

Implementasi Metode Naïve Bayes untuk Memperbaiki Instance Noise (missing value) pada Data Mining

*Diajukan Sebagai Syarat untuk Menyelesaikan
Pendidikan Program Strata-1
Di Jurusan Teknik Informatika*



Oleh :

PUTRI SEPTRIA
NIM: 09121002005

**JURUSAN TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS SRIWIJAYA
2019**

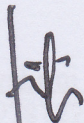
LEMBAR PENGESAHAN TUGAS AKHIR

**IMPLEMENTASI METODE NAÏVE BAYES UNTUK
MEMPERBAIKI INSTANCE NOISE (MISSING VALUE)
PADA DATA MINING**

Oleh :

Putri Septria
09121002005

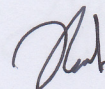
Pembimbing I,



Rifkie Primartha, M.T.

NIP. 197706012009121004

Palembang, Juli 2019
Pembimbing II,

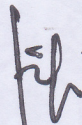
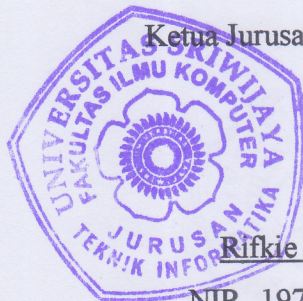


Kanda Januar Miraswan, M. T

NIP. 199001092019031012

Mengetahui,

Ketua Jurusan Teknik Informatika,



Rifkie Primartha, M.T.

NIP. 197706012009121004

TANDA LULUS UJIAN SIDANG TUGAS AKHIR

Pada hari Rabu tanggal 24 Juli 2019 telah dilaksanakan ujian sidang skripsi oleh Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

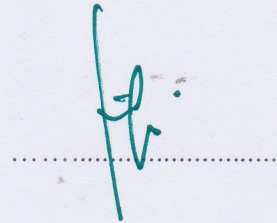
Nama : Putri Septria

NIM : 09121002005

Judul : Implementasi Metode Naïve Bayes Untuk Memperbaiki Instance Noise (Missing Value) Pada Data Mining

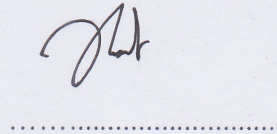
1. Ketua

Rifkie Primartha, M. T
NIP. 197706012009121004



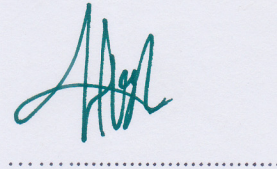
2. Sekretaris

Kanda Januar Miraswan, M. T
NIP. 199001092019031012



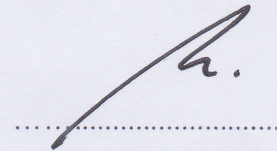
3. Penguji I

Alvi Syahrini Utami, M. Kom
NIP. 197812222006042003



4. Penguji II

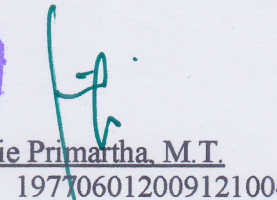
Rizki Kurniati, M.T.
NIP. 199107122019032016



Mengetahui,
Ketua Jurusan Teknik Informatika



Rifkie Primartha, M.T.
NIP. 197706012009121004



HALAMAN PERNYATAAN

Yang bertanda tangan di bawah ini :

Nama : Putri Septria
NIM : 09121002005
Program Studi : Teknik Informatika
Judul Skripsi : Implementasi Metode Naïve Bayes Untuk
Memperbaiki Instance Noise (Missing
Value) Pada Data Mining
Hasil Pengecekan Software *Turnitin* : 20 %

Menyatakan bahwa Laporan Proyek saya merupakan hasil karya sendiri dan bukan hasil penjiplakan/plagiat. Apabila ditemukan unsur penjiplakan/plagiat dalam laporan proyek ini, maka saya bersedia menerima sanksi akademik dari Universitas Sriwijaya sesuai dengan ketentuan yang berlaku.

Demikian, pernyataan ini saya buat dengan sebenarnya dan tidak ada paksaan oleh siapapun.



