

Perbandingan Metode Naive Bayes dan C4.5 dalam Klasifikasi Status Pengguna Media Sosial Facebook

Diajukan Sebagai Syarat untuk Menyelesaikan

Pendidikan Program Strata-I pada

Jurusan Teknik Informatika



Oleh :

DIAN LUKITA SARI

09021281320025

Jurusan Teknik Informatika

FAKULTAS ILMU KOMPUTER UNIVERSITAS SRIWIJAYA

2020

LEMBAR PENGESAHAN TUGAS AKHIR

**Perbandingan Metode Naive Bayes dan C4.5 dalam Klasifikasi
Status Pengguna Media Sosial Facebook**

Oleh :

DIAN LUKITA SARI

09021281320025

Indralaya, November 2020

Pembimbing I,


Rifkie Primartha, M.T.
NIP.197706012009121004

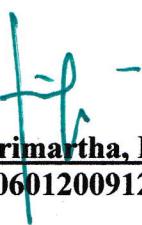
Pembimbing II


Osvari Arsalan,M.T.
NIP. 198806282018031001

Mengetahui,

Ketua Jurusan Teknik Informatika




Rifkie Primartha, M.T
NIP.197706012009121004

TANDA LULUS UJIAN SIDANG TUGAS AKHIR

Pada hari Kamis tanggal 30 Juli 2020 telah dilaksanakan ujian sidang tugas akhir oleh jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

Nama : Dian Lukita Sari
NIM : 09021281320025
Judul TA : Perbandingan Metode Naive Bayes dan C4.5 dalam Klasifikasi Status Pengguna Media Sosial Facebook

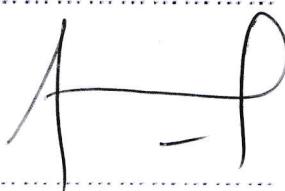
1. Pembimbing I

Rifkie Primartha, M.T.
NIP.197706012009121004



2. Pembimbing II

Osvari Arsalan, M.T.
NIP.198806282018031001



3. Penguji I

M.Fachrurrozi, M.T.
NIP.198005222008121002



4. Penguji II

Muhammad Qurhanul Rizqie, S.Kom, M.T.
NIP.



Mengetahui,

Ketua Jurusan Teknik Informatika

Rifkie Primartha, M.T.
NIP.197706012009121004



HALAMAN PERNYATAAN

Yang bertanda tangan di bawah ini :

Nama : Dian Lukita Sari
NIM : 09021281320025
Program Studi : Teknik Informatika
Judul Skripsi : Perbandingan Metode Naïve Bayes dan C4.5 dalam Klasifikasi Status Pengguna Media Sosial Facebook

Hasil Pengecekan Software *iThenticate/Turnitin* : 19 %

Menyatakan bahwa Laporan Projek saya merupakan hasil karya sendiri dan bukan hasil penjiplakan/plagiat. Apabila ditemukan unsur penjiplakan/plagiat dalam laporan projek 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,



(Dian Lukita Sari)

09021281320025

Motto :

Semua orang berjalan dengan waktunya sendiri,
Begitupula dengan dirimu.

Pencapaian hidupmu berbeda dengan pencapaian hidup orang lain.
Jangan merasa dirimu tertinggal. Kamu hanya berjuang di jalanmu saja.

Tenang. Hanya karena kamu lebih lambat. Bukan berarti dirimu lebih buruk.
Karena cepat bukan berarti lebih baik. Dan lambat bukan berarti lebih buruk.

Mungkin, dirimu sendiri dan orang lain mengira bahwa dirimu terlambat.
Padahal, kamu bukan terlambat. Kamu berada di waktu yang tepat.

Lakukan saja yang terbaik untuk hari ini.
Jangan lupa untuk tetap baik-baik saja.

