

**PERBANDINGAN PENERAPAN ALGORITMA *C4.5*, *NAÏVE*
BAYES DAN *K-NEAREST NEIGHBOR (K- NN)* DALAM
KLASIFIKASI KELULUSAN MAHASISWA**

SKRIPSI



OLEH :

**YUHENDRI PUTRA
1855201154**

**TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS LANCANG KUNING
PEKANBARU
2022**

HALAMAN PERSETUJUAN

Nama : Yuhendri Putra

NIM : 1855201154

Program Studi : Teknik Informatika

Judul : Perbandingan Penerapan Algoritma *C4.5*, *Naïve Bayes*

dan *K- Nearest Neighbor (K-NN)* Dalam Klasifikasi

Kelulusan Mahasiswa.

Pekanbaru, 10 Februari 2021

Disetujui oleh,
Pembimbing


Guntoro, S.T., M.Kom., MTA.
NIDN. 1012018803

Mengetahui,

Dekan

Fakultas Ilmu Komputer



H. Farizal, SP., M.Kom., MTA.
NIDN. 1024076201

Ketua Program Studi
Teknik Informatika



Dr. Yogi Yusef, M.Kom., MTA., MCE
NIDN. 1022068803

HALAMAN PENGESAHAN

Nama : Yuhendri Putra
NIM : 1855201154
Program Studi : Teknik Informatika
Judul : Perbandingan Penerapan Algoritma *C4.5*, *Naïve Bayes* dan
K-Nearest Neighbor (K-NN) Dalam Klasifikasi Kelulusan
Mahasiswa.

Skripsi ini telah dipertahankan di depan Dewan Penguji pada tanggal 12 Februari
2022 sebagai persyaratan untuk mendapatkan gelar Sarjana Komputer (S.Kom)

Dewan Penguji
Ketua

Guntoro, S.T., M.Kom., MTA.
NIDN. 1012018803

Anggota

Lisnawita., M.Kom., MTA
NIDN.1005098503

Anggota

Ahmad Zamsuri, M.Kom., MTA
NIDN. 1007078602

Mengetahui,

Dekan



H. Fairizal, S.P., M.Kom., MTA
NIDN 1024076201

Ketua Program Studi



Pr. Yogi Yunefri, M.Kom., MTA., MCE
NIDN 1022068803

2022/03/24 09:21

PERNYATAAN KEASLIAN KARYA ILMIAH

Dengan ini saya nyatakan bahwa dalam Tugas Akhir Semester ini tidak terdapat karya yang pernah di ajukan untuk memperoleh gelar ke serjanaan di suatu Perguruan Tinggi, dan sepanjang pengetahuan saya, juga tidak terdapat Karya/Pendapat yang pernah ditulis oleh orang lain kecuali secara tertulis diacu dalam naskah ini dan disebutkan dalam Daftar Pustaka.

Apabila ditemukan suatu jiplakan atau plagiat maka saya bersedia menerima akibat berupa sanksi yang dierikan oleh pihak yang berwewenang sesuai ketentuan peraturan Undang-Undangan yang berlaku.

Pekanbaru, 6 Febuari 2022

Yang membuat pernyataan



Yuhendri Putra
1855201154

HALAMAN PERSEMBAHAN

Alhamdulillah puji dan rasa syukur tak hentinya Saya Ucapkan kepada Allah SWT yang telah memberikan kenikmatan dan kemudahan sehingga skripsi dapat selesai pada waktunya. Tidak lupa saya ucapkan Terimakasih sebesar-besarnya kepada kedua orang tua yang selalu ada dalam keadaan dan kondisi apapun, selalu berjuang, dan selalu mendo'akan sehingga saya bisa ada pada titik yang sekarang ini dengan harapan bisa menyelesaikan skripsi ini. Karna berkat do'a kedua orang tua dan dorongan yang selalu mereka berikan sehingga saya bisa duduk dibangku kuliah ini. Semua yang dilakukan oleh kedua orang tua saya, itu tidak bisa ditulis dengan kata-kata. Dan karna merekalah saya bersemangat untuk menyelesaikan tugas akhir saya ini. Banyak rintangan yang dilalui tetapi dengan melihat semua yang telah dilakukan oleh kedua orang tua terhadap anaknya, hanya untuk memberikan yang terbaik buat masa depan anaknya nanti.

Buat FASILKOM serta keluarga yang ada FASILKOM saya sangat-sangat berterima kasih atas pelayanan dan fasilitas yang disediakan untuk pendidikan mahasiswa FASILKOM. Dan juga saya ucapkan terimakasih banyak terhadap dosen dan sftaf FASILKOM Universitas Lancang Kuning Riau atas ilmu yang diberikan dalam pembelajaran perkuliahan maupun dalam bentuk pengalaman yang telah diberikan kepada kami mahasiswa.

