



**Perbandingan, Algoritma C.45, Naïve Bayes dan K-Nearest Neighbor untuk  
Klasifikasi Penerima Bantuan Sosial Tunai (BST)  
(Studi Kasus : Kecamatan Tebet Jakarta Selatan)**

*TUGAS AKHIR*

Qonita Azillatin  
41518010088

UNIVERSITAS  
MERCU BUANA

**PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS ILMU KOMPUTER  
UNIVERSITAS MERCU BUANA  
JAKARTA  
2021**



**Perbandingan Algoritma C.45, Naïve Bayes dan K-Nearest Neighbor untuk  
Klasifikasi Penerima Bantuan Sosial Tunai (BST)  
(Studi Kasus : Kecamatan Tebet Jakarta Selatan)**

*Tugas Akhir*

Diajukan Untuk Melengkapi Salah Satu Syarat  
Memperoleh Gelar Sarjana Komputer

Oleh:  
Qonita Azillatin  
41518010088

UNIVERSITAS  
PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS ILMU KOMPUTER  
UNIVERSITAS MERCU BUANA  
JAKARTA  
2021

## LEMBAR PERNYATAAN ORISINALITAS

Yang bertanda tangan dibawah ini:

NIM : 41518010088

Nama : Qonita Azillatin

Judul Tugas Akhir : Perbandingan Algoritma C.45, Naïve Bayes dan K-Nearest Neighbor untuk Klasifikasi Penerima Bantuan Sosial Tunai (BST) (Studi Kasus : Kecamatan Tebet Jakarta Selatan)

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, 20 Januari 2022



Qonita Azillatin

UNIVERSITAS  
MERCU BUANA

## SURAT PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR

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

Nama Mahasiswa : Qonita Azillatin  
NIM : 41518010088  
Judul Tugas Akhir : Perbandingan, Algoritma C.45, Naïve Bayes dan K-Nearest Neighbor untuk Klasifikasi Penerima Bantuan Sosial Tunai (BST)  
(Studi Kasus : Kecamatan Tebet Jakarta Selatan)

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, 20 Januari 2022

UNIVERSITAS  
MERCU BUANA



Qonita Azillatin

### SURAT PERNYATAAN LUARAN TUGAS AKHIR

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

Nama Mahasiswa : Qonita Azillatin  
 NIM : 41518010088  
 Judul Tugas Akhir : Perbandingan Algoritma C.45, Naïve Bayes dan K-Nearest Neighbor untuk Klasifikasi Penerima Bantuan Sosial Tunai (BST)  
 (Studi Kasus : Kecamatan Tebet Jakarta Selatan)

Menyatakan bahwa :

1. Luaran Tugas Akhir saya adalah sebagai berikut :

No	Luaran	Jenis	Status
1	Publikasi Ilmiah	Jurnal Nasional Tidak Terakreditasi	
		Jurnal Nasional Terakreditasi	✓
		Jurnal International Tidak Bereputasi	
		Jurnal International Bereputasi	Diterima
Disubmit/dipublikasikan di :		Nama Jurnal : Jurnal Ilmu Teknik Dan Komputer	
		ISSN : 2621-1491	
		Link Jurnal : <a href="https://publikasi.mercubuana.ac.id/index.php/jitkom">https://publikasi.mercubuana.ac.id/index.php/jitkom</a>	
		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

Vina Ayumi S. Kom., M. Kom.

Jakarta, 20 Januari 2022



Qonita Azillatin

## LEMBAR PERSETUJUAN

Nama Mahasiswa : Qonita Azillatin  
NIM : 41518010088  
Judul Tugas Akhir : Perbandingan Algoritma C.45, Naïve Bayes dan K-Nearest Neighbor untuk Klasifikasi Penerima Bantuan Sosial Tunai (BST)  
(Studi Kasus : Kecamatan Tebet Jakarta Selatan)

Tugas Akhir ini telah diperiksa dan disetujui

Jakarta, 20 Januari 2022

Menyetujui,

  
A blue circular stamp with a white signature is placed over the text of the supervisor's name.