Kupersembahkan Tugas Akhir ini

Kepada:

- Allah SWT
- Papa dan Mama tercinta
- Adik-adikku tersayang Nico dan Ekal
- Ibuk, Om Romi, Tante ma, Tante Ta, dan seluruh keluarga besarku.
- Seluruh dosen Teknik Informatika UNSRI
- Sahabat-sahabatku
- IF REG '13
- Almamaterku

Perbandingan Metode Naive Bayes dan C4.5 dalam Klasifikasi Status Pengguna Media Sosial Facebook

Oleh :
DIAN LUKITA SARI
09021281320025

ABSTRAK

Klasifikasi merupakan metode *text mining* yang menjadi alternatif untuk mengolah data dengan jumlah besar dan banyak. Klasifikasi digunakan untuk memprediksi kelas dari objek yang kelasnya belum diketahui. Pada penelitian ini dilakukan perbandingan kinerja metode *naïve bayes* dan C4.5 bertujuan untuk mengukur tingkat akurasi dan lama waktu proses dari masing-masing metode untuk mendapatkan metode terbaik yang akan diterapkan dalam membantu proses klasifikasi status pengguna media sosial facebook. Pengujian pada penelitian ini dilakukan dengan cara membandingkan hasil keluaran perangkat lunak dengan manual. Hasil penelitian menunjukkan metode naïve bayes lebih unggul dengan menghasilkan *precision* sebesar 69,045%, nilai *recall* sebesar 66,665%, dan nilai *accuracy* sebesar 70% dibandingkan dengan metode C4 yang menghasilkan nilai *precision* sebesar 60%, nilai *recall* sebesar 60%, dan nilai *accuracy* sebesar 60%, sedangkan waktu komputasi menggunakan metode C4.5 bernilai 15774 Milisecond yang lebih cepat dibandingkan metode *naïve bayes* menghasilkan nilai 20313 Milisecond.

Kata kunci : Klasifikasi, *Naïve bayes*, C4.5.

Indralaya, November 2020

Pembimbing I,

Pembimbing II



Rifkie Primartha, M.T.
NIP.197706012009121004

Osvari Arsalan,M.T.
NIP. 198806282018031001

Mengetahui,
Ketua Jurusan Teknik Informatika



Rifkie Primartha, M.T
NIP.197706012009121004

Comparison of Naive Bayes and C4.5 Methods in Facebook Social Media User Status Classification

**By:
DIAN LUKITA SARI
09021281320025**

ABSTRACT

Classification is a text mining method that becomes an alternative to processing large and large amounts of data. Classification is used to predict the class of objects whose class is not yet known. In this study conducted a comparison of the performance of naive bayes and C4.5 methods aimed at measuring the accuracy and length of process time of each method to get the best method that will be applied in helping the process of classification of the status of social media users facebook. Testing in this study was done by comparing the results of software output manually. The results showed that naïve bayes method is superior by producing precision by 69.045%, recall value of 66.665%, and accuracy value of 70% compared to C4 method that produces precision value of 60%, recall value of 60%, and accuracy value of 60%, while computation time using C4.5 method is 15774 Milisecond which is faster than naïve bayes method resulting in 20313 Milisecond value.

Keywords: Classification, Naïve bayes, C4.5.

Indralaya, November 2020

Pembimbing I,

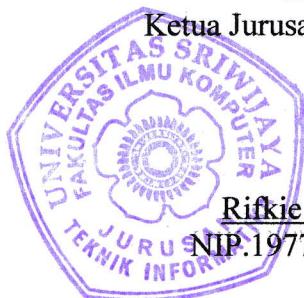
Rifkie Primartha, M.T.
NIP.197706012009121004

Pembimbing II

Osvari Arsalan,M.T.
NIP. 198806282018031001

Mengetahui,

Ketua Jurusan Teknik Informatika



Rifkie Primartha, M.T
NIP.197706012009121004

KATA PENGANTAR

Alhamdulillahirabbil’alamin, segala puji dan syukur dihaturkan kepada Allah SWT, atas segala nikmat kesehatan dan kesempatan, serta kekuatan yang telah diberikan sehingga Penulis dapat menyelesaikan laporan Tugas Akhir yang berjudul “**perbandingan Metode Naïve Bayes dan C4.5 dalam Klasifikasi Status Pengguna Media Sosial Facebook**”. Tugas Akhir ini disusun untuk memenuhi salah satu syarat guna menyelesaikan pendidikan program Strata-1 pada Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