Dan tidak lupa juga saya ucapkan terimakasih buat orang-orang yang selalu mendukung dan selalu ada dikala suka maupun duka yaitu kepada rekan-rekan saya yang sangat saya sayangi dan saya banggakan. Terutama kepada rekan-rekan saya tim teletabies yang selalu berjuang dari awal sampai sekarang, terimakasih

buat kalian semua semoga kalian sehat selalu Dan bisa membuktikan kepada semua apa yang kalian punya bisa membanggakan orang-orang yang kalian sayangi, terutama buat kedua orang tua rekan-rekan sekalian.

KATA PENGANTAR

Segala puji dan syukur kami sampaikan kehadiran Allah SWT Tuhan Yang Maha Esa, yang telah memberikan rahmat, taufik dan hidayah-Nya kepada kita semua. Sholawat dan salam tidak lupa kita sampaikan kepada Nabi Muhammad SAW. Bahwa atas izinnya Proposal Skripsi yang berjudul **PERBANDINGAN PENERAPAN ALGORITMA C4.5, NAÏVE BAYES DAN K- NN DALAM KLASIFIKASI KELULUSAN MAHASISWA**

Penulisan skripsi ini ditunjukan untuk melengkapi salah satu syarat kelulusan dan mencapai jenjang sarjana di Fakultas Ilmu Komputer, Universitas Lancang Kuning, Pekanbaru. Skripsi ini dapat disusun dengan baik karna banyak masukan dan dukungan dari berbagai pihak yang berupa informasi, arahan, dan bimbingan, maka oleh karena itu penulis mengucapkan terimakasih kepada:

1. Bapak H. Fajrizal, S.P., M.Kom., MTA. sebagai Dekan Fakultas Ilmu Komputer Universitas Lancang Kuning.
2. Bapak Ahmad Zamsuri, M.Kom., MTA. selaku Wakil Dekan I Fakultas Ilmu Komputer Universitas Lancang Kuning.
3. Ibu Lucky Lhaura Van FC, M.Kom., MTA. selaku Wakil Dekan II Fakultas Ilmu Komputer Universitas Lancang Kuning.
4. Bapak Muhamad Sadar, SE, M.Kom., MTA. selaku Wakil Dekan III Fakultas Ilmu Komputer Universitas Lancang Kuning.
5. Bapak Dr. Yogi Yunefri, M.Kom., MTA., MCE. selaku Ketua Prodi Teknik Informatika Fakultas Ilmu Komputer Universitas Lancang Kuning. Yang

telah memberikan kesempatan untuk melakukan penelitian Skripsi di Prodi Teknik Informatika FASILKOM.

6. Bapak Guntoro, S.T., M.Kom., MTA. Pembimbing yang telah memberikan masukan terhadap penulisan laporan Skripsi.
7. Bapak dan Ibu Dosen Fakultas Ilmu Komputer yang telah memberikan bekal ilmu pengetahuan kepada penulis
8. Segenap staf / karyawan Fasilkom yang telah melayani dan memberikan fasilitas bagi kelancaran laporan Skripsi ini.
9. Orang tua dan keluarga penulis yang selalu memberikan dukungan, doa, bimbingan, serta kasih sayang yang telah diberikan sehingga penulis dapat menyelesaikan laporan Skripsi ini..
10. Teman-teman seperjuangan yang telah memberikan motivasi, masukan dan dukungan serta semangat kepada penulis dalam menyelesaikan laporan Skripsi ini.

Pekanbaru, 10 Febuari 2021

Penilis

PERBANDINGAN PENERAPAN ALGORITMA C4.5, NAÏVE BAYES DAN K-NEAREST NEIGHBOR (K- NN) DALAM KLASIFIKASI KELULUSAN MAHASISWA

Yuhendri Putra¹, Guntoro²

Program Studi Teknik Informatika Fakultas Ilmu Komputer
Universitas Lancang Kuning

