



**PERBANDINGAN KLASIFIKASI ALGORITMA C4.5, NAÏVE BAYES
DAN K-NEAREST NEIGHBOR UNTUK MENENTUKAN CALON
PENERIMA BEASISWA**

TUGAS AKHIR

ISMAWATI MEILANA SARI
41518010146

UNIVERSITAS
MERCU BUANA
PROGRAM STUDY TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS MERCU BUANA
JAKARTA
2022



**PERBANDINGAN KLASIFIKASI ALGORITMA C4.5, NAÏVE BAYES
DAN K-NEAREST NEIGHBOR UNTUK MENENTUKAN CALON
PENERIMA BEASISWA**

Tugas Akhir

Diajukan Untuk Melengkapi Salah Satu Syarat
Memperoleh Gelar Sarjana Komputer

Oleh:

ISMAWATI MEILANA SARI

41518010146

PROGRAM STUDY TEKNIK INFORMATIKA

FAKULTAS ILMU KOMPUTER

UNIVERSITAS MERCU BUANA

JAKARTA

2022

LEMBAR PERNYATAAN ORISINALITAS

LEMBAR PERNYATAAN ORISINALITAS

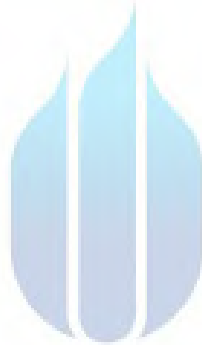
Yang bertanda tangan dibawah ini:

NIM : 41518010146

Nama : Ismawati Meilana Sari

Judul Tugas Akhir : Perbandingan Klasifikasi Algoritma C4.5, Naïve Bayes dan K-Nearest Neighbor untuk Menentukan Calon Penerima Beasiswa

Menyatakan bahwa Laporan Tugas Akhir saya adalah hasil karya sendiri dan bukan plagiat. Apabila ternyata ditemukan di dalam laporan Tugas Akhir saya terdapat unsur plagiat, maka saya siap untuk mendapatkan sanksi akademik yang terkait dengan hal tersebut.



UNIVERSITAS
MERCU BUANA

Jakarta, 01 Juli 2022



Ismawati Meilana Sari

SURAT PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR

SURAT PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR

Sebagai mahasiswa Universitas Mercu Buana, saya yang bertanda tangan di bawah ini :

Nama Mahasiswa : Ismawati Meilana Sari
NIM : 41518010146
Judul Tugas Akhir : Perbandingan Klasifikasi Algoritma C4.5, Naïve Bayes dan K-Nearest Neighbor untuk Menentukan Calon Penerima Beasiswa

Dengan ini memberikan izin dan menyetujui untuk memberikan kepada Universitas Mercu Buana **Hak Bebas Royalti Noneksklusif** (*None-exclusive Royalty Free Right*) atas karya ilmiah saya yang berjudul di atas beserta perangkat yang ada (jika diperlukan).

Dengan Hak Bebas Royalti/Noneksklusif ini Universitas Mercu Buana berhak menyimpan, mengalihmedia/formatkan, mengelola dalam bentuk pangkalan data (*database*), merawat dan mempublikasikan tugas akhir saya.

Selain itu, demi pengembangan ilmu pengetahuan di lingkungan Universitas Mercu Buana, saya memberikan izin kepada Peneliti di Lab Riset Fakultas Ilmu Komputer, Universitas Mercu Buana untuk menggunakan dan mengembangkan hasil riset yang ada dalam tugas akhir untuk kepentingan riset dan publikasi selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

UNIVERSITAS
MERCU BUANA

Jakarta, 01 Juli 2022



Ismawati Meilana Sari

SURAT PERNYATAAN LUARAN TUGAS AKHIR

SURAT PERNYATAAN LUARAN TUGAS AKHIR

Sebagai mahasiswa Universitas Mercu Buana, saya yang bertanda tangan di bawah ini :

Nama Mahasiswa : Ismawati Meilana Sari
NIM : 41518010146
Judul Tugas Akhir : Perbandingan Klasifikasi Algoritma C4.5, Naïve Bayes dan K-Nearest Neighbor untuk Menentukan Calon Penerima Beasiswa

Menyatakan bahwa :

1. Luaran Tugas Akhir saya adalah sebagai berikut :

No	Luaran	Jenis	Status
1	Publikasi Ilmiah	Jurnal Nasional Tidak Terakreditasi	Diajukan ✓
		Jurnal Nasional Terakreditasi ✓	
		Jurnal International Tidak Bereputasi	Diterima
		Jurnal International Bereputasi	
Disubmit/dipublikasikan di :	Nama Jurnal : Jurnal Informatika : Jurnal Pengembangan IT		
	ISSN : 2477-5126 (print), 2548-9356 (online)		
	Link Jurnal : https://ejournal.poltektegal.ac.id/index.php/informatika		
	Link File Jurnal Jika Sudah di Publish : -		

2. Bersedia untuk menyelesaikan seluruh proses publikasi artikel mulai dari submit, revisi artikel sampai dengan dinyatakan dapat diterbitkan pada jurnal yang dituju.
3. Diminta untuk melampirkan scan KTP dan Surat Pernyataan (Lihat Lampiran Dokumen HKI), untuk kepentingan pendaftaran HKI apabila diperlukan.

Demikian pernyataan ini saya buat dengan sebenarnya.

Mengetahui
Dosen Pembimbing TA

Dr. Harwikarya, MT

Jakarta, 01 Juli 2022

Ismawati Meilana Sari

LEMBAR PERSETUJUAN PENGUJI

NIM : 41518010146
Nama : Ismawati Meilana Sari
Judul Tugas Akhir : Perbandingan Klasifikasi Algoritma C4.5, Naïve Bayes dan K-Nearest Neighbor untuk Menentukan Calon Penerima Beasiswa

Tugas Akhir ini telah diperiksa dan disidangkan sebagai salah satu persyaratan untuk memperoleh gelar Sarjana pada Program Study Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana.

Jakarta, 28 Juli 2022



(Sabar Rudiarto, M.Kom)

UNIVERSITAS
MERCU BUANA

LEMBAR PERSETUJUAN PENGUJI

NIM : 41518010146
Nama : Ismawati Meilana Sari
Judul Tugas Akhir : Perbandingan Klasifikasi Algoritma C4.5, Naïve Bayes dan K-Nearest Neighbor untuk Menentukan Calon Penerima Beasiswa

Tugas Akhir ini telah diperiksa dan disidangkan sebagai salah satu persyaratan untuk memperoleh gelar Sarjana pada Program Study Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana.