Pada kesempatan ini, penulis mengucapkan terima kasih kepada pihak-pihak yang telah membantu, membimbing, dan terus mendukung penulis dalam menyelesaikan laporan tugas akhir ini diantaranya:

1. Yang pertama kepada Allah SWT, yang telah memberikan jalan dan mengabulkan do'a penulis selama ini.
2. Kedua orang tuaku tercinta, Junaidi, S.H dan Yuli Hartisia yang terus selalu memberikan kasih sayang, senantiasa mendo'akan, memberi semangat, dukungan baik moril maupun materil serta memberikan restunya agar selalu dapat meraih cita-cita, sukses dan selalu ingat kepada Allah SWT.
3. Kedua saudaraku tersayang, Nico Firmansyah dan Zulfikar yang terus mendo'akan, memberikan waktu, saran, dan motivasi.

4. Ibu Alvi Syahrini Utami M.Kom, Ph.D. selaku Ketua Jurusan Teknik Informatika Univesitas Sriwijaya.
5. Bapak Rifkie Primartha, M.T. dan Bapak Osvari Arsalan M.T. sekalu dosen pembimbing yang telah memberikan bimbingan, arahan, motivasi dan saran selama proses penyelesaian Tugas Akhir ini.
6. Bapak M. Fachrurrozi, M.T. dan Bapak Muhammad Qurhannul Rizqie, S.Kom,M.T.. selaku dosen penguji yang telah memberikan koreksi, saran dan masukan untuk Tugas Akhir ini.
7. Seluruh dosen Jurusan Teknik Informatika Fakultas Ilmu Komputer Univesitas Sriwijaya atas ilmu dan bimbangannya selama masa perkuliahan penulis.
8. Seluruh staf pegawai jurusan Teknik Informatika yang telah membantu dan memperlancar dalam hal urusan akademik dan administrasi selama perkuliahan penulis.
9. Sahabat penulis, yaitu Nova Hasti S.Kom, Ade Rahma Youanda, S.Kom, Septa Nopita Sari, S.Kom, Annisa Septi S.Kom, Dian Lestari, S.Kom, Monica, S.Kom dan seluruh sahabat seperjuangan di Teknik Informatika Reguler 2013.
10. Semua teman-teman Teknik Informatika angkatan 2013 terkhusus IF Reguler C beserta kakak dan adik tingkat.
11. Dan untuk semua pihak yang telah terlibat dalam membantu penyelesaian Tugas Akhir ini yang tidak dapat penulis sebutkan satu per satu.

Penulis menyadari bahwa dalam penyusunan tugas akhir ini masih terdapat banyak kekurangan, oleh karna itu kritik dan saran yang membangun amat sangat diharapkan demi penyempurnaan di masa mendatang. Akhir kata, semoga tugas akhir ini dapat berguna dan bermanfaat bagi kita semua.

Palembang, November 2020

Penulis

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PENGESAHAN	ii
TANDA LULUS UJIAN SIDANG TUGAS AKHIR.....	iii
HALAMAN PERNYATAAN BEBAS PLAGIAT	iv
MOTTO	v
ABSTRAK	vi
ABSTRACT	vii
KATA PENGANTAR	viii
DAFTAR ISI	xi
DAFTAR GAMBAR	xvi
DAFTAR TABEL.....	xviii

BAB I PENDAHULUAN

1.1 Latar Belakang	I-1
1.2 Rumusan Masalah	I-3
1.3 Tujuan Penelitian	I-3
1.4 Manfaat Penelitian	I-3
1.5 Batasan Masalah	I-4

BAB II TINJAUAN PUSTAKA

2.1 Pendahuluan	II-1
2.2 Landasan Teori	II-1
2.2.1 Media Sosial	II-1
2.2.2 <i>Klasifikasi</i>	II-2
2.2.3 Opini	II-3

2.2.4 Sentimen Analysis.....	II-4
2.2.5 Praproses	II-5
2.2.6 Pembobotan Kata (<i>Term Weighting</i>)	II-6
2.2.7 Metode Pembobotan TF-IDF	II-7
2.2.8 <i>Naïve Bayes</i>	II-8
2.2.9 Pohon Keputusan (<i>Decision Tree</i>)	II-12
2.2.10 Algoritma C4.5	II-13
2.2.11 <i>Cross Validation</i>	II-15
2.2.12 <i>Confusion Matrix</i>	II-16
2.2.13 <i>Rational Unified Process</i>	II-17
2.3 Penelitian Terkait	II-20