(Jl. Yos Sudarso KM. 8 Rumbai, Pekanbaru, Riau, telp. 0811 753 2015)

e-mail: yuhendri204@gmail.com, guntoro@unilak.ac.id

ABSTRAK

Klasifikasi termasuk suatu cara untuk mengelompokan teks, jadi teks dengan isi teks yang sama akan dikelompokan dengan tipe teks yang sama. Metode *data mining* untuk klasifikasi mampu digunakan untuk membantu meningkatkan kecepatan dan ketepatan dalam klasifikasi kelulusan mahasiswa. Di Fakultas Teknik Informatika Universitas Lancang Kuning Pekanbaru, ada dua prodi yaitu Teknik Informatika dan Teknik Informasi. Pada penelitian ini dilakukan klasifikasi pada data mahasiswa Teknik Informatika pada angkata 2015 s/d 2017 menggunakan algoritma C4.5, *Naïve Bayes classifier* dan *K-Nearest Neighbor* dengan pengujian menggunakan *Rapidminer*. Tiga algoritma ini termasuk algoritma yang paling populer digunakan dalam klasifikasi. Di penelitian ini akan melakukan perbandingan dari tiga algoritma ini. Perbandingan ini dilakukan guna untuk mengukur akurasi dari tiap-tiap algoritma untuk mendapatkan metode terbaik yang digunakan pada klasifikasi kelulusan mahasiswa. Pengujian akurasi metode pada penelitian ini dilakukan dengan *tes option cross validation* dan eavaluasi data uji menggunakan *confusion matrix*. Dari penelitian yang telah dilakukan, algoritma C4.5 yang paling tinggi tingkat akurasinya. Pada algoritma C4.5 didapatkan akurasi tertinggi dari tiga kalai pengujian setiap algoritma, C4.5 mendapatka 100% akurasi pertama, kedua, dan ketiga. *Naïve Bayes classifier* mendapatkan nilai akurasi seberas 99.34% pada pengujian pertama dan 100% akurasi untuk pengujian kedua dan ketiga. Dan *K-Nearest Nighbor* mendapatkan niali akurasi 97.37% pada pengujian kedua, 98.69% niali akurasi, dan pada pengujian ke tiga mendapatkan nilai akurasi 100%.

Kata kunci : Perbandingan Algoritma, *DataMining*, Kalsifikasi

COMPARISON OF APPLICATION OF C4.5, NAVE BAYES AND K-NEAREST NEIGHBOR (K-NN) ALGORITHM IN STUDENT GRADUATION CLASSIFICATION

Yuhendri Putra¹, Guntoro²

*Informatics Engineering Study Program, Faculty of Computer Science
Lancang Kuning University*

(Jl. Yos Sudarso KM. 8 Rumbai, Pekanbaru, Riau, tel. 0811 753 2015)

e-mail: yuhendri204@gmail.com, guntoro@unilak.ac.id

ABSTRACT

Classification includes a way to group text, so text with the same text content will be grouped with the same text type. Data mining methods for classification can be used to help increase speed and accuracy in the classification of student graduation. At the Faculty of Informatics Engineering, Lancang Kuning University, Pekanbaru, there are two study programs, namely Informatics Engineering and Information Engineering. In this study, classification was carried out on Informatics Engineering student data from 2015 to 2017 using the C4.5 algorithm, Naïve Bayes classifier and K-Nearest Neighbor by testing using Rapidminer. These three algorithms are among the most popular algorithms used in classification. In this study, we will compare these three algorithms. This comparison is carried out in order to measure the accuracy of each algorithm to get the best method used in the classification of student graduation. Testing the accuracy of the method in this study was carried out by using a cross validation option test and evaluating the test data using a confusion matrix. From the research that has been done, the C4.5 algorithm has the highest level of accuracy. In the C4.5 algorithm, the highest accuracy was obtained from three times of testing each algorithm, C4.5 got 100% accuracy for the first, second, and third funds. The Naïve Bayes classifier got an accuracy of 99.34% in the first test and 100% accuracy in the second and third tests. And K-Nearest Night got 97.37% accuracy value in the second test, 98.69% accuracy value, and in the third test it got 100% accuracy value.

Keywords: *Comparison of Algorithms, Data Mining, Calcification*

DAFTAR ISI

HALAMAN PENGESAHAN	Error! Bookmark not defined.
PERNYATAAN KEASLIAN KARYA ILMIAH	i
HALAMAN PERSEMBAHAN	ii
KATA PENGANTAR	iv
ABSTRAK	vi
ABSTRACT	vii
DAFTAR ISI	viii
DAFTAR TABEL	xi
DAFTAR GAMBAR	xiv
BAB I PENDAHULUAN	1
A. Latar Belakang	1
B. Rumusan Masalah	4
C. Batasan Masalah.....	4
D. Tujuan Penelitian	5
E. Manfaat Penelitian	5
F. Sistematika Penelitian	5
BAB II TINJAUAN PUSTAKA	7
A. Penelitian Terdahulu	7
B. Teori Dasar yang Digunakan	10
1. Data Mining.....	10
2. Klasifikasi Data Mining	13
3. Algoritma C4.5	14
4. Naive Bayes.....	16