Palembang, Juli 2019



Putri Septria
NIM 09121002005

Motto :

“Yang terbaik adalah yang mampu bertahan sampai akhir”

-PS-

“Barang siapa yang meringankan kesulitan seorang mukmin dari kesulitan dunia, maka Allah akan meringankan kesulitannya dari kesulitan di hari kiamat. Barang siapa yang memudahkan orang yang tertimpa kesulitan, maka Allah akan memudahkan kepadanya di dunia & akhirat. Barang siapa menutupi aib seorang muslim, maka Allah akan menutupi aibnya di dunia & akhirat. Allah akan membantu hamba-Nya selagi hamba tersebut membantu saudaranya”

(HR. Muslim)

Kupersembahkan karya tulis ini

Kepada :

- ❖ Abah dan Umik tercinta**
- ❖ Kakak-kakak & Ayuk-ayukku terlove**
- ❖ Almamaterku**
- ❖ IF Reg 2012**
- ❖ Parjim Squad**
- ❖ Teman-teman dan Sahabat**

Implementasi Metode Naïve Bayes untuk Memperbaiki Instance Noise (missing value) dalam Data Mining

Oleh:

Putri Septria

09121002005

ABSTRAK

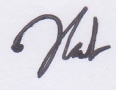
Performa algoritma pada data mining tergantung pada kualitas *dataset*, data *training* berkualitas rendah menyebabkan klasifikasi yang lemah. Dataset yang besar dengan banyak kelas memiliki *noise* atau mengandung *error*, hal ini menyebabkan berkurangnya akurasi pada klasifikasi. Data dikatakan *noise* apabila memiliki nilai kosong dan/atau derau dan/atau pencilan dan/atau inkonsistensi. Proses pembersihan data atau *data cleaning* dapat dilakukan dengan cara: mengisi nilai-nilai yang kosong, menghaluskan data yang berderau, membuang pencilan dan memperbaiki inkonsistensi dengan menggunakan regresi atau inferensi pada metode klasifikasi *Naïve Bayes* atau *Decision Tree*. Proses *Cleaning Data* dilakukan dengan menggunakan data masukan yang sudah ditentukan sebelumnya. Percobaan perangkat lunak ini menggunakan *instance* data sebanyak 3110. Hasil percobaan menunjukkan akurasi dengan *K-fold Validation* pada metode *Decision Tree* dengan *Naïve Bayes* dilakukan lebih baik sebesar 85.0% dibandingkan hasil akurasi metode *Decision Tree* yaitu sebesar 68.0%.

Kata Kunci: *Data Mining*, Klasifikasi, *Naïve Bayes*, *Decision Tree*, *Cleaning Data*, *K-fold Validation*.

Pembimbing I,

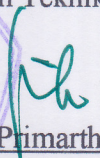

Rifkie Primartha, M.T.
NIP. 197706012009121004

Palembang, Juli 2019
Pembimbing II,


Kanda Januar Miraswan, M. T
NIP. 199001092019031012

Mengetahui,
Ketua Jurusan Teknik Informatika,




Rifkie Primartha, M.T.
NIP. 197706012009121004

Implementation of Naïve Bayes Method to Fix Instance Noise (Missing Value) on Data Mining

By:

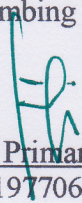
Putri Septria
09121002005

ABSTRACT

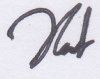
The performance of algorithms in data mining depends on the quality of dataset. Low quality of training data causes a weak classification. A big dataset with many classes possibly contains noises or errors results in reduced the accuracy of classification. Data containing noise when it has empty values and / or noise and / or outliers and / or inconsistencies. The process of data cleaning can be done by filling in the blank values; smoothing noisy data; removing outlier; correcting inconsistencies using regression or inference on the Naïve Bayes or Decision Tree classification method. The process of data cleaning was done using data imputation that had been determined. This software experiment used 3110 instant data. The present study showed the accuracy using K-fold Validation in Naïve Bayes method (85%) was better than the accuracy of Decision Tree method (68%).

Keywords: Data Mining, Classification, Naive Bayes, Decision Tree, Cleaning Data, K-fold Validation.

Pembimbing I,

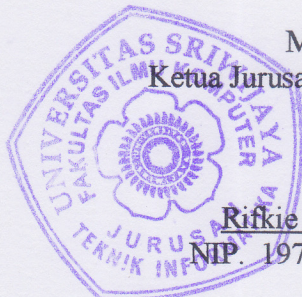

Rifkie Primartha, M.T.
NIP. 197706012009121004

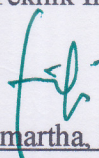
Palembang, Juli 2019
Pembimbing II,


Kanda Januar Miraswan, M. T
NIP. 199001092019031012

Mengetahui,

Ketua Jurusan Teknik Informatika,




Rifkie Primartha, M.T.
NIP. 197706012009121004

KATA PENGANTAR

Alhamdulillah, puji dan syukur kehadirat Allah SWT atas segala nikmat, rahmat dan karunia-Nya, sehingga penulis dapat menyelesaikan Tugas Akhir yang berjudul “Implementasi Metode *Naïve Bayes* Untuk Memperbaiki *Instance Noise (Missing Value)* Pada *Data Mining*”. Tugas akhir ini disusun untuk memenuhi salah satu persyaratan kelulusan tingkat sarjana pada Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

