



KOMPARASI ALGORITMA *K-NEAREST NEIGHBOR* DAN *NAÏVE BAYES* DALAM PENERIMAAN KARTU INDONESIA PINTAR DI DKI JAKARTA MENGGUNAKAN *K-FOLD CROSS VALIDATION*

TUGAS AKHIR



**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS MERCU BUANA
JAKARTA
2022**



KOMPARASI ALGORITMA *K-NEAREST NEIGHBOR* DAN *NAÏVE BAYES* DALAM PENERIMAAN KARTU INDONESIA PINTAR DI DKI JAKARTA MENGGUNAKAN *K-FOLD CROSS VALIDATION*

Tugas Akhir

Diajukan Untuk Melengkapi Salah Satu Syarat
Memperoleh Gelar Sarjana Komputer

Oleh:

Muhammad Yusuf Rizqon Rangkuti
41518010138

UNIVERSITAS
MERCU BUANA

PROGRAM STUDI TEKNIK INFORMATIKA

FAKULTAS ILMU KOMPUTER

UNIVERSITAS MERCU BUANA

JAKARTA

2022

LEMBAR PERNYATAAN ORISINALITAS

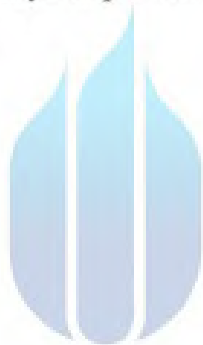
Yang bertanda tangan dibawah ini:

NIM : 41518010138

Nama : Muhammad Yusuf Rizqon Rangkuti

Judul Tugas Akhir : Komparasi Algoritma K-Nearest Neighbor dan Naïve Bayes
Dalam Penerimaan Kartu Indonesia Pintar di DKI Jakarta
Menggunakan K-Fold Cross Validation

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



Jakarta, 02 Juni 2022



UNIVERSITAS
MERCU BUANA

Muhammad Yusuf Rizqon Rangkuti

SURAT PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR

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

Nama Mahasiswa : 41518010138
NIM : Muhammad Yusuf Rizqon Rangkuti
Judul Tugas Akhir : Komparasi Algoritma K-Nearest Neighbor dan Naïve Bayes Dalam Penerimaan Kartu Indonesia Pintar di DKI Jakarta Menggunakan K-Fold Cross Validation

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 diatas 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.

Jakarta, 02 Juni 2022



Muhammad Yusuf Rizqon Rangkuti

SURAT PERNYATAAN LUARAN TUGAS AKHIR

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

Nama Mahasiswa : Muhammad Yusuf Rizqon Rangkuti
 NIM : 41518010138
 Judul Tugas Akhir : Komparasi Algoritma K-Nearest Neighbor dan Naïve Bayes Dalam Penerimaan Kartu Indonesia Pintar di DKI Jakarta Menggunakan K-Fold Cross Validation

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	:JUNAL RESTI (Rekayasa Sistem dan Teknologi Informasi)			
	ISSN	:2580-0760			
	Link Jurnal	:https://jurnal.iaii.or.id/index.php/RESTI/submissions			
	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.

Jakarta, 02 Juni 2022



Muhammad Yusuf Rizqon Rangkuti

LEMBAR PERSETUJUAN PENGUJI

NIM : 41518010138

Nama : Muhammad Yusuf Rizqon Rangkuti

Judul Tugas Akhir : Komparasi Algoritma *K-Nearest Neighbor* dan *Naïve Bayes* Dalam Penerimaan Kartu Indonesia Pintar di DKI Jakarta Menggunakan *K-Fold Cross Validation*

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

Jakarta, 7 Juli 2022



UNIVERSITAS
MERCU BUANA
Sabar Rudiarto
(Sabar Rudiarto, M.Kom)

LEMBAR PERSETUJUAN PENGUJI

LEMBAR PERSETUJUAN PENGUJI

NIM : 41518010138
Nama : Muhammad Yusuf Rizqon Rangkuti
Judul Tugas Akhir : Komparasi Algoritma *K-Nearest Neighbor* dan *Naïve Bayes* Dalam Penerimaan Kartu Indonesia Pintar di DKI Jakarta Menggunakan *K-Fold Cross Validation*

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

Jakarta, 7 Juli 2022



(Ir. Emil R. Kaburuan, Ph.D., IPM.)

UNIVERSITAS
MERCU BUANA

LEMBAR PERSETUJUAN PENGUJI

NIM : 41518010138

Nama : Muhammad Yusuf Rizqon Rangkuti

Judul Tugas Akhir : Komparasi Algoritma *K-Nearest Neighbor* dan *Naïve Bayes* Dalam Penerimaan Kartu Indonesia Pintar di DKI Jakarta Menggunakan *K-Fold Cross Validation*

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

Jakarta, 7 Juli 2022



(Dhanny Permatasari Putri, S.kom, MT)

LEMBAR PENGESAHAN


LEMBAR PENGESAHAN

NIM : 41518010138
Nama : Muhammad Yusuf Rizqon Rangkuti
Judul Tugas Akhir : *Komparasi Algoritma K-Nearest Neighbor dan Naïve Bayes Dalam Penerimaan Kartu Indonesia Pintar di DKI Jakarta Menggunakan K-Fold Cross Validation*

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

Jakarta, 7 Juli 2022

Menyetujui,


Rushezka S.Kom, M.TS
Dosen Pembimbing

UNIVERSITAS
MERCU BUANA

Mengetahui,


(Wawan Gunawan, S.Kom, MT)

Koord. Tugas Akhir Teknik Informatika


(Ir. Emil R. Kaburuan, Ph.D., IPM.)