5. K-Nearest Neighbor (K-NN).....	17
6. RapidMiner.....	17
BAB III METODOLOGI PENELITIAN	19
A. Tahapan-Tahapan Penelitian.....	19
B. Lokasi dan waktu penelitian.....	22
C. Data yang digunakan.....	22
D. Teknik pengumpulan data	22
E. Metode Yang Digunakan	23
BAB IV	26
GAMBARAN UMUM TEMPAT PENELITIAN	26
A. Sejarah Umum Fakultas Ilmu Komputer Universitas Lancang Kuning .	26
B. Visi, Misi, dan Tujuan FSILKOM	27
C. Struktur Organisasi Fakultas Ilmu Komputer	27
D. Tugas Jabatan Fakultas Ilmu Komputer (FASILKOM).....	28
E. Struktur Pejabat Fakultas Ilmu Komputer.....	32
F. Logo Fakultas Ilmu Komputer	33
BAB V.....	33
HASIL PEMBAHASAN.....	34
A. <i>Dataset</i> Yang Digunakan	34
B. Perhitungan Algoritma C4.5	38
C. Pengujian Algoritma C4.5 Menggunakan Rapidminer	41
D. Perhitungan Algoritma Naïve Bayes.....	46
E. Pengujian Algoritma Naïve Bayes Menggunakan Rapidminer.....	52
F. Perhitungan Algoritma <i>K-Nearest Neighbor</i> (K-NN).....	56
G. Pengujian Menggunakan <i>Rapidminer</i>	61

H. Hasil Perbandingan	65
BAB VI.....	68
KESIMPULAN DAN SARAN	68
DAFTAR PUSTAKA.....	70
LAMPIRAN.....	72

DAFTAR TABEL

Tabel 4. 1. Jabatan Faskultas Ilmu Komputer	32
Tabel 5. 1. <i>Dataset</i> yang digunakan.....	34
Tabel 5. 2. Tabel infor,asi IPS & IPK	38
Tabel 5. 3. Perhitungan algoritma untuk mencari nilai <i>entropy</i> dan <i>gain</i>	40
Tabel 5. 4. Table data sampel.....	40
Tabel 5. 5. Hasil dari sampel yang didapatkan.....	41
Tabel 5. 6. Jumlah Nilai Probabilitas Dari Atribut Jenis Kelamin.....	47
Tabel 5. 7. Jumlah Nilai Probabilitasn Atribut Target Kelulusan.	47
Tabel 5. 8. Tabel perhitungan Nilai Terhadap Jenis Kelamin pada IPS 1.....	48
Tabel 5. 9. Tabel perhitungan Nilai Terhadap Jenis Kelamin pada IPS 2.....	48
Tabel 5. 10. Tabel perhitungan Nilai Terhadap Jenis Kelamin pada IPS 3.....	48
Tabel 5. 11. Tabel perhitungan Nilai Terhadap Jenis Kelamin pada IPS 4.....	48
Tabel 5. 12. Tabel perhitungan Nilai Terhadap Jenis Kelamin pada IPS 5.....	48
Tabel 5. 13. Tabel perhitungan Nilai Terhadap Jenis Kelamin pada IPS 6.....	49
Tabel 5. 14. Tabel perhitungan Nilai Terhadap Jenis Kelamin pada IPS 1.....	49
Tabel 5. 15. Tabel perhitungan Nilai Terhadap Target Kelulusan pada IPS 1.....	49
Tabel 5. 16. Tabel perhitungan Nilai Terhadap Target Kelulusan pada IPS 2.....	49
Tabel 5. 17. Tabel perhitungan Nilai Terhadap Target Kelulusan pada IPS 3.....	50
Tabel 5. 18. Tabel perhitungan Nilai Terhadap Target Kelulusan pada IPS 4.....	50
Tabel 5. 19. Tabel perhitungan Nilai Terhadap Target Kelulusan pada IPS 5.....	50
Tabel 5. 20. Tabel perhitungan Nilai Terhadap Target Kelulusan pada IPS 6.....	50
Tabel 5. 21. Tabel perhitungan nilai terhadap target kelulusan pada IPK	51

Tabel 5. 22. Tabel Sampel Sebagai Data Testing.....	51
Tabel 5. 23. Hasil perhitungan atribut jenis kelamin dari tiap-tiap kategorinya ..	51
Tabel 5. 24. Hasil dari perhitugan data sampel	52
Tabel 5. 25. Hasil perhitungan <i>Euclidean Distance</i>	57
Tabel 5. 26. Hasil Perbandingan Akurasi	67

