



**PERBANDINGAN ALGORITMA C4.5 DAN NAÏVE BAYES DALAM
PENENTUAN KELULUSAN SISWA KE PTN MELALUI JALUR
SNMPTN**

(Studi Kasus : SMA HANG TUAH 1 JAKARTA)

TUGAS AKHIR

Citra Dewi Jivanti
41518010060

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



**PERBANDINGAN ALGORITMA C4.5 DAN NAÏVE BAYES DALAM
PENENTUAN KELULUSAN SISWA KE PTN MELALUI JALUR
SNMPTN
(Studi Kasus : SMA HANG TUAH 1 JAKARTA)**

TUGAS AKHIR

Diajukan Untuk Melengkapi Salah Satu Syarat
Memperoleh Gelar Sarjana Komputer

Oleh:

Citra Dewi Jivanti

41518010060

UNIVERSITAS
MERCU BUANA

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

LEMBAR PERNYATAAN ORISINALITAS

Yang bertanda tangan dibawah ini:

NIM : 41518010060

Nama : Citra Dewi Jivanti

Judul Tugas Akhir : Perbandingan Algoritma C4.5 dan Naive Bayes Dalam
Penentuan Kelulusan Siswa ke PTN Melalui Jalur
SNMPTN (Studi Kasus : SMA Hang Tuah 1 Jakarta)

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, 05 Juli 2022



Citra Dewi Jivanti



UNIVERSITAS
MERCU BUANA

SURAT PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR

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

Nama Mahasiswa : Citra Dewi Jivanti
NIM : 41518010060
Judul Tugas Akhir : Perbandingan Algoritma C4.5 dan Naïve Bayes
Dalam Penentuan Kelulusan Siswa ke PTN Melalui Jalur SNMPTN (Studi Kasus : SMA Hang Tuah 1 Jakarta)

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.

UNIVERSITAS
MERCU BUANA

Jakarta, 05 Juli 2022


Citra Dewi Jivanti

SURAT PERNYATAAN LUARAN TUGAS AKHIR

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

Nama Mahasiswa : Citra Dewi Jiyanti
 NIM : 41518010060
 Judul Tugas Akhir : Perbandingan Algoritma C4.5 dan Naïve Bayes
 Dalam Penentuan Kelulusan Siswa ke PTN Melalui Jalur SNMPTN (Studi Kasus : SMA Hang Tuah 1 Jakarta)

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	: JATISI (Jurnal Teknik Informatika dan Sistem Informasi)			
	ISSN	: 25032933			
	Link Jurnal	: https://jurnal.mdp.ac.id/index.php/jatisi/			
	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 HAKI), untuk kepentingan pendaftaran HAKI apabila diperlukan

Demikian pernyataan ini saya buat dengan sebenarnya.

Mengetahui,
 Jakarta 03 Agustus 2022



 Citra Dewi Jiyanti

LEMBAR PERSETUJUAN PENGUJI

NIM : 41518010060
Nama : Citra Dewi Jivanti
Judul Tugas Akhir : Perbandingan Algoritma C4.5 dan Naïve Bayes
Dalam Penentuan Kelulusan Siswa ke PTN Melalui
Jalur SNMPTN (Studi Kasus : SMA Hang Tuah 1
Jakarta)

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, 03 Agustus 2022


(Rushendra, S.Kom, M.T)

UNIVERSITAS
MERCU BUANA

LEMBAR PERSETUJUAN PENGUJI

NIM : 41518010060
Nama : Citra Dewi Jivanti
Judul Tugas Akhir : Perbandingan Algoritma C4.5 dan Naïve Bayes
Dalam Penentuan Kelulusan Siswa ke PTN Melalui
Jalur SNMPTN (Studi Kasus : SMA Hang Tuah 1
Jakarta)

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, 03 Agustus 2022



(Anis Cherid, SE, MTI)

UNIVERSITAS
MERCU BUANA

LEMBAR PERSETUJUAN PENGUJI

NIM : 41518010060
Nama : Citra Dewi Jivanti
Judul Tugas Akhir : Perbandingan Algoritma C4.5 dan Naïve Bayes
Dalam Penentuan Kelulusan Siswa ke PTN Melalui
Jalur SNMPTN (Studi Kasus : SMA Hang Tuah 1
Jakarta)

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, 03 Agustus 2022



(Eliyani, Dr. Ir.)

UNIVERSITAS
MERCU BUANA


LEMBAR PENGESAHAN

NIM : 41518010060
Nama : Citra Dewi Jivanti
Judul Tugas Akhir : Perbandingan Algoritma C4.5 dan Naïve Bayes Dalam Penentuan Kelulusan Siswa ke PTN Melalui Jalur SNMPTN (Studi Kasus : SMA Hang Tuah 1 Jakarta)

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, 03 Agustus 2022

an. Menyetujui,


(Leonard Goeirmanto, Dr., ST., M.Sc)
Dosen Pembimbing

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 syukur kehadirat Tuhan Yang Maha Esa, karena atas limpahan rahmat dan hidayah-Nya penulis dapat menyelesaikan laporan pelaksanaan Tugas Akhir dengan judul “Perbandingan Algoritma C4.5 dan Naïve Bayes Dalam Penentuan Kelulusan Siswa ke PTN Melalui Jalur SNMPTN (Studi Kasus : SMA Hang Tuah 1 Jakarta)”. Laporan Tugas Akhir ini merupakan salah satu persyaratan wajib untuk menyelesaikan Program Studi Strata Satu (S1) pada jurusan Informatika, Universitas Mercu Buana.

