

**EFEKTIFITAS ALGORITMA *NAÏVE BAYES* DALAM MENDETEKSI PESAN SPAM
DALAM *SHORT MESSAGE SERVICE SMARTPHONE***

SKRIPSI

Oleh :
MUHAMMAD FACHRIZA
NIM. 15650050



**JURUSAN TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2022**

**EFEKTIFITAS ALGORITMA *NAÏVE BAYES* DALAM MENDETEKSI
PESAN *SPAM* DALAM *SHORT MESSAGE SERVICE SMARTPHONE***

SKRIPSI

Diajukan kepada:
Universitas Islam Negeri Maulana Malik Ibrahim Malang
Untuk memenuhi Salah Satu Persyaratan dalam
Memperoleh Gelar Sarjana Komputer (S.Kom)

Oleh :
MUHAMMAD FACHRIZA
NIM. 15650050

**JURUSAN TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2022**

HALAMAN PERSETUJUAN

**EFEKTIFITAS ALGORITMA *NAÏVE BAYES* DALAM
MENDETEKSI PESAN SPAM DALAM *SHORT
MESSAGE SERVICE SMARTPHONE***

SKRIPSI

Oleh :
MUHAMMAD FACHRIZA
NIM. 15650050

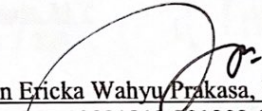
Telah Diperiksa dan Disetujui untuk Diuji
Tanggal : 15 Juni 2022

Dosen Pembimbing I



Dr. Cahyo Crysdian, MCS
NIP. 19740424 200901 1 008


Dosen Pembimbing II



Johan Ericka Wahyu Prakasa, M.Kom
NIP. 19831213 201903 1 004

Mengetahui,
Ketua Jurusan Teknik Informatika
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang




Dr. Fachrul Kurniawan ST., M.MT., IPM
NIP. 19771020 200912 1 001

HALAMAN PENGESAHAN

EFEKTIFITAS ALGORITMA *NAÏVE BAYES* DALAM MENDETEKSI PESAN SPAM DALAM *SHORT* *MESSAGE SERVICE*

SKRIPSI

Oleh :

MUHAMMAD FACHRIZA

NIM. 15650050

Telah Dipertahankan di Depan Dewan Penguji Skripsi
dan Dinyatakan Diterima Sebagai Salah Satu Persyaratan
Untuk Memperoleh Gelar Sarjana Komputer (S.Kom)
Tanggal : 16 Juni 2022

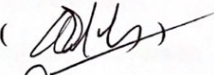
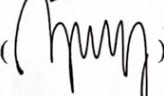

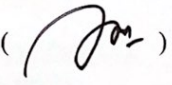
Susunan Dewan Penguji

Penguji Utama : Fajar Rohman Hariri, M.Kom
NIP. 19890515 201801 1 001

Ketua Penguji : Agung Teguh Wibowo Almais, M.T
NIDT. 19860103 20180201 1 235


Sekretaris Penguji : Dr. Cahyo Crysdian
NIP. 19740424 200901 1 008

Anggota Penguji : Johan Ericka Wahyu Prakasa, M.Kom
NIP. 19831213 201903 1 004

()
()
()
()

Mengetahui,
Ketua Jurusan Teknik Informatika
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang




Dr. Fachrudin Kurniawan ST., M.MT., IPM
NIP. 19771020 200912 1 001

PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan di bawah ini:

Nama : Muhammad Fachriza
NIM : 15650050
Fakultas / Jurusan : Sains dan Teknologi / Teknik Informatika
Judul Skripsi : Efektifitas Algoritma *Naïve Bayes* Dalam Mendeteksi Pesan
Spam Dalam *Short Message Service Smartphone*

Menyatakan dengan sebenarnya bahwa Skripsi yang saya tulis ini benar-benar merupakan hasil karya saya sendiri, bukan merupakan pengambil alihan data, tulisan, atau pikiran orang lain yang saya akui sebagai hasil tulisan atau pikiran saya sendiri, kecuali dengan mencantumkan sumber cuplikan pada daftar pustaka.

Apabila dikemudian hari terbukti atau dapat dibuktikan skripsi ini merupakan hasil jiplakan, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Malang, 23 Juni 2022

Yang membuat pernyataan,



Muhammad Fachriza
NIM. 15650050

MOTTO

“Berakit-rakit ke hulu, berenang-renang ke tepian.”

HALAMAN PERSEMBAHAN

الْحَمْدُ لِلَّهِ رَبِّ الْعَالَمِينَ

Puji syukur atas kehadiran Allah SWT, dengan mengucap Alhamdulillah penulis mempersembahkan sebuah karya untuk orang-orang yang sangat berarti.

Terima kasih penulis ucapkan kepada kedua Orang Tua yang selalu memberikan cinta, kasih sayang, motivasi, doa, harapan dan pengorbanan yang luar biasa untuk memberikan yang terbaik dalam segala hal di kehidupan penulis, Ibu Syumarni Lubis dan Bapak Edwin, serta ketiga saudara dan keluarga besar yang selalu memberikan arahan yang terbaik.

Terima kasih pula penulis ucapkan kepada Bapak Cahyo Crysdiyan selaku dosen pembimbing pertama dan Johan Ericka Wahyu Prakasa selaku pembimbing kedua yang selalu sabar, serta sat set dalam memberikan respon dalam bimbingan hingga penelitian ini dapat terselesaikan dengan lancar. Serta kepada Bapak Fajar dan Bapak Agung selaku dewan penguji dalam sidang skripsi yang mana telah memberikan koreksi dalam penelitian ini sehingga dapat menjadikan penelitian ini lebih baik.

Tidak lupa terima kasih penulis ucapkan kepada rekan-rekan seperjuangan jurusan Teknik Informatika 2015 UIN Maulana Malik Ibrahim Malang yang telah menemani dan mau untuk direpoti dengan pertanyaan-pertanyaan seputar penelitian ini.

Terima kasih juga untuk orang-orang yang tidak dapat disebutkan satu per satu yang telah memberikan motivasi, semangat dan doa yang tiada henti hingga penelitian ini dapat terselesaikan.

KATA PENGANTAR

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

Alhamdulillah rabbil 'alamiin, segala puji bagi Allah SWT, yang telah memberikan karunia, rahmat dan hidayahnya. Sehingga memberikan kemudahan dalam proses penyusunan skripsi dengan judul “**Efektifitas Algoritma Naïve Bayes Dalam Mendeteksi Pesan Spam Dalam Short Message Service Smartphone**” dengan lancar dan baik. Sholawat serta salam semoga senantiasa tersampaikan kepada Nabi Muhammad SAW yang memberikan syafaat dari zaman jahiliyah menuju zaman yang penuh berkah.

Penulis menyadari banyak keterbatasan yang penulis miliki, sehingga banyak pihak yang telah memberikan bantuan baik moril maupun materil dalam proses menyelesaikan penelitian ini. Maka dari itu dengan segenap kerendahan hati penulis mengucapkan terima kasih kepada :

1. Prof. Dr. H. M. Zainuddin, MA selaku Rektor UIN Maulana Malik Ibrahim Malang.
2. Dr. Sri Harini M.Si selaku Dekan Fakultas Sains dan Teknologi UIN Maulana Malik Ibrahim Malang.
3. Dr. Fachrul Kurniawan ST., M.MT ., IPM selaku Ketua Jurusan Teknik Informatika Fakultas Sains dan Teknologi UIN Maulana Malik Ibrahim Malang.
4. Dr. Cahyo Crysdian, MCS selaku dosen pembimbing pertama dan Bapak Johan Ericka Wahyu Prakasa selaku dosen pembimbing kedua yang senantiasa sabar dan berkenan meluangkan waktunya untuk membimbing dan memberikan arahan.

5. Seluruh dosen dan staf jurusan Teknik Informatika Fakultas Sains dan Teknologi UIN Maulana Malik Ibrahim Malang yang telah memberikan ilmu dan pengalaman yang bermanfaat.
6. Segenap civitas akademik Jurusan Teknik Informatika Fakultas Sains dan Teknologi UIN Maulana Malik Ibrahim Malang.
7. Kedua Orang Tua dan seluruh keluarga besar yang senantiasa mendukung dan mendoakan.
8. Rekan-rekan dan sahabat seperjuangan Jurusan Teknik Informatika 2015 Fakultas Sains dan Teknologi UIN Maulana Malik Ibrahim Malang.

Penulis menyadari dalam karya ini masih banyak kekurangan. Oleh karena itu penulis selalu menerima segala kritik dan saran dari pembaca. Semoga karya ini dapat bermanfaat dan dipergunakan mestinya bagi seluruh pihak.