Jakarta, 28 Juli 2022



(Afiyati, S.Si, MT)

UNIVERSITAS
MERCU BUANA

LEMBAR PERSETUJUAN PENGUJI

NIM : 41518010146
Nama : Ismawati Meilana Sari
Judul Tugas Akhir : Perbandingan Klasifikasi Algoritma C4.5, Naïve Bayes dan K-Nearest Neighbor untuk Menentukan Calon Penerima Beasiswa

Tugas Akhir ini telah diperiksa dan disidangkan sebagai salah satu persyaratan untuk memperoleh gelar Sarjana pada Program Study Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana.

Jakarta, 28 Juli 2022


(Wawan Sunawan, S.Kom, MT)

UNIVERSITAS
MERCU BUANA

LEMBAR PENGESAHAN

LEMBAR PENGESAHAN

NIM : 41518010146
Nama : Ismawati Meilana Sari
Judul Tugas Akhir : Perbandingan Klasifikasi Algoritma C4.5, Naïve Bayes dan K-Nearest Neighbor untuk Menentukan Calon Penerima Beasiswa

Tugas Akhir ini telah diperiksa dan disidangkan sebagai salah satu persyaratan untuk memperoleh gelar Sarjana pada Program Study Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana.

Jakarta, 28 Juli 2022

Menyetujui,



(Dr. Harwikarya, MT)
Dosen Pembimbing


Mengetahui,

UNIVERSITAS

MERCU BUANA



(Wawan Gunawan, S.Kom, MT)
Koord. Tugas Akhir Teknik Informatika



(Ir. Emil R. Kaburuan, Ph.D., IPM.)
Ka. Prodi Teknik Informatika

KATA PENGANTAR

Segala puji dan syukur kehadirat Allah SWT atas berkah, rahmat dan hidayah-Nya yang senantiasa dilimpahkan kepada penulis, sehingga bisa menyelesaikan skripsi dengan judul “*Perbandingan Klasifikasi Algoritma C4.5, Naïve Bayes dan K-Nearest Neighbor untuk Menentukan Calon Penerima Beasiswa*” sebagai syarat untuk menyelesaikan Program Sarjana (S1) pada Program Sarjana Fakultas Ilmu Komputer Jurusan Informatika Universitas Mercu Buana.

Dalam penyusunan skripsi ini banyak hambatan serta rintangan yang penulis hadapi namun pada akhirnya dapat melaluinya berkat adanya bimbingan dan bantuan dari berbagai pihak baik secara moral maupun spiritual. Untuk itu pada kesempatan ini penulis menyampaikan ucapan terimakasih kepada:

1. Kedua orang tua yang telah memberikan support serta kasih sayang hingga saya bisa menyelesaikan tugas akhir ini
2. Bapak Dr. Harwikarya, MT selaku dosen pembimbing saya yang telah membantu dan memberikan arahan kepada saya dalam menyelesaikan tugas akhir ini.
3. Bapak Emil R. Kaburuan Ph.D selaku Kepala Program Study Informatika yang telah memberikan arahan dan bimbingannya selama perkuliahan.
4. Bapak Wawan Gunawan S.Kom.,MT selaku Koordinator Tugas Akhir Prodi Informatika yang telah memberikan arahan selama periode tugas akhir
5. Ibu Saruni Dwiasnati, S.Kom, M.Kom selaku dosen pembimbing akademik saya yang telah membimbing selama perkuliahan hingga sampai pada tugas akhir.
6. Hilwa Lelisa yang telah membantu memberikan arahan dalam menyelesaikan tugas akhir sehingga dapat selesai dengan baik.
7. Teman-teman Informatika 2018, sahabat dan kerabat yang selalu

memberikan support dan waktu untuk selalu mengingatkan dalam menyelesaikan tugas akhir ini

Penulis menyadari bahwa penyusunan Jurnal Tugas Akhir ini masih jauh dari kesempurnaan dan masih terdapat kekurangan dan kelemahan, walaupun demikian, penulis telah berusaha semaksimal mungkin untuk mendapatkan hasil yang terbaik. Akhir kata, penulis berharap jika ada kritik atau saran apapun yang sifatnya membangun bagi saya dengan senang hati akan saya terima.

Jakarta, 01 Juli 2022
Penulis



DAFTAR ISI

HALAMAN SAMPUL.....	i
HALAMAN JUDUL	i
LEMBAR PERNYATAAN ORISINALITAS	ii
SURAT PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR... iii	
SURAT PERNYATAAN LUARAN TUGAS AKHIR.....	iv
LEMBAR PERSETUJUAN PENGUJI.....	v
LEMBAR PENGESAHAN	viii
ABSTRAK	ix
ABSTRACT	x
KATA PENGANTAR.....	xi
DAFTAR ISI.....	xiii
NASKAH JURNAL	1
KERTAS KERJA.....	10
BAB 1. LITERATURE REVIEW	11
BAB 2. SOURCE CODE.....	22
BAB 3. DATASET.....	33
BAB 4. TAHAPAN EKSPERIMEN	38
BAB 5. HASIL SEMUA EKSPERIMEN.....	45
DAFTAR PUSTAKA	61
LAMPIRAN DOKUMEN HAKI.....	63
LAMPIRAN KORESPONDENSI	65

NASKAH JURNAL

Perbandingan Klasifikasi Algoritma C4.5, Naïve Bayes dan K-Nearest Neighbor untuk Menentukan Calon Penerima Beasiswa

Ismawati Meilana Sari^{1*}, Dr. Harwikarya, MT²

^{1,2}Jurusan Teknik Informatika, Fakultas Komputer, Universitas Mercu Buana, Jakarta

^{1,2}Jl. Raya, Meruya Sel., Kec. Kembangan, Jakarta, Daerah Khusus Ibukota Jakarta, 11650, Indonesia

email: ¹41518010146@student.mercubuana.ac.id, ²harwikarya@mercubuana.ac.id

Abstract – The development of science and technology is inseparable from the demands of education. Scholarships are government programs that provide assistance in the form of tuition fees for students who have achievements in the academic field and also students with economic limitations to study at the tertiary level. Classification in data mining is a data learning method to predict the value of a group of attributes. In this case the author applies data mining to determine the classification of prospective scholarship recipients. In this study, the authors took a case study at STIT Muhammadiyah Bojonegoro because the university was still conducting manual selection and making decisions that were less accurate, causing scholarship acceptance to be inappropriate. From the results of testing on scholarship applicant data that has been carried out, it can be seen that the average value of t The purpose of this study is to compare the C4.5, Naïve Bayes and K-Nearest Neighbor algorithms so that they can provide classification results to determine whether students are accepted or rejected who have registered for the scholarship program. The accuracy of the C4.5 algorithm is the best, which is 72.01%, while the Naive Bayes algorithm has an average accuracy of 71.90% and an average value of 71.90%. The average accuracy of the K-NN algorithm is 70.40%. The most influential factors to determine scholarship acceptance are GPA, Number of Credits, Parents' Income, Dependents and Information about Poor.