Dalam penyelesaian laporan Tugas Akhir ini, penulis mendapatkan banyak dukungan, bantuan, serta sumbangan ide maupun pikiran dari berbagai pihak. Untuk itu, penulis menyampaikan ucapan terimakasih kepada :

1. Bapak Leonard Goeirmento, ST., M.Sc selaku dosen Pembimbing Tugas Akhir yang telah membimbing dan memberikan masukan kepada penulis sehingga laporan tugas akhir ini dapat diselesaikan dengan baik.
2. Bapak Dr. Harwikarya, MT selaku dosen Pembimbing Akademik yang selalu membimbing dan menasehati selama masa kuliah.
3. Bapak Emil Robert Kabuaran, ST, MA, PhD selaku Kepala Program Studi Teknik Informatika Fakultas Ilmu Komputer Universitas Mercu Buana.
4. Orang tua, kakak dan adik yang telah memberi dukungan serta motivasi dalam berjalannya pengerjaan tugas akhir
5. Galih, Eren, Lestari dan Qonita Azzillatin S.Kom yang telah membantu menyelesaikan laporan tugas akhir

Akhir kata, penulis berharap tugas akhir ini dapat bermanfaat bagi pembaca guna menambah pengetahuan dan wawasan serta pembaca dapat memberikan kritik maupun saran. Penulis menyadari bahwa penyusunan laporan Tugas Akhir ini masih jauh dari kesempurnaan dan masih terdapat banyak kekurangan. Walaupun demikian, penulis telah berusaha semaksimal mungkin untuk mendapatkan hasil yang terbaik.

Jakarta, 05 Juli 2022

Citra Dewi Jivanti



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	viii
ABSTRACT.....	x
KATA PENGANTAR.....	xi
DAFTAR ISI.....	xi
NASKAH JURNAL	1
KERTAS KERJA.....	13
BAB 1. LITERATUR REVIEW.....	14
BAB 2. ANALISIS DAN PERANCANGAN.....	21
BAB 3. SOURCE CODE.....	24
BAB 4. DATASET.....	31
BAB 5. TAHAPAN EKSPERIMEN.....	33
BAB 6. HASIL SEMUA EKSPERIMEN.....	37
DAFTAR PUSTAKA	43
LAMPIRAN DOKUMEN HAKI.....	45
LAMPIRAN KORESPONDENSI	48

NASKAH JURNAL

■1

Perbandingan Algoritma C4.5 & Naive Bayes Dalam Penentuan Kelulusan Siswa Jalur SNMPTN (Studi Kasus : SMA Hang Tuah 1 Jakarta)

Citra Dewi Jivanti*¹, Leonard Goeirianto²

^{1,2} Universitas Mercu Buana; Jl. Raya, RT.4/RW.1, Meruya Sel., Kec. Kembangan, Jakarta,
Daerah Khusus Ibukota Jakarta 11650, (021) 5857722

Jurusan Ilmu Komputer, Teknik Informatika, Jakarta

e-mail: 415180100260@student.mercubuana.ac.id, leonard@mercubuana.ac.id

Abstrak

SMA Hang Tuah 1 Jakarta merupakan salah satu sekolah swasta yang berakreditasi A sejak tahun 2009 hingga sekarang. SMA Hang Tuah 1 Jakarta memperkirakan siswa kelas 12 masih dengan cara manual sehingga banyak siswa yang nilai rata-rata rapotnya tinggi tetapi tidak lolos di jalur snmptn. Hal ini merugikan siswa kelas 12 yang nilai rata-rata rapot tinggi namun tidak lolos dalam jalur SNMPTN. Penelitian ini melakukan pengujian model dengan membandingkan dua metode yaitu, algoritma C4.5(Decision Tree) dan algoritma Naive Bayes dengan menggunakan data nilai rata-rata siswa dari SMA Hang Tuah 1 Jakarta. Untuk mengukur kinerja dari kedua algoritma tersebut digunakan metode pengujian dengan split validatin dan cross validation dan Dapat diketahui untuk penggunaan algoritma C4.5 memiliki akurasi senilai 94,14%. Untuk pengujian menggunakan cross validation pada algoritma C4.5. menghasilkan nilai akurasi yang sama sebesar 98,67%. Dan untuk penggunaan algoritma Naive Bayes mendapatkan akurasi senilai 98,40%. algoritma Naive Bayes terbaik menghasilkan nilai akurasi yang sama sebesar 97,59%. Jika dilihat dari perbandingan algoritma yang telah dilakukan untuk menggunakan metode split validation Algoritma memiliki akurasi yang lebih baik dengan algoritma Naive Bayes, tetapi dalam penggunaan metode cross validation kedua algoritma memiliki akurasi yang bisa dibilang sedikit sama, tidak berbeda jauh dari akurasi yang didapat.