Ka. Prodi Teknik Informatika

KATA PENGANTAR

Puji dan syukur penulis haturkan kehadiran Allah SWT, karena berkat rahmat dan hidayah-Nya-lah saya dapat menyelesaikan Tugas akhir yang berjudul “Komparasi Algoritma *K-Nearest Neighbor* dan *Naïve Bayes* Dalam Penerimaan Kartu Indonesia Pintar di DKI Jakarta Menggunakan *K-Fold Cross Validation*” dengan baik.

Tugas akhir ini dibuat dalam memenuhi persyaratan untuk memperoleh gelar Sarjana Strata 1 Universitas Mercu Buana serta dibuat sebagai salah satu wujud implementasi dari ilmu yang didapatkan selama perkuliahan di Program Studi Teknik Informatika Fakultas Ilmu Komputer Universitas Mercu Buana. Tugas akhir ini tentunya tidak lepas dari bantuan ketersediaan data, bimbingan, masukan, dan arahan dari berbagai pihak. Oleh karena itu, pada kesempatan ini saya ingin mengucapkan terima kasih yang sebesar-besarnya kepada :

1. Rushendra, S.Kom, M.T selaku Dosen Pembimbing Tugas Akhir
2. Bapak Emil Robert Kaburuan, PhD selaku Kepala Program Studi Teknik Informatika.
3. Bapak Wawan Gunawan, S.Kom, MT selaku Koordinator Tugas Akhir Jurusan Teknik Informatika.
4. Saruni Dwiasnati, ST.MM., M.Kom selaku Dosen Pembimbing Akademik
5. Kedua orang tua yang selalu senantiasa memberikan doa serta dukungan.
6. Afina Tu'ulya yang senantiasa mendukung saya.
7. Teman-teman Serta seluruh pihak yang telah membantu dalam proses penulisan Tugas akhir ini baik secara langsung maupun tidak langsung.

Penulis menyadari bahwa penelitian ini jauh dari kesempurnaan. Oleh sebab itu, kritik maupun saran selalu penulis harapkan demi menghasilkan hasil terbaik dari penelitian ini. Besar harapan penulis, semoga penelitian ini dapat memberikan manfaat sekaligus menambah pengetahuan bagi berbagai pihak. Amin

Jakarta, 29 Maret 2022

Penulis

(Muhammad Yusuf Rizqon Rangkti)



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	xi
KATA PENGANTAR.....	xiii
DAFTAR ISI.....	xv
NASKAH JURNAL	1
KERTAS KERJA.....	9
BAB 1. LITERATUR REVIEW	10
BAB 2. ANALISIS DAN PERANCANGAN.....	27
BAB 3. SOURCE CODE	34
BAB 4. DATASET.....	41
BAB 5. TAHAPAN EKSPERIMEN	43
BAB 6. HASIL SEMUA EKSPERIMEN.....	48
DAFTAR PUSTAKA	54
LAMPIRAN DOKUMEN HAKI.....	58
LAMPIRAN KORESPONDENSI	60

NASKAH JURNAL

Terakreditasi SINTA Peringkat 2

Surat Keputusan Direktur Jenderal Pendidikan Tinggi, Riset, dan Teknologi, Nomor: 158/E/KPT/2021 masa berlaku mulai Volume 5 Nomor 2 Tahun 2021 sampai Volume 10 Nomor 1 Tahun 2026

Terbit online pada laman web jurnal: <http://jurnal.iaii.or.id>



JURNAL RESTI

(Rekayasa Sistem dan Teknologi Informasi)

Vol. 6 No. x (2022) x - x

ISSN Media Elektronik: 2580-0760

Komparasi Algoritma K-Nearest Neighbor dan Naïve Bayes Dalam Penerimaan KIP di DKI Jakarta Menggunakan K-Fold Cross Validation

Comparison of K-Nearest Neighbor and Naïve Bayes Algorithms in Accepting KIP in DKI Jakarta Using K-Fold Cross-Validation

Muhammad Yusuf Rizqon Rangkuti¹, Rushendra²

Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana

¹rangkuti289@gmail.com, ²rushendra@mercubuana.ac.id*

Abstract

Education is a real government effort in improving human resources and advancing the country. One of the efforts of the Indonesian government to improve education in Indonesia is the existence of the Smart Indonesia Card (KIP) by Presidential Instruction number 7 of 2014. The KIP program provides educational funding assistance to all school-age children 6-21 years who come from poor families. (underprivileged) or registered as participants in the Family Hope Program (PKH) or Prosperous Family Card (KKS). But in reality, the KIP program itself is not the proper target. Some students can afford KIP cards, which should be accepted by students who cannot. This is of course the government's efforts to improve human resources and make the country progress. Therefore we need a model that can predict KIP acceptance so that the KIP revenue target is appropriate and facilitates the achievement of government efforts in improving human resources. This study aims to predict KIP acceptance using the K-Nearest Neighbor and Naïve Bayes algorithms and the algorithm testing model using K-Fold Cross Validation. With the application of the KNN and Naïve Bayes algorithms to predict KIP acceptance, it is hoped that it can help increase human resources and advance the country. The results showed that the KNN algorithm had better results with an average accuracy rate of 96.9% with an average level of precision, recall, and f1-score of 97.87%, 92.87%, and 95%. By using the value of K with a small error value of K=3 while the Naïve Bayes algorithm only obtains an accuracy of 86.87% with an average level of precision, recall, and f1-score of 79.62%, 86.25%, 82%.