Keywords : Classification, scholarship, C4.5, Naïve Bayes, K-NN

Abstrak – Perkembangan ilmu pengetahuan dan teknologi tidak terlepas dari tuntutan Pendidikan. Beasiswa merupakan program pemerintah yang memberikan bantuan berupa biaya pendidikan untuk para mahasiswa yang memiliki prestasi di bidang akademik dan juga para mahasiswa dengan keterbatasan ekonomi untuk menempuh pendidikan di jenjang perguruan tinggi. Klasifikasi dalam data mining merupakan metode pembelajaran data untuk memprediksi nilai dari sekelompok atribut. Dalam hal ini penulis menerapkan data mining untuk mengetahui klasifikasi calon penerima beasiswa. Pada penelitian ini penulis mengambil study kasus di STIT Muhammadiyah Bojonegoro dikarenakan perguruan tinggi tersebut masih melakukan penyeleksian secara manual dan penentuan keputusan yang kurang akurat sehingga menyebabkan penerimaan beasiswa menjadi tidak tepat sasaran. Tujuan dari penelitian ini adalah melakukan perbandingan algoritma C4.5, Naïve Bayes dan K-Nearest Neighbor sehingga dapat memberikan hasil klasifikasi untuk menentukan diterima dan ditolak mahasiswa yang telah mendaftar program beasiswa. Dari hasil pengujian pada data pelamar beasiswa yang telah dilakukan dapat dilihat bahwa nilai rata-rata akurasi algoritma C4.5 adalah yang paling baik yaitu 72,01%, sedangkan algoritma naïve bayes nilai rata-rata akurasi sebesar 71,90% dan nilai rata-rata akurasi algoritma K-NN sebesar 70,40%. Faktor yang paling berpengaruh untuk menentukan penerimaan beasiswa yaitu IPK, Jumlah SKS, Pendapatan Orang Tua, Tanggungan dan keterangan Tidak Mampu.

Kata Kunci : Klasifikasi, Beasiswa, C4.5, Naïve Bayes, K-NN

I. PENDAHULUAN

Perkembangan ilmu pengetahuan dan teknologi tidak terlepas dari tuntutan Pendidikan. Beberapa orang menganggap bahwa hal utama dalam mendukung kemajuan suatu negara yaitu dengan memberikan Pendidikan yang layak dan berkualitas bagi putra-putri bangsa. Dalam mewujudkan Pendidikan yang layak dan berkualitas menjadi tantangan tersendiri bagi pihak perguruan tinggi untuk meningkatkan mutu di bidang layanan pendidikan. Dibalik Pendidikan yang berkualitas dan bermutu tentunya tidak luput dari biaya yang terbilang cukup mahal, hal tersebut tentunya tidak dapat dijangkau oleh semua orang terutama kalangan menengah kebawah. Melihat hal tersebut maka pemerintah berinisiatif memberikan beasiswa bagi mereka yang berprestasi dan memiliki kendala dalam hal biaya untuk dapat meneruskan pendidikannya di perguruan tinggi.

Beasiswa merupakan program pemerintah yang memberikan bantuan berupa biaya pendidikan untuk para mahasiswa yang memiliki prestasi di bidang akademik dan juga para mahasiswa dengan keterbatasan ekonomi untuk menempuh pendidikan di jenjang perguruan tinggi. Beasiswa dapat dikatakan pembiayaan yang tidak bersumber dari pendanaan sendiri atau orang tua, akan tetapi diberikan oleh pemerintah, perusahaan swasta, kedutaan, Universitas, serta lembaga pendidik atau peneliti, atau juga dari kantor tempat bekerja yang karena prestasi seorang karyawan dapat diberikan kesempatan untuk meningkatkan kapasitas sumber daya manusianya melalui pendidikan. Biaya tersebut diberikan kepada yang berhak menerima sesuai dengan kriteria yang ditentukan [1]. Sebagai generasi penerus bangsa, banyak remaja terutama di kalangan mahasiswa berlomba-lomba untuk dapat menerima bantuan berupa beasiswa untuk dapat melanjutkan pendidikan di jenjang perguruan tinggi. Hal tersebut dapat dilihat dari prestasi-prestasi akademik yang mereka raih sebagai salah satu syarat untuk dapat menerima beasiswa.

Data mining telah banyak menarik perhatian di masyarakat dalam beberapa tahun ini,

karena mampu mengubah data yang luas dan jumlah yang besar menjadi informasi yang berguna dan pengetahuan [2]. Data mining memiliki beberapa algoritma untuk klasifikasi data, tiga diantaranya yaitu Algoritma C4.5, naïve bayes, dan K-Nearest Neighbor Algoritma C4.5 adalah metode pohon keputusan, yang didasarkan pada algoritma ID3 dan diusulkan oleh Quinlan pada 1993. Algoritma C4.5 ditingkatkan lebih dari algoritma ID3, dan telah digunakan dalam industri keuangan, medis dan lainnya secara luas [3]. Algoritma C4.5 banyak digunakan dengan menggunakan metode klasifikasi dan estimasi. Algoritma naïve bayes adalah salah satu algoritma yang digunakan dalam klasifikasi atau pengelompokan data yang didasarkan pada teorema bayes dengan pengklasifikasian yang dilakukan menggunakan training set beberapa data dengan efisien [4]. Algoritma K-Nearest Neighbors adalah pendekatan untuk mencari kasus dengan menghitung kedekatan antara kasus yang baru dengan kasus lama berdasarkan pencocokan bobot dari sejumlah fitur yang ada [5].