Malang, 29 Juni 2022

Penulis

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PENGAJUAN	ii
HALAMAN PERSETUJUAN	iii
HALAMAN PENGESAHAN	iv
PERNYATAAN KEASLIAN TULISAN	v
MOTTO	vi
HALAMAN PERSEMBAHAN	vii
KATA PENGANTAR	viii
DAFTAR ISI	x
DAFTAR GAMBAR	xi
DAFTAR TABEL	xii
ABSTRAK	xiii
ABSTRACT	xiv
نبذة مختصرة	xv
BAB I PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Pernyataan Masalah	3
1.3. Tujuan Penelitian	3
1.4. Manfaat Penelitian	3
1.5. Batasan Penelitian	3
BAB II STUDI PUSTAKA	5
BAB III DESAIN SISTEM	12
3.1. Sumber Data	12
3.2. Desain Sistem	12
3.2.1. Input	13
3.2.2. Proses dan Output	13
BAB IV UJI COBA DAN PEMBAHASAN	20
4.1. Skenario Uji Coba	20
4.2. Hasil Uji Coba	21
4.3. Pembahasan	39
4.3.1 Perspektif Informatika	39
4.3.2. Integrasi Islam	44
BAB V PENUTUP	46
5.1. Kesimpulan	45
5.2. Saran	45
DAFTAR PUSTAKA	

DAFTAR GAMBAR

Gambar 3.1.	Desain Alur Sistem	13
Gambar 3.2.	Pseudocode Trimming	14
Gambar 3.3.	Contoh Hasil Algoritma Trimming.....	14
Gambar 3.4.	Pseudocode Case Folding	15
Gambar 3.5.	Contoh Hasil Algoritma Case Folding.....	15
Gambar 3.6.	Pseudocode Stemming	15
Gambar 3.7.	Contoh Hasil Algoritma Stemming.....	16
Gambar 3.8.	Pseudocode Tokenizing	16
Gambar 3.9.	Contoh Hasil Algoritma Tokenizing.....	17
Gambar 4.1	Gambar Tabulasi Confusion Matrix.....	20

DAFTAR TABEL

Tabel 2.1. Dokumentasi Kegiatan Fakultas TIK Universitas Semarang	5
Tabel 4.1. Tabel Jumlah Kata 300 Pesan Teruji	22
Tabel 4.2. Tabel Jumlah Kata 600 Pesan Teruji	22
Tabel 4.3. Tabel Hasil Uji Coba Pesan Masuk	22
Tabel 4.4. Tabel Bobot Kata Pesan Pertama Himpunan 300.....	24
Tabel 4.5. Tabel Bobot Kata Pesan Pertama Himpunan 600.....	25
Tabel 4.6. Tabel Bobot Kata Pesan Kedua Himpunan 300.....	26
Tabel 4.7. Tabel Bobot Kata Pesan Kedua Himpunan 600.....	27
Tabel 4.8. Tabel Bobot Kata Pesan Kelima Himpunan 300	28
Tabel 4.9. Tabel Bobot Kata Pesan Kelima Himpunan 600	28
Tabel 4.10. Tabel Bobot Kata Pesan Kedelapan Himpunan 300.....	29
Tabel 4.11. Tabel Bobot Kata Pesan Kedelapan Himpunan 600.....	30
Tabel 4.12. Tabel Bobot Kata Pesan Kesepuluh Himpunan 300.....	31
Tabel 4.13. Tabel Bobot Kata Pesan Kesepuluh Himpunan 600.....	32
Tabel 4.14. Tabel Bobot Kata Pesan Ketiga Himpunan 300	33
Tabel 4.15. Tabel Bobot Kata Pesan Masuk Himpunan 600	34
Tabel 4.16. Tabel Bobot Kata Pesan Keempat Himpunan 300.....	35
Tabel 4.17. Tabel Bobot Kata Pesan Keempat Himpunan 600.....	35
Tabel 4.18. Tabel Bobot Kata Pesan Keenam Himpunan 300.....	35
Tabel 4.19. Tabel Bobot Kata Pesan Keenam Himpunan 600.....	36
Tabel 4.20. Tabel Bobot Kata Pesan Ketujuh Himpunan 300	37
Tabel 4.21. Tabel Bobot Kata Pesan Ketujuh Himpunan 600	38
Tabel 4.22. Tabel Penilaian Pesan	39

ABSTRAK

Fachriza, Muhammad. 2022. **Efektifitas Algoritma *Naïve Bayes* Dalam Mendeteksi Pesan Spam Dalam *Short Message Service SMS***. Skripsi. Jurusan Teknik Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang. Pembimbing: (I) Dr. Cahyo Crysdiyan, MCS (II) Johan Ericka Wahyu Prakasa, M.Kom.

Kata Kunci : *Naïve Bayes, Text Mining, Deteksi Spam, Short Message Service, Laplace Smoothing*

Dengan perkembangan teknologi yang pesat, akses informasi saat ini juga semakin mudah serta lebih cepat diakses. Kemudahan penyebaran informasi memiliki dampak yang baik untuk masyarakat. Seiring dengan dampak positif yang dimiliki oleh kemudahan akses informasi, perkembangan tersebut juga memiliki dampak negative bagi masyarakat. Salah satunya adalah penyebaran informasi yang bertujuan menyimpang. Contoh dari penyebaran menyimpang tersebut adalah penyebaran pesan spam khususnya pada perangkat *smartphone* yang dimiliki oleh kebanyakan masyarakat.. Penyebaran informasi spam tersebut juga dapat dihindari dengan teknologi informasi. Penelitian ini memanfaatkan teknologi informasi dalam melakukan klasifikasi pesan-pesan mana yang merupakan pesan yang berisi informasi spam. Teknik Klasifikasi menggunakan algoritma *Naïve Bayes* yang nantinya akan mengkategorikan pesan-pesan ke dalam 3 kategori, yaitu pesan normal, pesan promosi, dan juga pesan spam. Pada pengujian di penelitian ini, 50 buah pesan masuk diuji dengan membandingkan masing-masing pesan kepada himpunan 300 pesan yang sudah teruji dan 600 pesan yang sudah teruji, lalu hasilnya dibandingkan. Untuk pengujian terhadap himpunan 300 pesan teruji, akurasi yang didapatkan sebesar 89% sedangkan untuk pengujian pesan terhadap 600 pesan teruji menghasilkan akurasi sebesar 86%.

ABSTRACT

Fachriza, Muhammad. 2022. **Efektifitas Algoritma *Naïve Bayes* Dalam Mendeteksi Pesan Spam Dalam Short Message Service SMS**. Skripsi. Jurusan Teknik Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang. Pembimbing: (I) Dr. Cahyo Crysdian, MCS (II) Johan Ericka Wahyu Prakasa, M.Kom.

Keywords : *Naïve Bayes, Text Mining, Spam Detection, Short Message Service, Laplace Smoothing*

Along with rapid development of technology, information also spread easier and faster. The rapid spread of information has some good impact towards society. Along with those good impact, some negative also comes along with it. One of them is the spread of false information or some straight up false news. One of the example is the spread of spam messages especially in smartphone. The spread of spam information can be avoided with information technology. This research use *naïve bayes* as classification technique. The purpose of the research is to be able to categorize incoming message into three specifics category, which is normal message, promotion message, and scam message. From the research of 50 incoming message towards 300 set of tested message, the accuration of algorithm reaches 89% while experiment towards 600 set of tested meesage, the accuaration reaches 86%.

نبذة مختصرة

SMS. فعالية خوارزمية ساذج بايز في الكشف عن الرسائل غير المرغوب فيها في خدمة الرسائل القصيرة

زكريا , محمد لطفي. 2022.

البحث العلمي. قسم تقنية المعلومات، كلية العلوم والتكنولوجيا، جامعة مولانا مالك إبراهيم الإسلامية الحكومية بمالانج.

المشرفة : (1) طبيب كاهيو كريسديان ، (2) يوهان إريكا الوحي براكاسا.

الكلمات الرئيسية : الخلجان الساذجة ، استخراج النصوص ، اكتشاف الرسائل غير المرغوب فيها ، الرسائل القصيرة ،

تجانس لا بلاس

التطور السريع للتكنولوجيا يجعل الوصول إلى المعلومات أسهل. سهولة الوصول إلى المعلومات تجعل الرسائل غير المرغوب فيها أسهل في الانتشار

البحث يخلق وسيلة لتصنيف الرسائل الواردة لتجنب الرسائل غير المرغوب فيها مع *Naive Bayes Classifier* سجلت أبحاث الرسائل الواردة ما يصل إلى 89٪

BAB I

PENDAHULUAN

1.1. Latar Belakang

Sejak diberlakukannya Peraturan Menteri Komunikasi Nomor 1 Tahun 2019, penggunaan media telekomunikasi diperketat sehingga segala jenis bentuk telekomunikasi yang dikategorikan *spam* mulai dilarang di Indonesia. Peraturan tersebut memungkinkan berkurangnya peredaran penyalahgunaan penyebaran media komunikasi di Indonesia, terutama dalam bentuk pesan singkat elektronik.

Pesan singkat elektronik atau SMS (Short Message Service) merupakan median penyebaran informasi dan komunikasi yang hingga kini masih sangat banyak digunakan. SMS dapat ditemukan di perangkat elektronik telepon genggam. Tujuan penggunaannya memudahkan komunikasi dua arah antara perangkat satu dengan yang lainnya, dengan menggunakan teks sebagai media informasi.

Salah satu peran SMS pada era digital adalah promosi dan iklan. Iklan yang dikirim lewat sms menjadi efisien daripada cara konvensional seperti pemasangan iklan di jalan raya. Target iklan dengan media SMS juga menjadi lebih cepat dengan target yang lebih eksklusif. Meski begitu berbagai kemudahan yang didapatkan lewat media SMS dapat menjadi gangguan bagi salah satu pihak, yaitu dengan adanya spam.