Kata kunci— Algoritma C4.5, Naive Bayes, Klasifikasi, SNMPTN, Akreditasi

Abstract

Hang Tuah 1 Jakarta High School is one of the private schools that has been accredited A since 2009 until now. SMA Hang Tuah 1 Jakarta ranks 12th grade students manually, so that many students have high average grades but do not qualify for the SNMPTN route. This is detrimental to grade 12 students whose average grades in report cards are high but do not qualify for the SNMPTN pathway. This study tested the model by comparing two methods, namely, the C4.5 algorithm (Decision Tree) and the Naive Bayes algorithm using the data on the average value of students from Hang Tuah 1 Jakarta High School. To measure the performance of the two algorithms, a test method with split validation and cross validation is used and it can be seen that the use of the C4.5 algorithm has a value of 94.14%. For testing using cross validation on the C4.5 algorithm, produces the same accuracy value of 98.67%. And for the use of the Naive Bayes algorithm, the value is 98.40%. The best Naive Bayes algorithm produces the same accuracy value of 97.59%. When viewed from the comparison of algorithms that have been carried out using the split validation method, the algorithm has better accuracy with the Naive Bayes algorithm, but in using the cross-validation method the two algorithms have accuracy that is practically the same, not far from the accuracy obtained.

Keywords— C4.5, Naive Bayes, Classification, SNMPTN, Accreditation

I. PENDAHULUAN

Perguruan Tinggi Negeri (*PTN*) merupakan pendidikan formal terakhir yang dikelola oleh pemerintah dibawah kuasa kemendikbud. Lulusan Sekolah Menengah Atas (*SMA*) para siswa banyak yang mempunyai keinginan untuk melanjutkan pendidikannya ke Perguruan Tinggi Negeri (*PTN*). Hal ini dikarenakan ada keunggulan dari masuk *PTN* yaitu biaya kuliah yang terjangkau dan juga perusahaan menginginkan para tenaga kerja yang lulusan dari Perguruan Tinggi Negeri (*PTN*) yang mencari tenaga kerja lulusan dari *PTN* karena untuk masuk ke *PTN* tidak mudah dan juga persaingan yang sangat ketat. Sehingga membuat perusahaan tidak meragukan lagi kemampuan seseorang yang lulus dari *PTN*. Tiap tahunnya *PTN* membuka kesempatan untuk siswa kelas 12 yang ingin masuk ke *PTN* melalui 3 jalur sebagai jalur seleksi para siswa untuk masuk *PTN*, yaitu *SNMPTN*, *SBMPTN*, dan *Mandiri*.

Ketiga jalur tersebut memiliki proses yang berbeda-beda dalam mengikuti tahapannya. Pada jalur *SNMPTN* setiap sekolah memiliki kuotanya masing-masing dilihat dari akreditasi sekolah tersebut. Untuk sekolah yang ber akreditasi *A* memiliki kuota sebesar 40%, untuk sekolah yang ber akreditasi *B* memiliki kuota sebesar 25% dan untuk sekolah yang akreditasi *C* memiliki kuota 5% dari keseluruhan siswa yang ada di sekolah tersebut. Tahap selanjutnya setiap sekolah diberikan kewenangan sendiri untuk memperingkatkan siswa kelas 12 dengan melihat nilai rata-rata *raport* dari *semester 1* sampai *semester 5* dan memilih siswa kelas 12 yang akan mendapatkan jalur *PTN*. Sedangkan untuk *SBMPTN* siswa diminta untuk memilih 3 *PTN* yang diminati dan mengikuti ujian hingga skor yang dicapai dinyatakan lolos oleh salah satu *PTN* yang sudah dipilih. Dan jalur *Mandiri* mengikuti tahapan sesuai Perguruan Tinggi Negeri masing-masing.

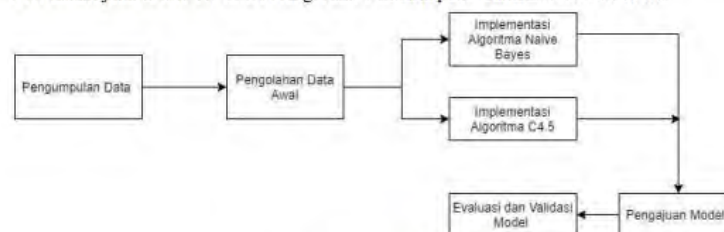
SMA Hang Tuah 1 Jakarta ialah salah satu sekolah *SMA* swasta yang akreditasi *A* sejak tahun 2009 hingga sekarang. *SMA Hang Tuah 1 Jakarta* mendapatkan kuota tiap tahunnya sebesar 40% pada jalur *SNMPTN* untuk siswa kelas 12. Siswa kelas 12 *SMA Hang Tuah 1 Jakarta* yang ingin masuk *PTN* melalui jalur *SNMPTN* dibantu oleh guru bimbingan konseling dengan meminta nilai rata-rata *raport* siswa kelas 12. *SMA Hang Tuah 1 Jakarta* memperingkatkan siswa kelas 12 masih dengan cara manual sehingga banyak siswa yang nilai rata-rata *rapotnya* tinggi tetapi tidak lolos di jalur *snmptn*. Tentu hal ini merugikan siswa kelas 12 yang nilai rata – rata *raport* tinggi namun tidak lolos dalam jalur *SNMPTN*.