Seringnya terjadi kesalahan dalam pemberian beasiswa yang tidak sesuai target menjadi salah satu faktor peneliti melakukan penelitian ini, terlebih lagi di era digital saat ini yang semua proses penyeleksian sudah dilakukan menggunakan sistem. Penelitian yang dilakukan oleh Azahari pada tahun 2021 dengan judul “Rekomendasi Penerimaan Beasiswa Yayasan Untuk Siswa Baru SMK TI Airlangga dengan Algoritma C4.5” [6]. Pada penelitian tersebut peneliti hanya menggunakan beberapa atribut dan satu algoritma saja, menurut jurnal Rini pada tahun 2021 penggunaan atribut yang salah (tidak berpengaruh) dan algoritma klasifikasi yang tidak cocok dapat menyebabkan hasil akurasi menjadi rendah atau tidak bekerja dengan optimal [7]. Penggunaan algoritma yang bekerja kurang optimal ditemukan pada penelitian Ayudhitama pada tahun 2020 yaitu dengan menggunakan algoritma klasifikasi Neural Network dengan hasil akurasi yang cukup rendah yaitu 55,42% [8]. Hal ini dikarenakan algoritma neural network tidak efektif dalam operasi numerik dengan presisi tinggi, tidak efisien dalam operasi algoritma aritmatika serta logika dan simbolik dan waktu pelatihan akan sangat lama apabila jumlah datanya besar. Sama halnya dengan

Universitas Mercu Buana

penelitian yang dilakukan oleh sumiah pada tahun 2020 dengan judul “Perbandingan Metode K-Nearest Neighbor dan Naive Bayes untuk Rekomendasi Penentuan Mahasiswa Penerima Beasiswa pada Universitas Kuningan” [4]. Sayangnya pada penelitian tersebut hanya membandingkan 2 algoritma klasifikasi saja dan menyebabkan proses validasi menjadi kurang baik sama seperti penelitian Hendrian pada tahun 2018 yang melakukan validasi data dengan percobaan hanya satu kali saja [2].

Oleh karena itu peneliti akan mengimplementasikan 3 algoritma klasifikasi untuk menentukan algoritma yang paling cocok pada atribut penting (berpengaruh) yang digunakan dalam menentukan pemberian beasiswa. Berdasarkan uraian permasalahan dari penelitian-penelitian sebelumnya, maka penulis tertarik untuk menggunakan algoritma C4.5, Naïve Bayes dan K-Nearest Neighbor untuk diimplementasikan dalam pengklasifikasian data calon penerima beasiswa. Penelitian ini juga akan menentukan algoritma klasifikasi mana yang cocok dalam menentukan penerima beasiswa. Metode ini dipilih karena ketiga algoritma merupakan algoritma yang populer digunakan dalam proses pengklasifikasian data sehingga penulis tertarik untuk melihat tingkat akurasi data dari ketiga algoritma terhadap data calon penerima beasiswa. Metode tersebut memiliki kelebihan yaitu dapat mencari informasi tersembunyi di dalam sekumpulan data, membagi data tersebut ke dalam bagian yang lebih kecil dan hasil analisa yang mudah dimengerti. Dengan membandingkan algoritma C4.5, Naïve Bayes dan K-Nearest Neighbor diharapkan dapat memberikan hasil berupa rekomendasi untuk perguruan tinggi terhadap suatu proses prediksi calon penerima beasiswa dengan peminat yang sangat banyak. Selain itu peneliti juga akan melakukan validasi data menggunakan metode split validation dan menambahkan metode cross validation dengan masing-masing dilakukan 5 kali percobaan guna mendapatkan akurasi yang paling baik.

*) **penulis korespondensi:** Ismawati Meilana Sari
Email: 41518010146@student.mercubuana.ac.id

II. PENELITIAN YANG TERKAIT

Beberapa penelitian telah menunjukkan bahwa beasiswa tidak hanya diperuntukkan bagi mahasiswa berprestasi saja, namun beasiswa juga diberikan kepada mahasiswa kurang mampu untuk dapat membantu mereka melanjutkan pendidikannya. Penelitian pada tahun 2020 yang dilakukan untuk menentukan penerimaan beasiswa dengan melakukan pengujian data mahasiswa sebanyak 125 data dari tahun 2019. Hasil yang dicapai menggunakan algoritma C4.5 diketahui tingkat akurasi mencapai 92%, spesifisitas 92.3%, dan sensitifitas 91.6%. Akan tetapi data yang digunakan pada penelitian tersebut masih tergolong sedikit sehingga tingkat akurasi yang dihasilkan belum maksimal [6]. Penelitian lain menggunakan data pelamar beasiswa sebanyak 460 data yang terdiri dari 350 data dengan kelas “YA” dan 110 data yang memiliki kelas “TIDAK”. Dari hasil pengujian pada algoritma C4.5 didapatkan nilai akurasi sebesar 85,55% dengan menggunakan Teknik post pruning sebagai acuan klasifikasi pendaftar baru [7]. Penelitian lain juga dilakukan untuk memprediksi penerimaan Bantuan Pangan Non Tunai (BPNT) dengan menggunakan metode naïve bayes. Pada penelitian ini dibagi menjadi 2 class yaitu Layak dan Tidak Layak. Dari pengujian menggunakan sebanyak 350 data pada tahun 2019 dengan metode naïve bayes didapatkan tingkat akurasi sebesar 96% [8]. Masalah yang sama dalam pemberian beasiswa juga dipaparkan oleh Ramakrishna Hegde, dkk (2021) dalam judul penelitiannya yaitu “*A Review on Data Mining and Machine Learning Methods for Student Scholarship Prediction*”. Penelitian ini melakukan perbandingan antara metode Naïve Bayes, Decision Tree dan K-Nearest Neighbor (K-NN) untuk rekomendasi penentuan mahasiswa penerima beasiswa. Hasil dari penelitian dari data penerima beasiswa sebanyak 8.218 data tersebut membuktikan bahwa tingkat akurasi dengan menggunakan metode decision tree lebih baik dibandingkan dengan menggunakan metode naïve bayes dan K-NN yaitu sebesar 82,18 [9]. Penelitian tentang klasifikasi beasiswa juga dipaparkan oleh Kurniadi dkk. Dari hasil penelitian dengan menggunakan algoritma K-NN

Universitas Mercu Buana

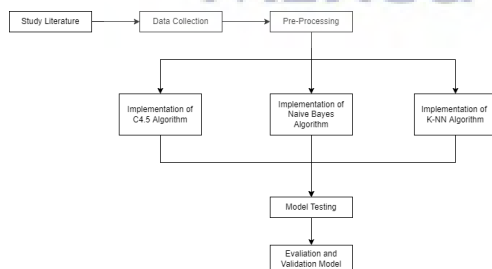
menunjukkan bahwa akurasi yang didapat cukup baik yaitu sebesar 95,83%. Data yang digunakan pada penelitian ini sebanyak 1018 siswa dengan atribut yang digunakan yaitu semester, pendapatan orang tua, jumlah tanggungan keluarga dan IPK [10].