Penulis menyadari bahwa dalam penulisan Tugas Akhir ini tidak terlepas dari dukungan, doa, bantuan, pengarahan maupun bimbingan dari berbagai pihak. Untuk itu penulis mengucapkan terima kasih setulus-tulusnya kepada :

1. Kedua orang tua saya dan Keluarga Besar saya yang selalau memberikan semangat dalam penulisan Tugas Akhir ini.
2. Pemerintah dan Universitas Sriwijaya yang telah memberikan saya kesempatan dan berbagai fasilitas dalam perkuliahan sehingga saya dapat menyelesaikan Tugas Akhir ini.
3. Bapak Jaidan Jauhari, S. Pd. M. T. selaku Dekan Fakultas Ilmu Komputer Universitas Sriwijaya.
4. Bapak Rifkie Primartha, M. T. selaku Ketua Jurusan Teknik Informatika.
5. Bapak Rifkie Primartha, M. T. dan Bapak Kanda Januar Miraswan, M. T. selaku Pembimbing Tugas Akhir atas bimbingan dan masukannya penulis dapat menyelesaikan Tugas Akhir ini dengan baik.

6. Bapak Rifkie Primartha, M. T. selaku Pembimbing Akademik atas bimbingan dan masukannya sehingga penulis dapat menyelesaikan Tugas Akhir ini dengan baik.
7. Ibu Alvi Syahrini Utami, M. Kom. dan Rizki Kurniati, M. T. selaku dosen penguji yang telah memberikan koreksi dan masukan untuk Tugas Akhir ini.
8. Segenap staf pengajar di Fakultas Ilmu Komputer Universitas Sriwijaya yang telah mengajar, membimbing dan memberikan pemahaman tentang ilmu komputer.
9. Segenap karyawan Fakultas Ilmu Komputer Universitas Sriwijaya, terutama Kak Ricy Firnando selaku Admin Jurusan Teknik Informatika atas bantuannya selama ini.
10. Teman berjuang Dara Karnindo, Lisa Desta Sari, Dwi Erviana, Elbananda Permana, Rahmi Fadillah Busyra dan terkhusus Wenty Octoviani yang telah memberi semangat, membantu dan menemani saya dalam menyelesaikan penulisan Tugas Akhir ini.
11. Semua teman sebangsa, setanah, dan seair sahabat kecil saya Nitak, Eda, Evi, Kiki, Tuty, Anton, Tewe, Rio, Yahya dan Wahyu yang tidak pernah lelah bertanya “Kapan Wisuda?” kepada saya selama menyelesaikan Tugas Akhir ini.
12. Semua anggota Teknik Informatika Reguler 2012 yang telah memberi semangat kepada saya dalam menyelesaikan penulisan Tugas Akhir ini.

13. Semua teman-teman dan sahabat yang telah memberikan semangat kepada saya dalam penulisan Tugas Akhir ini.

Akhir kata, penulis menyadari bahwa laporan tugas akhir ini masih jauh dari kesempurnaan karena keterbatasan ilmu yang dimiliki penulis. Oleh karena itu, penulis mengharapkan kritik dan saran yang membangun untuk kesempurnaan tugas akhir ini. Semoga tugas akhir yang sederhana ini dapat memberikan manfaat bagi yang membutuhkan.

Palembang, Juli 2019

Penulis

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PENGESAHAN	ii
HALAMAN TANDA LULUS UJIAN.....	iii
HALAMAN PERNYATAAN BEBAS PLAGIAT	iv
MOTTO	v
ABSTRAK	vi
KATA PENGANTAR.....	viii
DAFTAR ISI.....	xi
DAFTAR GAMBAR.....	xvii
DAFTAR TABEL	xix
BAB I.....	1-1
PENDAHULUAN.....	1-1
1.1. Pendahuluan	1-1
1.2. Latar Belakang Masalah.....	1-1
1.3. Rumusan Masalah	I-3
1.4. Tujuan Penelitian	I-4
1.5. Manfaat Penelitian	1-4