DAFTAR GAMBAR

Gambar 2. 1. Tahapan-tahapan KDD.....	11
Gambar 3. 1 Tahapan-tahapan penelitian.....	19
Gambar 4. 1 Struktur Organisasi FASILKOM UNILAK.....	28
Gambar 5.1. 1. Pohon keputusan.....	41
Gambar 5.1. 2. <i>Import</i> data kedalam <i>rapidminer</i>	42
Gambar 5.1. 3. Memasukan data yang kedalam proses <i>rapidminer</i>	42
Gambar 5.1. 4. Menghubungkan data dengan dlgoritma.....	43
Gambar 5.1. 5. <i>Seting Set Role</i>	43
Gambar 5.1. 6. <i>Seting Algoritma</i>	44
Gambar 5.1. 7. Pohon Keputusan.....	44
Gambar 5.1. 8. Proses <i>cross validation</i>	45
Gambar 5.1. 9. Proses <i>Validation Data</i>	45
Gambar 5.1. 10. Hasil perhitungan Algoritma C4.5	45
Gambar 5.1. 11. Hasil perhitungan Algoritma C4.5	46
Gambar 5.1. 12. Hasil perhitungan Algoritma C4.5	46
Gambar 5.2. 1. Proses <i>Rapidminer</i> Pada Algoritma <i>Naïve Bayes</i>	53
Gambar 5.2. 2. Perhitungan <i>Rapidminer</i> Dengan Algoritma <i>Naïve Bayes</i>	53
Gambar 5.2. 3. Hasil Perhitungan <i>Naïve Bayes</i> Dengan <i>Rapidminer</i>	54
Gambar 5.2. 4. Hasil Perhitungan <i>Naïve Bayes</i> Dengan <i>Rapidminer</i>	54
Gambar 5.2. 5. Hasil Akurasi Yang Didapat Menggunakan <i>Rapidminer</i>	55
Gambar 5.2. 6. Hasil Perhitungan <i>Naïve Bayes</i> Menggunakan <i>Rapidminer</i>	55
Gambar 5.2. 7. Hasil akurasi pengujian ketiga	56

Gambar 5.2. 9. Distance Yang Sudah di Urutkan Ranknya	61
Gambar 5.2. 10. Hasil Dari Data Sampel.....	61
Gambar 5.3. 1. Proses Rapidminer yang dilakukan pada algoritma <i>K-NN</i>	61
Gambar 5.3. 2. Hasil perhitungan <i>K-NN</i>	62
Gambar 5.3. 3. Hasil Akurasi.....	62
Gambar 5.3. 4. Hasil Perhitunga	63
Gambar 5.3. 5. Hasil Akurasi Pada Pengujian Kedua.....	63
Gambar 5.3. 6. Hasil Perhitungan	64
Gambar 5.3. 7. Hasil Akurasi Yang Didapat	64

LAMPIRAN

Lampiran 1. Surat Izin Penelitian dari Fakultas.....	72
Lampiran 2. Foto Ditempat Penelitian.	73
Lampiran 3. Bukti Wawancara	75
Lampiran 4 Bukti Lulus Turnitin	77
Lampiran 5 Surat Keterangan Turnitin	78
Lampiran 6. Data Mahasiswa.....	79

BAB I

PENDAHULUAN

A. Latar Belakang

Pada masa informasi dan era teknologi canggih seperti beberapa tahun terakhir ini dapat kita lihat bahwa sudah banyak sistem komputer yang dibangun dengan menggunakan desain database untuk data berskala besar. Apalagi didukung dengan tempat penyimpanan data yang sangat besar pula sehingga mendukung manusia untuk leluasa menyimpan banyaknya data.

Semakin banyak data tersimpan maka semakin banyak pula penumpukan yang terjadi pada data sangat banyak pada tempat penyimpanan data ke dalam satu atau beberapa kelas yang sudah didefinisikan sebelumnya. Kondisi data yang bertumpuk terus menerus ini akan sia-sia jika tidak dimanfaatkan kembali untuk kebutuhan informasi dimana kebutuhan informasi dari tahun ke tahun terus meningkat.(Saputra & Primadasa, 2018)

Untuk mengatasi supaya tidak terjadi penumpukan data berskala besar, digunakanlah suatu teknik penggalian informasi dari data yang sudah bertumpuk tersebut. Teknik yang disebut dengan *Data mining*. *Data mining* merupakan teknik yang sering digunakan untuk menggali informasi yang tersembunyi dalam data yang besar. Sehingga *Data mining* bisa menjadi solusi untuk mengatasi tumpukan data yang ada pada tempat penyimpanan, dengan menggunakan teknik data mining tersebut kita dapat

menemukan informasi yang berupa pola, ciri, dan aturan atau dikenal sebagai istilah *knowledge*.