III. METODE PENELITIAN

Penelitian ini merupakan penelitian kuantitatif karena penulis akan melakukan penelitian eksperimen dengan menggunakan pengujian observasi antar variabel dan pada penelitian ini akan dilakukan analisis statistik serta klasifikasi beasiswa dengan memperhatikan tingkat akurasi agar mendapatkan hasil klasifikasi dengan akurasi terbaik untuk mengetahui pola dan faktor pemberian beasiswa di perguruan tinggi. Klasifikasi dalam data mining merupakan metode pembelajaran data untuk memprediksi nilai dari sekelompok atribut [11]. Adapun tahapan dalam penelitian ini adalah sebagai berikut.

A. Alur penelitian dengan diagram alir

Alur pada penelitian ini diawali dengan study literature kemudian dilanjutkan dengan pengumpulan data, setelah mendapatkan data lalu dilakukan pre-processing, lalu implementasi algoritma (C4.5, Naïve Bayes dan K-NN), model testing dan yang terakhir evaluasi dan validasi model. Adapun diagram alir penelitian ini dapat dilihat pada gambar 1.



Gbr. 1 Diagram Alir Alur Penelitian

B. Rincian Alur Penelitian

Alur penelitian ini dapat dilihat pada gambar. Berikut adalah penjelasan dari setiap tahapannya sebagai berikut:

1. Study Literature

Pada tahap ini penulis melakukan study literature dengan mereview penelitian terdahulu untuk mendapatkan referensi yang berkaitan dengan penelitian yang akan

dilakukan. Study literature yang dilakukan pada penelitian ini yaitu mempelajari proses pengumpulan dan penarikan data, konsep data mining, implementasi algoritma data mining dan konsep pemberian beasiswa pada perguruan tinggi. Dari hasil study literature yang telah dilakukan maka penulis memutuskan untuk menggunakan algoritma C4.5, Naïve Bayes dan K-NN untuk digunakan dalam pengklasifikasian pada penelitian ini. Selain itu pada penelitian ini juga menggunakan cross validation untuk mendapatkan hasil akurasi yang akurat.

2. Data Collection

Berdasarkan study literature, data yang diperlukan dalam penelitian ini yaitu berupa data bantuan berupa beasiswa yang ditujukan untuk mahasiswa berprestasi dan kurang mampu di perguruan tinggi. Data yang digunakan pada penelitian ini didapatkan dari STIT Muhammadiyah Bojonegoro. Pada penelitian ini menggunakan data yang diambil dari tahun 2019-2022 dengan total data sebanyak 1.519 data.

3. Preprocessing

Pada tahap Preprocessing dilakukan pembersihan data yang terdiri dari :

a. Data cleansing : Penelitian ini menggunakan data sebanyak 1.519 data. Proses cleansing data dilakukan secara manual. Pada proses cleansing data yang dilakukan yaitu menghapus atau membersihkan seluruh data pada baris yang memiliki kolom kosong atau data kosong. Dari proses cleansing yang telah dilakukan maka didapatkan data sebanyak 1.316 data.

b. Data reduction : Dari data yang telah didapatkan memiliki 11 variabel, tetapi pada penelitian ini penulis tidak menggunakan seluruh variabel yang diberikan. Untuk menentukan variabel apa saja yang akan digunakan maka penulis melakukan wawancara kepada pihak STIT Muhammadiyah Bojonegoro dan melakukan riset kriteria penerima beasiswa. Oleh karena itu, didapatkan 6 variabel yang akan digunakan dalam implementasi algoritma diantaranya yaitu IPK, Jumlah SKS, Pendapatan Ortu, Tanggungan, Rekomendasi PCM dan Keterangan Tidak Mampu.

c. Data Transformasi : Pada tahap ini penulis melakukan transformasi data pada 4 variabel dengan mengkategorikan

Universitas Mercu Buana

masing-masing variabel ke dalam kelompok yang telah ditentukan untuk mempermudah proses pengolahan data. Variabel data yang dilakukan transformasi diantaranya yaitu merubah data IPK, Jumlah SKS, Pendapatan Ortu dan Tanggungan. Setelah itu seluruh variabel data diubah menjadi data integer agar dapat dibaca pada saat proses preprocessing. Hal ini dilakukan karena jika data yang digunakan berbentuk string/varchar maka tidak dapat masuk ke proses penerapan aritmatika algoritma.

4. Implementasi Algoritma C4.5, Naïve Bayes dan K-NN

Setelah preprocessing data dan mendapatkan data bersih maka tahap selanjutnya yaitu implementasi algoritma yang terdiri dari algoritma C4.5, Naïve Bayes dan K-NN. Dalam pengimplementasian algoritma ditentukan nilai x dan y yang merupakan variabel independent dan variabel dependent. Variabel yang digunakan untuk nilai x adalah IPK, Jumlah SKS, Pendapatan Ortu, Tanggungan dan Keterangan Tidak Mampu. Sedangkan untuk nilai y menggunakan variabel Rekomendasi PCM yang merupakan label. Adapun penjelasan dari masing-masing algoritma yang digunakan adalah sebagai berikut:

a. Algoritma C4.5

Algoritma C4.5 adalah algoritma yang sangat populer yang digunakan dalam menyusun pohon keputusan (decision tree). Algoritma C4.5 banyak digunakan dengan menggunakan metode klasifikasi dan estimasi. Algoritma C4.5 memiliki manfaat dalam mengeksplorasi data, menentukan hubungan tersembunyi pada sejumlah calon variabel input dengan variabel target. Algoritma C4.5 adalah metode pohon keputusan, yang didasarkan pada algoritma ID3 dan diusulkan oleh Quinlan pada 1993. Algoritma C4.5 ditingkatkan lebih dari algoritma ID3, dan telah digunakan dalam industri keuangan, medis dan lainnya secara luas. Algoritma C4.5 diimplementasikan dalam perolehan informasi cara membuat klasifikasi [3]. Pernyataan tentang klasifikasi algoritma C4.5 (decision tree) yang terdapat pada cabang dan kelasnya atau segmen yang terdapat pada daun. Akar akan diambil

dari atribut yang dipilih, dengan menghitung nilai gain dari setiap atribut, lalu nilai gain yang paling tinggi yang akan menjadi akar pertama. Sebelum menghitung gain nilai atribut, terlebih dahulu menghitung nilai entropi.