(Vina Ayumi S.Kom.,M.Kom)  
Dosen Pembimbing

UNIVERSITAS  
MERCU BUANA

## LEMBAR PERSETUJUAN PENGUJI

NIM : 41518010088  
Nama : Qonita Azillatin  
Judul Tugas Akhir : Perbandingan, Algoritma C.45, Naïve Bayes dan K-Nearest Neighbor untuk Klasifikasi Penerima Bantuan Sosial Tunai (BST)  
(Studi Kasus : Kecamatan Tebet Jakarta Selatan)

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, 20 Januari 2022



(Sabar Rudiarto, M.Kom)

Dosen Penguji 1

UNIVERSITAS  
MERCU BUANA

## LEMBAR PERSETUJUAN PENGUJI

NIM : 41518010088  
Nama : Qonita Azillatin  
Judul Tugas Akhir : Perbandingan, Algoritma C.45, Naïve Bayes dan K-Nearest Neighbor untuk Klasifikasi Penerima Bantuan Sosial Tunai (BST)  
(Studi Kasus : Kecamatan Tebet Jakarta Selatan)

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, 20 Januari 2022



(Saruni Dwiasnati, ST, MM, M.Kom)  
Dosen Penguji 2

UNIVERSITAS  
MERCU BUANA



## LEMBAR PERSETUJUAN PENGUJI

NIM : 41518010088  
Nama : Qonita Azillatin  
Judul Tugas Akhir : Perbandingan, Algoritma C.45, Naïve Bayes dan K-Nearest Neighbor untuk Klasifikasi Penerima Bantuan Sosial Tunai (BST)  
(Studi Kasus : Kecamatan Tebet Jakarta Selatan)

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, 20 Januari 2022



UNIVERSITAS  
(Afiyati, S.Si, M.Kom)  
Dosen Penguji 3  
MERCU BUANA

## LEMBAR PENGESAHAN

### LEMBAR PENGESAHAN

NIM : 41518010088  
Nama : Qonita Azillatin  
Judul Tugas Akhir : Perbandingan Algoritma C 45, Naive Bayes dan K-Nearest Neighbor untuk Klasifikasi Penerima Bantuan Sosial Tunai (BST)  
(Studi Kasus : Kecamatan Tebet Jakarta Selatan)

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, Des 2021

Menyetujui,



(Vina Ayumi S. Kom, M. Kom)  
Dosen Pembimbing

UNIVERSITAS

MERCU BUANA

Mengetahui,



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



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

## KATA PENGANTAR

Segala puji dan syukur kehadirat Allah SWT atas berkah, rahamat dan hidayah-Nya yang senantiasa dilimpahkan kepada penulis, sehingga bisa menyelesaikan skripsi dengan judul “*Perbandingan Algoritma C.45, Naïve Bayes dan K-Nearest Neighbor untuk Klasifikasi Penerima Bantuan Sosial Tunai (BST) (Studi Kasus : Kecamatan Tebet Jakarta Selatan)*” 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 maupu spiritual. Untuk itu pada kesempatan ini penulis menyampaikan ucapan terimakasih kepada:

1. Kedua orang tua yang telah memberika support serta kasih sayang hingga saya bisa menyelesaikan tugas akhir ini
2. Ibu Vina Ayumi S,Kom.,M.Kom 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 Studi 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. Bapak Drs. Achmad Kodar, MT selaku dosen pembimbing akademik saya yang telah membimbing selama perkuliahan hingga sampai pada tugas akhir.
6. Jessica Putri Sabrina S.Kom 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
8. Senior atas berbagai saran yang diberikan dan junior yang dengan semangat mendukung dikala bertemu.

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 Akhir kata, penulis berharap jika ada kritik atau saran apapun yang sifatnya membangun bagi saya dengan senang hati akan saya terima.

Jakarta, Des 2021  
Penulis



## DAFTAR ISI

HALAMAN SAMPUL.....	i
HALAMAN JUDUL .....	i
LEMBAR PERNYATAAN ORISINALITAS .....	ii
SURAT PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR... iii	
LEMBAR PERSETUJUAN .....	v
LEMBAR PERSETUJUAN PENGUJI .....	vi
LEMBAR PENGESAHAN .....	ix
ABSTRAK .....	x
ABSTRACT .....	xi
KATA PENGANTAR.....	xii
DAFTAR ISI.....	xiv
NASKAH JURNAL .....	1
KERTAS KERJA.....	13
BAB 1. LITERATUR REVIEW.....	14
BAB 2. SOURCE CODE .....	20
BAB 3 DATASET.....	29
BAB 4. TAHAPAN EKSPERIMEN.....	33
BAB 5 HASIL SEMUA EKSPERIMEN.....	37
DAFTAR PUSTAKA .....	44
LAMPIRAN KORESPONDENSI .....	46