1.6. Batasan Masalah	I-5
1.7. Sistematika Penulisan	I-5
BAB II	II-1
KAJIAN LITERATUR.....	II-1
2.1. Pendahuluan	II-1
2.2. Landasan Teori.....	II-1
2.2.1. <i>Data Mining</i>	II-1
2.2.2. Pengelompokan Teknik <i>Data Mining</i>	II-1
2.2.3. Tahapan KDD.....	II-2
2.3. Pembersihan Data.....	II-4
2.3.1. Membersihkan nilai kosong	II-5
2.4. Klasifikasi	II-6
2.5. <i>Naïve Bayes</i>	II-7
2.6. <i>Decision Tree</i>	II-12
2.7. Algoritma C45.....	II-14
2.8. Metode Pengembangan Perangkat Lunak.....	II-19
2.9. Penelitian yang relevan	II-21
2.10. Kesimpulan	II-26

BAB III.....	III-1
METODOLOGI PENELITIAN	III-1
3.1. Pendahuluan	III-1
3.2. Unit Penelitian	III-1
3.3. Metode Pengumpulan Data	III-1
3.3.1. Jenis Data.....	III-1
3.3.2. Sumber Data	III-1
3.3.3. Teknik Pengumpulan Data	III-2
3.3.4. Teknik Pengolahan Data.....	III-2
3.4. Tahapan Penelitian.....	III-2
3.4.1. Menentukan Ruang Lingkup dan Unit Penelitian	III-3
3.4.2. Menetapkan Kriteria Pengujian	III-4
3.4.3. Praproses.....	III-5
3.4.3.1 Proses cleaning data dengan <i>Naïve Bayes</i>	III-5
3.4.4. Menentukan alat yang digunakan dalam penelitian	III-6
3.5. Metode Pengembangan Perangkat Lunak.....	III-7
3.6. Teknik Pengujian	III-11
3.7. Manajemen Proyek Penelitian	III-12

BAB IV	IV-1
PENGEMBANGAN PERANGKAT LUNAK.....	IV-1
4.1. Pendahuluan	IV-1
4.2. Fase Insepsi	IV-1
4.2.1. Pemodelan Bisnis	IV-1
4.2.2. Kebutuhan.....	IV-2
4.2.3. Analisis dan Desain	IV-3
4.2.3.1 Analisis Data	IV-4
4.2.3.2 Desain Perangkat Lunak	IV-6
1. Model <i>Use Case</i>	IV-7
2. <i>Diagram Activity</i>	IV-10
4.3. Fase Elaborasi	IV-11
4.3.1. Pemodelan Bisnis	IV-11
4.3.1.1 Perancangan Data.....	IV-12
4.3.1.2 Perancangan Antarmuka	IV-15
4.3.2. Kebutuhan Sistem.....	IV-16
4.3.3. Sequence Diagram.....	IV-17
4.4. Fase Konstruksi.....	IV-20
4.4.1. Pemodelan Bisnis	IV-20
4.4.2. Diagram Kelas	IV-20

4.4.3. Implementasi	IV-22
4.4.3.1 Implementasi Kelas.....	IV-22
4.4.3.2 Implementasi Antarmuka.....	IV-24
4.5. Fase Transisi	IV-27
4.5.1. Pemodelan Bisnis	IV-28
4.5.2. Kebutuhan Sistem.....	IV-28
4.5.3. Rencana Pengujian	IV-29
4.5.3.1 Rencana Pengujian <i>Use Case DT+NB</i>	IV-29
4.5.3.2 Rencana Pengujian <i>Use Case Decision Tree</i>	IV-30
4.5.4. Kasus Uji	IV-31
BAB V.....	V-1
HASIL DAN ANALISIS PENELITIAN.....	V-1
5.1. Pendahuluan.....	V-1
5.2. Data Hasil Penelitian.....	V-1
5.2.1. Pembagian data dengan Kfold Validation.....	V-2
5.2.2. Hasil Pengujian <i>Decision Tree</i>	V-3
5.2.3. Hasil Pengujian <i>Decision Tree + Naïve Bayes</i>	V-3
5.2.4. Hasil Pengujian <i>Decision Tree</i> dan <i>DT+Naïve Bayes</i>	V-5
5.3. Analisis Hasil Penelitian	V-7

BAB VI	VI-1
KESIMPULAN DAN SARAN	VI-1
6.1. Pendahuluan.....	VI-1
6.2. Kesimpulan	VI-1
6.2. Saran.....	VI-2
DAFTAR PUSTAKA	
LAMPIRAN	