b. Algoritma Naïve Bayes

Algoritma Naïve Bayes adalah salah satu algoritma yang digunakan dalam klasifikasi atau pengelompokan data yang didasarkan pada teorema bayes dengan pengklasifikasian yang dilakukan menggunakan training set beberapa data dengan efisien [4]. Naive Bayes adalah pengklasifikasi probabilistik sederhana dengan perhitungan rentang probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari kumpulan data yang diberikan. Naïve bayes berasumsi bahwa nilai dari suatu inputan atribut di dalam kelas yang diberikan tidak tergantung pada nilai atribut lainnya. Hasil akurasi dan kecepatan pada algoritma naïve bayes tergolong tinggi Ketika diaplikasikan pada database dengan jumlah yang cukup besar.

c. Algoritma K-Nearest Neighbor (KNN)

Algoritma K-Nearest Neighbors adalah pendekatan untuk mencari kasus dengan menghitung kedekatan antara kasus yang baru dengan kasus lama berdasarkan pencocokan bobot dari sejumlah fitur yang ada [5]. Pendekatan nearest neighbor adalah pendekatan pencarian kasus, yang menghitung kedekatan antar kasus baru dan lama yang didasarkan pada pembobotan beberapa fitur yang ada. Algoritma klasifikasi K-Nearest Neighbor dilatih untuk mengidentifikasi pola klik dari pengguna, mengelompokkannya ke dalam kelompok pengguna, dan membuat rekomendasi untuk melihat informasi berdasarkan karakteristik pengguna.

5. Model Testing

Setelah diimplementasikan pada ke-3 algoritma maka tahap selanjutnya adalah model testing. Data akan diuji pada model yang telah di proses melalui 2 metode yaitu split validation dan cross validation.

a. Split validation

Split validation adalah metode validasi yang memiliki 2 proses dalam mengolah data yang diantaranya adalah proses pelatihan dan proses pengujian data.

Universitas Mercu Buana

Proses pelatihan data digunakan untuk mempelajari dan membangun model. Setelah melakukan proses pelatihan maka model tersebut kemudian diterapkan dalam proses pengujian. Kinerja model juga diukur selama fase pengujian. Pada penelitian ini persentase validasi dibagi menjadi 5 skenario eksperimen dengan pembagian data sebagai berikut:

TABEL I
PERSENTASE SPLIT VALIDATION

Experiment ke-	Data Training	Data Testing
1	90%	10%
2	80%	20%
3	70%	30%
4	60%	40%
5	50%	50%

b. Cross Validation

Cross validation atau bisa disebut estimasi rotasi merupakan metode validasi model untuk mengetahui nilai dari hasil statistic analisis dalam perataan kumpulan data independent. Pada umumnya Teknik ini digunakan dalam melakukan model prediksi dan digunakan untuk menguji keakuratan sebuah model prediktif saat diterapkan pada pengujiannya. Salah satu metode dari cross validation yaitu k-fold cross-validation, yang digunakan untuk membagi data kedalam K bagian set data dengan porsi yang sama. Pada penelitian ini dilakukan 5 kali eksperimen menggunakan metode cross validation dengan nilai k pada k-fold sebagai berikut :

TABEL II
NILAI K-FOLD YANG DIGUNAKAN

Experiment ke-	K-fold
1	5-fold
2	10-fold
3	15-fold
4	20-fold
5	25-fold

6. Evaluation and Validation Model

Tahap terakhir setelah penerapan metode klasifikasi yaitu menghitung performa masing-masing algoritma yang terdiri dari evaluasi dan validasi model. Evaluasi dan validasi model digunakan untuk menganalisa keakuratan model yang telah diterapkan dan untuk melihat hasilnya. Pada penelitian ini

menggunakan confusion matrix untuk membantu menghitung nilai akurasi, presisi dan recall.

a. Akurasi

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

b. Presisi

$$precision = \frac{TP}{TP + FP}$$

c. Recall

$$recall = \frac{TP}{TP + FN}$$

IV. HASIL DAN PEMBAHASAN

Pada penelitian ini menggunakan data yang didapat dari salah satu instansi Pendidikan yang ada di Jawa Timur yaitu STIT Muhammadiyah Bojonegoro. Data yang digunakan adalah data pemohon beasiswa pada periode 2019-2022. Jumlah data yang digunakan pada uji coba penelitian ini sebanyak 1.519 record dengan 11 atribut yang dimana data tersebut sesuai dengan jumlah pemohon beasiswa pada 3 tahun terakhir. Adapun atribut yang terdapat dalam data disajikan dalam tabel berikut ini.

TABEL III
ATRIBUT DATA BEASISWA

No.	Description	Data Type	Data
1	Nama	Varchar	ABC
2	Jenis Kelamin	Int	0/1
3	IPK	Int	4.00
4	Jumlah SKS	Int	22
5	Status	Varchar	Aktif
6	Pendapatan Orang Tua	Int	2.000.000
7	Pekerjaan Orang Tua	Varchar	Petani
8	Tanggungan	Int	2
9	Rekomendasi PCM	Varchar	Ya
10	Keterangan Tidak Mampu	Varchar	Ya
11	Keterangan Terdampak Covid-19	Varchar	Ya

Pada bagian ini akan menampilkan hasil semua eksperimen yang telah dilakukan.

a. Algoritma C4.5

1. Pohon Keputusan



Gbr. 2 Pohon Keputusan

2. Hasil Accuracy, Precision and Recall menggunakan split validation

TABEL IV
HASIL ALGORITMA C4.5 SPLIT VALIDATION

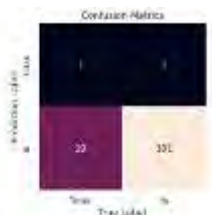
Experiment	Accuracy	Precision	Recall
1	77,27%	50%	3%
2	70,45%	50%	4%
3	70,63%	28%	5%
4	70,01%	37%	7%
5	71,73%	50%	1%

3. Hasil Accuracy, Precision and Recall menggunakan cross validation

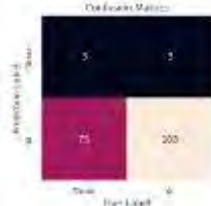
TABEL V
HASIL ALGORITMA C4.5 CROSS VALIDATION

Experiment	Accuracy	Precision	Recall
5-fold	69,75%	71%	100%
10-fold	69,00%	78%	99%
15-fold	67,85%	73%	95%
20-fold	66,67%	72%	95%
25-fold	69,30%	71%	98%