## NASKAH JURNAL

Jurnal Ilmu Teknik dan Komputer  
Vol. X No. X Jan 2017

### Perbandingan Algoritma C.45, Algoritma Naïve Bayes dan Algoritma K-Nearest Neighbor untuk Klasifikasi Penerima Bantuan Sosial Tunai (BST) (Studi Kasus : Kecamatan Tebet Jakarta Selatan)

Qonita Azillatin  
Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana  
Jl. Raya Meruya Selatan, Kembangan, Jakarta, Indonesia 11650  
qonitaazillatin23@gmail.com

#### ABSTRACT

*Bantuan Sosial Tunai (BST) merupakan salah satu bantuan dari pemerintah berupa uang tunai kepada masyarakat miskin yang terkena dampak dari covid - 19. Penerima BST dilihat dari id basis data terpadu (BDT) yang merupakan kumpulan data orang tidak mampu dan kartu lansia. Dalam penerimaan BST pemerintah juga memberikan kewenangan terhadap Dinas Social untuk menentukan peserta yang berhak menerima BST. Dalam hal ini peneliti mengambil studi kasus dari salah satu kelurahan yang ada di Provinsi DKI Jakarta yaitu Kelurahan Bukit Duri Kecamatan Tebet Jakarta Selatan dimana banyak warga yang kurang mampu yang tidak bisa menerima BST karena tidak ada data di BDT tersebut. Hal ini dikarenakan data tidak di update, penyeleksian data masih secara manual, penyebaran informasi yang tidak merata dan penentuan keputusan yang kurang akurat sehingga peserta penerima BST tidak tepat sasaran. Dari rata - rata accuracy yang ada dapat dilihat bahwa nilai rata rata accuracy algoritma K-NN adalah 79,74%, sedangkan algoritma C.45 nilai rata rata accuracy 76,66% dan untuk algoritma naïve bayes nilai rata - rata accuracy sebesar 75,48%.*

*KATA KUNCI : Klasifikasi, BST, Naïve Bayes, C.45, K-NN.*

#### PENDAHULUAN

Bantuan Sosial Tunai (BST) merupakan salah satu bantuan dari pemerintah berupa uang tunai kepada masyarakat miskin yang terkena dampak dari covid - 19. Seperti yang kita ketahui sekarang Indonesia bahkan seluruh dunia sedang terserang covid - 19 dimana dari seluruh aspek terkena dampak dalam sistem pemerintahan salah satunya dibidang perekonomian. Perekonomian di Indonesia menurun, banyak pekerja yang

diberhentikan dari pekerjaannya. Upaya pemerintah dalam membantu masyarakat yang terkena dampak covid - 19 salah satunya yaitu memberi bantuan berupa uang tunai sebesar Rp. 300.000 tiap bulannya terhitung dari bulan Februari 2021 kepada setiap kepala keluarga yang di distribusikan oleh Ketua Rukun Tetangga (RT).

Penerima BST dilihat dari id basis data terpadu (BDT) yang merupakan kumpulan data orang tidak mampu dan kartu lansia. Dalam penerimaan BST

pemerintah juga memberikan kewenangan terhadap Dinas Sosial untuk menentukan peserta yang berhak menerima BST. Provinsi DKI Jakarta merupakan provinsi yang paling kacau dalam data BDT untuk penerima bantuan sosial menurut Wakil Menteri Desa, Pengembangan Tertinggal dan Transmigrasi (Wamendes PDTT) yaitu Budi Arie Setiadi. Dalam hal ini peneliti mengambil studi kasus dari salah satu kelurahan yang ada di Provinsi DKI Jakarta yaitu Kelurahan Bukit Duri Kecamatan Tebet Jakarta Selatan dimana banyak warga yang kurang mampu yang tidak bisa menerima BST karena tidak ada data di BDT tersebut. Hal ini dikarenakan data tidak di *update*, penyeleksian data masih secara manual, penyebaran informasi yang tidak merata dan penentuan keputusan yang kurang akurat sehingga peserta penerima BST tidak tepat sasaran. Tentu hal ini merugikan berbagai pihak terutama kalangan menengah kebawah yang seharusnya bisa merasakan bantuan BST ini.