Spam atau spamming merupakan penggunaan fitur pesan dengan tujuan untuk mengirim pesan dengan jumlah banyak. Seringkali spam digunakan untuk menyebarkan iklan, serta penyebaran informasi komersil lainnya. Meski penggunaan pesan berantai tidak dikategorikan dalam tindakan kejahatan, penggunaan spam secara khusus meresahkan banyak pengguna teknologi informasi karena jumlah, asal, serta tujuan dari pesan tersebut sering tidak jelas. Banyak pengguna teknologi informasi juga meresahkan jumlah spam yang didapat serta bagaimana setiap pengguna tersebut terjebak dalam rantai pesan spam. Tidak sedikit konten dari spam tersebut menyentuh hal-hal ilegal, diantaranya adalah penipuan dan penjualan obat-obatan terlarang.

Spam merupakan masalah yang terjadi di seluruh dunia. Pada tahun 2020, *Truecaller* mencatat setidaknya ada 31.3 Miliar panggilan spam yang tersebar di seluruh dunia. Jenis-jenis spam tersebut terangkai dari penipuan, jasa keuangan, operator, hingga jasa telemarketing.

Keberadaan spam membuat perlunya ketersediaan jasa untuk menghindari spam-spam tersebut. Salah satu fitur spam yang paling banyak digunakan adalah lewat fitur sms. Spam dalam fitur sms dapat dihindari dengan menggunakan *filter* yang terdapat pada *smartphone*. Sayangnya tidak semua *smartphone* memiliki fitur filter, terlebih dalam bahasa selain inggris yang merupakan bahasa global.

Menurut penjelasan latar belakang ini, penelitian dilakukan untuk membuat layanan filter spam yang mumpuni, dengan mempertimbangkan kata-kata yang sering digunakan di dalam pesan spam berbahasa Indonesia. Penelitian didasarkan pada konsep *Naïve Bayes Classifier* untuk melakukan klasifikasi kata-kata spam dan pada akhirnya layanan filter dapat menghindari pesan spam yang masuk pada *smartphone*.

Dengan metode *Naïve Bayes*, pengguna dapat menghindari mendapatkan spam dengan melakukan penyaringan pesan-pesan yang masuk ke *smartphone* milik pengguna. Metode *Naive Bayes* memungkinkan pengguna mengkategorikan pesan spam dengan pesan yang bukan spam. Pengkategorian dilakukan dengan menghitung probabilitas setiap kata yang terdapat pada masing-masing jenis kategori pesan. Probabilitas dari masing-masing kata tersebut nantinya menjadi penentu apakah pesan yang didapat merupakan pesan spam atau bukan.

Allah telah berfirman dalam Al-Qur'an Surat Ali Imran ayat 161 tentang balasan bagi orang-orang yang berkhianat.

وَمَا كَانَ لِنَبِيٍّ أَنْ يَغُلَّ وَمَنْ يَغُلْ يَأْتِ بِمَا غَلَّ يَوْمَ الْقِيَمَةِ ۖ ثُمَّ تُوَفَّى كُلُّ نَفْسٍ مَّا كَسَبَتْ وَهُمْ لَا يُظْلَمُونَ

“Dan tidak mungkin seorang nabi berkhianat (dalam urusan harta rampasan perang). Barangsiapa berkhianat, niscaya pada hari Kiamat dia akan datang membawa

apa yang dikhianatkannya itu. Kemudian setiap orang akan diberi balasan yang sempurna sesuai dengan apa yang dilakukannya, dan mereka tidak dizalimi.” (Q.S. Ali 'Imran: 161)

Orang-orang yang berkhianat diumpamakan sebagai orang-orang yang menipu dengan melakukan penyebaran pesan spam yang kadang berisi konten penipuan. Jenis orang-orang tersebut akan mendapat balasan yang setimpal dengan apa yang diperbuatnya.

Berdasarkan latar belakang yang telah diuraikan di atas, maka tujuan penelitian ini difokuskan pada bagaimana membuat filter spam yang diterapkan pada *smartphone* pengguna. Dengan penelitian ini, maka akan diperoleh jenis-jenis pesan yang dapat dikategorikan sebagai spam dan pesan yang bukan spam.

1.2. Pernyataan Masalah

- a. Seberapa besar akurasi serta presisi dari algoritma Naïve Bayes dalam melakukan klasifikasi pesan spam.
- b. Hal apa saja yang mempengaruhi akurasi serta presisi dalam menentukan pesan spam pada algoritma Naïve Bayes.

1.3. Tujuan Penelitian

- a. Mengukur akurasi serta presisi dari algoritma Naïve Bayes dalam melakukan klasifikasi pesan spam.
- b. Menganalisa hal-hal apa saja yang mempengaruhi akurasi dan presisi dalam menentukan pesan spam pada algoritma Naïve Bayes.

1.4. Manfaat Penelitian

Penelitian dapat dimanfaatkan oleh berbagai pihak seperti perusahaan informasi dan komunikasi untuk melawan hoax dan misinformasi, serta berguna bagi akademisi informatika bidang *machine learning* sebagai material pembelajaran, dan juga pengguna *smartphone* agar terhindar dari hoax dan misinformasi.

1.5. Batasan Penelitian

Batasan pada penelitian adalah penggunaan himpunan data yang berisi kumpulan pesan berserta dengan kategori dari pesan tersebut berjumlah sebanyak 600 pesan yang berasal dari sumber bernama Yudi Wibisono, seorang pakar informatika ilmu komputer UPI. Sumber-sumber ini didapatkan dari dokumentasi kumpulan kata-kunci yang banyak terdapat pada pesan spam.

BAB II

STUDI PUSTAKA

Terdapat beberapa penelitian yang terkait dengan aplikasi yang akan dibuat dalam penelitian ini. Penelitian Putri (2017) meneliti tentang penerapan Naïve Bayes pada sistem perankingan kegiatan di fakultas TIK Universitas Semarang. Penelitian membandingkan program studi, jenis kegiatan, serta hasil kegiatan dengan tujuan untuk mencari ranking dari kegiatan yang ingin dicari nilainya. Data acuan sumber penelitian yang mendokumentasikan kegiatan-kegiatan yang ada pada Fakultas TIK di Universitas Semarang dapat dilihat pada tabel 2.1.

Tabel 2.1 Dokumentasi Kegiatan Fakultas TIK Universitas Semarang

No	Program Studi	Jenis Kegiatan	Hasil Kegiatan	Ranking
1	Teknik Informatika	Workshop Install U-Buntu Release Party 11.04	Memuaskan	Favorit
2	Teknik Informatika	Workshop Framework PHP	Memuaskan	Favorit
3	Teknik Informatika	Workshop Linux	Cukup	Tidak Favorit
4	Teknik Informatika	Workshop Sistem Sensor	Memuaskan	Favorit
5	Sistem Informasi	Workshop Website	Cukup	Tidak Favorit
6	Sistem Informasi	Festival IT Fest	Memuaskan	Favorit
7	Sistem Informasi	Workshop Mikrotik	Cukup	Tidak Favorit
8	Ilmu Komunikasi	Pelatihan Lab. TV	Cukup	Tidak Favorit
9	Ilmu Komunikasi	Festival Charity	Cukup	Tidak Favorit
10	Ilmu Komunikasi	Festival Komukino	Memuaskan	Favorit
11	Ilmu Komunikasi	Pelatihan Acara Retorika	Memuaskan	Favorit
12	Ilmu Komunikasi	Festival Review TV	Memuaskan	Favorit
13	Ilmu Komunikasi	Festival Public Relation	Cukup	

Penelitian mencari nilai ranking berdasarkan informasi dari kegiatan-kegiatan sebelumnya. Dalam penelitian ini, yang dicari adalah ranking dari program studi Ilmu Komunikasi dengan jenis kegiatan Festival dan dengan hasil kegiatan bernilai cukup. Hasil dari penelitian ini adalah probabilitas dari masing-masing variabel kategori, yaitu apakah kegiatan yang ingin dicari merupakan kegiatan favorit atau kegiatan tidak favorit.

Penelitian ini hanya menggunakan dokumentasi yang sangat terbatas, sehingga informasi yang didapatkan kemungkinan tidak sesuai dengan keadaan yang seharusnya. Fitur-fitur yang membentuk kegiatan juga terbentuk dari kategori-kategori kecil, yaitu tiga kelas pada program studi dan jenis kegiatan serta 2 kelas pada hasil kegiatan. Berbeda dengan tujuan dari peneliti yang membentuk fitur dari kelas kumpulan-kumpulan kata yang terdapat pada ponsel masing-masing pengguna.

Anugroho *et al.* (2015) bertujuan untuk melakukan klasifikasi spam yang terdapat pada media elektronik email dengan menggunakan metode Naïve Bayes Classifier. Penelitian ini dibangun dengan cara mendapatkan kata kunci yang mewakili konten dari masing-masing email. Dalam penelitian, spam dikategorikan sebagai:

- Iklan, yaitu spam untuk mempromosikan suatu produk serta layanan;
- Malware;
- Phishing, yaitu bertindak sebagai lembaga palsu dengan tujuan untuk mencuri data dan informasi penting dari korban;
- *Scam*, yaitu pesan dengan tujuan penipuan;
- dan pesan yang tidak berarti, seperti pesan yang salah kirim.