Keywords: Kartu Indonesia Pintar, K-Nearest Neighbor, Naïve Bayes, K-Fold Cross Validation, Machine Learning

Abstrak

Pendidikan merupakan upaya yang nyata pemerintah dalam meningkatkan sumber daya manusia dan memajukan negara. Salah satu upaya pemerintah Indonesia dalam meningkatkan pendidikan di Indonesia adalah dengan adanya Kartu Indonesia Pintar (KIP) sesuai dengan instruksi Presiden nomor 7 tahun 2014. Program KIP sendiri adalah pemberian bantuan tunai pendidikan kepada seluruh anak usia sekolah 6-21 tahun yang berasal dari keluarga miskin (kurang mampu) ataupun yang terdaftar sebagai peserta Program Keluarga Harapan (PKH) maupun Kartu Keluarga Sejahtera (KKS). Namun pada kenyataannya program KIP sendiri terkadang tidak sesuai dengan target yang semestinya. Ada siswa/i yang mampu tetapi mendapatkan kartu KIP yang seharusnya diterima oleh siswa/i yang tidak mampu hal ini tentu saja mempengaruhi upaya pemerintah dalam meningkatkan sumber daya manusia dan memajukan negara. Oleh sebab itu diperlukan suatu model yang dapat memprediksi penerimaan KIP agar target penerimaan KIP sesuai dan mempermudah tercapainya upaya pemerintah dalam meningkatkan SDM. Penelitian ini bertujuan memprediksi penerimaan KIP menggunakan algoritma K-Nearest Neighbor dan Naïve Bayes serta pengujian model algoritma menggunakan K-Fold Cross Validation. Dengan adanya penerapan algoritma KNN dan Naïve Bayes untuk memprediksi penerimaan KIP, diharapkan dapat membantu dalam meningkatkan SDM dan memajukan negara. Hasil penelitian ini menunjukkan bahwa algoritma KNN memiliki hasil yang lebih baik dengan tingkat akurasi rata-rata sebesar 96,9% dengan tingkat rata-rata precision, recall dan f1-score sebesar 97,87%, 92,87%, 95%. Dengan menggunakan nilai K dengan nilai error yang kecil yaitu K=3 Sedangkan algoritma Naïve Bayes hanya memperoleh akurasi sebesar 86.87% dengan tingkat rata-rata precision, recall, dan f1-score sebesar 79.62%,

86.25%, 82%.
Neighbor, Naïve Bayes, K-Fold Cross Validation, Machine Learning

Kata kunci: : Kartu Indonesia Pintar, K-Nearest



1. Pendahuluan

Pendidikan merupakan upaya yang nyata pemerintah dalam meningkatkan sumber daya manusia dan memajukan negara. Pendidikan juga merupakan hal yang terpenting pada manusia dan pendidikan secara umum memiliki arti suatu proses kehidupan dalam mengembangkan diri tiap individu untuk dapat hidup dan melangsungkan kehidupan. Sehingga menjadi seorang yang terdidik itu sangat penting [1]. Dikarenakan pendidikan itu sangat penting, maka pendidikan tidak akan ada habisnya. Pendidikan akan selalu berubah dan berkembang seiring dengan perubahan zaman.

Hal ini menjadikan pemerintah harus lebih meningkatkan kualitas dan kuantitas pendidikan di Indonesia. Salah satu upaya pemerintah Indonesia dalam meningkatkan pendidikan di Indonesia adalah dengan adanya Kartu Indonesia Pintar (KIP) sesuai dengan instruksi Presiden nomor 7 tahun 2014. Program KIP sendiri ialah pemberian bantuan tunai pendidikan kepada seluruh anak usia sekolah 6-21 tahun yang berasal dari keluarga miskin (kurang mampu) ataupun yang terdaftar sebagai peserta Program Keluarga Harapan (PKH) maupun Kartu Keluarga Sejahtera (KKS) [2]. KIP sendiri merupakan bagian dari penyempurnaan program Bantuan Siswa Miskin (BSM) dan diharapkan dapat memaksimalkan wajib belajar 12 tahun.

Namun pada kenyataannya program KIP sendiri terkadang tidak sesuai dengan target yang semestinya. Ada siswa/i yang mampu tetapi mendapatkan kartu KIP yang seharusnya diterima oleh siswa/i yang tidak mampu. Dalam hal ini penulis mendapatkan beberapa referensi dari penelitian terkait seperti pada penelitian [3] mengenai efektivitas program kartu Indonesia pintar bagi siswa smk di kecamatan jeruk legi kabupaten cilacap, dimana penelitian ini menganalisis tingkat efektifitas kartu Kartu Indonesia Pintar. Dimana hasil penelitian ini ini diketahui efektifitas program Kartu Indonesia Pintar bagi siswa SMK di Kecamatan Jeruklegi Kabupaten Cilacap lebih dari 70% [3].

Penelitian mengenai program Kartu Indonesia Pintar ini berbeda dari sebelumnya, dimana penelitian ini penulis ingin mengkomparasikan 2 algoritma untuk mengetahui tingkat keefektifan penerima KIP dilihat dari tingkat akurasi dari data Kesejahteraan Masyarakat yang didapatkan dari Kementerian Sosial RI. Penelitian ini menggunakan algoritma K-Nearest Neighbor dan Naïve Bayes.

Algoritma K-Nearest Neighbor sendiri ialah metode klasifikasi dengan mencari jarak terdekat antar data yang akan dievaluasi dengan K tetangga (neighbor) terdekatnya [4]. Algoritma K-Nearest Neighbor

termasuk Supervised Machine Learning dimana algoritma K-Nearest Neighbor melakukan pembelajaran yang berguna untuk menyelesaikan masalah pada klasifikasi [5]. Naïve Bayes ialah algoritma klasifikasi dengan rumus yang sederhana dan mudah di implementasikan [6], Naive Bayes menggunakan teorema Bayes untuk menghitung probabilitas menggunakan pengalaman masa lalu [7]. Skenario seperti komposisi data, nilai K tetangga, dan pengujian algoritma [8][9] menggunakan K-Fold Cross Validation juga mempengaruhi hasil dan performa algoritma.

Maka dari itu penelitian ini dilakukan untuk mencari algoritma terbaik dari K-Nearest Neighbor dan Naïve Bayes yang di implementasikan kedalam penerimaan Kartu Indonesia Pintar di DKI Jakarta.