*Naive Bayes* merupakan teknik prediksi berbasis probabilistik sederhana yang berdasar pada penerapan teorema *Bayes* (aturan *Bayes*) dengan asumsi independensi (ketidaktergantungan) yang kuat (naif). Dengan kata lain, dalam *Naive Bayes* model yang digunakan adalah model

fitur *independent* (Prasetyo, 2012). (Wijaya & Dwiasnati, 2020) banyak kelebihan *naive bayes* dibandingkan model klasifikasi lainnya seperti dengan menggunakan *naive bayes* Akan lebih efisien dari segi kecepatan ataupun ruang dan untuk pengestimasi parameter hanya memerlukan jumlah yang kecil dalam pengklasifikasiannya (Rasyida, 2020). Algoritma *K-Nearest Neighbor* merupakan algoritma yang memecahkan masalah dengan sederhana dengan hasil yang kompetitif dan signifikan (Nikmatun & Waspada, 2019). Algoritma C4.5 merupakan algoritma *decision tree* terbaik dan yang paling efektif untuk melakukan klasifikasi (Kurniawan, 2018) (Ermawati, 2019).

Dengan keunggulan masing masing algoritma diatas peneliti ingin menggunakan ketiga algoritma diatas untuk diimplementasikan dalam pengklasifikasian data penerima BST. Dan juga penelitian ini akan menentukan algoritma klasifikasi mana yang cocok dalam menentukan peserta penerima BST di Kelurahan Bukit Duri Kecamatan Tebet Jakarta Selatan.

#### Rumusan masalah

Berdasarkan latar belakang diatas, maka rumusan masalah didalam penelitian ini yaitu :

1. Bagaimana implementasi algoritma *naive bayes*, algoritma K-NN, dan algoritma C.45 dalam mengklasifikasikan data di penerima Bantuan Sosial Tunai (BST) ?

2. Algoritma apakah yang cocok dalam menentukan penerimaan Bantuan Sosial Tunai (BST) ?

#### Tujuan dan manfaat

Tujuan Pada Penelitian Ini adalah :

1. Menentukan faktor – faktor yang mempengaruhi dalam menentukan Penerima BST dengan menggunakan Algoritma C.45, algoritma *naive bayes* dan algoritma K- NN.
2. Mengklasifikasikan layak menerima atau tidak layak menerima BST di provinsi DKI Jakarta.
3. Menentukan Algoritma yang paling cocok dalam penentuan penerima BST.

Manfaat penelitian ini adalah :

1. Memperoleh informasi mengenai algoritma yang cocok dalam penentuan penerima BST.

#### STUDI LITERATUR

##### 1. Klasifikasi

Klasifikasi adalah suatu teknik pembentukan model dari data yang belum

terklasifikasi, untuk digunakan mengklasifikasi data baru. Klasifikasi termasuk ke dalam tipe *supervised learning*, artinya dibutuhkan data pelatihan untuk membangun suatu model klasifikasinya (Ermawati, 2019). Klasifikasi merupakan topik yang penting dalam penelitian data mining. Dimana terdiri dari sekumpulan data yang masing masing sudah dikelompokkan kedalam kelas tertentu, masalah klasifikasi yang diperhatikan yaitu dengan penentuan aturan yang memungkinkan adanya klasifikasi walaupun data yang ada belum menunjukkan hal tersebut (Ermawati, 2019).

##### 2. Algoritma C.45

Algoritma C4.5 adalah algoritma pohon keputusan yang terkenal. Algoritma C4.5 pertama sekali diperkenalkan oleh Quinlan (1996) sebagai versi perbaikan ID3 (M. H. Rifqo, 2016). Algoritma C4.5 digunakan untuk melakukan klasifikasi atau pengelompokan yang bersifat prediktif. Pernyataan klasifikasi pada pohon keputusan (*decision tree*) terdapat pada cabang-cabangnya dan kelas-kelas atau segmen-segmennya terdapat pada daun-daunnya.

Ada beberapa tahap dalam membuat sebuah pohon keputusan dengan algoritma C4.5 (M. H. Rifqo, 2016).



1. Menyiapkan *data training*. *Data training* biasanya diambil dari data histori yang pernah terjadi sebelumnya dan sudah dikelompokkan ke dalam kelas-kelas tertentu.
2. Menentukan akar dari pohon. Akar akan diambil dari atribut yang terpilih, dengan cara menghitung nilai *gain* dari masing-masing atribut, nilai *gain* yang paling tinggi yang akan menjadi akar pertama. Sebelum menghitung nilai *gain* dari atribut, hitung dahulu nilai *entropy*. Untuk menghitung nilai *entropy* digunakan rumus:

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

Keterangan:

S = himpunan kasus

n = jumlah partisi S

pi = proporsi Si terhadap S

3. Ulangi langkah ke-2 hingga semua *record* terpartisi.
4. Proses partisi pohon keputusan akan berhenti saat:
  - 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. Algoritma Naïve Bayes

Menurut teorema Bayes, Naïve Bayes adalah Algoritma yang menggabungkan *prior probability* dengan probabilitas bersyarat pada rumus digunakan untuk menghitung kemungkinan tiap klasifikasi pada probabilitas (Septiani, 2017).