Penelitian membedakan pesan spam dan pesan tidak spam berdasarkan struktur dari email, yaitu pada:

- Header, menunjukkan *address* dari pengirim dan penerima email;
- Subject, yaitu judul topic yang mewakili isi email, dan

- Body, isi dari pesan email.

Penelitian mencari kata kunci yang terdapat pada setiap struktur email tersebut dan melakukan klasifikasi kata-kata yang sering dijumpai pada email dengan jenis spam. Hasil dari penelitian tersebut cukup baik, namun jumlah kata kunci yang ada pada data training berpengaruh besar terhadap baik tidaknya sistem klasifikasi pada penelitian. Penelitian juga hanya berpengaruh pada kata-kata dalam bahasa Inggris, sehingga tidak akan berpengaruh terhadap email yang memuat kata-kata selain bahasa Inggris.

Ginting & Trinanda (2013) melakukan penelitian teknik *data mining* dengan metode Naïve Bayes untuk optimalisasi pencarian pada aplikasi perpustakaan. Penelitian menggunakan teknik data mining untuk mendapatkan informasi yang penting yang berkaitan pada sistem aplikasi perpustakaan. Informasi tersebut diantaranya adalah judul buku, kategori buku, pengarang buku, serta deskripsi dari masing-masing buku tersebut. Penelitian ini menggunakan proses *Knowledge Discovery in Database* (KDD) pada teknik data mining.

Algoritma Naïve Bayes pada penelitian mencakup dua langkah, yaitu *Learning* (Pembelajaran) serta *Classify* (Pengklasifikasian). Tujuan dari learning adalah untuk menemukan *keyword* (kata kunci) serta membentuk *vocabulary*, yang nantinya akan menjadi pembanding antara satu kata dengan kata lainnya. Proses *learning* juga akan menghitung probabilitas pada setiap kategori dan menentukan frekuensi dari setiap kata yang muncul pada setiap kategori tersebut untuk nantinya akan diklasifikasikan.

Proses *classifier* bertujuan untuk menghitung probabilitas pada setiap dokumen terhadap sekumpulan dokumen yang ada pada informasi data buku yang telah dibuat dan kemudian menentukan probabilitas kemunculan kata yang terbesar pada suatu dokumen. Kata yang memiliki persentase terbesar lah yang akan dimunculkan pada hasil pencarian. *Output* yang diharapkan pada penelitian ini adalah hasil yang buku yang diinginkan oleh user sendiri, dengan menampilkan informasi lengkap buku yaitu

judul buku, kategori buku, pengarang buku, deskripsi buku, serta dimana buku tersebut disimpan.

Syarli & Muin (2016) menggunakan metode Naïve Bayes untuk memprediksi kelulusan masuknya mahasiswa baru ke perguruan tinggi. Beberapa variabel yang dijadikan fitur oleh penelitian diantaranya adalah jurusan sekolah, pilihan pertama perguruan tinggi, pilihan kedua perguruan tinggi, nilai rata-rata dari setiap mahasiswa terkait, keterangan lulus, pilihan perguruan tinggi yang lulus, dan pilihan jurusan yang lulus. Hasil dari penelitian ini menunjukkan hanya sebanyak 6.37% data yang merupakan klasifikasi yang tidak benar.

Bustami (2014) menerapkan algoritma Naïve Bayes pada data nasabah asuransi. Tujuan dari penelitiannya adalah untuk mengetahui lacer tidaknya seorang nasabah dalam melakukan pembayaran asuransi sehingga memungkinkan perusahaan asuransi untuk menerima atau menolak calon nasabah tersebut. Informasi penelitian didapatkan dengan teknik data mining dengan proses *Knowledge Discovery in Database* (KDD). Variabel-variabel penting yang membentuk informasi yang dibutuhkan pada penelitian termasuk diantaranya adalah:

1. Jenis Kelamin, merupakan variabel jenis kelamin nasabah yang dikelompokkan dalam dua kategori yaitu laki-laki dan perempuan.
2. Usia, merupakan variabel usia nasabah yang dikelompokkan dalam tiga kategori yaitu 20-29 tahun, 30-40 tahun, dan diatas 40 tahun.
3. Status, merupakan variabel status nasabah yang dikelompokkan dalam dua kategori yaitu kawin dan belum kawin.
4. Pekerjaan, merupakan variabel pekerjaan nasabah yang dikelompokkan dalam tiga kategori yaitu PNS, Pegawai Swasta, Wiraswasta.
5. Penghasilan, merupakan variabel penghasilan dari nasabah yang dikelompokkan dalam tiga kategori yaitu 0-25 juta, 25-50 juta, dan diatas 50 juta.

6. Cara pembayaran premi, yaitu variabel cara pembayaran premi yang dikelompokkan dalam empat kategori yaitu bulanan, triwulan, semesteran, dan tahunan.
7. Masa pembayaran premi, merupakan variabel masa pembayaran premi yang dikelompokkan dalam tiga kategori yaitu 5-10 tahun, 11-15 tahun, dan diatas 15 tahun.

Variabel-variabel tersebut menjadi fitur atau atribut dalam algoritma klasifikasi Naïve Bayes. Tujuan dari penelitian adalah variabel kelancaran yang berarti tingkat kelancaran calon nasabah dalam melakukan pembayaran asuransi. Tingkat kelancaran tersebut dibagi kedalam tiga kelompok yaitu lancar, kurang lancar, dan tidak lancar.

Listiowarni & Ramadhani (2019) meneliti implementasi Naïve Bayes dengan Laplacian Smoothing untuk peminatan dan lintas minat siswa di SMAN 5 Pamekasan. Metode Naïve Bayes digabungkan dengan konsep Laplacian Smoothing dengan tujuan untuk menghindari adanya nilai 0 pada persamaan nilai fitur yang nantinya akan membuat hasil akhir dari suatu pencarian kelas bernilai 0. Data yang digunakan pada penelitian ini merupakan sebuah kesatuan data yang didapat dari instansi terkait yaitu SMAN 5 Pamekasan dengan atribut-atribut pembentuknya adalah nama siswa, nilai raport dari siswa terkait, nilai UNBK dari siswa terkait, nilai IQ dari siswa terkait, serta nilai tes pendukung lainnya. Tujuan dari penelitian agar dapat bisa mengklasifikasi murid-murid SMAN 5 Pamekasan pada program-program studi pada sekolah. Diantaranya adalah program studi IPA dan program studi IPS yang selanjutnya dari program studi tersebut dikelompokkan lagi menjadi IPS ekonomi dan IPS geografi(kebumian) dan IPA kimia dan IPA fisika. Penelitian menggunakan data siswa selama 4 tahun dari tahun 2014 hingga tahun 2017 berjumlah 720 data siswa.

Selanjutnya untuk menilai keabsahan dari model klasifikasi yang termasuk di dalamnya model klasifikasi Naïve Bayes, harus dapat diukur akurasi serta presisi dari model tersebut. Salah satu cara untuk melakukan pengukuran tersebut adalah dengan

metode *confusion matrix*. Karimi (2021) menyatakan *confusion matrix* adalah alat yang digunakan untuk analisa prediksi pada *machine learning*. Menurutnya, *confusion matrix* adalah tabel ringkas dari nilai benar dan nilai tidak benar yang dihasilkan oleh model klasifikasi.

Larner (2021) lalu menambahkan bahwa *confusion matrix* merupakan aplikasi dari tabel dengan ukuran 2 x 2 dimana bagian kolom mengindikasikan contoh dari kelas sebenarnya, sedangkan bagian baris mengindikasikan contoh dari kelas prediksi. Istilah “*confusion matrix*” berasal dari observasi bahwa tabulasi silang yang terbentuk dapat menunjukan apakah sistem keliru (*confuse*) dalam memperhatikan sistem dua kelas.

Arif *et al.* (2020) mengukur popularitas *rating* pada wilayah destinasi di Kota Malang. Penelitian ini menggunakan algoritma Naïve Bayes serta algoritma *selection sort* untuk melakukan penyortiran. Peneletian menggunakan *preprocessing text* yang terdiri dari *cleaning*, *case folding*, *stopword removal*, *stemming*, serta *tokenizing*. Beberapa fitur yang terdapat pada penelitian diantaranya adalah kategori dari destinasi tersebut, popularitas dari *tweet* yang menyangkut destinasi, serta apakah *tweet* tersebut memiliki kesan positif pada masing-masing destinasi wisata.

Santoso *et al.* (2018) menggunakan algoritma *naïve bayes classifier* untuk melakukan klasifikasi objek ruang terbuka dengan berbasis citra *google earth*. Penelitian memfokuskan bagaimana melakukan optimalisasi struktur multi model NBC yang melibatkan beberapa model NBC sedemikian sehingga dapat meningkatkan akurasi dalam klasifikasi. Tujuan dari penelitian ini adalah agar dapat membangun struktur model multi model NBC dalam melakukan klasifikasi ruang terbuka hijau. Dasar.

Permana *et al.* (2021) melakukan analisa perbandingan algoritma *decision tree*, kNN, dan *Naïve Bayes* untuk prediksi kesuksesan *start-up*. Hasil pada penelitian membuktikan algoritma *decision tree* sebagai algoritma yang memiliki nilai akurasi

paling besar dengan nilai 79,29%. Sedangkan untuk algoritma kNN memiliki nilai sebesar 66.69% serta algoritma *naïve bayes* yang memiliki nilai sebesar 79.29%.