2. Metode Penelitian

Metode yang dilakukan pada penelitian ini bisa dilihat pada Gambar 1.



Gambar 1. Metode Penelitian

Berikut merupakan penjelasan dari Gambar 1. Yang dilakukan untuk menguji algoritma K-Nearest Neighbor dan Naïve Bayes pada Penerimaan Kartu Indonesia Pintar di DKI Jakarta :

2.1. Pengumpulan Data

Pada penelitian ini dimulai dengan tahap pengumpulan data, data yang digunakan merupakan Data Kesejahteraan Sosial yang berasal dari Kementerian Sosial Republik Indonesia. Data tersebut memiliki 50 atribut data antara lain ada_kks, ada_pkh, dan ada_kip serta memiliki 1.048.575 record data, data ini merupakan data tahun 2020. Akan tetapi dalam melakukan pengujian algoritma training dan testing hanya menggunakan variable atribut ada_kip. Selain mencari data yang ingin digunakan, dilakukan juga pencarian studi literatur yang digunakan sebagai referensi pada penelitian ini, sumber referensi berbentuk artikel, buku dan jurnal yang berkaitan dengan algoritma K-Nearest Neighbor dan Naïve Bayes.

2.2 Preparing Data

Data yang sudah didapatkan kemudian diseleksi menjadi 250.000 data dan dibagi menjadi 2 bagian yaitu data training (data latih) dan data testing (data uji). Berikut ini perbandingan pembagian data latih dan data uji per keseluruhan data seperti pada tabel 1.

Tabel 1. Tabel Perbandingan Data Latih dan Data Uji

Perbandingan Data Latih (%)	Perbandingan Data Uji (%)
80	20
70	30
60	40
50	50

2.3 Cleaning Data

Tahap cleaning data merupakan tahap penting, dimana tahap ini data dibersihkan untuk menghilangkan noise dan data yang tidak konsisten atau data yang tidak relevan [10]. Tahap ini data dibersihkan sebaik mungkin agar saat pemrosesan data menghasilkan informasi yang akurat. Pada tahap ini data yang semula mempunyai 50 atribut data dihilangkan menjadi 10 atribut data. Data tersebut kemudian dicek apakah ada record data yang kosong, jika ada record data yang kosong maka baris data tersebut akan dihilangkan[11], kemudian nantinya data tersebut akan diimplementasikan menggunakan algoritma K-Nearest Neighbor dan Naïve Bayes. Sedikit potongan program dengan bahasa Python pada saat cleaning data terdapat pada Gambar 2.

```
#Menghapus Kolom yang tidak digunakan
data = test_ta.drop(['status_bangunan','status_lahan','ada_t
abung_gas','ada_lemari_es','ada_ac','ada_pemanas','ada_telep
on','ada_tv','ada_emas','ada_laptop','ada_sepeda','ada_motor
','ada_mobil','ada_perahu','ada_motor_tempel','ada_perahu_mo
tor','ada_kapal','aset_tak_bergerak','luas_atb','rumah_lain'
,'jumlah_sapi','jumlah_kerbau','jumlah_kuda','jumlah_babi','
jumlah_kambing','sta_art_usaha','JnsKelamin','Tmplahir','Tgl
Lahir','Hub_KRT','NUK','lapangan_usahaart','omset_usaha','l
okasi_usaha','Lapangan_usaha','kepemilikan_kartu','Status_pe
kerjaan','NMPROP','NMKAB','Hubkel'], axis = 1)
#Menghapus baris data yang kosong
data.dropna(inplace=True)
data.isnull().sum()
```

Gambar 2. Potongan Program Cleaning Data

2.4 Processing Data

Tahap ini data yang sudah di bersihkan kemudian di implementasikan kedalam algoritma klasifikasi K-Nearest Neighbor dan Naïve Bayes dan kemudian kedua algoritma tersebut di uji menggunakan K-Fold Cross Validation untuk melihat performa kedua algoritma.

Algoritma K-Nearest Neighbor mempelajari pola baru dalam data dengan cara menghubungkan pola data lama dengan data baru[12]. Algoritma K-Nearest Neighbor mudah di implementasikan, cepat dalam melatih data, cocok untuk data yang noisy dan banyak kelas serta tidak konsisten[13]. K-Nearest Neighbor memiliki persamaan bisa dilihat pada rumus 1 Eucliden Distance [14].

$$D_i = \sqrt{\sum_{i=1}^n (p_i - q_i)^2} \quad (1)$$

Dengan n sama dengan jumlah dimensi data, p_i data training, q_i data testing dan i variabel data.

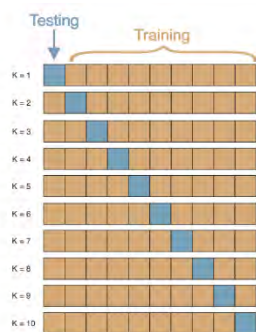
Algoritma Naïve Bayes adalah sebuah metode klasifikasi berdasarkan teorema bayes[15]. Cara kerja Naïve Bayes memprediksi probabilitas dimasa depan berdasarkan pengalaman sebelumnya[15]. Algoritma Naive Bayes cocok diterapkan pada bentuk data yang berskala ordinal[16]. Algoritma Naïve Bayes memiliki persamaan seperti pada rumus 2 Teorema Bayes.

$$P(H|X) = \frac{P(X|H).P(H)}{P(X)} \quad (2)$$

Dengan X sama dengan data kelas yang belum di ketahui, H adalah hipotesis data merupakan class spesifik, P(H|X) merupakan probabilitas hipotesis H berdasarkan kondisi X (posteriori probabilitas), P(H) merupakan probabilitas hipotesis H (prior probabilitas), P(X|H) merupakan probabilitas X berdasarkan kondisi pada hipotesis H, dan P(X) merupakan probabilitas X.