Salah satu misi Universitas Lancang Kuning ialah Meningkatkan Kualitas Pendidikan untuk Menghasilkan Lulusan Kompetitif Berskala Nasional. Kualitas pendidikan disini terkait langsung dengan kemampuan mahasiswa yang merupakan aspek penting dalam menentukan keberhasilan penyelenggaraan program studi pada suatu perguruan tinggi. Salah satu indikator dari keberhasilan suatu perguruan tinggi dapat dilihat dari banyaknya tingkat kelulusan mahasiswa setiap tahunnya.

Pada peneliti sebelumnya sudah melakukan penelitian untuk mengukur tingkat kelulusan mahasiswa Fakultas Ilmu Komputer Universitas Lancang Kuning melalui klasifikasi lulus tepat waktu atau lulus tidak tepat waktu menggunakan algoritma *Naïve Bayes Classifier* yang dilakukan oleh (N. Nasution et al., 2015) berdasarkan evaluasi kinerja akademik di tingkat 3 atau semester 6 dengan hasil dari penelitian dengan pembangunan model *Naïve Bayes* terbaik berada pada akurasi 76%, kemudian model tersebut diterapkan untuk prediksi kelulusan mahasiswa pada tingkat 3, dengan perolehan hasil 48 mahasiswa lulus tepat waktu dan 13 mahasiswa lulus tidak tepat waktu. Pada penelitian ini atribut yang digunakan adalah jenis kelamin; program studi; asal sekolah; daerah asal; tempat tanggal lahir; indeks prestasi semester 1 sampai dengan 5; indeks prestasi kumulatif; dan target kelulusan

Dalam penelitian yang dilakukan oleh (Rohman & Rochcham, 2019) perbandingan metode *data mining* yaitu algoritma *neural network*, *k-nearest neighbor* dan *Decision Tree* yang diaplikasikan pada data kelulusan mahasiswa. Pada penelitian ini parameter yang digunakan adalah usia, jenis kelamin, indeks prestasi semester satu sampai dengan indeks prestasi semester empat. Dari hasil pengujian dengan mengukur kinerja ketiga metode tersebut menggunakan *cross validation*, *confusion matrix* dan kurva *ROC*, diketahui bahwa *neural network* memiliki nilai akurasi yang paling tinggi yakni sebesar 87,32%, diikuti oleh metode algoritma *Decision Tree* dengan nilai akurasi 83,57%, kemudian metode *k-nearest neighbor* dengan akurasi 83,66%. Nilai *AUC* untuk metode *neural network* menunjukkan nilai tertinggi yakni 0,917 dan yang terendah adalah metode *Decision Tree* yaitu 0,844.

Dari beberapa penelitian di atas, dapat disimpulkan bahwa atribut yang signifikan dalam penentuan kelulusan secara umum adalah Indeks Prestasi Kumulatif (IPK) dengan menggunakan algoritma *Naïve Bayes*, *k-nearest neighbor*, *Decision Tree*, *neural network*. Sehingga pada penelitian ini akan dilakukan uji coba perbandingan algoritma klasifikasi *C4.5*, *Naïve Bayes classifier* dan *K-Nearest Neighbor* dengan beberapa atribut terpilih seperti NIM, Nama, Jenis Kelamin, Indeks Prestasi Semester (IPS) 1-6, Atribut nilai dan Indeks Prestasi Kumulatif (IPK) yang dapat meningkatkan akurasi.

Berdasarkan latar belakang diatas, maka penulis mengambil judul **“Perbandingan Penerapan Algoritma C4.5, Naïve Bayesian, dan K-Nearest Neighbor dalam Klasifikasi Kelulusan Mahasiswa (studi kasus: Fakultas Ilmu Komputer Universitas Lancang Kuning)”**. Penelitian ini akan menggunakan software *Rapidminer* sebagai alat bantu untuk mengevaluasi kinerja tiga algoritma tersebut.

B. Rumusan Masalah

Berdasarkan penelitian diatas, maka rumusan masalah dalam penelitian ini, adalah:

1. Bagaimana melakukan perbandingan Algoritma klasifikasi kelulusan mahasiswa berdasarkan data mahasiswa Fasilkom..?
2. Bagaimana tingkat akurasi Algoritma C4.5, Naïve Bayes dan K-Nearest Neighbor(K-NN)..?