DAFTAR GAMBAR

Gambar II-1. Tahapan Proses <i>Knowledge Discovery in Database</i>	II-4
Gambar II-2. Bentuk <i>Decision Tree</i>	II-12
Gambar II-3. Arsitektur RUP.....	II-19
Gambar III-1. Ilustrasi <i>4-fold Cross Validation</i>	III-11
Gambar IV-1. Distribusi Variabel Kontinyu.....	IV-6
Gambar IV-2. Diagram <i>Use Case</i> Perangkat Lunak.....	IV-7
Gambar IV-3. Diagram Aktivitas <i>Use Case</i> melakukan Klasifikasi <i>Decision</i> Tree dengan <i>Naïve Bayes</i>	IV-10
Gambar IV-4. Diagram Aktivitas <i>Use Case</i> melakukan Klasifikasi <i>Decision</i> <i>Tree</i>	IV-11
Gambar IV-5. Contoh Tabel data_awal	IV-13
Gambar IV-6. Contoh Tabel data_bersih.....	IV-13
Gambar IV-7. Contoh Tabel data_missing	IV-14
Gambar IV-8. Contoh Tabel data_bobot.....	IV-15
Gambar IV-9. Contoh Tabel data_testing	IV-15
Gambar IV-10. Rancangan Antaramuka Menu Utama.....	IV-16
Gambar IV-11. Diagram Sekuen melakukan Klasifikasi <i>Decision Tree</i> dengan <i>Naïve Bayes</i>	IV-18

Gambar IV-12. Diagram Sekuen melakukan Klasifikasi <i>Decision Tree</i>	IV-19
Gambar IV-13. Diagram Sekuen Kelas Perangkat Lunak	IV-21
Gambar IV-14. Implementasi Antarmuka <i>main form</i> Perangkat Lunak.....	IV-25
Gambar IV-15. Implementasi Antarmuka hasil klasifikasi <i>Decision Tree</i> dengan <i>Naïve Bayes</i>	IV-25
Gambar IV-16. Implementasi Antarmuka hasil klasifikasi <i>Decision Tree</i>	IV-26
Gambar IV-17. Implementasi Antarmuka hasil perbandingan klasifikasi <i>Decision Tree</i> dan <i>Decision Tree</i> dengan <i>Naïve Bayes</i>	IV-27
Gambar V-1. Diagram perbandingan rata-rata Akurasi Metode <i>Decision Tree</i> dan <i>Decision Tree</i> dengan <i>Naïve Bayes</i>	V-6

DAFTAR TABEL

Tabel III-1. Tahapan Pengembangan Perangkat Lunak.....	III-8
Tabel III-2. Jadwal Penelitian	III-14
Tabel IV-1. Kebutuhan Fungsional.....	IV-3
Tabel IV-2. Kebutuhan Non-Fungsional.....	IV-3
Tabel IV-3. Definisi Aktor.....	IV-8
Tabel IV-4. Definisi <i>Use Case</i>	IV-8
Tabel IV-5. Skenario <i>Use Case</i> Klasifikasi <i>Decision Tree</i> dan <i>Naïve Bayes</i>	IV-9
Tabel IV-6. Skenario <i>Use Case</i> Klasifikasi <i>Decision Tree</i>	IV-9
Tabel IV-7. Implementasi Kelas	IV-22
Tabel IV-8. Rencana Pengujian <i>Use Case</i> Klasifikasi <i>Decision Tree</i> dan <i>Naïve Bayes</i>	IV-30
Tabel IV-9. Rencana Pengujian <i>Use Case</i> Klasifikasi <i>Decision Tree</i>	IV-30
Tabel IV-10. Pengujian <i>Use Case</i> Klasifikasi <i>Decision Tree</i> dan <i>Naïve Bayes</i>	IV-32
Tabel IV-11. Rencana Pengujian <i>Use Case</i> Klasifikasi <i>Decision Tree</i>	IV-33
Tabel V-1. Pembagian Data Latih dan Data Uji	V-2
Tabel V-2. Hasil Klasifikasi Metode <i>Decision Tree</i>	V-3
Tabel V-2. Hasil Klasifikasi Metode <i>Decision Tree</i> dengan <i>Naïve Bayes</i>	V-4
Tabel V-5. Perbandingan Hasil Akurasi Klasifikasi Metode <i>Decision Tree</i> dan <i>Decision Tree</i> dengan <i>Naïve Bayes</i>	V-5

BAB I

PENDAHULUAN

1.1. Pendahuluan

Bab ini menjelaskan secara detail tentang latar belakang, perumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah, dan sistematika penulisan dokumen penelitian.