BAB III METODOLOGI PENELITIAN

3.1 Pendahuluan	III-1
3.2 Unit Penelitian	III-1
3.3 Data dan Metode Pengumpulan Data	III-1
3.4 Tahapan Penelitian	III-2
3.4.1 Menentukan Ruang Lingkup dan Unit Penelitian	III-3
3.4.2 Menetapkan Kriteria Pengujian	III-3
a. Praproses	III-4
b. Pembobotan kata	III-4
3.4.3 Menentukan Alat yang digunakan dalam Pelaksanaan Penelitian	III-4
3.4.4 Melaksanakan Pengujian Penelitian	III-5
a. Pengujian Algoritma <i>Naïve Bayes</i>	III-5
b. Pengujian Algoritma C4.5	III-5

c. Evaluasi Menggunakan <i>Cross Validation</i>	III-5
3.4.5 Melakukan Analisa Hasil Pengujian dan Membuat Kesimpulan Penelitian.....	III-8
3.5 Metode Pengembangan Perangkat Lunak	II-8
3.5.1 Fase Insepsi	III-9
3.5.2 Fase Elaborasi	III-9
3.5.3 Fase Kontruksi	III-10
3.5.4 Fase Transisi	III-10
3.6 Penjadwalan Penelitian	III-10

BAB IV PENGEMBANGAN PERANGKAT LUNAK

4.1 Pendahuluan	IV-1
4.2 Fase Insepsi	IV-1
4.2.1 Pemodelan Bisnis	IV-2
4.2.2 Kebutuhan Sistem	IV-3
a. Fitur Praproses	IV-3
b. Fitur Klasifikasi	IV-3
c. Fitur Evaluasi	IV-3
4.2.3 Analisis dan Desain	IV-5
4.2.3.1 Analisis Perangkat Lunak	IV-5
4.2.3.1.1 Analisis Kebutuhan Perangkat Lunak	IV-5
4.2.3.1.2 Analisis Data	IV-6
4.2.3.1.3 Analisis Praproses	IV-7
4.2.3.1.4 Analisis Pembobotan Kata	IV-9
4.2.3.1.5 Analisis Algoritma <i>Naïve Bayes</i>	IV-10

4.2.3.1.6 Analisis Algoritma C4.5	IV-14
4.2.3.2 Desain Perangkat Lunak	IV-19
4.3 Fase Elaborasi	IV-28
4.3.1 Pemodelan Bisnis	IV-28
4.3.1.1 Perancangan Data	IV-28
4.3.1.2 Perancangan Antarmuka	IV-28
4.3.2 Kebutuhan Sistem	IV-29
4.3.3 <i>Sequence Diagram</i>	IV-30
4.4 Fase Konstruksi	IV-34
4.4.1 Kebutuhan Sistem	IV-34
4.4.2 Diagram Kelas	IV-34
4.4.3 Implementasi	IV-35
4.4.3.1 Implementasi Kelas	IV-35
4.4.3.2 Implementasi Antarmuka	IV-37
4.5 Fase Transisi	IV-34
4.5.1 Pemodelan Bisnis	IV-39
4.5.2 Kebutuhan Sistem	IV-39
4.5.3 Analisis dan Desain	IV-40
4.5.3.1 Rencana Pengujian <i>Use Case</i> Memasukkan Data	IV-40
4.5.3.2 Rencana Pengujian <i>Use Case</i> Klasifikasi <i>Naïve Bayes</i>	IV-40
4.5.3.3 Rencana Pengujian <i>Use Case</i> Klasifikasi C4.5	IV-41
4.5.4 Implementasi	IV-41
4.5.4.1 Pengujian <i>Use Case</i> Memasukkan Data	IV-41
4.5.4.2 Pengujian <i>Use Case</i> Klasifikasi <i>Naïve Bayes</i>	IV-41
4.5.4.3 Pengujian <i>Use Case</i> Klasifikasi C4.5	IV-41