BAB III

DESAIN DAN IMPLEMENTASI

3.1. Sumber Data

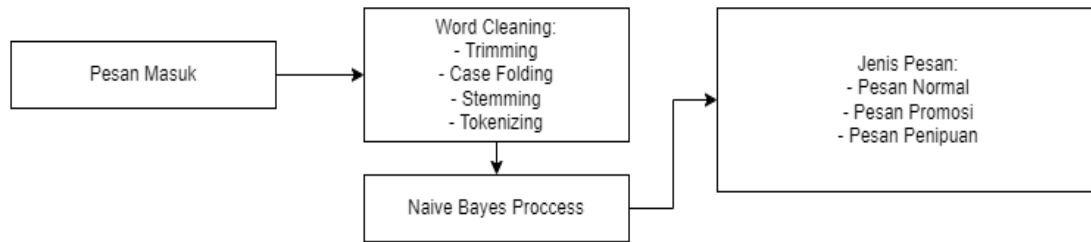
Sumber data yang digunakan untuk penelitian ini didapatkan dari kumpulan pesan-pesan singkat yang ada pada *smartphone* milik pengguna. Kumpulan pesan-pesan tersebut diambil lewat aplikasi berbasis mobile yang berintegrasi dengan sistem pesan *smartphone*. Data yang dikumpulkan membentuk dua variabel penting yang akan diolah, yaitu isi pesan serta jenis pesan. Isi pesan adalah konten dari pesan tersebut dan jenis pesan adalah kategori dari pesan dan akan menjadi acuan dari penelitian.

Sumber data telah diolah terlebih dahulu oleh pemilik data dengan dibagi ke dalam tiga kategori. Tiga kategori tersebut adalah pesan normal, pesan promo, dan pesan penipuan. Demi penelitian, distribusi dari penggunaan himpunan masing-masing kategori pesan disamakan. Dengan kata lain, sumber data yang digunakan terbagi ke dalam 200 pesan normal, 200 pesan promo, serta 200 pesan penipuan.

3.2. Desain Sistem

Aplikasi merupakan aplikasi berbasis mobile yang berfungsi untuk membedakan pesan sesuai dengan kategori pesan tersebut. Pengguna akan menerima pesan masuk pada *smartphone* milik pengguna, dan aplikasi akan secara otomatis mengkategorikan jenis pesan yang masuk berdasarkan 3 kategori. Kategori tersebut yaitu pesan normal, pesan promosi, dan pesan spam. Pengguna aplikasi dapat melihat pesan serta kategori dari masing-masing pesan tersebut.

Aplikasi dibuat dengan menggunakan bahasa pemrograman Javascript dengan menggunakan NodeJS dengan versi 14.17.0 dan diimpor ke menjadi aplikasi berbasis mobile. Data disimpan dalam format *data object*. Desain alur dari sistem yang dipakai pada aplikasi dapat dilihat pada Gambar 3.1.



Gambar 3.1 Desain Alur Sistem

3.2.1. Input

Input merupakan pesan singkat (sms) yang masuk ke smartphone pengguna lewat nomor milik pengguna smartphone. Data yang diambil dari pesan singkat tersebut adalah nomor pengirim pesan, nama pengirim pesan, serta isi dari pesan tersebut. Isi pesan tersebut nantinya akan diolah di proses pengolahan pesan sehingga pada akhirnya mendapatkan hasil kategori dari pesan tersebut.

3.2.2. Proses dan Output

Proses adalah pengolahan data pesan agar bisa diolah ke dalam komputasi algoritma Naïve Bayes. Data pesan berupa kumpulan kalimat-kalimat yang membentuk pesan tersebut. Kalimat-kalimat tersebut kemudian proses sehingga dapat diolah oleh algoritma Naïve bayes. Pengolahan kalimat-kalimat tersebut adalah proses *text mining*.

Text mining adalah kumpulan proses yang dilakukan dengan tujuan untuk mengambil pola-pola signifikan dan menarik dengan maksud untuk mencari pengetahuan yang terkait pada penelitian, dalam hal ini kaitannya dengan kategori pesan spam. Yang termasuk pada proses text mining yang dipakai pada peneltian ini diantaranya :

a. Trimming

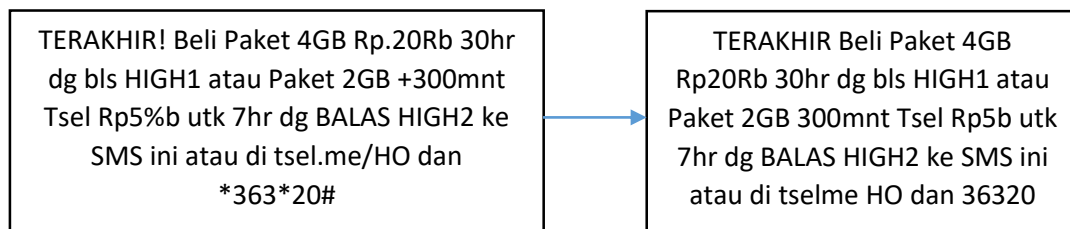
Word Trimming proses pengolahan pesan dimana setiap karakter yang tidak dibutuhkan pada pesan tersebut dihilangkan. Hal ini nantinya bertujuan untuk

memperbaiki perbandingan kata-kata dimana terdapat kata yang sama untuk dibandingkan, namun memiliki nilai yang berbeda karena adanya perbedaan karakter simbolik yang ada di setiap kata tersebut. Misalnya kata “[serius]” memiliki nilai yang berbeda dengan kata “(serius)” karena adanya perbedaan karakter simbolik pada setiap kata tersebut. *Pseudocode* untuk *trimming* dapat dilihat pada Gambar 3.2.

```
function trim_word(message) {
    message.trim(kata_kata_yang_tidak_dibutuhkan())
    return message
}
```

Gambar 3.2. Pseudocode Trimming

Contoh dari hasil penggunaan algoritma trimming dapat dilihat pada gambar 3.3.



Gambar 3.3. Contoh Hasil Algoritma Trimming

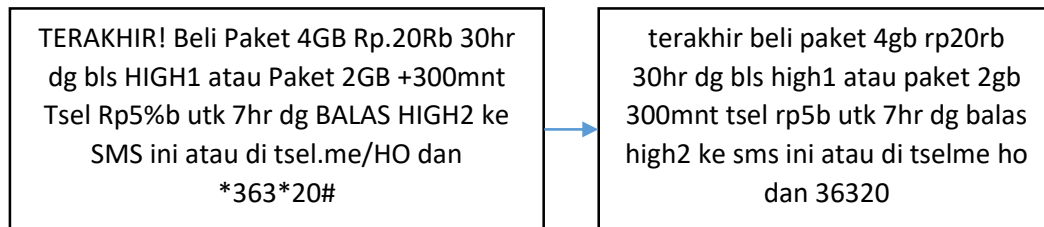
b. Case Folding

Case folding adalah proses pengolahan data dimana setiap huruf capital pada tiap pesan diubah menjadi huruf tidak capital (huruf kecil). Hal ini juga bertujuan untuk memberikan nilai yang sama pada perbandingan kata yang sama. Proses ini diperlukan karena komputer mendeteksi huruf kapital sebagai nilai yang berbeda dengan huruf yang sama tapi tidak kapital. Sebagai contoh, “SERIUS” memiliki nilai perbandingan yang berbeda dengan kata “serius” walaupun kata tersebut serupa. *Pseudocode* untuk *case folding* dapat dilihat pada Gambar 3.4.

```
function case_folding(message) {
    message.toLowerCase()
    return message
}
```

Gambar 3.4. Pseudocode Case Folding

Contoh dari hasil penggunaan algoritma trimming dapat dilihat pada gambar 3.5.



Gambar 3.5. Contoh Hasil Algoritma Case Folding

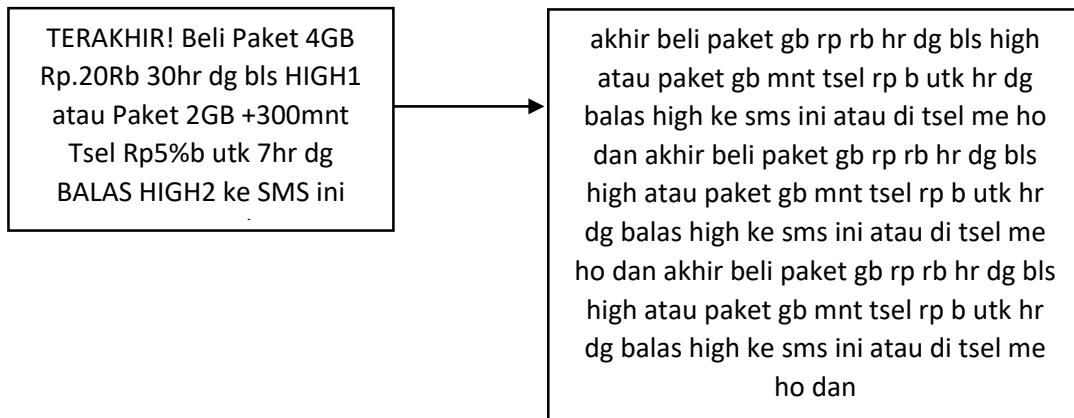
c. Stemming

Stemming adalah proses menghilangkan imbuhan pada kata pada setiap kalimat pada pesan. Sebagai contoh, kata “memakan” diubah menjadi kata makan. *Pseudocode* untuk *stemming* dapat dilihat pada Gambar 3.6.

```
function stem_word(message) {
  var stemmed = []           // menyimpan kata ke dalam array
  for (word of message) {
    stemmed.push(stemmer.stem(word))
  }
  return stemmed
}
```