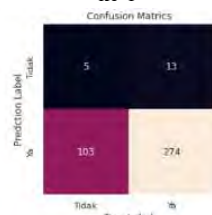
4. Confusion Matrix ke-5 Percobaan



Gbr. 3 Matrix Percobaan ke-1



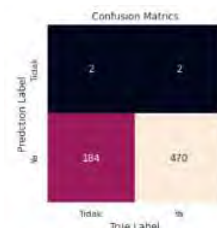
Gbr. 4 Matrix Percobaan ke-2



Gbr. 5 Matrix Percobaan ke-3



Gbr. 6 Matrix Percobaan ke-4



Gbr. 7 Matrix Percobaan ke-5

b. Algoritma Naïve Bayes

1. Hasil Accuracy, Precision and Recall menggunakan split validation

TABEL VI
HASIL ALGORITMA NAÏVE BAYES SPLIT VALIDATION

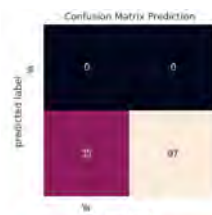
Experiment	Accuracy	Precision	Recall
1	73,48%	0%	0%
2	74,62%	0%	0%
3	70,12%	0%	0%
4	71,53%	0%	0%
5	69,75%	25%	1%

2. Hasil Accuracy, Precision and Recall menggunakan cross

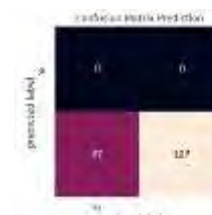
TABEL VII
HASIL ALGORITMA NAÏVE BAYES CROSS VALIDATION

Experiment	Accuracy	Precision	Recall
5-fold	70,36%	70%	99%
10-fold	68,91%	75%	100%
15-fold	69,92%	70%	99%
20-fold	68,69%	72%	99%
25-fold	69,68%	73%	100%

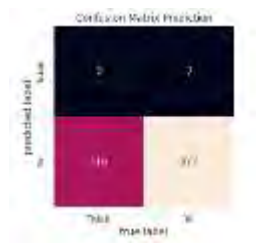
3. Confusion Matrix ke-5 Percobaan



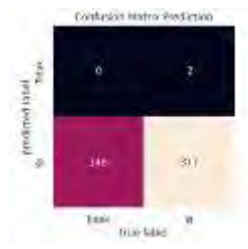
Gbr. 8 Matrix Percobaan ke-1



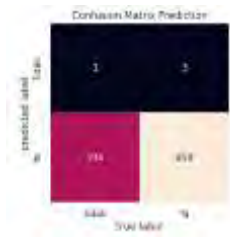
Gbr. 9 Matrix Percobaan ke-2



Gbr. 10 Matrix Percobaan ke-3



Gbr. 11 Matrix Percobaan ke-4



Gbr. 12 Matrix Percobaan ke-5

c. Algoritma K-NN

1. Hasil Accuracy, Precision and Recall menggunakan split validation.

TABEL VIII
HASIL ALGORITMA K-NN SPLIT
VALIDATION

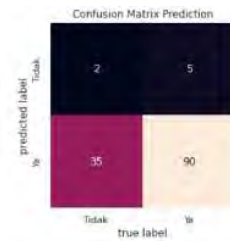
Experiment	The Best K	Accuracy	Precision	Recall
1	23	69,69%	29%	5%
2	19	70,45%	50%	3%
3	9	70,63%	33%	3%
4	15	70,58%	27%	3%
5	21	70,66%	33%	2%

2. Hasil Accuracy, Precision and Recall menggunakan cross validation

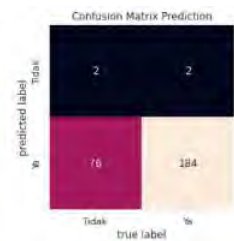
TABEL IX
HASIL ALGORITMA K-NN CROSS
VALIDATION

Experiment	The Best K	Accuracy	Precision	Recall
5-fold	19	68,44%	71%	99%
10-fold	23	69,25%	72%	95%
15-fold	9	65,68%	72%	98%
20-fold	15	68,32%	72%	97%
25-fold	21	68,70%	71%	98%

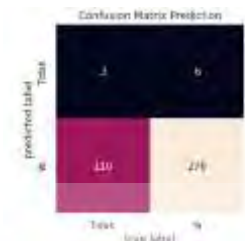
3. Confusion Matrix ke-5 Percobaan



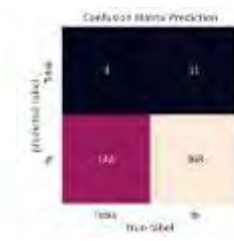
Gbr. 13 Matrix Percobaan ke-1



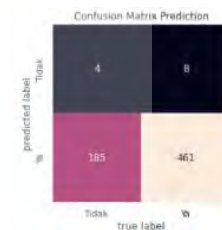
Gbr. 14 Matrix Percobaan ke-2



Gbr. 15 Matrix Percobaan ke-3

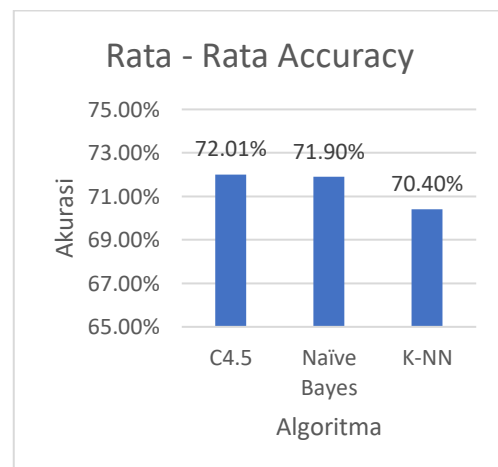


Gbr. 16 Matrix Percobaan ke-4



Gbr. 17 Matrix Percobaan ke-5

Rata-rata akurasi dari setiap percobaan pada masing-masing algoritma ditunjukkan pada diagram di bawah ini.



Gbr. 18 Rata-rata Akurasi Algoritma

Hasil penelitian pengujian klasifikasi pada data penerima beasiswa dapat dilihat bahwa nilai rata-rata akurasi algoritma C4.5 paling tinggi yaitu 72,01%, sedangkan algoritma naïve bayes nilai rata-rata akurasi sebesar 71,90% dan nilai rata-rata akurasi algoritma K-NN sebesar 70,40%.

KESIMPULAN

Dari pengukuran kinerja masing-masing algoritma yang telah dilakukan berdasarkan jumlah data dapat disimpulkan bahwa algoritma C4.5 memiliki kemampuan dalam pengambilan keputusan untuk menentukan penerimaan beasiswa pada perguruan tinggi. Perhitungan nilai gain dan ratio berpengaruh pada algoritma C4.5 sehingga algoritma ini memiliki model yang cukup baik untuk klasifikasi pemberian beasiswa.