BAB V HASIL DAN ANALISIS PENELITIAN

5.1	Pendahuluan	V-1
5.2	Data Hasil Penelitian	V-1
5.2.1	Hasil Pengujian Metode <i>Naïve Bayes</i>	V-3
5.2.2	Hasil Pengujian Metode C4.5	V-5
5.2.3	Analisis Hasil Pengujian	V-9

BAB VI KESIMPULAN DAN SARAN

6.1	Pendahuluan	VI-1
6.2	Kesimpulan	VI-1

DAFTAR PUSTAKA	xxi
LAMPIRAN 1 Kode Program Perangkat Lunak	L-1

DAFTAR GAMBAR

Gambar II-1	Jumlah Pengguna Aktif Media Sosial	II-1
Gambar II-2	Arsitektur RUP	II-18
Gambar III-1	Diagram Blok Tahapan Penelitian.....	III-2
Gambar III-2	Tahapan Kriteria Pengujian Penelitian.....	III-7
Gambar III-3	Penjadwalan untuk Tahap Menentukan Ruang Lingkup dan Unit Penelitian	III-18
Gambar III-4	Penjadwalan untuk Mengumpulkan Dasar Teori yang Berkaitan dengan Penelitian, Menentukan Kriteria Pengujian	III-18
Gambar III-5	Penjadwalan untuk Tahap Menentukan Alat yang Digunakan untuk Pelaksanaan Penelitian pada Fase Insepsi	III-19
Gambar III-6	Penjadwalan untuk Tahap Menentukan Alat yang Digunakan untuk Pelaksanaan Penelitian pada Fase Elaborasi	III-19
Gambar III-7	Penjadwalan untuk Tahap Menentukan Alat yang Digunakan untuk Penelitian pada Fase Konstruksi	III-20
Gambar III-8	Penjadwalan untuk Tahap Menentukan Alat yang Digunakan untuk Pelaksanaan Penelitian pada Fase Transisi	III-20
Gambar III-9	Penjadwalan untuk Tahap Melakukan Pengujian Penelitian, Analisa dan Menarik Kesimpulan dari hasil pengujian	III-21
Gambar IV-1	Hasil Perhitungan Nilai Gain Simpul Akar	IV-16
Gambar IV-2	Hasil Perhitungan Nilai Gain cabang Simpul sukses nilai 0	IV-17

Gambar IV-3	Hasil Perhitungan Nilai Gain cabang Simpul curang nilai 1.....	IV-18
Gambar IV-4	Diagram <i>Use Case</i>	IV-19
Gambar IV-5	Diagram Aktivitas Memasukkan Data Training	IV-25
Gambar IV-6	Diagram Aktivitas Memasukkan Data Uji	IV-26
Gambar IV-7	Diagram Aktivitas Proses Naïve Bayes	IV-27
Gambar IV-8	Diagram Aktivitas Proses C4.5	IV-27
Gambar IV-9	Rancangan Antar Muka <i>Naïve Bayes</i> san C4.5	IV-29
Gambar IV-10	<i>Sequence Diagram</i> Memasukkan data training	IV-30
Gambar IV-11	<i>Sequence Diagram</i> Memasukkan data uji	IV-31
Gambar IV-12	<i>Sequence Diagram</i> Memproses <i>Naïve Bayes</i>	IV-32
Gambar IV-13	<i>Sequence Diagram</i> Memproses C4.5.....	IV-33
Gambar IV-14	Diagram Kelas.....	IV-35
Gambar IV-15	Antarmuka Main Perangkat Lunak	IV-38
Gambar IV-16	Antarmuka Hasil Klasifikasi <i>Naïve Bayes</i> dan C4.5	IV-38
Gambar V-1	Hasil Pelatihan dengan 20 data latih	V-1
Gambar V-2	Perbandingan Waktu Komputasi <i>Naïve Bayes</i> dan C4.5	V-10
Gambar V-3	Perbandingan Performa <i>Naïve Bayes</i> dan C4.5	V-10

