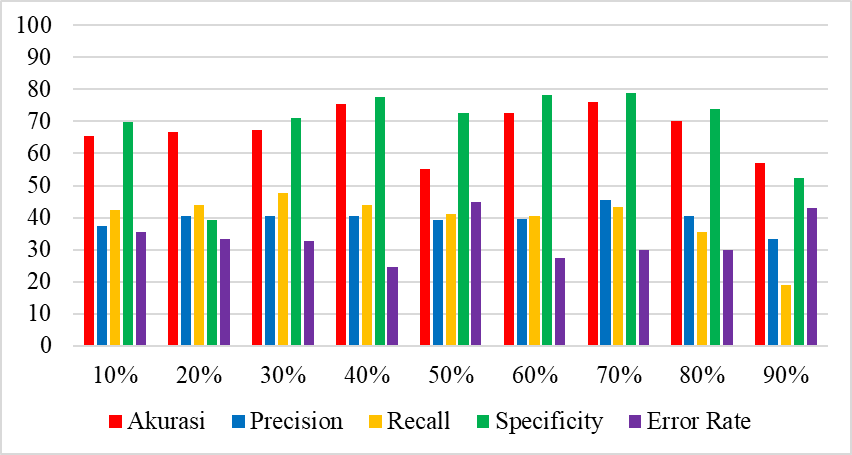
Pada perhitungan per *ratio* *Data Training,* yaitu 10% hingga 90%terdapat beberapa data yang dijadikan sebagai *Data Training* untuk diolah dalam perhitungan, data tersebut adalah data paling atas dari tabel data penelitian *(Linear Sampling),* sehingga tidak perlu mengacak urutan datanya kembali. Data inilah yang akan menjadi dasar perhitungan terhadap *Data Testing*. Yang mana *Data Testing* akan di mulai dari nomor urut selanjutnya hingga ke data paling bawah, dalam hal ini data nomor 158, sehingga didapatkan pada tahap evaluasi hasil dari *Confussion Matrix* dengan label cepat, tepat waktu dan terlambat menggunakan metode *K-Nearest Neighbor*. Beberapa hasil dari evaluasi algoritma yang telah dilakukan dapat dilihat pada tabel 1.

Tabel 1. Evaluasi Algoritma *K-Nearest Neighbor*

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Total Data** | **Evaluasi Algoritma** | | | | | | | |
| **Rasio (%)** | **Jumlah *Data Training*** | **Jumlah *Data Testing*** | **Akurasi (%)** | ***Precision* (%)** | ***Recall* (%)** | ***Specificity***  ***(%)*** | ***Error Rate* (%)** |
| 158 Data | 10% | 16 | 142 | 65,49% | 37,51% | 42,47% | 69,94% | 34,51% |
| 20% | 32 | 126 | 66,70% | 40,42% | 43,98% | 39,33% | 33.30% |
| 30% | 48 | 110 | 67,30% | 40,42% | **47,62%** | 70,93% | 32,70% |
| 40% | 64 | 94 | 75,53% | 40,42% | 43,98% | 77,60% | 24,47% |
| 50% | 80 | 78 | 55,26% | 39,31% | 41,13% | 72,64% | 44,74% |
| 60% | 96 | 62 | 72,58% | 39,48% | 40,55% | 78,08% | 27.42% |
| **\*70%** | **112** | **46** | **76,09%** | **45,57%** | 43,30% | **78,87%** | **23,91%** |
| 80% | 128 | 30 | 70,00% | 40,47% | 35,45% | 73,80% | 30,00% |
| 90% | 144 | 14 | 57,14% | 33,33% | 19,04% | 52,38% | 42.86% |

Berdasarkan tabel 1. hasil akurasi, *precision*, *recall*, *specificity* dan *error rate* dari algoritma *K-Nearest Neighbor* untuk menganalisa kinerja akademik mahasiswa berdasarkan masing-masing pembagian *Data Training* yang dimulai dari 10% hingga 90% didapat akurasi dan precision testinggi pada rasio 70%, dimana akurasi menunjukan angka 76,09%, *precision* 45,57%, *specificity* 78,87% dan *error rate* terendah dengan 23,91% sedangkan *recall* testinggi didapat pada rasio 30% dengan jumlah 47,62%. Hal ini menandakan bahwa model algoritma K-NN dengan hasil terbaik dalam menganalisa kinerja akademik mahasiswa adalah pada rasio 70%.