K-Fold Cross Validation ialah sebuah teknik validasi model untuk menilai bagaimana hasil statistik analisis akan menggeneralisasi kumpulan data independen[17]. K-Fold Cross Validation

memecah data menjadi k bagian set data dengan ukuran yang sama split data[17]. Konsep K-Fold Cross Validation bisa dilihat pada gambar 3.



Gambar 3. 10 Cross Validation

2.5 Evaluasi dan Komparasi

Setelah dataset diproses menggunakan algoritma K-Nearest Neighbor dan Naïve Bayes serta di uji menggunakan K-Fold Cross Validation. Hasil yang dihasilkan dari tahap processing kemudian di evaluasi menggunakan Confusion Matrix dan Classification Report. Confusion Matrix ialah tabel yang terdiri dari jumlah baris data uji yang di prediksi benar dan salah dari model klasifikasi yang digunakan[12]. Confusion Matrix dapat dilihat pada tabel 2.

Tabel 2. Confusion Matrix

Predicted Values	Actual Values	
	Positive (1)	Negative (0)
Positive (1)	TP	FP
Negative (0)	FN	TN

Classification Report bertujuan untuk mengetahui hasil evaluasi pada accuracy, precision, recall, dan f1-score. Tahap ini dilakukan perhitungan yang diuji menggunakan Akurasi di definisikan sebagai tingkat kedekatan antara nilai prediksi dengan nilai actual [7].

Akurasi digunakan untuk mengevaluasi banyaknya label prediksi yang sesuai dengan label aslinya. Semakin besar nilai akurasinya, maka performa algoritma tersebut semakin baik.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (3)$$

Precision adalah perbandingan data yang diprediksi benar dengan jumlah keseluruhan data.

$$Precision = \frac{TP}{(FP+TP)} \quad (4)$$

Recall adalah jumlah data yang diprediksi benar dengan jumlah rasio data yang diprediksi benar dan negatif palsu.

$$Recall = \frac{TP}{(FN+TP)} \quad (5)$$

F1-Score adalah rata-rata nilai yang didapatkan dari Recall dan Precision yang dibobotkan.

$$F1\ Score = 2 \times \frac{(Recall \times Precision)}{(Recall+Precision)} \quad (6)$$

3. Hasil dan Pembahasan

Penelitian ini dilakukan dengan beberapa komposisi data, antara lain 80%:20%, 70%:30%, 60%:40%, dan 50%:50% yang merupakan skenario pengujian.

Perbandingan Data Latih (%)	Perbandingan Data Uji (%)	Label	Precision K-Nearest Neighbor (%)	Recall K-Nearest Neighbor (%)	F1-Score K-Nearest Neighbor (%)	Accuracy K-Nearest Neighbor (%)	Precision Naïve Bayes (%)	Recall Naïve Bayes (%)	F1-Score Naïve Bayes (%)	Accuracy Naïve Bayes (%)
80	20	0 (Tidak)	97	100	98	97	96	87	91	86,8
		1 (Ya)	100	86	92		63	85	73	
70	30	0 (Tidak)	96	100	98	96,9	96	88	91	86,9
		1 (Ya)	100	86	92		63	85	73	
60	40	0 (Tidak)	96	100	98	96,9	96	88	92	87,2
		1 (Ya)	99	85	92		64	85	73	
50	50	0 (Tidak)	96	100	98	96,8	96	87	91	86,6
		1 (Ya)	99	86	92		63	85	72	
Rata-Rata			97,875	92,875	95	96,9	79,625	86,25	82	86,875

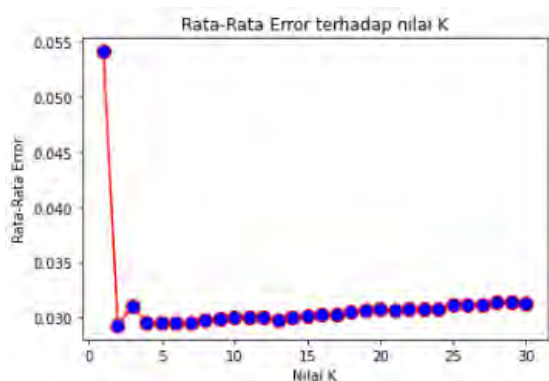
Gambar 4. Hasil Pengujian Algoritma K-Nearest Neighbor dan Naïve Bayes

Dilihat dari hasil pengujian, algoritma K-Nearest Neighbor memiliki hasil lebih baik dengan tingkat akurasi rata-rata sebesar 96,9% dengan tingkat rata-rata precision, recall dan f1-score sebesar 97,87%, 92,87%, 95%. Sedangkan algoritma Naïve Bayes hanya memperoleh akurasi sebesar 86,87% dengan tingkat rata-rata precision, recall, dan f1-score sebesar 79,62%, 86,25%, 82%. Hasil yang dihasilkan oleh kedua algoritma juga di pengaruhi oleh besarnya jumlah data latih dan data uji bisa dilihat pada gambar 4 dan gambar 5.



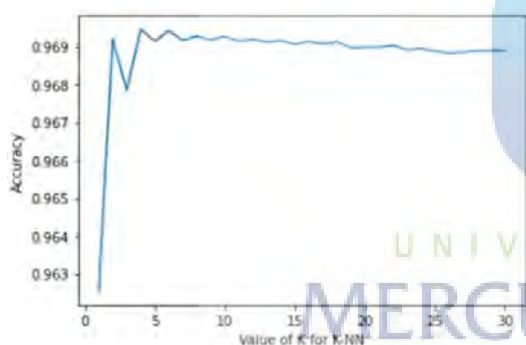
Gambar 5. Perbandingan Akurasi Kedua Algoritma

Pengujian algoritma K-Nearest Neighbor nilai K juga mempengaruhi hasil dari pengujian K-Nearest Neighbor. Untuk mengetahui performa nilai K dapat dilihat pada gambar 6.