DAFTAR TABEL

Tabel II-1	Contoh Hasil Praproses Kalimat	II-5
Tabel III-1	Hasil Pengujian Klasifikasi	III-6
Tabel III-2	Rancangan Tabel Hasil Rata-rata Pengujian Perbandingan Klasifikasi Algoritma C4.5 dan Naïve Bayes	III-6
Tabel III-3	Penjadwalan Penelitian dalam Bentuk <i>Work Breakdown Structure</i>	III-11
Tabel IV-1	Kebutuhan Fungsional	IV-4
Tabel IV-2	Kebutuhan Non-Fungsional	IV-4
Tabel IV-3	Hasil <i>Tokenizing</i> Contoh Komentar	IV-8
Tabel IV-4	Hasil <i>Stopword Removal</i> Contoh Komentar	IV-8
Tabel IV-5	Hasil <i>Stemming</i> Contoh Komentar.....	IV-9
Tabel IV-6	Hasil Pembobotan Kata Contoh Komentar	IV-9
Tabel IV-7	Contoh Data Komentar beserta Kelas Sentimen	IV-11
Tabel IV-8	Hasil Praproses Contoh Data <i>Naïve Bayes</i>	IV-11
Tabel IV-9	Perhitungan <i>Naïve Bayes</i>	IV-11
Tabel IV-10	Proses <i>Matching</i>	IV-12
Tabel IV-11	Contoh Data Komentar beserta Kelas Sentimen	IV-14
Tabel IV-12	Perhitungan Information Gain Untuk Simpul Akar	IV-16

Tabel IV-13	Perhitungan Information Gain Untuk Simpul Akar	IV-17
Tabel IV-14	Perhitungan Information Gain Untuk Simpul Akar	IV-17
Tabel IV-15	Perhitungan Information Gain Untuk Cabang Simpul	IV-18
Tabel IV-16	Definisi Aktor	IV-20
Tabel IV-17	Definisi <i>Use case</i>	IV-20
Tabel IV-18	Skenario <i>Use Case</i> Memasukkan Data Traiing	IV-21
Tabel IV-19	Skenario <i>Use Case</i> Memasukkan Data Uji	IV-22
Tabel IV-20	Skenario <i>Use Case</i> Memproses C4.5	IV-24
Tabel IV-21	Implementasi Kelas	IV-36
Tabel IV-22	Rencana Pengujian <i>Use Case</i> Memasukkan Data	IV-40
Tabel IV-23	Rencana Pengujian <i>Use Case</i> Klasifikasi <i>Naïve Bayes</i>	IV-40
Tabel IV-24	Rencana Pengujian <i>Use Case</i> Klasifikasi C4.5	IV-41
Tabel IV-25	Pengujian <i>Use Case</i> Memasukkan Data	IV-42
Tabel IV-26	Pengujian <i>Use Case</i> Klasifikasi <i>Naïve Bayes</i>	IV-42
Tabel IV-27	Pengujian <i>Use Case</i> Klasifikasi C4.5	IV-43
Tabel V-1	Waktu Komputasi Proses <i>naïve bayes</i> dan C4.5	V-3

Tabel V-2	Proses Pengujian Klasifikasi Metode <i>naïve bayes</i>	V-3
Tabel V-3	Proses Pengujian Klasifikasi Metode C4.5	V-6
Tabel V-4	Analisis Hasil Pengujian	V-9

BAB I

PENDAHULUAN

1.1 Pendahuluan

Pada bab ini membahas latar belakang masalah, tujuan, manfaat penelitian, dan batasan masalah. Bab ini akan memberikan penjelasan umum mengenai keseluruhan penelitian.

1.2 Latar Belakang

Analisis sentimen atau *opinion mining* merupakan proses memahami, mengekstrak, dan mengolah data textual secara otomatis untuk mendapatkan suatu informasi sentimen yang terkandung dalam suatu kalimat opini. Secara umum, analisis sentimen bertujuan untuk menentukan sikap pembicara atau penulis terhadap suatu topik atau keseluruhan polaritas kontekstual pada suatu dokumen atau kalimat. Sikap dapat berupa penilaian atau evaluasi, sisi emosional penulis pada saat menulis atau efek komunikasi emosional yang penulis inginkan terhadap pembacanya. Analisis sentimen akan mengelompokkan polaritas dari teks yang ada dalam kalimat atau dokumen untuk mengetahui pendapat yang dikemukakan dalam kalimat atau dokumen tersebut apakah bersifat positif, negatif atau netral (Pang & Lee, 2008). Pada media sosial, *opinion mining* digunakan untuk menganalisa dan mengelompokkan data-data yang dibuat oleh pengguna seperti, ulasan (*review*), blog, komentar, artikel, dan lain-lain.