Gambar 3.6. Pseudocode Stemming

Contoh dari hasil penggunaan algoritma trimming dapat dilihat pada gambar 3.7



Gambar 3.7. Contoh Hasil Algoritma Stemming

d. Tokenizing

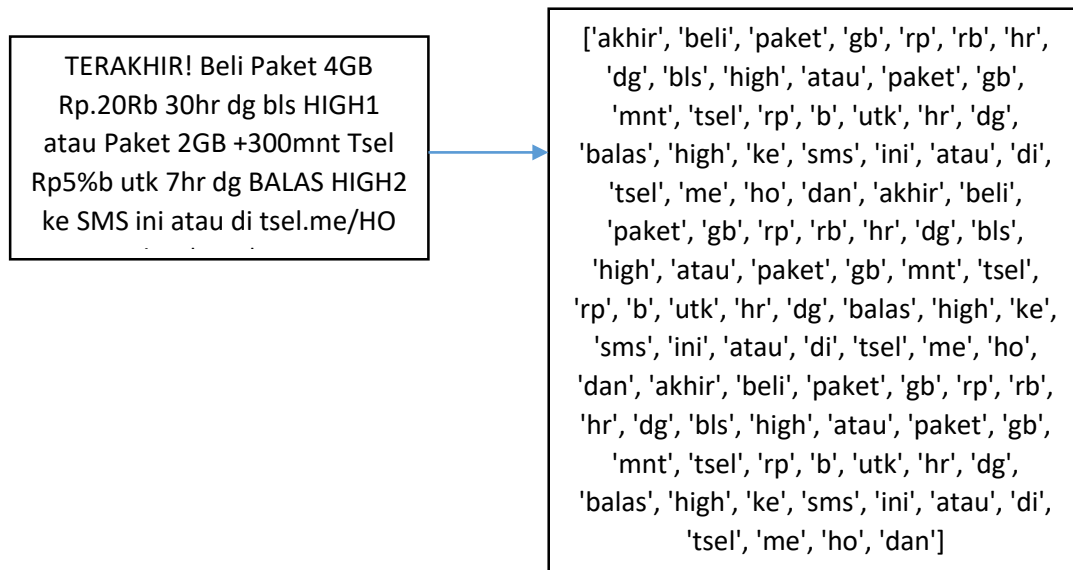
Word Tokenizing merupakan sebuah proses memisahkan kata per-kata pada setiap kalimat pada setiap pesan menjadi sebuah himpunan kata. Hal ini berguna untuk memudahkan dalam melakukan komparasi data karena data yang dikomparasikan dalam kasus ini adalah kata per-kata pada setiap kalimat pada pesan. Sebagai contoh, kalimat “saya menanam pohon” diolah menjadi himpunan dengan nilai yang berisi “saya”, “menanam”, dan “pohon”. *Pseudocode* untuk *tokenizing* dapat dilihat pada Gambar 3.8.

```

function word_tokenizing(message) {
    message.tokenize()
    return message
}
  
```

Gambar 3.8. Pseudocode Tokenizing

Contoh dari hasil penggunaan algoritma trimming dapat dilihat pada gambar 3.9.



Gambar 3.9. Contoh Hasil Algoritma Tokenizing

e. Stopword Removal

Saat berkerja pada bidang aplikasi *text mining*, sering didengar terminology seperti *stop words*. Pada bidang text mining, stop words mengacu pada kumpulan kata-kata umum yang seringkali terdapat pada data, tergantung bahasa dari penggunaan kata tersebut.

Alasan mengapa stop words begitu penting pada aplikasi yang menggunakan konsep text mining, adalah jika aplikasi menghilangkan kata-kata yang umum pada bahasa yang digunakan, maka aplikasi tersebut bisa lebih fokus kepada kata-kata yang lebih penting pada data yang hendak diolah. Khususnya pada penelitian ini, penggunaan konsep stop words dapat membuat aplikasi lebih akurat dalam melakukan pengelompokkan kategori dengan cara menghilangkan kata-kata yang umum yang terdapat pada semua kategori pesan. Kata-kata yang termasuk dalam stop words dilampirkan pada Lampiran ...

f. Naïve Bayes Process

Naïve Bayes merupakan sebuah algoritma untuk proses klasifikasi. Dalam penelitian ini, algoritma Naïve Bayes dipakai untuk melakukan klasifikasi pesan. Persamaan algoritma Naïve Bayes dapat dilihat pada Persamaan 3.1.

$$P(Y|x_i) = \prod_{i=1}^i P(x_i|Y) \cdot P(Y) \quad (3.1)$$

Pada persamaan 3.1, Y adalah kelas klasifikasi yang merepresentasikan masing-masing kategori dari pesan, yaitu pesan normal, pesan promosi, dan pesan spam. Simbol x_i adalah fitur atau atribut yang merepresentasikan kata-kata yang ada pada kumpulan pesan. Fitur atau atribut tersebut didapatkan setelah proses tokenisasi, yang bertujuan untuk mendapatkan kumpulan kata-kata yang memenuhi variabel pada persamaan tersebut. Implementasi variabel-variabel persamaan algoritma Naïve Bayes terhadap kasus penelitian dapat dilihat pada Persamaan 3.2.

$$\begin{aligned} P(\text{Pesan Normal}|Kata_i) &= \prod_{kata=1}^{kata} P(kata | \text{Pesan Normal}) \cdot P(\text{Pesan Normal}) \\ P(\text{Pesan Promosi}|Kata_i) &= \prod_{kata=1}^{kata} P(kata | \text{Pesan Promosi}) \cdot P(\text{Pesan Promosi}) \\ P(\text{Pesan Penipuan}|Kata_i) &= \prod_{kata=1}^{kata} P(kata | \text{Pesan Penipuan}) \cdot P(\text{Pesan Penipuan}) \end{aligned} \quad (3.2)$$

Setelah dicari masing-masing nilai probabilitas dari masing-masing persamaan, maka nilai terbesar diantara probabilitas tersebut merupakan jenis pesan yang merepresentasikan pesan yang dihitung. Output dari penelitian adalah pesan dengan nilai probabilitas paling besar yang sesuai dengan jenis-jenis pesan tersebut, yaitu antara pesan normal, pesan promo, ataupun pesan penipuan.

g. Zero Frequency Problem

Zero Frequency Problem (ZFP) adalah sebuah kondisi dimana terdapat nilai nol pada salah satu probabilitas fitur pada persamaan Naïve Bayes. Dalam kasus penelitian ini, berarti terdapat sebuah kata yang tidak terdapat pada salah satu kategori pesan normal, pesan promosi, ataupun pesan penipuan. Saat kondisi tersebut terjadi, maka nilai dari probabilitas kategori pesan akan menjadi nol. Untuk menghindari hal tersebut, maka digunakan metode *smoothing* yang disebut dengan *additive smoothing*. Metode ini menambahkan nilai satu pada seluruh atribut fitur pada setiap kelas, sehingga salah satu nilai atribut yang bernilai nol tadi sekarang memiliki nilai dan menghindari hasil akhir yang bernilai nol. Persamaan serta analisis untuk ZFP terhadap kasus penelitian dapat dilihat pada Persamaan 3.3.

$$\begin{aligned}
 P(x_i|Y) &= \frac{P((x + 1) \cap P(Y))}{P(Y)} \\
 P(kata|jenis\ pesan) &= P((kata + 1) \cap P(Y))/P(Y)
 \end{aligned}
 \tag{3.3}$$

BAB IV

UJI COBA DAN PEMBAHASAN

4.1. Skenario Uji Coba

Dalam penelitian ini, uji coba menggunakan 20 buah data pesan yang diambil dari data *smartphone*. Data-data ini nantinya akan diolah dan diklasifikasi dengan metode Naïve Bayes dan hasilnya diteliti serta dibandingkan *ground truth*. Hasil perbandingan nantinya akan diukur dengan menggunakan acuan *Confusion Matrix*. Acuan tersebut mencakup perhitungan *Accuracy*, *Precision*, *Recall*, serta *Specifity* dari desain sistem.

		True Class	
		Positive	Negative
Predicted Class	Positive	TP	FP
	Negative	FN	TN

Gambar 4. 1 Gambar Tabulasi Confusion Matrix

50 Pesan tersebut dianalisa dengan menggunakan acuan Confusion Matrix. Acuan Confusion Matrix menganalisa *Accuracy*, *Precision*, *Recall*, dan *Specifity* dan metode Naïve Bayes pada penelitian ini. Variabel-variabel yang menjadi nilai dari Confusion Matrix diantaranya adalah *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative*.

a. True Positive

True Positive (TP) adalah kondisi dimana nilai prediksi sama dengan nilai *ground truth* (data yang teruji). Pada penelitian, kondisi true merujuk pada keabsahan informasi dan kondisi positive merujuk kepada keabsahan prediksi. Pada penelitian, jika *ground truth* pesan sama dengan prediksi pesan, maka pesan mendapatkan nilai TP.

b. False Positive

False Positive (FP) adalah kondisi dimana nilai prediksi berbeda dengan nilai *ground truth*. Pada penelitian yang menggunakan tiga kategori kelas, dua jenis prediksi yang berbeda dengan *ground truth* dikategorikan sebagai FP.

c. True Negative

Sama dengan TP, True Negative adalah kondisi dimana nilai prediksi yang tidak benar sama dengan nilai *ground truth* yang memiliki nilai yang tidak benar (tidak absah).

d. False Negative

False Negative (FN) merujuk kepada kasus yang memiliki kondisi yang sama dengan FP. Pada penelitian yang menggunakan tiga kelas yang memiliki satu nilai positive. Prediksi pada nilai negative lainnya termasuk ke dalam kategori FN.