Naive Bayes merupakan metode algoritma yang digunakannya sebuah klasifikasi data yang dimana, algoritma tersebut dapat memprediksi klasifikasi suatu data yang nantinya masuk ke dalam kelas tertentu. Kelebihan dari *Naive Bayes* yaitu, data kuantitatif dapat mengatasi nilai yang hilang selama perhitungan estimasi dari peluang data yang ada. Serta *Algoritma C4.5* merupakan algoritma yang digunakan untuk klasifikasi yang nantinya akan terjadi untuk memperoleh hasil yang cukup maksimal. Keunggulannya ialah dapat menghasilkan suatu pohon keputusan (*Decision Tree*) yang mudah, dan juga memiliki tingkat akurasi yang dapat diterima, sehingga efisien dalam menangani setiap atribut data.

Pada penelitian terdahulu berjudul “Analisis Perbandingan *Algoritma Decision Tree (C4.5)* Dan *K-Naive Bayes* Untuk Mengklasifikasi Penerimaan Mahasiswa Baru Tingkat Universitas” yang ditulis oleh *A. Setyanto dan Hanif Al Fattah*. Penelitian dilakukan menggunakan metode yang sama dengan mengambil data melalui suatu universitas. [1]

2. METODE PENELITIAN

Metode penelitan yang akan digunakan dalam penelitian mengenai penentuan kelulusan PTN dalam jalur SNMPTN SMA Hang Tuah 1 Jakarta pada Gambar 1. di bawah ini.



Gambar 1. Metode Penelitian

2.1 Tahapan Review

Pada *data collection* diambil dari sekolah yang ada di Hang Tuah 1 Jakarta. Data yang digunakan diambil dari tahun 2017 hingga 2022 dengan total data sebanyak 1.109. Untuk variabel data diambil berdasarkan hasil rata-rata siswa dari masa *semester 1* hingga *semester 5*. Karena nilai rata-rata tersebut merupakan penentuan untuk siswa lulus dan tidak lulus melalui jalur SNMPTN.

2.2 Preprocessing

- Data Cleaning* : Penelitian ini menggunakan data sebanyak 1.109 data. Pada proses *cleansing* data peneliti menghapus seluruh data di baris yang ada 1 kolom yang kosong. Pada tahap ini data yang dihasilkan sebanyak 941 data.
- Data Reduction* : Data yang ada terdapat 9 variabel, namun peneliti tidak menggunakan keseluruhan variabel, variabel yang digunakan hanya variabel yang sudah disesuaikan berdasarkan hasil dari dataset yang didapat. Variabel yang digunakan mendapatkan 7 variabel yang digunakan untuk tahap evaluasi yaitu *Semester 1 – Semester 5*.
- Data Transformasi* : Dalam proses ini, penulis melakukan transformasi data sebanyak 1 kali, yaitu mengubah variabel atribut SNMPTN yang awalnya *string* menjadi *int*. Karena jika menggunakan *string / varchar* tidak masuk ke proses perhitungan aritmatika algoritma.

2.3 Metode Algoritma yang diusulkan

Setelah semua data telah di proses dan mendapatkan data bersih maka selanjutnya data akan di implementasikan menggunakan 2 algoritma yaitu *algoritma C.45* dan *algoritma Naive Bayes*. Dalam implementasinya *nilai x* dan *y* ditentukan yaitu variabel *independent* dan variabel *dependent*.

2.3.1 Klasifikasi

Klasifikasi adalah teknik pembentukan model dari data yang tidak terklasifikasi, klasifikasi dilakukan untuk menghasilkan data baru. *Classifier* termasuk ke dalam tipe *supervised learning*, artinya data latih diperlukan untuk membangun model klasifikasi. [2] Klasifikasi merupakan topik penting dalam penelitian *data mining*. Ketika berhadapan dengan sekumpulan data, yang masing-masing telah dikelompokkan ke dalam kelas tertentu, masalah klasifikasi yang dipertimbangkan adalah menentukan aturan yang memungkinkan klasifikasi meskipun data yang ada tidak menunjukkannya. [3]

2.3.2 Algoritma C4.5

Algoritma C4.5 adalah algoritma pohon keputusan yang cukup sering digunakan dalam suatu penelitian. *Algoritma C4.5* pertama kali diperkenalkan oleh *Quinlan (1996)* sebagai versi perbaikan *ID3*. *Algoritma C4.5* digunakan untuk melakukan klasifikasi atau pengelompokan yang bersifat prediktif. [4] Pernyataan klasifikasi pada pohon keputusan (*decision tree*) menghasilkan *node daun (terminal note)* yang memberitahu suatu kelas-kelas atau *segment* dan membentuk cabang-cabang pada pohon keputusan. [5]