C. Batasan Masalah

Pada penelitian ini peneliti memberikan batasan masalah agar tidak keluar dari pembahasan, yaitu :

1. Data yang dianalisis berdasarkan data mahasiswa teknik informatika angkatan 2015-2017 di FASILKOM Universitas Lancang Kuning.
2. Mencari perbandingan dan tingkat akurasi Algoritma C4.5, Naïve Bayes dan K-Nearest Neighbor(K-NN)

D. Tujuan Penelitian

Adapun tujuan dari penelitian ini adalah sebagai berikut:

1. Untuk mengetahui klasifikasi yang terbentuk dari perbandingan algoritma klasifikasi *Data mining*
2. Mengetahui tingkat akurasi Algoritma *C4.5*, *Naïve Bayes* dan *K-Nearest Neighbor(K-NN)*

E. Manfaat Penelitian

Adapun manfaat dari penelitian ini di antara lain:

1. Bagi Penulis
Sebagai pengalaman dan penambah wawasan yang sudah didapatkan selama masa pendidikan di Fakultas Ilmu Komputer Universitas Lancang Kuning.
2. Bagi Tempat Penelitian
Mengetahui hasil klasifikasi kelulusan mahasiswa dari perbandingan algoritma *Data mining*.
3. Bagi Pembaca
Sebagai referensi tentang analisa perbandingan algoritma klasifikasi *Data mining*.

F. Sistematika Penelitian

Sistematika penulisan proposal skripsi diperlukan untuk melihat gambaran – gambaran pokok yang telah dibuat oleh penulis, adapun sistematika penulisan dari skripsi ini adalah:

BAB I PENDAHULUAN

Bab ini berisikan tentang latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan.

BAB II TINJAUAN PUSTAKA

Bab ini berisikan tinjauan penelitian terdahulu, dan teori dasar yang digunakan berkaitan dengan penyusunan laporan penelitian.

BAB III METODOLOGI PENELITIAN

Bab ini berisikan mengenai tahapan-tahapan penelitian, lokasi dan waktu penelitian, data yang digunakan, teknik pengumpulan data, dan metode atau pemodelan yang akan digunakan.

BAB IV GAMBARAN UMUM OBJEK PENELITIAN

Bab ini berisikan gambaran umum tempat objek penelitian.

BAB IV HASIL DAN PEMBAHASAN

Bab ini berisikan hasil dan pembahasan berupa analisa perbandingan dengan metode yang dipakai serta pengujian hasil menggunakan *software*.

BAB IV KESIMPULAN DAN SARAN

Bab ini berisikan kesimpulan dari pelaksanaan penelitian yang penulis lakukan dan saran untuk penelitian selanjutnya.

DAFTAR PUSTAKA

LAMPIRAN

BAB VI

KESIMPULAN DAN SARAN

A. Kesimpulan

Dari hasil pada penjelasan diatas bisa disimpulkan bahwa penggunaan Algoritma *C4.5*, *Naïve Bayes*, *K-Nearest Nighbor* dalam perbandingan tidak jauh beda untuk akurasi masing-masing algoritma. Dalam perbandingan algoritma dengan perhitungan manual ketiga algoritma ini nilai keakuratannya sama-sama 100%. Tetapi setelah melakukan pengujian menggunakan *rapidminer* nilai akurasi tiap-tiap algoritma cukup terlihat berbeda dalam tiga kali pengujian dari masing-masing algoritma. Nilai akurasi *rapidminer* pada algoritma *C4.5* dengan tiga kali pengujian sama-sama menghasilkan 100% nilai akurasi dari 40:60% persentase pengujian pertama, persentase 70:30% untuk pengujian kedua, persentase 90:10% untuk pengujian ketiga. Dan nilai akurasi untuk algoritma *Naïve Bayes* pada pengujian pertama dengan persentase 40:60% bernilai 99,34%, dan untuk pengujian kedua 70:30% dan ketiga 90:10% nilai akurasinya didapatkan 100%. Algoritma *K-NN* pada pengujian pertama dengan persentase 40:60% di dapat nilai akurasi sebesar 97,37%, pengujian kedua dengan persentase 70:30% di dapatkan akurasi 98,68% dan untuk pengujian ketiga dengan persentase 90:10% dengan akurasi sebesar 100%. Jadi untuk pengujian ketiga algoritma pada *rapidminer* ini yang paling akurat adalah algoritma *C4.5* dengan tiga kali pengujian nilai akurasinya 100%.

B. Saran

Untuk kedepannya agar penelitian selanjutnya bisa memberikan hasil yang lebih memuaskan maka penulis memiliki saran sebagai berikut:

1. Peneliti selanjutnya menentukan tempat penelitian yang berbeda sehingga metode yang digunakan bisa memberikan hasil sesuai dengan apa yang diharapkan.
2. Jika penulis menggunakan topik yang sama maka lebih baik menggunakan metode yang berbeda agar dapat menghasilkan tingkat akurasi yang bisa dibandingkan dengan penelitian sebelumnya.