1.2. Latar Belakang Masalah

Sejak tahun 1960an, *database* dan teknologi informasi telah berkembang secara sistematis yang memungkinkan untuk menyimpan data dalam jumlah yang besar. Saat ini manusia hidup di era dimana sejumlah data besar dikumpulkan setiap harinya. Proses menemukan pola dalam data untuk mengubah data menjadi sebuah informasi disebut *Data mining*. *Data mining* adalah ekstraksi informasi implisit yang tidak diketahui sebelumnya, dengan mengekstrak data mentah dalam *database* guna menggali informasi yang mudah dipahami dan dapat digunakan untuk berbagai kebutuhan. *Data mining* merupakan proses pemecahan masalah dengan menganalisis data yang sudah ada di *database* dengan menemukan pola dalam data dan data selalu hadir dalam jumlah yang besar (Han, Kamber, & Pei, 2012).

Data mining adalah proses analisis data untuk menemukan pola dan hubungan dalam data yang dapat digunakan untuk membuat prediksi yang valid. Teknik yang paling banyak digunakan dalam *data mining* adalah klasifikasi. Teknik klasifikasi dapat dikelompokkan menjadi lima kategori, yang didasarkan

pada konsep matematika yang berbeda. Kategori ini *statistical-based*, *distance-based*, *decision tree-based*, *neural network-based*, dan *rule-based*. Setiap kategori terdiri dari beberapa algoritma, namun yang paling populer dari setiap kategori yang digunakan secara ekstensif adalah *C4.5*, *Naïve Bayes*, *K-Nearest Neighbors*, dan Jaringan *Neural Backpropagation* (Karim & Rahman, 2013).

Decision Tree dan *Naïve Bayes* merupakan metode yang paling sering digunakan untuk mengklasifikasi sebuah data. *C4.5* merupakan algoritma yang paling populer dan efisien dalam pendekatan berbasis *decision tree*. Algoritma *decision tree* membuat model pohon dengan menggunakan nilai hanya satu atribut pada satu waktu. Pohon yang diciptakannya dimana masing-masing simpul di pohon mewakili tempat di mana keputusan harus dibuat berdasarkan masukan, dan pindah ke simpul berikutnya sampai mencapai sebuah daun yang memberi tahu prediksi atau perkiraan keluaran. Sedangkan metode *Naïve Bayes* adalah algoritma yang mengukur probabilitas posterior untuk menentukan keluaran *instance* data baru berdasarkan pengalaman data yang sebelumnya (Karim & Rahman, 2013).

Data dicirikan sebagai fakta yang tercatat, maka informasi adalah kumpulan pola, atau harapan, yang mendasari data. Sejumlah besar informasi yang dikurung dalam basis data memiliki informasi yang berpotensi penting tetapi belum ditemukan atau diartikulasikan. Performa algoritma pada data mining tergantung pada kualitas *dataset*, karena data *training* berkualitas rendah menyebabkan klasifikasi yang lemah. Banyak *dataset* berukuran besar dan memiliki banyak kelas atau multi kelas memiliki *noise* atau mengandung *error*,

hal ini dapat menyebabkan berkurangnya akurasi pada klasifikasi (Witten, Frank, & Hall, 2011).

Sebuah data dikatakan tidak bersih jika mengandung kotoran yang berupa **nilai kosong** dan/atau **derau** dan/atau **pencilan** dan/atau **inkonsistensi**. Sebelum melakukan *data mining*, data yang masih kotor harus dibersihkan dengan cara: mengisi nilai-nilai yang kosong dan/atau menghaluskan data yang berderau dan/atau membuang pencilan dan/atau memperbaiki inkonsistensi. Hingga saat ini belum ada teknik *data mining* yang memberikan hasil terbaik untuk data kotor. Kalau pun ada biasanya teknik tersebut tidak stabil dan tidak selalu tahan (*robust*) untuk semua jenis data kotor (Suyanto, 2017). Prapemrosesan data bisa dilakukan dengan **pembersihan** dan/atau **integrasi** dan/atau **reduksi** dan/atau **penambahan** dan/atau **transformasi** (Han et al., 2012).