Perhitungan yang dilakukan dalam *algoritma C4.5* adalah *nilai entropy, gain* dan *split* dari masing-masing atribut pada data *trainign* yang ada, sehingga nantinya menghasilkan *nilai gain ratio* pada atribut. [6] Atribut *ratio* terbesar dipilih untuk membuat simpul pada akar pohon keputusan. Lalu untuk menghitung *nilai entropy, gain* dan *split* dari masing-masing atribut dengan menghilangkan atribut yang dipilih sebelumnya. Atribut yang memiliki *nilai gain ratio* terbesar dipilih untuk membuat *simpul (cabang)* pohon keputusan. Ulangi perhitungan tersebut hingga semua atribut telah memiliki kelas maka akan menampilkan bentuk pohon keputusan awal dan *generate* aturan keputusan awal. [7] Dari penjelasan tersebut juga dapat membantu suatu tahapan dalam membentuk sebuah pohon keputusan dengan *algoritma C4.5* :

1. Mempersiapkan *data training*, yang biasanya diambil melalui data histori yang pernah terjadi sebelumnya dan sudah dikelompokkan ke dalam kelas-kelas tertentu
2. Menentukan akar pohon keputusan. Akar akan diambil dari atribut yang terpilih dengan menghitung *nilai gain* yang tertinggi dari masing-masing atribut dengan *nilai entropy* terendah. Untuk menghitung *nilai entropy* menggunakan rumus :

$$Entropy(S) = \sum_{i=1}^n - p_i \log_2 p_i \quad (1)$$

Keterangan :

S : Himpunan kasus
n : Jumlah partisi S
pi : Jumlah kasus pada partisi ke-i

3. Menghitung *nilai gain* dengan rumus :

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^r Entropy(S_i) \quad (2)$$

Keterangan:

S : Himpunan kasus
A : Atribut
n : Jumlah partisi atribut A
|Si| : Jumlah kasus pada partisi ke-i
|S| : Jumlah kasus dalam S

4. Mengulangi tahap kedua hingga seluruh *record* terparti. Proses partisi pohon keputusan akan berhenti jika :
 - a. Semua *record* dalam *simpul N* mendapat kelas yang sama

- b. Tidak ada atribut di dalam *record* yang dipartisi lagi
- c. Tidak ada *record* di dalam cabang yang kosong.

2.3.3 Algoritma Naive Bayes

Menurut *teorema Bayes*, *Naive Bayes* adalah algoritma yang menggabungkan *probability* dengan probabilitas bersyarat pada rumus digunakan menghitung kemungkinan tiap klasifikasi pada suatu probabilitas. [8] Proses yang dilakukan dalam algoritma dengan memasukkan data yang akan diitung label atau kelas. Kemudian dari kelas tersebut dihitung probabilitas dari setiap kelas yang ada pada data. [9] Untuk perhitungan pada algoritma *Naive Bayes* menggunakan rumus :

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)} \quad (3)$$

Keterangan :

$P(A|B)$: Probabilitas *A* terjadi dengan bukti bahwa *B* telah terjadi (probabilitas superior).

$P(B|A)$: Probabilitas *B* terjadi dengan bukti bahwa *A* telah terjadi.

$P(A)$: Peluang terjadinya *A*

$P(B)$: Peluang terjadinya *B*

2.4 Model Testing

Model akan diuji pada data yang telah diproses melalui 2 (dua) metode yaitu *Split Validasi* dan *Cross Validasi*.

a. Split Validasi

Split Validasi (*Split Validation*) merupakan teknik validasi dengan membagi data menjadi 2 bagian, nantinya akan menghasilkan *data train* serta *data test* yang dapat digunakan dalam uji model algoritma.

b. Cross Validasi

Cross Validasi (*CV*) merupakan metode statistik yang digunakan untuk melakukan evaluasi kinerja model algoritma, yang dimana data akan dibagi menjadi 2 subdata, yaitu data proses serta data validasi/evaluasi. [10] Pada penelitian ini juga melakukan 5(lima) kali experiment menggunakan metode *cross validasi* dengan nilai *k* pada *k-fold* yang dijelaskan pada tabel 1 :

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

Tabel 1. Skenario Cross Validasi

2.5 Evaluasi Model

Evaluasi diperlukan untuk menganalisa dan mengukur sejauh mana keakuratan hasil yang telah dicapai oleh model dengan menggunakan *Confusion Matrix*. [11] Lalu hasil evaluasi akan diberikan dengan menggunakan *classification report* yang terdiri dari *Accuracy*, *Precision* dan *Recall*. [12]

1. Akurasi

$$accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (4)$$

2. Precision

$$precision = \frac{TP}{TP+FP} \quad (5)$$

3. Recall

$$recall = \frac{TP}{TP+FN} \quad (6)$$

3. HASIL DAN PEMBAHASAN

Berdasarkan yang telah dilakukan oleh penulis, pada bagian ini akan dilakukan pengujian eksperimen untuk mencari tahu keakuratan dalam pengklasifikasian yang dilakukan menggunakan algoritma yang ditentukan.

3.1 Data Set

Metode pengumpulan data pada penelitian ini di dapatkan dari SMA Hang Tuah 1 Jakarta dengan format data file csv. Data yang di ambil merupakan data nilai rata-rata raport pada tahun 2018-2022, dengan jumlah data yang di dapatkan 941 data dan memiliki 5 di antaranya, nama, semester 1, semester 2, semester 3, semester 4, semester nilai, total rata-rata, dan *snmptn lulus/tidak lulus*. Seluruh informasi tersebut disajikan pada tabel 2.