Diperlukan suatu metode untuk mengelola informasi dari sekumpulan polaritas teks yang jumlahnya sangat besar sehingga dapat mempermudah dalam

pencarian informasi yang relevan dengan kebutuhan. Metode yang dapat mengorganisir polaritas teks secara otomatis diantaranya adalah klasifikasi.

Klasifikasi digunakan untuk memprediksi kelas dari objek yang kelasnya belum diketahui (Raviya & Gajjar, 2013). Metode klasifikasi yang umum digunakan antara lain Decision Tree, K-Nearest Neighbor, Naïve Bayes, neural network dan support vector machines (Sahu dkk, 2011). Naive Bayes dan C4.5 juga termasuk pada 10 peringkat yang sering digunakan untuk pengolahan data karena memiliki algoritma yang sederhana, cepat dan memiliki nilai akurasi tinggi (Wu dkk, 2007).

Facebook merupakan sosial media yang paling banyak digunakan di Indonesia dan merupakan salah satu pengguna tertinggi di dunia yaitu 111 juta. Total pengguna Facebook mencapai 48% paling tinggi dibandingkan dengan Instagram yang penggunanya mencapai 39% dan Twitter sebanyak 38% dari populasi penduduk Indonesia.

Penelitian yang berkaitan dengan *opinion mining* telah banyak dilakukan. Diantaranya adalah penelitian yang dilakukan oleh Antonius Rachmat (2016) pernah meneliti penggunaan metode Naive Bayes untuk melakukan klasifikasi sentimen positif atau negatif terhadap komentar dari status kampanye politik dari Facebook Page. Studi kasus yang digunakan pada penelitian tersebut adalah status dan komentar terhadap Facebook Page calon presiden Republik Indonesia pada Pemilu tahun 2014. Tahapan penelitian dilakukan dengan pengumpulan data 68 status (3400 komentar) selama masa kampanye, dengan kegiatan preprosesing pembobotan token, klasifikasi, dan pengujian menggunakan confusion matrix.

Dari hasil implementasi dan pengujian, metode Naive Bayes memiliki tingkat akurasi klasifikasi sentimen mencapai lebih dari 83%.

Penelitian menurut Manjusha K. K (2014) menyatakan bahwa metode Naïve Bayes juga digunakan untuk memprediksi penyakit Dermatology yang diabaikan tetapi dapat menyebabkan kematian dimana metode Naïve Bayes digunakan untuk mengenal pola data dan mengungkap kemungkinan penyakit dermatology. Selain itu (S.L. Ting, 2011) metode Naïve Bayes juga dinilai berpotensi baik dalam mengklasifikasi dokumen dibandingkan metode pengklasifikasian yang lain dalam hal ini akurasi dan efisiensi saat komputasi..

Penelitian yang dilakukan Kumara dan Supriyanto untuk klasifikasi data mining untuk Penerimaan Seleksi Calon Pegawai Negeri Sipil 2014 Menggunakan Algoritma Decision Tree C4.5 Hasil tingkat keberhasilan prediksi calon pegawai baru secara keseluruhan yang telah diukur menggunakan metode ten -fold cross validation adalah sebesar 71%.

Berdasarkan uraian diatas, penelitian ini akan dikembangkan sebuah perangkat lunak untuk *opinion mining* pada status terhadap suatu produk di media sosial Facebook menggunakan algoritma *naïve bayes* dan C 4.5 dan mengetahui algoritma mana yang menghasilkan klasifikasi status yang cepat dan akurat.

1.3 Rumusan Masalah

Berdasarkan dari latar belakang diatas, maka rumusan masalah pada penelitian ini adalah bagaimana hasil perbandingan antara algoritma *naïve bayes* dan algoritma C4.5 dalam klasifikasi sentimen status pengguna *facebook* .