Berdasarkan hasil penelitian yang dipaparkan di atas, maka akan dilakukan penelitian lebih lanjut mengenai perbandingan metode *Decision Tree* dan metode *Decision Tree* dengan *Naïve Bayes* untuk mengurangi data *noise* pada *dataset* multi kelas guna membantu menyelesaikan permasalahan yang sering terjadi dalam menentukan klasifikasi pada teknik *data mining*.

1.1 Rumusan Masalah

Berdasarkan latar belakang tersebut, masalah yang dibahas pada penelitian ini adalah mengimplementasikan metode *Naïve Bayes* pada proses pembersihan data atau *data cleaning* guna mengatasi *instance* yang mengandung kotoran atau *noise* berupa nilai kosong (*missed value*) dalam *dataset* dengan memperbaiki

dan/atau mengisi nilai data yang hilang untuk digunakan pada metode klasifikasi *Decision Tree*.

1.4 Tujuan Penelitian

1. Mengimplementasikan metode *Naïve Bayes* untuk proses pembersihan data atau *data cleaning* pada *dataset*.
2. Membangun model perbandingan klasifikasi *Decision Tree* dan *Decision Tree* dengan *Naïve Bayes*.
3. Mengevaluasi hasil model klasifikasi data yang menerapkan metode *Decision Tree* dengan *Naïve Bayes* dengan mengukur nilai *accuracy*.

1.5 Manfaat Penelitian

Manfaat yang diperoleh dari penelitian ini adalah sebagai berikut :

1. Menghasilkan sistem perbandingan klasifikasi *Decision Tree* dengan *noise* dan *Decision Tree* tanpa *noise* yang telah diperbaiki menggunakan metode *Naïve Bayes*.
2. Mengatasi *instance* yang mengandung *noise* berupa nilai kosong (*missed value*) dalam *dataset* dengan memperbaiki dan/atau mengisi nilai data yang hilang pada *dataset* multi kelas menggunakan metode *Naïve Bayes* untuk meningkatkan hasil pohon keputusan pada metode *Decision Tree*.
3. Hasil penelitian ini dapat dijadikan perbandingan hasil dengan metode klasifikasi lainnya.

1.6 Batasan Masalah

Dalam pembuatan sistem, untuk mengatasi permasalahan yang ada maka penyusun membatasi permasalahan sebagai berikut:

1. Penelitian ini hanya berfokus pada perbandingan metode klasifikasi *Decision Tree* dengan *noise* dan klasifikasi *Decision Tree* tanpa *noise* pada *dataset* yang telah diperbaiki menggunakan metode *Naïve Bayes*.
2. Penelitian ini menggunakan *dataset* yang telah tersedia dalam UCI *Machine Learning Repository*.
3. Penelitian ini menerapkan metode *Decision Tree* dan *Naïve Bayes*.
4. *Instance* yang bermasalah pada *dataset* berupa *instance* yang mengandung nilai “?” atau *missing value*.
5. Penelitian ini menggunakan teknik validasi *K-fold Cross Validation* untuk membagi *dataset* menjadi data *training* dan data *testing*.

1.7 Sistematika Penulisan

Sistematika penulisan tugas akhir ini mengikuti standar penulisan tugas akhir Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya, antara lain :

BAB I. PENDAHULUAN

Pada bab ini diuraikan mengenai latar belakang, perumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah, dan sistematika penulisan.

BAB II. KAJIAN LITERATUR

Pada bab ini akan dibahas dasar-dasar teori yang digunakan dalam penelitian, seperti definisi *Data Mining*, *Naïve Bayes*, *Decision Tree*, *Data Cleaning*, *Cross Validation*, *Rational Unified Process* dan penelitian lain yang relevan dengan penelitian yang sedang dikembangkan.

BAB III. METODOLOGI PENELITIAN

Pada bab ini akan dibahas mengenai tahapan yang akan dilaksanakan pada penelitian ini. Masing-masing rencana tahapan penelitian dideskripsikan dengan rinci dengan mengacu pada suatu kerangka kerja. Di akhir bab ini berisi perancangan manajemen proyek pada pelaksanaan penelitian.

BAB IV. PENGEMBANGAN PERANGKAT LUNAK

Pada bab ini akan dibahas mengenai perancangan, lingkungan implementasi, dan hasil pengujian perangkat lunak.

BAB V. HASIL DAN ANALISIS PENELITIAN

Pada bab ini, hasil pengujian berdasarkan langkah-langkah yang telah direncanakan disajikan. Analisis diberikan sebagai basis dari kesimpulan yang diambil dalam penelitian ini

BAB VI. KESIMPULAN DAN SARAN

Pada bab ini berisi kesimpulan dari semua uraian pada bab-bab sebelumnya dan juga berisi saran-saran yang diharapkan berguna dalam penerapan pengembangan perangkat lunak ini.