Tabel 2. Keterangan Data yang diterima

Atribut	Tipe	Keterangan
No	<i>int64</i>	Nomor data
Nama	<i>object</i>	Nama siswa
Semester 1	<i>float64</i>	Nilai rata-rata semester 1
Semester 2	<i>float64</i>	Nilai rata-rata semester 2
Semester 3	<i>float64</i>	Nilai rata-rata semester 3
Semester 4	<i>Float64</i>	Nilai rata-rata semester 4
Semester 5	<i>float64</i>	Nilai rata-rata semester 5

Total Rata-Rata	<i>float64</i>	Total rata-rata seluruh nilai tiap semester
SNMPTN	<i>object</i>	Hasil lulus dan tidak lulus

3.2 Pengujian Algoritma C4.5 (Decision Tree)

Untuk pengujian *algoritma C4.5* yang telah dilakukan, penulis mendapatkan beberapa informasi yang disajikan diantaranya, gambaran pohon keputusan, serta hasil evaluasi mengenai penggunaan metode *split validasi* dan *cross validasi*.

1. Decisoion Tree (Pohon Keputusan)

Pada Gambar 2. Terdapat gambaran pohon keputusan yang dibuat untuk mendapatkan bukti dalam pengujian apakah sesuai pada hasil atau tidaknya.



Gambar 2. Visualisasi Hasil Pohon Keputusan

Hasil dari grafik visualisasi yang ada diatas, menunjukkan hasil dari data nilai dari Semester 1 hingga Semester 5 pada tahun 2018-2021, grafik diatas adalah sebuah grafik pohon keputusan yang khusus hanya dimiliki oleh algoritma decision tree baik untuk kebutuhan klasifikasi ataupun prediksi.

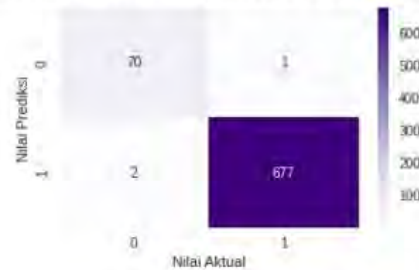
2. Hasil *Akurasi*, *Precision*, dan *Recall* menggunakan metode *split validation* pada data train serta data test yang disajikan pada tabel 3.

Eksperimen	Akurasi	Precision	Recall
Train	99,6%	99,85%	99,70%
Test	94,14%	93,60%	99,08%

Tabel 3. Hasil Split Validation pada Algoritma C4.5

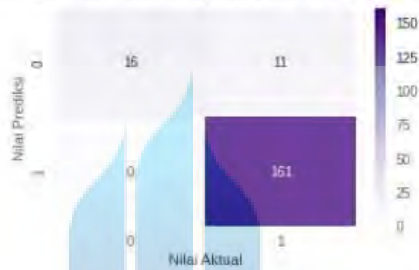
3. Gambar 3 dan 4 memberi informasi mengenai *Confusion Matrix Algoritma C4.5* menggunakan percobaan *split validation*

Nilai Perbandingan Hasil Klasifikasi pada Data Training di Algoritma C4.5



Gambar 3. Confusion Matrix Data Training C4.5

Nilai Perbandingan Hasil Klasifikasi pada Data Testing di Algoritma C4.5



Gambar 4. Confusion Matrix Data Test C4.5

4. Hasil *Accuracy* menggunakan *cross validation* yang disajikan pada tabel 4.

Experiment	Accuracy				
	5-fold	10-fold	15-fold	20-fold	25-fold
Train	97,68%	98,75%	98,74%	98,76%	99,11%
Test	96,54%	96,80%	96,27%	97,59%	97,07%

Tabel 4. Hasil Cross Validation pada Algoritma C4.5

3.3 Pengujian Algoritma Naives Bayes

Selanjutnya untuk pengujian algoritma Naive Bayes, sama seperti pengujian algoritma sebelumnya, penulis mendapatkan informasi hasil evaluasi dari algoritma.

1. Hasil Akurasi, Precision, dan Recall menggunakan metode split validation pada data train serta data test disajikan pada tabel 5.

Eksperimen	Akurasi	Precision	Recall
Train	96,93%	99,69%	96,91%
Test	98,40%	98,17%	99,80%

Tabel 5. Hasil Split Validation pada Algoritma Naive Bayes

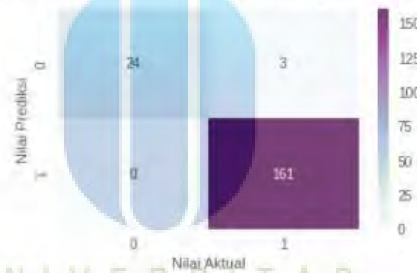
2. Gambar 5 dan 6 memberi informasi mengenai *Confusion Matrix* Algoritma Naive Bayes menggunakan percobaan *split validation*