Dengan melihat tabel 1. Evaluasi Algoritma *K-Nearest Neighbor*, selain dalam bentuk tabel peneliti juga membuat dalam bentuk diagram grafik agar mudah dalam melihat tinggi rendahnya persentase yang didapat. diagram grafik tersebut dapat dilihat pada gambar 4.



Gambar 4. Diagram Hasil *Confussion Matrix* Metode *K-NN*

# 4 KESIMPULAN

Berdasarkan penelitian yang telah dilakukan, maka penulis dapat menarik kesimpulan bahwa hasil dari penelitian ini menunjukan Ratio 70% adalah yang terbaik, ditandai dengan pola dasar 70:30 dengan 112 Data Training dan 46 Data Testing mendapatkan nilai akurasi sebesar 76,09%, precision 45,57%, recall 43,30%, specificity 78,87%, dan error rate 23,91%. Sehingga dapat diketahui bahwa kinerja akademik mahasiswa dengan IP Semester 2 menjadi faktor yang paling berpengaruh.

# 5. SARAN

Dari hasil anailisa yang dilakukan maka disarankan agar perlu adanya pertimbangan kembali dalam menentukan variabel yang memiliki pengaruh yang significant dalam waktu kelulusan mahasiswa Program Studi Teknik Informatika Fakultas Ilmu Komputer dan Teknologi Informasi.Sampel data harusnya lebih tinggi agar data yang diolah lebih banyak sehingga dapat mempengaruhi hasil prediksi.

UCAPAN TERIMA KASIH

Dengan mengucapkan puji dan syukur atas kehadirat Tuhan yang Maha Kuasa yang telah senantiasa melimpahkan rahmat dan karunia-Nya. Bersyukur dan salam penulis sanjungkan kepada Tuhan, yang telah memberikan penulis pengharapan untuk bisa berjuang dalam masa kuliah sampai pada saat ini, sehingga penulis dapat menyelesaikan penelitian ini. Terima kasih kepada orang tua saya yang selalu memberi dukungan materil maupun moril serta semangat. Kedua pembimbing Ibu Masna Wati, S.Si., MT, dan Ibu Novianti Puspitasari, S.Kom., M.Eng yang telah membimbing dengan sabar dan baik dalam penelitian ini. Terima kasih kepada teman baik saya Dwicky Ari Pandawa, S.Kom dan Alfan Nurfaizin, S.Kom yang senantiasa memberi dukungan dan bantuan dalam penelitian ini.

## DAFTAR PUSTAKA

1. Senat Universitas Mulawarman. (2017). Peraturan Akademik Universitas Mulawarman. Samarinda: Universitas Mulawarman.
2. Han, Jiawei., Kamber, Micheline., & Pei, Jian. (2011). DATA MINING CONCEPTS AND TECHNIQUES (3rd Edition).
3. Mutrofin, S., Izzah, A., Kurniawardhani, A., & Masrur, M. (2015). Optimasi teknik klasifikasi modified k nearest neighbor menggunakan algoritma genetika. Jurnal Gamma, 10(1).
4. Shearer, Colin. (2000) The CRISP-DM Model: The New Blueprint for Data Mining.
5. Sugiyono, P. D. (2016). Metode Penelitian Kuantitatif, Kualitatif, dan R&D. Bandung: Alfabeta.
6. Rohman, Abdul. (2017), Model Algoritma K-Nearest Neighbor (K-NN) Untuk Prediksi Kelulusan Mahasiswa.
7. Indrawan, Wahyu. (2017), Komparasi Algoritma Naive Bayes Classifier Dan Tree C4.5 Untuk Evaluasi Kinerja Akademik Mahasiswa Program Studi Teknik Informatika Universitas Mulawarman.
8. Ridwan, M., Suyono, H., & Sarosa, M. (2013). Penerapan Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier. jurnal EECCIS, 7(1), 59-64.
9. Mustafa, M. S., Ramadhan, M. R., & Thenata, A. P. (2018). Implementasi Data Mining untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier. Creative Information Technology Journal, 4(2), 151-162
10. Priati. (2016). Kajian Perbandingan Teknik Klasifikasi Algoritma C4.5, Naïve Bayes Dan CART (Classifcation and Regression Tree) Untuk Prediksi Kelulusan Mahasiswa (Studi Kasus : Stmik Rosma Karawang)