Perhitungan pada algoritma naive bayes adalah sebagai berikut :

$$P(x|y) = \frac{P(y|x)P(x)}{P(y)}$$

Keterangan:

y = data dengan kelas yang belum diketahui.

x = hipotesis data y merupakan suatu kelas spesifik.

P(x|y) = probabilitas hipotesis x berdasarkan kondisi y.

P(x) = probabilitas hipotesis x.

P(y|x) = probabilitas y berdasarkan kondisi pada hipotesis x.

P(y) = probabilitas dari y.

## 3. Algoritma K- Nearest Neighbor

Algoritma KNN merupakan metode klasifikasi dengan mencari jarak terdekat antara data yang akan dievaluasi dengan K tetangga (*neighbor*) terdekatnya dalam data pelatihan (Hasanah et al., 2019). Fitur – fitur yang

ada sangat mempengaruhi algoritma ini serta hubungan antara bobot yang ada di fitur ini dengan klasifikasi yang tidak setara (Dewi, 2016).

Model persamaan algoritma KNN dibawah ini (Bode, 2017):

$$D = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2} \quad (1)$$

Keterangan:

$x$  = sampel data

$y$  = data uji

$D$  = jarak

$$f_{knn}(x') = \frac{1}{K} \sum_{i \in N} y_i k(x') \quad (2)$$

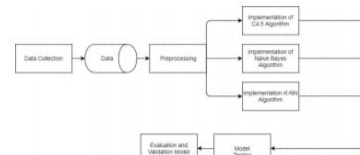
$x'$  = Perkiraan atau estimasi

$K$  = Jumlah tetangga terdekat

$Nk(x')$  = Tetangga terdekat

$y_i$  = Output tetangga terdekat

## METODOLOGI



Gambar 1. Diagram Metodologi

### 1. Data Collection

Pada *data collection* diambil dari Basis Data Terpadu (BDT) yang ada di kelurahan bukit duri. BDT ini juga merupakan acuan untuk penerima bantuan sosial dari pemerintah. Data yang digunakan diambil dari tahun 2021 dengan total data sebanyak 8.958 data. Untuk variabel pendapatan peserta diambil dari Data Lapangan Usaha dan Data Pendapatan Bersih Golongan Pekerjaan yang diberikan oleh Data Badan Pusat Statistika (BPS) Tahun 2018 karena merupakan penentuan penerima bantuan sosial tunai.

### 2. Preprocessing

- a. *Data Cleaning* : Penelitian ini menggunakan data sebanyak 8.958 data. Pada proses *cleansing* data peneliti menghapus seluruh data di baris yang ada 1 kolom yang kosong. Pada tahap ini data yang dihasilkan sebanyak 1.499 data.
- b. *Data Reduction* : Data yang ada terdapat 14 variabel, namun peneliti tidak menggunakan keseluruhan variabel, variabel yang digunakan hanya variabel yang sudah disesuaikan berdasarkan hasil wawancara dan juga melihat data

dari kriteria masyarakat miskin sehingga variabel yang digunakan hanya 8 variabel saja untuk digunakan pada pengimplementasian algoritma. Adapun variabel yang digunakan yaitu Jumlah\_ART, Status\_bangunan, Umur, Pendidikan\_tertinggi, Sta\_Bekerja, Gol\_Pendapatan, Status\_pekerjaan, ada\_BST.

- d. Data Transformasi : Dalam proses ini, penulis melakukan transformasi data sebanyak 2 kali, pertama merubah variabel pendidikan\_tertinggi lalu merubah variabel Gol\_Pendapatan. Setelah itu seluruh variabel diubah menjadi Integer.

### 3. Implementasi algoritma C.45, Naïve Bayes, dan K-NN

Setelah semua data telah di proses dan mendapatkan data bersih maka selanjutnya data akan di implementasikan menggunakan 3 algoritma yaitu algoritma C.45, Algoritma *Naïve Bayes*, dan Algoritma K-NN. Dalam implementasinya nilai  $x$  dan  $y$  ditentukan yaitu variabel *independent* dan variabel *dependent*. Untuk nilai  $x$  variabel yang digunakan adalah Jumlah\_ART, Status\_bangunan, Umur, Pendidikan\_tertinggi, Sta\_Bekerja, Gol\_Pendapatan, Status\_pekerjaan, sedangkan untuk nilai