Nilai Perbandingan Hasil Klasifikasi pada Data Training di Algoritma Naive Bayes



Gambar 5. Confusion Matrix Data Training NB

Nilai Perbandingan Hasil Klasifikasi pada Data Testing di Algoritma Naive Bayes



Gambar 6. Confusion Matrix Data Test NB

3. Hasil *Accuracy* menggunakan *cross validation* disajikan pada tabel 6

Experiment	Accuracy				
	5-fold	10-fold	15-fold	20-fold	25-fold
Train	96,26%	95,91%	96,26%	96,26%	96,28%
Test	98,67%	98,67%	98,40%	98,14%	98,13%

Tabel 6. Hasil Cross Validation Algoritma Naive Bayes

Dari hasil penelitian yang telah dilakukan dapat dilihat bahwa penggunaan *Algoritma C4.5* memiliki akurasi yang baik untuk metode *split validasi*, dibandingkan *Algoritma Naive Bayes*. Untuk akurasi tertinggi yang didapat pada *Algoritma C4.5* senilai 94,14% dan akurasi tertinggi yang didapat pada *algoritma Naive Bayes* senilai 98,40%.

Tetapi berbanding terbalik jika dilihat dalam metode *cross validasi*, yang dimana *Naive Bayes* memiliki akurasi yang lebih baik dibandingkan dengan *Algoritma C4.5*. Yang dimana akurasi yang dimiliki *algoritma naive bayes* tertinggi senilai 98,67% dan akurasi tertinggi yang diperoleh *algoritma C4.5* ialah 97,59%. Setiap algoritma punya kelebihan dan kekurangan masing-masing, pada *algoritma C4.5* sendiri mempunyai konsep yang cukup jelas yaitu, pohon keputusan (*Decision Tree*) yang dihasilkan dapat memberikan penjelasan model algoritma klasifikasi data. *Algoritma C4.5* sendiri bersifat mengeksplorasi data yang diterima dan berusaha mendapatkan kaitan yang tersembunyi atau tidak terbaca pada sebuah variabel data. [13] Sedangkan *Algoritma Naive Bayes* memiliki teknik prediksi berlandas probabilitas berbentuk sederhana dengan asumsi ketidaktergantungan (*independent*) dengan variabel pada sebuah data. [14]

Setelah melakukan melakukan uji coba yang telah dilakukan, dapat diketahui beberapa perbedaan yang memberikan informasi dari penelitian algoritma yang digunakan, yang dijelaskan pada tabel 7.

No.	<i>Algoritma C4.5</i>	<i>Algoritma Naive Bayes</i>
1.	Memiliki pohon keputusan yang memberikan hasil dari pengujian data yang dilakukan.	Tidak memiliki suatu metode ataupun hasil seperti <i>algoritma C4.5</i>
2.	Pengujian menggunakan metode <i>split validasi</i> memberikan hasil akurasi yang tinggi	Pengujian metode <i>split validasi</i> memberikan hasil akurasi tinggi tetapi tidak melebihi <i>algoritma C4.5</i>
3.	Metode <i>cross validasi algoritma C4.5</i> memiliki hasil akurasi yang menurun	Metode <i>cross validasi</i> pada <i>algoritma naive bayes</i> memiliki akurasi tinggi dibandingkan menggunakan metode sebelumnya
4.	Pengimplentasian algoritma cukup sulit dan memiliki waktu pengujian lebih lama	Pengimplementasian algoritma cukup dapat dipahami, karena cukup banyak referensi yang memukakan algoritma ini

Tabel 7. Perbandingan Hasil Klasifikasi Algoritma C4.5 dan Naive Bayes

4. KESIMPULAN

Kesimpulan dari penelitian ini adalah dengan membandingkan dua algoritma pada mengklasifikasi data mengenai penentuan siswa dalam kelulusan *SNMPTN* pada Sekolah SMA Hang Tuah 1 Jakarta memiliki jumlah data sebesar 941 data yang menghasilkan kinerja kedua algoritma dengan cukup baik. Pada penelitian memiliki akurasi yang cukup tinggi, mungkin dikarenakan beberapa faktor, dapat berasal dari pemisahan pada data training maupun data test, ataupun berdasarkan visualisasi pada confusion matrix sendiri yang dapat diketahui apakah algoritma yang didapat tinggi atau tidaknya.

Lalu untuk *Algoritma C4.5* memberikan hasil akurasi tertinggi sebesar 94,14% pada pengujian menggunakan metode *Split Validation* dan akurasi tertinggi sebesar 97,07% pada

pengujian metode *Cross Validation*. Algoritma *Naive Bayes* memberikan hasil akurasi tertinggi sebesar 98,40% pada pengujian menggunakan metode *Split Validation* dan akurasi tertinggi sebesar 98,67% pada pengujian metode *Cross Validation*. Setiap algoritma memiliki Penelitian ini juga dapat kita buktikan jika dalam perhitungan nyata mengenai tentang kenaikan kelas pada suatu sekolah.

5. SARAN

Saran terhadap penelitian ini, masih perlu dilakukannya uji coba seperti penggunaan algoritma lain, serta rancangan data yang didapat harus mendapatkan informasi yang lebih lanjut.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Universitas Mercu Buana yang telah memberi dukungan terhadap penelitian ini.