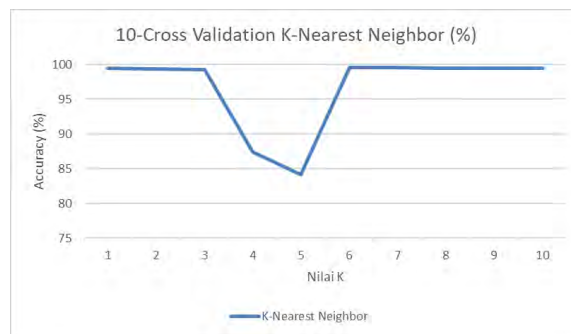
Gambar 6. Rata-rata Error Nilai K

Rata-rata error nilai K berpengaruh terhadap tingkat akurasi algoritma K-Nearest Neighbor, hal ini dikarenakan prediksi K-Nearest Neighbor berdasarkan pada asumsi intuitif objek yang berdekatan antara tetangga terdekat [18]. Pada penelitian ini semakin besar nilai K maka rata-rata error semakin besar, hal ini mempengaruhi akurasi yang semakin menurun namun tidak menurun secara signifikan hal ini juga bisa disebut sebagai Hyper Parameter Tuning yang berfungsi mencari akurasi terbaik bisa dilihat pada gambar 7.

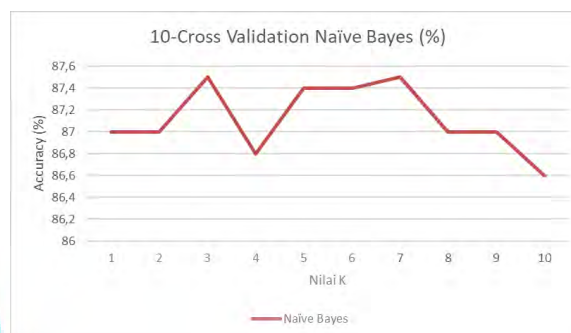


Gambar 7. Pengaruh Nilai K Terhadap Akurasi

Setelah hasil kedua algoritma di dapatkan, dilakukan pengujian algoritma menggunakan 10 K-Fold Cross Validation. Dengan menggunakan teknik ini membagi data secara acak ke dalam tiap bagian dimana terdiri dari 10 bagian untuk setiap bagian akan dilakukan proses klasifikasi terlebih dahulu [16] yang bertujuan untuk menilai kinerja proses sebuah metode algoritma [19] bisa dilihat pada gambar 3.



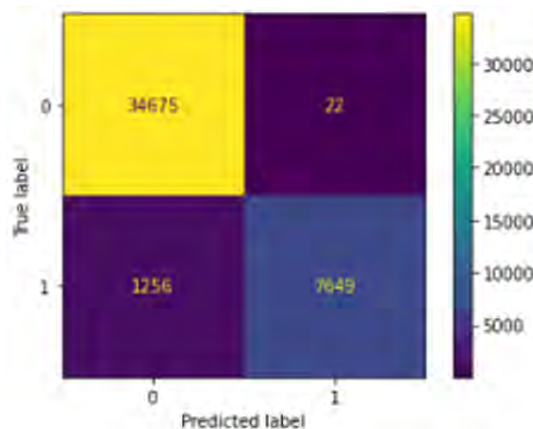
Gambar 8. 10 Cross Validation KNN



Gambar 9. 10 Cross Validation Naïve Bayes

Hasil K-Fold Cross Validation menunjukkan akurasi algoritma K-Nearest Neighbor lebih baik dibandingkan dengan algoritma Naïve Bayes, namun jarak rata-rata akurasi algoritma Naïve Bayes lebih dekat dibandingkan menggunakan algoritma K-Nearest Neighbor. Hal ini juga disebabkan beberapa faktor salah satunya adalah komposisi dari data latih dan data uji.

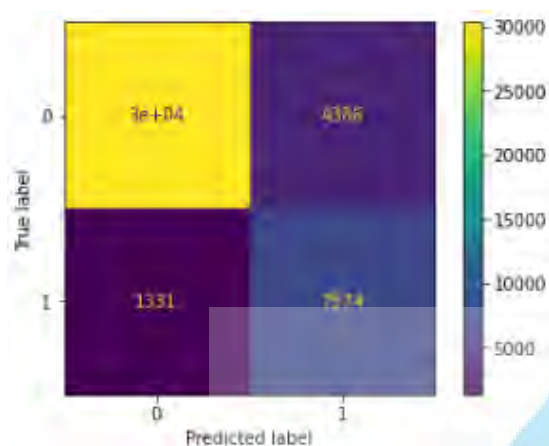
Kedua algoritma tersebut kemudian di evaluasi menggunakan Confusion Matrix. Confusion Matrix digunakan untuk mengevaluasi model klasifikasi untuk memperkirakan objek yang benar atau salah[20].



Gambar 10. Confusion Matrix K-Nearest Neighbor

Confusion matrix algoritma K-Nearest Neighbor memiliki True positif sebesar 34675 yang artinya

terdapat 34675 data positif yang di prediksi benar positif, terdapat 22 False positif yang artinya terdapat 22 data negatif yang di prediksi sebagai positif (positif palsu), lalu terdapat 1256 False negatif yang artinya sebanyak 1256 data positif di prediksi sebagai negative (negative palsu), terdapat 7649 True negative yang artinya sebanyak 7649 data negatif di prediksi benar sebagai data negative. Sedangkan untuk Confusion Matrix Naïve Bayes bisa dilihat pada gambar 11.