$y$  menggunakan variabel ada\_BST yang merupakan label. Adapun tahapan implementasi algoritma yang akan digunakan adalah sebagai berikut :

#### a. Algoritma C4.5

Perhitungan yang dilakukan dalam algoritma ini adalah nilai *entropy*, *gain* dan *split* dari masing-masing atribut pada data training yang ada sehingga menghasilkan nilai *gain ratio*. Atribut yang memiliki *gain ratio* terbesar dipilih untuk membuat simpul akar pohon. Lalu menghitung nilai *entropy*, *gain* dan *split* dari masing-masing atribut dengan menghilangkan atribut yang telah dipilih sebelumnya. Atribut yang memiliki nilai *gain ratio* terbesar dipilih untuk membuat simpul pohon. Ulangi perhitungan tersebut hingga semua atribut memiliki kelas. Jika semua atribut telah memiliki kelas maka tampilkan pohon keputusan awal dan generate aturan keputusan awal.

#### b. Algoritma Naive Bayes

Proses yang dilakukan dalam algoritma *naive bayes* adalah data yang dimasukkan akan dihitung label atau kelasnya. Kemudian dari kelas tersebut akan dihitung probabilitas dari setiap kelas. Setelah didapat maka hasil kelas akan dibandingkan. Jika nilai probabilitas dari setiap kelas pada label

ya lebih besar dari tidak maka data tersebut digolongkan ke dalam penerima BST dan jika sebaliknya, maka data tersebut digolongkan bukan penerima BST.

c. Algoritma K-Nearest Neighbor

Proses yang dilakukan dalam algoritma *K-Nearest Neighbor* adalah data set yang digunakan akan dijadikan sebagai data pembelajaran dimana terdapat label ya atau tidak. Lalu tentukan jumlah tetangga terdekat. Setelah menentukan K maka data testing akan dihitung jaraknya terhadap data training tersebut. Lalu dilihat jarak paling dekat dengan K yang telah ditentukan, lalu kelompokkan data testing tersebut berdasarkan label mayoritas pada K.

#### 4. Model Testing

Model akan diuji pada data yang telah diproses melalui metode *Split Validation*.

a. Split Validation

*Split Validation* adalah teknik validasi yang membagi data menjadi dua bagian secara acak, sebagian sebagai data training dan sebagian lainnya sebagai data testing. Pada penelitian ini pembagian persentase dibagi menjadi 5 (lima) skenario eksperimen dengan pembagian data sebagai berikut :

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

Tabel 1. Presentase *split validation*

#### 5. Evaluation and Validation Model

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

1. Accuracy

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

2. Precision

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

3. Recall

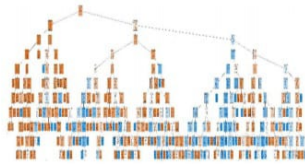
$$recall = \frac{TP}{TP + FN}$$

#### HASIL DAN DISKUSI

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

**a. Algoritma C.45**

1. Pohon keputusan



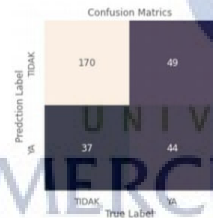
Gambar 2. Pohon keputusan

2. Hasil Accuracy, Precision and Recall menggunakan split validation

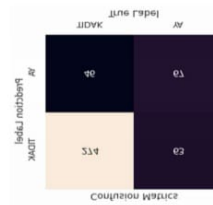
Experi ment	Accurac y	Precisi on	Recal l
1	71,33%	78%	82%
2	80,66%	84%	90%
3	75,77%	81%	86%
4	77,16%	83%	86%
5	78,4%	85%	86%

Tabel 2. Hasil Accuracy, Precision and Recall C.45

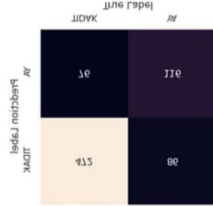
3. Confusion Matrix ke-5 Percobaan



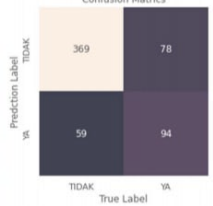
Gambar 3. Matrix Percobaan ke 1



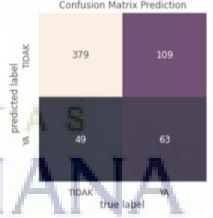
Gambar 4. Matrix Percobaan ke 2



Gambar 5. Matrix Percobaan ke 3



Gambar 6. Matrix Percobaan ke 4



Gambar 7. Matrix Percobaan ke 5

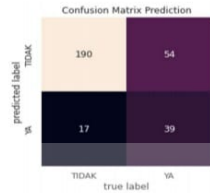
**b. Algoritma Naïve bayes**

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

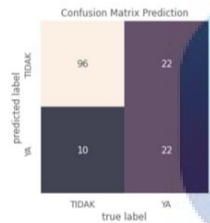
Experiment	Accuracy	Precision	Recall
1	76,33%	78%	92%
2	78,66%	81%	91%
3	75,33%	79%	89%
4	73,66%	78%	89%
5	73,46%	79%	87%