1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah :

1. Untuk mengetahui tingkat akurasi algoritma *naive bayes* dan C4.5;
2. Untuk mengetahui waktu yang digunakan dalam proses klasifikasi pada masing-masing algoritma.

1.5 Manfaat Penelitian

Manfaat dari penelitian ini adalah :

1. Dapat membantu pengguna dalam mengetahui status mana yang mengandung hal positif, negatif, dan netral menggunakan algoritma *naive bayes* dan algoritma c.45;
2. Menambah ilmu pengetahuan, menjadi bahan referensi dan perbandingan untuk penelitian yang berkaitan dengan penggunaan metode *naive bayes*, c4.5 dan sentimen analisis.
3. Mengetahui perbandingan metode *naive bayes* dan C4.5 sehingga mendapatkan metode yang lebih baik untuk melakukan klasifikasi.

1.6 Batasan Masalah

Batasan masalah dalam penelitian ini adalah :

1. Data yang akan diolah dibatasi hanya status pengguna facebook.;
2. Pembobotan kata menggunakan metode TF-IDF;
3. Status yang akan diolah adalah status dalam bahasa indonesia.

DAFTAR PUSTAKA

- Beel, Joeran and Langer, Stefan and Gipp, B. (2017). TF-IDuF: A Novel Term-Weighting Scheme for User Modeling based on Users' Personal Document Collections. *Proceedings of the IConference 2017*, 1–7. Retrieved from <http://mr-dlib.org>
- Han, J., Pei, J., & Kamber, M. (2011). *Data mining: concepts and techniques*: Elsevier.
- HSSINA, B., MERBOUHA, A., EZZIKOURI, H., & ERRITALI, M. (2014). A comparative study of decision tree ID3 and C4.5. *International Journal of Advanced Computer Science and Applications*, 4(2), 13–19. <https://doi.org/10.14569/SpecialIssue.2014.040203>
- Indranandita, A., Susanto, B., & Rahmat, A. (2011). Sistem Klasifikasi dan Pencarian Jurnal dengan Menggunakan Metode Naive Bayes dan Vector Space Model. *Jurnal Informatika*, 4(2). Engineering and Technology (ICAIET), 2014 4th International Conference on.
- Patil, T. R. (2013). Performance Analysis of Naive Bayes and J48 Classification Algorithm for Data Classification. *International Journal Of Computer Science And Applications*, ISSN: 0974-1011, 6(2), 256–261. <https://doi.org/ISSN: 0974-1011>
- Quinlan, J. R. (1993). J. Ross Quinlan_C4.5_ Programs for Machine Learning.pdf. *Morgan Kaufmann*. <https://doi.org/10.1007/BF00993309>
- Ridwan, M., Suyono, H., dan Sarosa, M. 2013. Penerapan Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier. Eeccis, Vol.7 No.1, 59–64. Rodiyansyah, S.F., dan Winarko, E., 2013. Klasifikasi Posting Twitter Kemacetan Lalu Lintas Kota Bandung Menggunakan Naive Bayesian Classification. IJCCS, Vol.7, No.1, January 2013, ISSN: 1978-1520
- Sartika, D., dan Sensuse, D. I. 2017. Perbandingan Algoritma Klasifikasi Naive Bayes, Nearest Neighbour, dan Decision Tree pada Studi Kasus Pengambilan Keputusan Pemilihan Pola Pakaian. Jatisi, Vol.1 No.2, 151– 161.

Sebastiani, F. (2005). Text Categorization. *Text Mining and Its Applications to Intelligence, CRM and Knowledge Management*, (MI), 109–123.
<https://doi.org/10.1.1.105.1540>

Sumathi, S., & Esakkirajan, S. (2007). *Fundamentals of Relational Database Management Systems* (Vol. 47). <https://doi.org/10.1007/978-3-540-48399-1>

Taheri, S., Mammadov, M., & Bagirov, A. M. (2010). Improving Naive Bayes Classifier Using Conditional Probabilities, 63–68.

Wong, T., & Yang, N. (2017). Dependency Analysis of Accuracy Estimates in k-fold Cross Validation, 4347(c), 1–12.
<https://doi.org/10.1109/TKDE.2017.2740926>