Gambar 11. Confusion Matrix Naïve Bayes

Confusion matrix algoritma Naïve Bayes memiliki True positif sebesar 30311 yang artinya terdapat 30311 data positif yang di prediksi benar positif, terdapat 4386 False positif yang artinya terdapat 4386 data negatif yang di prediksi sebagai positif (positif palsu), lalu terdapat 1331 False negatif yang artinya sebanyak 1331 data positif di prediksi sebagai negative (negative palsu), terdapat 7574 True negative yang artinya sebanyak 7574 data negatif di prediksi benar sebagai data negative. Hasil dari Confusion Matrix bisa digunakan untuk menghitung Accuracy, Precision, Recall, dan F1-Score seperti pada rumus 3,4,5, dan 6.

4. Kesimpulan

Berdasarkan hasil pengujian kedua algoritma K-Nearest Neighbor dan Naïve Bayes dalam implementasi penerimaan KIP di DKI Jakarta bisa di simpulkan bahwa algoritma K-Nearest Neighbor memiliki hasil yang lebih baik dengan tingkat akurasi rata-rata sebesar 96,9% dengan tingkat rata-rata precision, recall dan f1-score sebesar 97,87%, 92,87%, 95%. Dengan menggunakan nilai K dengan nilai error yang kecil yaitu K=3 Sedangkan algoritma Naïve Bayes hanya memperoleh akurasi sebesar 86.87% dengan tingkat rata-rata precision, recall, dan f1-score sebesar 79.62%, 86.25%, 82%. Skenario yang ada dalam penelitian ini juga mempengaruhi hasil dari kedua algoritma seperti jumlah komposisi data dan jumlah nilai K tetangga

KNN. Dapat disimpulkan dari penelitian ini, bahwa dengan data yang memiliki banyak noise dan jumlah data yang banyak algoritma KNN dan Naïve Bayes dapat dijadikan model alternatif untuk memprediksi penerimaan KIP dan untuk menghindari kesalahan dalam menargetkan penerima KIP sesuai dengan kriteria Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi (KEMENDIKBUD). Selanjutnya percobaan mengenai penerimaan KIP harus lebih banyak di lakukan lagi agar target yang dicapai sesuai tujuan awal KEMENDIKBUD dan dalam pengimplementasian selanjutnya dapat menggunakan metode lain ditambah dengan jumlah data dan variabel yang digunakan lebih banyak serta skenario yang lebih beragam agar informasi yang dihasilkan lebih akurat lagi.

Daftar Rujukan

- [1] Yayan Alpian, Sri Wulan Anggraeni, Unika Wiharti, and Nizmah Maratos Soleha, "PENTINGNYA PENDIDIKAN BAGI MANUSIA," *J. BUANA Pengabdian*, vol. 1, no. 1, pp. 66–72, Aug. 2019, doi: 10.36805/jurnalbuanapengabdian.v1i1.581.
- [2] R. F. Sinaga, S. R. Andani, and S. Suhada, "Penentuan Penerima KIP Dengan Menggunakan Metode Moora Pada Sd Negeri 124395 Pematang Siantar," *KOMIK (Konferensi Nas. Teknol. Inf. dan Komputer)*, vol. 2, no. 1, pp. 278–285, 2018, doi: 10.30865/komik.v2i1.938.
- [3] S. Setyawati, "Efektivitas Program Kartu Indonesia Pintar (Kip) Bagi Siswa Smk Di Kecamatan Jeruklegi Kabupaten Cilacap (Studi Permendikbud No. 12 Tahun 2015 Tentang Program Indonesia Pintar)," *Skrripsi IAIN Purwokerto*, no. 12, 2018.
- [4] M. A. Bahrudin, R. K. Niswatin, and L. S. Wahyuniar, "Sistem Pendukung Keputusan Penentuan Ekstrakurikuler Siswa SMK Al- Asy ' ariyah Prambon Dengan Metode K-Nearest Neighbor (KNN)," pp. 185–192, 2021.
- [5] A. Kumar, A. Verma, G. Shinde, Y. Sukhdeve, and N. Lal, "Crime Prediction Using K-Nearest Neighboring Algorithm," *Int. Conf. Emerg. Trends Inf. Technol. Eng. ic-ETITE 2020*, pp. 4–7, 2020, doi: 10.1109/ic-ETITE47903.2020.155.
- [6] A. Jadhav, A. Pandita, A. Pawar, and V. Singh, "Classification of Unstructured Data using Naïve Bayes Classifier and Predictive Analysis for RTI Application," *ABHIYANTRIKI An Int. J. Eng. Technol.*, vol. 3, no. 6, pp. 1–6, 2016.
- [7] E. R. Setyaningsih and I. Listiowarni, "Categorization of Exam Questions based on Bloom Taxonomy using Naïve Bayes and Laplace Smoothing," *3rd 2021 East Indones. Conf. Comput. Inf. Technol. EIConCIT 2021*, pp. 330–333, 2021, doi: 10.1109/EIConCIT50028.2021.9431862.
- [8] Rushendra, K. Ramli, N. Hayati, E. Ihsanto, T. S. Gunawan, and A. H. Halbouni, "Development of Intrusion Detection System using Residual Feedforward Neural Network Algorithm," *2021 4th Int. Semin. Res. Inf. Technol. Intell. Syst. ISRITI 2021*, pp. 539–543, 2021, doi: 10.1109/ISRITI47903.2021.9431862.