Tabel 3. Hasil *Accuracy, Precision and Recall Naïve Bayes*

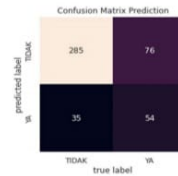
2. *Confusion Matrix* ke-5 Percobaan



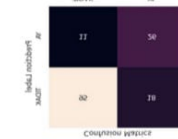
Gambar 8. *Matrix* Percobaan ke 1



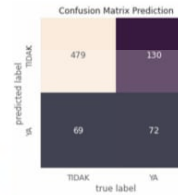
Gambar 9. *Matrix* Percobaan ke 2



Gambar 10. *Matrix* Percobaan ke 3



Gambar 11. *Matrix* Percobaan ke 4



Gambar 12. *Matrix* Percobaan ke 5

**c. Algoritma K- Nearest Neighbor (K-NN)**

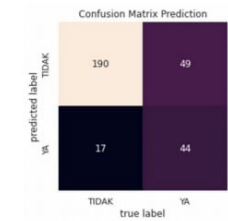
1. Hasil *Accuracy, Precision and Recall* menggunakan *split validation*

Experiment ke	K	Accuracy	Precision	recall
1	9	78%	79%	92%
2	3	82,66%	82%	97%
3	9	78,88%	81%	92%
4	9	79,16%	81%	92%
5	9	80,00%	82%	93%

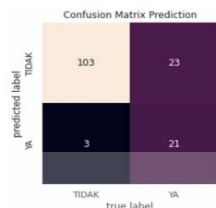
Tabel 4. Hasil *Accuracy, Precision and Recall*



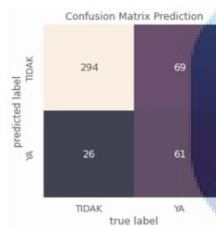
2. Confusion Matrix ke-5 Percobaan



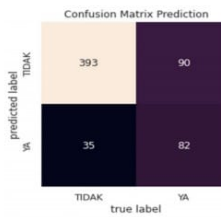
Gambar 13. Matrix Percobaan ke 1



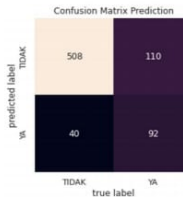
Gambar 14. Matrix Percobaan ke 2



Gambar 15. Matrix Percobaan ke 3



Gambar 16. Matrix Percobaan ke 4



Gambar 17. Matrix Percobaan ke 5

Rata-rata akurasi masing-masing algoritma disajikan dalam Gambar di bawah ini.

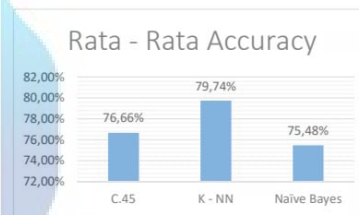


Diagram 1. Rata – rata Accuracy

Dari rata – rata *accuracy* yang ada dapat dilihat bahwa nilai rata rata *accuracy* algoritma K-NN adalah 79,74%, sedangkan algoritma C.45 nilai rata rata *accuracy* 76,66% dan untuk algoritma naive bayes nilai rata – rata *accuracy* sebesar 75,48%.



**KESIMPULAN**

Algoritma K-NN nilai rata – rata *accuracy* nya lebih tinggi dibandingkan algoritma lain hal ini karena dari karakteristik data, algoritma K-NN dapat bekerja dengan baik setelah data melalui *preprocessing* dan menghasilkan data nominal. Keterkaitan data satu sama lain membuat performa K-NN meningkat. Hal ini dikarenakan algoritma K-NN yang bersifat *non liner* dan bekerja berdasarkan data *training* yang jaraknya paling dekat dengan *object* tersebut. Selain itu, data yang cukup besar menjadi daya tarik keunggulan yang dimiliki oleh algoritma K-NN.