DAFTAR PUSTAKA

- Anam, C., ENERGY, H. S.-, & 2018, undefined. (2018). Perbandingan Kinerja Algoritma C4. 5 dan Naive Bayes untuk Klasifikasi Penerima Beasiswa. *Ejournal.Upm.Ac.Id*, 8(1), 2088–4591.
<https://ejournal.upm.ac.id/index.php/energy/article/view/111>
- Annur, H. (2018). Klasifikasi Masyarakat Miskin Menggunakan Metode Naive Bayes. *ILKOM Jurnal Ilmiah*, 10(2), 160–165.
<https://doi.org/10.33096/ilkom.v10i2.303.160-165>
- Budiman, I., & Ramadina, R. (2015). Penerapan Fungsi Data Mining Klasifikasi untuk Prediksi Masa Studi Mahasiswa Tepat Waktu pada Sistem Informasi Akademik Perguruan Tinggi. *Ijccs*, x, No.x(1), 1–5.
- Budiman, & Niqotaini, Z. (2021). *PERBANDINGAN ALGORITMA KLASIFIKASI DATA MINING UNTUK Abstrak*. 15(3), 1–5.
- Faid, M., Jasri, M., & Rahmawati, T. (2019). Perbandingan Kinerja Tool Data Mining Weka dan Rapidminer Dalam Algoritma Klasifikasi. *Teknika*, 8(1), 11–16. <https://doi.org/10.34148/teknika.v8i1.95>
- Fay, D. L. (2018). Klasifikasi Dalam Data Mining. *Angewandte Chemie International Edition*, 6(11), 951–952., 6–16.
- Hendrian, S. (2018). Algoritma Klasifikasi Data Mining Untuk Memprediksi Siswa Dalam Memperoleh Bantuan Dana Pendidikan. *Faktor Exacta*, 11(3), 266–274. <https://doi.org/10.30998/faktorexacta.v11i3.2777>
- Hendrickx, T., Cule, B., Meysman, P., Naulaerts, S., Laukens, K., & Goethals, B. (2015). Mining association rules in graphs based on frequent cohesive itemsets. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 9078(3), 637–648. https://doi.org/10.1007/978-3-319-18032-8_50
- Kadafi, A. R. (2018). Perbandingan Algoritma Klasifikasi Untuk Penjurusan Siswa SMA. *Jurnal ELTIKOM*, 2(2), 67–77.
<https://doi.org/10.31961/eltikom.v2i2.86>

- Mardi, Y. (2017). Data Mining : Klasifikasi Menggunakan Algoritma C4.5. *Edik Informatika*, 2(2), 213–219. <https://doi.org/10.22202/ei.2016.v2i2.1465>
- Nasution, D. A., Khotimah, H. H., & Chamidah, N. (2019). Perbandingan Normalisasi Data untuk Klasifikasi Wine Menggunakan Algoritma K-NN. *Computer Engineering, Science and System Journal*, 4(1), 78. <https://doi.org/10.24114/cess.v4i1.11458>
- Nasution, N., Djahara, K., & Zamsuri, A. (2015). Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naïve Bayes (Studi Kasus : Fasilkom Unilak). *Jurnal Fakultas Ilmu Komputer*, 1(1), 1–11. <https://repository.unilak.ac.id/305/1/91-Article-Text-163-1-10-20180211%281%29.pdf>
- Rohman, A., & Rochcham, M. (2019). Komparasi Metode Klasifikasi Data Mining Untuk Prediksi Kelulusan Mahasiswa. *Neo Teknika*, 5(1), 23–29. <https://doi.org/10.37760/neoteknika.v5i1.1379>
- Saputra, A. Y., & Primadasa, Y. (2018). Penerapan Teknik Klasifikasi Untuk Prediksi Kelulusan Mahasiswa Menggunakan Algoritma K-Nearest Neighbor. *Techno.Com*, 17(4), 395–403. <https://doi.org/10.33633/tc.v17i4.1864>
- Sari, Y. R., Sudewa, A., Lestari, D. A., & Jaya, T. I. (2020). Penerapan Algoritma K-Means Untuk Clustering Data Kemiskinan Provinsi Banten Menggunakan Rapidminer. *CESS (Journal of Computer Engineering, System and Science)*, 5(2), 192. <https://doi.org/10.24114/cess.v5i2.18519>
- Sibuea, M. L., & Safta, A. (2017). Pemetaan Siswa Berprestasi Menggunakan Metode K-Means Clustering. *Jurteksi*, 4(1), 85–92. <https://doi.org/10.33330/jurteksi.v4i1.28>
- Sugianto, C. A. (2015). Analisis Komparasi Algoritma Klasifikasi Untuk Menangani Data Tidak Seimbang Pada Data Kebakaran Hutan. *Techno.Com*, 14(4), 336–342. <http://publikasi.dinus.ac.id/index.php/technoc/article/view/992>