- 10.1109/ISRITI54043.2021.9702773.
- [9] M. Yusuf, R. Hidayat, Y. Liklikwatil, and D. Surya Subrata, "Rancang Bangun Sistem Deteksi Dini Ketinggian Air Banjir Berbasis IoT dengan Sensor Ultrasonik," *J. ICT Inf. Commun. Technol.*, vol. 18, no. 2, pp. 93–101, 2019, doi: 10.36054/jict-ikmi.v18i2.57.
- [10] Alvina Felicia Watratan, Arwini Puspita. B, and Dikwan Moeis, "Implementasi Algoritma Naive Bayes Untuk Memprediksi Tingkat Penyebaran Covid-19 Di Indonesia," *J. Appl. Comput. Sci. Technol.*, vol. 1, no. 1, pp. 7–14, 2020, doi: 10.52158/jacost.v1i1.9.
- [11] F. Harahap *et al.*, "Implementation of Naïve Bayes Classification Method for Predicting Purchase," *2021 Int. Conf. Inf. Commun. Technol. Sustain. Dev. ICT4SD 2021 - Proc.*, no. Citsm, pp. 1–5, 2019, doi: 10.1109/CITSM.2018.8674324.
- [12] M. R. Romadhon and F. Kurniawan, "A Comparison of Naive Bayes Methods, Logistic Regression and KNN for Predicting Healing of Covid-19 Patients in Indonesia," *3rd 2021 East Indones. Conf. Comput. Inf. Technol. EIconCIT 2021*, pp. 41–44, 2021, doi: 10.1109/EIconCIT50028.2021.9431845.
- [13] N. U. R. C. Aziz, P. S. Informatika, F. Komunikasi, D. A. N. Informatika, and U. M. Surakarta, "Implementasi algoritma knn untuk memprediksi potensi penyakit jantung dengan python flask," *Dr. Diss. Univ. Muhammadiyah Surakarta*, p. 23, 2021.
- [14] A. P. Permana, K. Ainayah, and K. F. H. Holle, "Analisis Perbandingan Algoritma Decision Tree, kNN, dan Naive Bayes untuk Prediksi Kesuksesan Start-up," *JISKA (Jurnal Inform. Sunan Kalijaga)*, vol. 6, no. 3, pp. 178–188, 2021, doi: 10.14421/jiska.2021.6.3.178-188.
- [15] D. Darwis, N. Siskawati, and Z. Abidin, "Penerapan Algoritma Naive Bayes Untuk Analisis Sentimen Review Data Twitter Bmkg Nasional," *J. Tekno Kompak*, vol. 15, no. 1, p. 131, 2021, doi: 10.33365/jtk.v15i1.744.
- [16] H. D. Wijaya and S. Dwiasnati, "Implementasi Data Mining dengan Algoritma Naïve Bayes pada Penjualan Obat," *J. Inform.*, vol. 7, no. 1, pp. 1–7, 2020, doi: 10.31311/ji.v7i1.6203.
- [17] F. Tempola, M. Muhammad, and A. Khairan, "Perbandingan Klasifikasi Antara KNN dan Naive Bayes pada Penentuan Status Gunung Berapi dengan K-Fold Cross Validation," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 5, p. 577, 2018, doi: 10.25126/jtiik.201855983.
- [18] A. Bayhaqy, S. Sfenrianto, K. Nainggolan, and E. R. Kaburuan, "Sentiment Analysis about E-Commerce from Tweets Using Decision Tree, K-Nearest Neighbor, and Naïve Bayes," *2018 Int. Conf. Orange Technol. ICOT 2018*, pp. 1–6, 2018, doi: 10.1109/ICOT.2018.8705796.
- [19] A. Rohani, M. Taki, and M. Abdollahpour, "A novel soft computing model (Gaussian process regression with K-fold cross validation) for daily and monthly solar radiation forecasting (Part: I)," *Renew. Energy*, vol. 115, pp. 411–422, 2018, doi: 10.1016/j.renene.2017.08.061.
- [20] N. Dengen, "Comparison Performance of C4 . 5 , Naïve Bayes and K-Nearest Neighbor in," *Int. Conf. Sci. Inf. Technol.*, pp. 112–117, 2019.

KERTAS KERJA

Ringkasan

Kertas kerja ini merupakan material kelengkapan artikel jurnal dengan judul “Komparasi Algoritma *K-Nearest Neighbor* dan *Naïve Bayes* Dalam Penerimaan Kartu Indonesi Pintar DKI Jakarta Menggunakan *K-Fold Cross Validation*”. Judul ini di ambil dikarenakan masih ada saja penyalahgunaan dana KIP dan kesalahan dalam menargetkan penerima KIP. Kertas kerja ini berisi semua material hasil penelitian Tugas Akhir. Di dalam kertas kerja ini disajikan beberapa bagian yang terdiri dari literature review, Analisa perancangan, dataset yang digunakan, tahapan eksperimen, dan hasil eksperimen secara keseluruhan. Literatur review berisi tentang jurnal-jurnal pendukung yang terkait. Pada bab ini dijelaskan masing-masing judul jurnal yang dikutip, topik yang dibahas pada setiap jurnal yang dikutip, metode yang digunakan dalam penelitian tersebut, dan hasil yang didapatkan dari pengujian algoritma yang digunakan. Pada bab *Dataset* dilampirkan variabel-variabel yang digunakan sebagai landasan pengujian algoritma, Data tersebut terbagi menjadi dua yaitu data latih dan data uji. Pada bab ini, dilampirkan pula sedikit sampel data yang digunakan dalam penelitian. Tahapan eksperimen merupakan bab yang membahas tentang eksperimen yang akan dilakukan untuk analisa penerimaan KIP menggunakan metode *K-Nearest Neighbor* dan *Naïve Bayes*. Pada bab source code disajikan beberapa potongan code yang digunakan untuk melakukan pengujian algoritma dengan menggunakan bahasa program Python dan beberapa library pendukungnya. Kemudian yang terakhir yaitu bab hasil eksperimen secara keseluruhan membahas hasil yang didapat saat melakukan pengujian algoritma baik dari segi akurasi, *precision*, *recall* dan *f1-score*. Selain itu, dijabarkan pula hasil dari perbandingan kedua algoritma yang digunakan tersebut dalam melakukan analisa penerimaan KIP sampai pada penarikan kesimpulan algoritma terbaik.