Perhitungan nilai *gain* dan *ratio* membuat algoritma C4.5 memiliki model yang cukup baik dalam klasifikasi penerima BST ini. Setiap atribut akan menunjukkan tingkat pengaruhnya dalam menentukan suatu *class* dalam sebuah data sehingga setiap variabel bersifat *dependent* satu sama lain. Kinerja algoritma C4.5 yang saling terhubung antar atribut dapat dimanfaatkan dengan baik. Namun, keragaman atribut membuat pembentukan pohon keputusan sedikit kurang maksimal.

Tidak termanfaatkannya probabilitas bersyarat variabel *input* yang dimiliki *naive bayes* membuat performa *naive*

*bayes* di bawah algoritma yang lain. *Naive bayes* akan berjalan baik pada data yang bersifat *independent*. Sedangkan data yang diolah memiliki variabel yang *dependent*.

**DAFTAR PUSTAKA**

- Bode, A. (2017). K-Nearest Neighbor Dengan Feature Selection Menggunakan Backward Elimination Untuk Prediksi Harga Komoditi Kopi Arabika. *ILKOM Jurnal Ilmiah*, 9(2), 188–195.  
<https://doi.org/10.33096/ilkom.v9i2.139.188-195>
- Dewi, S. (2016). Komperasi 5 Metode Algoritma Klasifikasi Data Mining Pada Prediksi Keberhasilan Pemasaran Produk Layanan Perbankan. *Techno Nusa Mandiri*, XIII(1), 60–66.
- Ermawati, E. (2019). Algoritma Klasifikasi C4.5 Berbasis Particle Swarm Optimization Untuk Prediksi Penerima Bantuan Pangan Non Tunai. *Sistemasi*, 8(3), 513.  
<https://doi.org/10.32520/stmsi.v8i3.576>
- Hasanah, R. L., Hasan, M., Pangesti, W. E., Wati, F. F., & Gata, W. (2019). Klasifikasi Penerima Dana Bantuan Desa Menggunakan Metode Knn (K-Nearest Neighbor). *Jurnal Techno Nusa Mandiri*, 16(1), 1–6.  
<https://doi.org/10.33480/techno.v16i1.25>
- Kurniawan, Y. I. (2018). Perbandingan Algoritma Naive Bayes dan C.45 dalam Klasifikasi Data Mining. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 5(4), 455.  
<https://doi.org/10.25126/jtiik.201854803>
- M. H. Rifqo. (2016). *BANK BRI CABANG CURUP*. 83–90.
- Nikmatun, I. A., & Waspada, I. (2019). Implementasi Data Mining untuk Klasifikasi Masa Studi Mahasiswa Menggunakan Algoritma K-Nearest Neighbor. *Jurnal SIMETRIS*, 10(2), 421–432.
- Rasyida, M. (2020). Naive Bayes Classification untuk Penentuan Status Penduduk Miskin. *Jurnal Informatika Kaputama*, 4(2), 175–180.
- Septiani, W. D. (2017). Komparasi Metode Klasifikasi Data Mining Algoritma C4.5 Dan



---

Jurnal Ilmu Teknik dan Komputer  
Vol. X No. X Jan 2017

Naïve Bayes Untuk Prediksi Penyakit  
Hepatitis. *None*, 13(1), 76–84.

<https://doi.org/10.33480/pilar.v13i1.149>

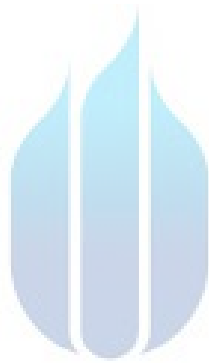
Wijaya, H. D., & Dwiasnati, S. (2020).

Implementasi Data Mining dengan Algoritma

Naïve Bayes pada Penjualan Obat. *Jurnal*

*Informatika*, 7(1), 1–7.

<https://doi.org/10.31311/ji.v7i1.6203>



UNIVERSITAS  
MERCU BUANA

## KERTAS KERJA

### Ringkasan

Kertas kerja ini merupakan material kelengkapan artikel jurnal dengan judul “Perbandingan Algoritma C.45, Naïve Bayes dan K-Nearest Neighbor untuk Klasifikasi Penerima Bantuan Sosial Tunai (BST) (Studi Kasus : Kecamatan Tebet Jakarta Selatan)”. Kertas kerja berisi semua material hasil penelitian Tugas Akhir yang tidak dimuat/atau disertakan di artikel jurnal. Di dalam kertas kerja ini disajikan: literature review, dataset yang digunakan, source code, dan hasil eksperimen secara keseluruhan.