Untuk scenario uji coba, 50 buah pesan masuk pada smartphonte akan diuji. Untuk penelitian, pesan-pesan tersebut akan diuji ke beberapa skenario jumlah pembanding pesan, yaitu ke 300 pesan teruji, 450 contoh pesan teruji, dan 600 contoh pesan teruji. Skenario ini dilakukan agar dapat mencari tahu apakah jumlah pesan pembanding mempengaruhi akurasi dari prediksi algoritma pada penelitian.

4.2. Hasil Uji Coba

Berikut adalah hasil uji coba metode Naïve Bayes yang diaplikasikan pada 50 buah pesan dari data smartphone yang dibandingkan dengan 300 pesan teruji, dan 600 pesan teruji. Tabel Hasil Uji Coba Pesan Masuk serta tabel himpunan 300 pesan teruji dan himpunan 600 pesan teruji dapat dilihat pada lembar lampiran. Untuk jumlah kata pada pengolahan data pesan teruji, dapat dilihat pada Tabel 4.1.

Tabel 4. 1. Tabel Jumlah Kata 300 Pesan Teruji

Jenis Pesan	Jumlah
Pesan Normal	842
Pesan Promo	1656
Pesan Penipuan	1675
Jumlah Kata	4173

Tabel 4. 2. Tabel Jumlah Kata 600 Pesan Teruji

Jenis Pesan	Jumlah
Pesan Normal	1595
Pesan Promo	3207
Pesan Penipuan	3412
Jumlah Kata	8214

Algoritma naïve bayes memuat data tertinggi dari perhitungan kalkulasi persamaan. Nilai dari hasil algoritma Naïve Bayes baik untuk himpunan 300 pesan teruji dan himpunan 600 pesan teruji dapat dilihat pada lembar lampiran. Selanjutnya, dari hasil uji coba tersebut, terdapat 8 buah prediksi dari himpunan 300 pesan teruji yang tidak akurat dan 10 buah prediksi dari himpunan 600 pesan teruji yang tidak akurat. Berikut adalah tabel perbedaan kategori pesan yang dapat dilihat pada Tabel 4.3.

Tabel 4.3. Tabel Hasil Uji Coba Pesan Masuk

No	Pesan	Ground Truth	Prediksi	
			300 Pesan	600 Pesan

1	Aplikasi Uang Terbaru Di Indonesia ! Tebak Skor Salah Sepak Bola Investasi Keuangan yang Paling Bagus ! Buruan Download Dan Hub cs di https://bit.ly/3HkJ002	penipuan	normal	normal
2	Sebentar lagi Liga Champion loh !! Pilih tim favoritmu di AFAPo-Ker cutt.ly/AF4Po-Ker Nonton santai sambil profit !	penipuan	normal	normal
3	Datang dan dapatkan penghasilan tambahan Anda, hingga 200RP per hari, hubungi kami dan biarkan kami memberi tahu Anda cara menghasilkan cutt.ly/VUmCTr6	penipuan	normal	normal
4	nabilah maulida agar hub sayaYes, nah sebelum itu nanti ada gang sebelah kanan jalan	normal	penipuan	penipuan
5	MODEM TANPA KABEL! Praktis dibawa kemana aja. Harga paket mulai 20K. Dapatkan ORBIT di konter GT RELOAD LOWOKWARU. Info: bit.ly/TELKOMSELORBIT	promo	normal	normal
6	No HP 081217299196, COIN PULSA Hari Ini: +1000. Bisa Kamu Ambil di *500*220#. Bisa Kamu Tukar PULSA+KUOTA. SEGERA Hub *500*220# dari HP ini utk Ambil REWARD mu!	penipuan	promo	promo
7	Nomor HP 081217299196, COIN PULSA Anda: +2000. SEGERA Ambil di *500*75# Sebelum Hangus!	penipuan	penipuan	promo
8	Proses Depo & WD secepat kilat di T063LMANDIRI dik.si/Ob1NG Main Games b4ccar4t,P@K3R,Ce'Me,To_Gel japan, sydney,Bo'la sbo & Slot'Games online.	penipuan	normal	normal

9	J01N BR0T@63L & Main CardGames Do'mino B4cc4r4t DinG'DonG hkb,Slots,T*Gel Sydney,sbo C@S1N0 & Bo'la. cutt.ly/BR0899 Raih Bonusnya skrg jg	penipuan	penipuan	normal
10	Raih Jackpot dgn main Slots online di TEXASP0 KERCC s.id/TPK999 Ada jg B4cc4r4t,P@K3R & DinG'DonG,To_Gel Sydney,Bo'la & C@51N0 Sbo. Mainkan skrg!	penipuan	normal	normal

Dari tabel diatas pada pesan nomor 7 dan nomor 9, himpunan 300 pesan teruji memprediksi pesan masuk secara akurat sedangkan prediksi himpunan 600 pesan teruji memprediksi pesan masuk secara tidak akurat. Untuk pesan-pesan masuk lainnya, kedua himpunan pesan teruji memiliki prediksi yang sama pada nilai kategori dalam memprediksi pesan masuk.

Pada pesan pertama, baik himpunan 300 pesan teruji dan 600 pesan teruji sama-sama memprediksi pesan masuk sebagai kategori normal. Tabel perhitungan bobot kata pesan masuk pertama dapat dilihat pada Tabel 4.4 dan Tabel 4.5.

Tabel 4.4. Tabel Bobot Kata Pesan Pertama Himpunan 300

300 Kata			
kata	normal	promo	penipuan
aplikasi	0	0	0
uang	0	0	4
baru	0	0	0
di	0	0	0
indonesia	0	2	1
tebak	0	0	0
skor	1	0	0
salah	0	0	0
sepak	0	0	0
bola	0	0	1
investasi	0	0	0

uang	0	0	4
yang	0	0	0
paling	0	0	0
bagus	0	0	0
buru	2	8	0
download	0	8	0
dan	0	0	0
hub	0	5	26
cs	0	0	3
di	0	0	0

Tabel 4.5. Tabel Bobot Kata Pesan Pertama Himpunan 600

kata	normal	promo	penipuan
aplikasi	0	8	0
uang	0	0	12
baru	0	0	0
di	0	0	0
indonesia	0	3	10
tebak	0	0	0
skor	1	0	0
salah	0	0	0
sepak	0	0	0
bola	0	0	1
investasi	0	0	0
uang	0	0	12
yang	0	0	0
paling	0	0	0
bagus	2	1	0
buru	2	16	0
download	0	14	0
dan	0	0	0
hub	0	14	45
cs	0	0	9
di	0	0	0

Dari tabel diatas, dapat dilihat pada kedua himpunan pesan teruji pada himpunan kategori penipuan terlihat memiliki bobot kata yang lebih berat. Khususnya pada penggunaan kata “hub”. Meski begitu, algoritma penelitian memprediksi pesan masuk sebagai pesan dengan kategori normal. Hal ini dikarenakan kalkulasi *smoothing* yang menghitung jumlah kata dari masing-masing kategori pesan.

Pada kasus penelitian, jumlah kata pada pesan normal lebih kecil daripada jumlah kata pada masing-masing pesan dengan kategori promosi dan penipuan sehingga nilai yang ada pada kalkulasi *smoothing naïve bayes* pesan normal menjadi lebih besar.

Untuk pesan kedua yang merupakan pesan dengan kategori penipuan, baik himpunan 300 pesan teruji dan himpunan 600 pesan teruji sama-sama memprediksi pesan masuk sebagai pesan normal. Tabel bobot kata untuk pesan kedua dapat dilihat pada Tabel 4.6 dan Tabel 4.7.

Tabel 4.6. Tabel Bobot Kata Pesan Kedua Himpunan 300

kata	normal	promo	penipuan
sebentar	0	0	0
lagi	0	0	0
liga	0	0	0
champion	0	0	0
loh	0	3	1
pilih	0	6	7
tim	0	0	0
favorit	0	0	0
di	0	0	0
afapo	0	0	0
ker	0	0	0
cutt	0	0	0
ly	0	0	0
af	0	0	1
po	0	0	0
ker	0	0	0

nonton	0	1	0
santai	0	0	0
sambil	1	0	0
profit	0	0	1

Tabel 4.7. Tabel Bobot Kata Pesan Kedua Himpunan 600

kata	normal	promo	penipuan
sebentar	0	0	0
lagi	0	0	0
liga	0	0	0
champion	0	0	0
loh	0	4	1
pilih	0	7	14
tim	0	0	0
favorit	0	1	0
di	0	0	0
afapo	0	0	0
ker	0	0	0
cutt	0	0	0
ly	0	0	0
af	0	0	1
po	0	0	0
ker	0	0	0
nonton	0	2	0
santai	0	0	0
sambil	1	0	1
profit	0	0	1

Hal yang sama pada pesan pertama juga terjadi pada pesan kedua. Proses smoothing memberatkan bobot nilai kalkulasi pada himpunan dengan kategori normal.