Alur perhitungan yang tidak panjang membuat algoritma naïve bayes lebih mudah digunakan untuk pengklasifikasian. Namun, algoritma naïve bayes akan lebih cocok untuk digunakan pada data yang bersifat independent, sedangkan untuk data yang bersifat dependent algoritma ini akan memberikan hasil yang kurang maksimal.

Algoritma K-NN memiliki keterkaitan data satu sama lain yang membuat performanya meningkat. Selain itu algoritma K-NN juga sangat sederhana dan mudah dipelajari. Namun, nilai K yang digunakan terkadang memberikan kesalahan yang konsisten dalam memperkirakan sebuah nilai karena mudah tertipu dengan atribut yang tidak relevan.

Jumlah data training berpengaruh terhadap tingkat akurasi data. Faktor yang paling berpengaruh untuk menentukan penerimaan beasiswa yaitu IPK, Jumlah SKS, Pendapatan Orang Tua, Tanggungan dan keterangan Tidak Mampu.

UCAPAN TERIMA KASIH

Ucapan terima kasih penulis kepada pihak yang membantu ataupun memberikan dukungan terkait dengan penelitian yang dilakukan seperti bantuan fasilitas penelitian, dana hibah, dan lainnya.

DAFTAR PUSTAKA

- [1] D. Arbian, "Sistem Pendukung Keputusan (SPK) Pemberian Beasiswa Berbasis TOPSIS (Study Kasus Yayasan Pendidikan Al-Hikmah Bululawang Malang)," *J. Ilm. Teknol. Inf.*

- Asia*, vol. 11, no. 1, p. 29, 2017, doi: 10.32815/jitika.v11i1.40.
- [2] S. Hendrian, "Algoritma Klasifikasi Data Mining Untuk Memprediksi Siswa Dalam Memperoleh Bantuan Dana Pendidikan," *Fakt. Exacta*, vol. 11, no. 3, pp. 266–274, 2018, doi: 10.30998/faktorexacta.v11i3.2777.
- [3] J. Chen and D. Zhang, "Comparison analysis for classification algorithm in data mining and the study of model use," *AIP Conf. Proc.*, vol. 1955, 2018, doi: 10.1063/1.5033837.
- [4] A. Sumiah and N. Mirantika, "Perbandingan Metode K-Nearest Neighbor dan Naive Bayes untuk Rekomendasi Penentuan Mahasiswa Penerima Beasiswa pada Universitas Kuningan," *Buffer Inform.*, vol. 6, no. 1, pp. 1–10, 2020.
- [5] M. Windarti and A. Suradi, "Perbandingan Kinerja 6 Algoritme Klasifikasi Data Mining untuk Prediksi Masa Study Mahasiswa," *Telematika*, vol. 12, no. 1, p. 14, 2019, doi: 10.35671/telematika.v12i1.778.
- [6] A. S. Suweleh, D. Susilowati, and U. Bumigora, "Aplikasi Penentuan Penerima Beasiswa Menggunakan Algoritma C4 . 5 Jurnal BITe: Jurnal Bumigora Information Technology Jurnal BITe: Jurnal Bumigora Information Technology," *J.*, vol. 2, no. 1, pp. 12–21, 2020, doi: 10.30812/bite.v2i1.798.
- [7] Y. Kustiyahningsih, B. K. Khotimah, D. R. Anamisa, M. Yusuf, T. Rahayu, and J. Purnama, "Decision Tree C 4.5 Algorithm for Classification of Poor Family Scholarship Recipients," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 1125, no. 1, p. 012048, 2021, doi: 10.1088/1757-899x/1125/1/012048.
- [8] B. Hermanto and A. Jaelani, "Penerapan Data Mining Untuk Prediksi Penerima Bantuan Pangan Non Tunai (Bpnt) Di Desa Wanacala Menggunakan Metode Naïve Bayes," *SIGMA - J. Teknol. Pelita Bangsa*, vol. 18, no. 4, pp. 64–72, 2019.
- [9] R. Hegde, V. Anusha G, S. Madival, S. Sowjanya H, and U. Sushma, "A Review on Data Mining and Machine Learning Methods for Student Scholarship Prediction," *Proc. - 5th Int. Conf. Comput. Methodol. Commun. ICCMC 2021*, no. Iccmc, pp. 923–927, 2021, doi: 10.1109/ICCMC51019.2021.9418376.
- [10] D. Kurniadi, E. Abdurachman, H. L. H. S. Warnars, and W. Suparta, "The prediction of scholarship recipients in higher education using k-Nearest neighbor algorithm," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 434, no. 1, 2018, doi: 10.1088/1757-899X/434/1/012039.
- [11] S. Wahyuningsih and D. R. Utari, "Perbandingan Metode K-Nearest Neighbor , Naive Bayes dan Decision Tree untuk Prediksi Kelayakan Pemberian Kredit," *Konf. Nas. Sist. Inf. 2018 STMIK Atma Luhur Pangkalpinang*, 8 – 9 Maret 2018, pp. 619–623, 2018.

KERTAS KERJA

Ringkasan

Kertas kerja ini merupakan material kelengkapan artikel jurnal dengan judul “Perbandingan Klasifikasi Algoritma C4.5, Naïve Bayes dan K-Nearest Neighbor untuk Menentukan Calon Penerima Beasiswa”. Kertas kerja ini berisi semua material hasil penelitian Tugas Akhir. Di dalam kertas kerja ini disajikan beberapa bagian yang terdiri dari literaturee review, dataset yang digunakan, tahapan eksperimen dan hasil eksperimen secara keseluruhan.

Bagian I membahas mengenai literaturee review yang berisi artikel jurnal yang merupakan dasar atau landasan teori pendukung dalam penelitian ini. Bagian II berisikan mengenai source code yang digunakan pada penelitian ini. Bagian III menjelaskan mengenai dataset yang digunakan, meliputi penjelasan, cara pengolahan data, variabel data dan pemrosesan data sehingga data sesuai dengan kebutuhan penelitian serta siap untuk digunakan. Bagian IV memuat tahapan eksperimen yang disajikan dalam gambar diagram alir beserta penjelasan dari setiap tahapan. Bagian V merupakan bagian akhir dari kertas kerja yang menjelaskan hasil keseluruhan eksperimen yang telah dilakukan, meliputi hasil eksperimen, visualisasi serta penjelasannya.