1.8 Kesimpulan

Berdasarkan uraian di atas, maka dalam penelitian ini akan dikembangkan sebuah perangkat lunak untuk mengatasi *instance* yang mengandung *noise* berupa *missing value* pada proses *data cleaning* dengan cara mengisi nilai atribut kosong yang paling mungkin dapat ditentukan menggunakan algoritma *Naïve Bayes*. Diharapkan metode yang digunakan dapat diimplementasikan ke dalam perangkat lunak dengan baik.

DAFTAR PUSTAKA

- Fijorek, K., & Grotowski, M. (2012). Bankruptcy Prediction: Some Results from a Large Sample of Polish Companies. *International Business Research*, 5(9), 70–77. <https://doi.org/10.5539/ibr.v5n9p70>
- Han, J., Kamber, M., & Pei, J. (2012). *Data Mining: Concepts and Techniques*. San Francisco, CA, itd: Morgan Kaufmann. <https://doi.org/10.1016/B978-0-12-381479-1.00001-0>
- Karim, M., & Rahman, R. M. (2013). Decision Tree and Naïve Bayes Algorithm for Classification and Generation of Actionable Knowledge for Direct Marketing. *Journal of Software Engineering and Applications*, 6(4), 196–206. <https://doi.org/10.4236/jsea.2013.64025>
- Aprilla Dennis. (2013). Belajar Data Mining dengan RapidMiner. *Innovation and Knowledge Management in Business Globalization: Theory & Practice, Vols 1 and 2*, 5(4), 1–5. <https://doi.org/10.1007/s13398-014-0173-7.2>
- Farid, D. M., Zhang, L., Rahman, C. M., Hossain, M. A., & Strachan, R. (2014). Hybrid decision tree and naïve Bayes classifiers for multi-class classification tasks. *Expert Systems with Applications*, 41(4 PART 2), 1937–1946. <https://doi.org/10.1016/j.eswa.2013.08.089>
- Ginting, S. L. B., & Trinanda, R. P. (2015). Teknik Data Mining Menggunakan Metode Bayes Classifier Untuk Optimalisasi Pencarian Pada Aplikasi Perpusatakan, *d*, 1–14.
- Han, J., Kamber, M., & Pei, J. (2012). *Data Mining: Concepts and Techniques*.

San Francisco, CA, itd: Morgan Kaufmann. <https://doi.org/10.1016/B978-0-12-381479-1.00001-0>

Karim, M., & Rahman, R. M. (2013). Decision Tree and Naïve Bayes Algorithm for Classification and Generation of Actionable Knowledge for Direct Marketing. *Journal of Software Engineering and Applications*, 6(4), 196–206. <https://doi.org/10.4236/jsea.2013.64025>

Khadafy, A. R., & Wahono, R. S. (2015). Penerapan Naïve Bayes Untuk Mengurangi Data Noise Pada Klasifikasi Multi Kelas Dengan Decision Tree. *Journal of Intelligent Systems*, 1(2), 11–17.

Kohavi, R. (1996). Scaling Up the Accuracy of Naive-Bayes Classifiers : a Decision-Tree Hybrid Accuracy Scale-Up : the Learning. *Data Mining and Visualization*, (Utgo 1988), 1–6.

Mudry, A., & Tjellström, A. (2011). *Historical background of bone conduction hearing devices and bone conduction hearing aids. Advances in Oto-Rhino-Laryngology* (Vol. 71). <https://doi.org/10.1159/000323569>

Singh, D. M., Harbi, N., & Zahidur Rahman, M. (2010). Combining Naive Bayes and Decision Tree for Adaptive Intrusion Detection. *International Journal of Network Security & Its Applications*, 2(2), 12–25. <https://doi.org/10.5121/ijnsa.2010.2202>

Wang, L. M., Li, X. L., Cao, C. H., & Yuan, S. M. (2006). Combining decision tree and Naive Bayes for classification. *Knowledge-Based Systems*, 19(7), 511–515. <https://doi.org/10.1016/j.knosys.2005.10.013>

Witten, I. H., Frank, E., & Hall, M. a. (2011). *Data Mining: Practical Machine*

Learning Tools and Techniques (Google eBook). Complementary literature

None. <https://doi.org/0120884070>, 9780120884070

Suyanto, Dr. (2017). *Data Mining*: untuk klasifikasi dan klasterisasi data.