DAFTAR PUSTAKA

- [1] A. Setyanto dan Hanif Al Fattah, "Analisis Perbandingan Algoritma Decision Tree (C4.5) Dan K-Naive Bayes Untuk Mengklasifikasi Penerimaan Mahasiswa Baru Tingkat Universitas," 2017.
- [2] F. Rini, N. Kahar, and Juliana, "Penerapan Algoritma K-Means Pada Pengelompokan Data Siswa Baru Berdasarkan Jurusan Di Smk Negeri 1 Kota Jambi Berbasis Web," *Semin. Nas. APTIKOM*, pp. 94–99, 2016.
- [3] D. Sartika and D. I. Sensus, "Perbandingan Algoritma Klasifikasi Naive Bayes, Nearest Neighbour, dan Decision Tree pada Studi Kasus Pengambilan Keputusan Pemilihan Pola Pakaian," 2017.
- [4] W. D. Septiani, "Komparasi Metode Klasifikasi Data Mining Algoritma C4.5 Dan Naive Bayes Untuk Prediksi Penyakit Hepatitis," *None*, vol. 13, no. 1, pp. 76–84, 2017.
- [5] M. Sadikin and F. Alfiandi, "Comparative Study of Classification Method on Customer Candidate Data to Predict its Potential Risk," *Int. J. Electr. Comput. Eng.*, vol. 8, no. 6, p. 4763, 2018, doi: 10.11591/ijece.v8i6.pp4763-4771.
- [6] N. Khotimah and D. Istiawan, "Perbandingan Algoritma C4.5, Naive Bayes dan K-Nearest Neighbour untuk Prediksi Lahan Kritis di Kabupaten Pematang," *Urecol*, vol. 7, no. 1, pp. 41–50, 2018.
- [7] I. Yulianti, "Analisis Komparasi Klasifikasi Algoritma C4.5 Dan Naive Bayes Pada Prediksi Keberhasilan Software Reuse," *Swabumi*, vol. 7, no. 1, pp. 62–66, 2019, doi: 10.31294/swabumi.v7i1.5637.
- [8] S. Bahri, D. Marisa Midyanti, R. Hidayati, J. Sistem Komputer, and F. Mipa, "Perbandingan Algoritma Naive Bayes dan C4.5 Untuk Klasifikasi Penyakit Anak," 2018.
- [9] S. Dewi, "Komparasi 5 Metode Algoritma Klasifikasi Data Mining Pada Prediksi Keberhasilan Pemasaran Produk Layanan Perbankan," *Techno Nusa Mandiri*, vol. 13, no. 1, pp. 60–66, 2016.
- [10] C. Anam and H. B. Santoso, "Perbandingan Kinerja Algoritma C4.5 dan Naive Bayes untuk Klasifikasi Penerima Beasiswa," 2018.
- [11] M. Sadikin, M. I. Fanany, and T. Basaruddin, "A New Data Representation Based on Training Data Characteristics to Extract Drug Name Entity in Medical Text," *Comput. Intell. Neurosci.*, vol. 2016, 2016, doi: 10.1155/2016/3483528.

-
- [12] F. P. B. Muhamad, D. O. Siahaan, and C. Fatichah, "Software Fault Prediction Using Filtering Feature Selection in Cluster-Based Classification," *IPTEK J. Proc. Ser.*, vol. 4, no. 1, p. 59, 2018, doi: 10.12962/j23546026.y2018i1.3508.
- [13] M. F. Arifin and D. Fitriana, "Penerapan Algoritma Klasifikasi C4.5 Dalam Rekomendasi Penerimaan Mitra Penjualan Studi Kasus : PT Atria Artha Persada," *InComTech*, vol. 8, no. 2, pp. 87–102, 2018, doi: 10.22441/incomtech.v8i1.2198.
- [14] M. F. Fibrianda and A. Bhawiyuga, "Analisis Perbandingan Akurasi Deteksi Serangan Pada Jaringan Komputer Dengan Metode Naïve Bayes Dan Support Vector Machine (SVM)," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 11, no. 9, pp. 3112–3123, 2018.



KERTAS KERJA

Ringkasan

Kertas kerja ini merupakan material kelengkapan artikel jurnal yang telah terlampir sebelumnya dengan judul “Perbandingan Algoritma C4.5 dan Naïve Bayes Dalam Penentuan Kelulusan Siswa ke PTN Melalui Jalur SNMPTN (Studi Kasus : SMA Hang Tuah 1 Jakarta)”. Kertas kerja ini berisi semua material hasil penelitian Tugas Akhir. Di dalam kertas kerja ini disajikan beberapa bagian yang terdiri dari literature review, dataset yang digunakan, tahapan eksperimen, dan hasil eksperimen secara keseluruhan.

Bagian I membahas mengenai literature review yang berisi artikel jurnal sebelumnya yang menjadi dasar atau landasan dalam penelitian ini. Bagian II menjelaskan tentang source code yang digunakan pada penelitian ini. Bagian III menjelaskan mengenai dataset yang digunakan. Bagian IV memuat tahapan eksperimen yang disajikan dalam gambar beserta penjelasan dari tiap tahapan. Bagian V merupakan bagian terakhir dari kertas kerja ini yang menjelaskan hasil keseluruhan dari eksperimen yang telah dilakukan, meliputi penjelasannya.

UNIVERSITAS
MERCU BUANA