Bukan hanya pesan kedua, masalah yang sama juga terjadi dengan pesan kelima, kedelapan, dan kesepuluh. Pada kasus 5 pesan masuk ini, proses smoothing memberatkan nilai probabilitas pada himpunan pesan dengan kategori normal.

Tabel 4.8. Tabel Bobot Kata Pesan Kelima Himpunan 300

kata	normal	promo	penipuan
modem	0	0	0
tanpa	0	0	0
kabel	1	0	0
praktis	0	0	0
bawa	5	0	0
mana	0	0	0
aja	12	4	1
harga	1	4	6
paket	2	27	0
mulai	0	0	0
K	0	8	1
dapat	0	0	0
orbit	0	0	0
Di	0	0	0
konter	0	0	0
Gt	2	0	1
reload	0	0	0
lowokwaru	0	0	0
info	2	24	47
bit	0	0	0
Ly	0	0	0
telkomselorbit	0	0	0

Tabel 4.9. Tabel Bobot Kata Pesan Kelima Himpunan 600

kata	normal	promo	penipuan
modem	0	2	0
tanpa	0	0	0
kabel	1	0	0
praktis	0	0	0
bawa	7	2	0
mana	0	0	0

aja	24	10	4
harga	1	10	14
paket	2	74	2
mulai	0	0	0
k	1	18	14
dapat	0	0	0
orbit	0	0	0
di	0	0	0
konter	0	0	0
gt	2	0	1
reload	0	0	0
lowokwaru	0	0	0
info	2	46	95
bit	0	0	0
ly	0	0	0
telkomselorbit	0	0	0

Pada pesan kelima, dapat dilihat pada Tabel 4.8 dan Tabel 4.9, kata “info” memiliki kata yang berbobot lebih berat pada himpunan kategori pesan penipuan serta kata “paket” yang memiliki bobot lebih besar pada himpunan kategori pesan promo pada kedua himpunan pesan teruji.. Meski begitu, proses smoothing memberatkan nilai pada pesan yang tidak bernilai yang lebih besar pada himpunan kategori pesan normal sehingga algoritma penelitian memprediksi pesan masuk kelima sebagai pesan normal.

Tabel 4.10. Tabel Bobot Kata Pesan Kedelapan Himpunan 300

kata	normal	promo	penipuan
proses	0	0	0
depo	0	0	0
wd	0	0	0
cepat	2	1	4
kilat	0	0	0
di	0	0	0
t	0	4	4
lmandiri	0	0	0

dik	0	0	0
si	1	0	0
ob	0	0	0
ng	0	0	0
main	0	1	1
games	0	2	0
japan	0	0	0
sydney	0	0	0
bo	0	0	0
la	0	0	0
sbo	0	0	0
slot	0	0	0
games	0	2	0
online	1	0	11

Tabel 4.11. Tabel Bobot Kata Pesan Kedelapan Himpunan 600

kata	normal	promo	penipuan
proses	0	0	0
depo	0	0	0
wd	0	0	0
cepat	2	4	6
kilat	0	0	0
di	0	0	0
t	0	7	4
lmandiri	0	0	0
dik	0	0	0
si	2	0	0
ob	0	0	0
ng	0	0	0
main	0	4	1
games	0	4	1
japan	0	0	0
sydney	0	0	0
bo	0	0	0
la	0	0	0
sbo	0	0	0

slot	0	0	0
games	0	4	1
online	1	1	11

Pada pesan kedelapan, kata “online” memiliki bobot kata lebih besar pada himpunan kategori pesan penipuan. Namun proses smoothing memberatkan kata yang tidak memiliki bobot lebih besar pada kata-kata pada himpunan kategori pesan normal yang membuat algoritma penelitian memprediksi pesan masuk sebagai pesan normal.

Tabel 4.12. Tabel Bobot Kata Pesan Kesepuluh Himpunan 300

kata	normal	promo	penipuan
raih	0	3	4
jackpot	0	0	0
dgn	1	11	6
main	0	1	1
slots	0	0	0
online	1	0	11
di	0	0	0
texasp	0	0	0
kercc	0	0	0
s	2	24	4
id	0	2	0
tpk	0	0	0
ada	0	0	0
jg	6	0	0
ding	0	0	0
dong	0	0	0
to	0	4	1
gel	0	0	0
sydney	0	0	0
bo	0	0	0
la	0	0	0
sbo	0	0	0
main	0	1	1
skrg	1	3	1

Tabel 4.13. Tabel Bobot Kata Pesan Kesepuluh Himpunan 600

kata	normal	promo	penipuan
raih	0	4	10
jackpot	0	0	0
dgn	1	21	9
main	0	4	1
slots	0	0	0
online	1	1	11
di	0	0	0
texasp	0	0	0
kercc	0	0	0
s	2	47	7
id	0	5	1
tpk	0	0	0
ada	0	0	0
jg	8	0	0
ding	0	0	0
dong	2	0	0
to	0	5	1
gel	0	0	0
sydney	0	0	0
bo	0	0	0
la	0	0	0
sbo	0	0	0
main	0	4	1
skrg	2	4	2

Pada Tabel 4.12 yang memuat tabel bobot kata untuk pesan masuk himpunan 300 pesan teruji serta pada Tabel 4.13 untuk 600 himpunan pesan teruji, kata “s” dan kata “dgn” memiliki bobot kata yang lebih besar pada himpunan kategori pesan promosi. Meski begitu, algoritma penelitian melakukan prediksi bahwa pesan masuk

termasuk ke dalam golongan pesan dengan kategori normal karena algoritma smoothing yang memberatkan nilai pada kata-kata dengan bobot 0.

Hal yang sama juga terjadi pada pesan ketiga yang memiliki kategori pesan penipuan. Algoritma penelitian baik dari himpunan 300 pesan teruji dan 600 pesan teruji sama-sama memprediksi pesan masuk sebagai pesan dengan kategori normal. Namun dilihat dari bobot kata-kata pada pesan masuk, bobot dari kata-kata terlihat lebih berat pada himpunan pesan dengan kategori promo daripada kategori pesan sebenarnya. Tabel bobot kata untuk pesan ketiga dapat dilihat pada Tabel 4.14 dan Tabel 4.15.

Tabel 4.14. Tabel Bobot Kata Pesan Ketiga Himpunan 300

kata	normal	promo	penipuan
datang	0	0	0
dan	0	0	0
dapat	0	0	0
hasil	0	0	0
tambah	0	1	0
anda	0	0	0
hingga	0	0	0
rp	0	47	19
per	0	0	0
hari	0	0	0
hubung	0	2	6
kami	0	0	0
dan	0	0	0
biar	1	0	0
kami	0	0	0
beri	2	0	2
tahu	0	0	0
anda	0	0	0
cara	0	0	0
hasil	0	0	0
cutt	0	0	0
ly	0	0	0

vumctr	0	0	0
--------	---	---	---

Tabel 4.15. Tabel Bobot Kata Pesan Masuk Himpunan 600

kata	normal	promo	penipuan
datang	0	0	0
dan	0	0	0
dapat	0	0	0
hasil	0	0	0
tambah	0	6	0
anda	0	0	0
hingga	0	0	0
rp	0	81	42
per	0	1	0
hari	0	0	0
hubung	0	7	17
kami	0	0	0
dan	0	0	0
biar	4	0	0
kami	0	0	0
beri	2	1	2
tahu	0	0	0
anda	0	0	0
cara	0	0	0
hasil	0	0	0
cutt	0	0	0
ly	0	0	0
vumctr	0	0	0

Dari dua tabel diatas, dapat dilihat bobot kata lebih berat pada himpunan kategori pesan promosi. Kata “rp” mendominasi bobot kata dan jauh lebih berat daripada jumlah yang ada pada himpunan kategori pesan normal maupun pesan penipuan.

Untuk pesan masuk keempat yang memiliki kategori pesan normal, algoritma penelitian untuk kedua himpunan pesan teruji memprediksi pesan masuk sebagai pesan

dengan kategori penipuan. Tabel bobot kata pesan masuk keempat dapat dilihat pada Tabel 4.16 dan Tabel 4.17.

Tabel 4.16. Tabel Bobot Kata Pesan Keempat Himpunan 300

kata	normal	promo	penipuan
nabi	0	0	0
maulida	0	0	0
agar	0	0	0
hub	0	5	26
saya	0	0	0

Tabel 4.17. Tabel Bobot Kata Pesan Keempat Himpunan 600

kata	normal	promo	penipuan
nabi	0	0	0
maulida	0	0	0
agar	0	0	0
hub	0	14	45
saya	0	0	0

Dari dua tabel diatas, dapat dilihat kata “hub” memiliki bobot kata berlebih pada himpunan kategori pesan penipuan. Hal ini yang menjadi penentu algoritma penelitian pada dua himpunan pesan teruji memprediksi pesan masuk keempat sebagai pesan dengan kategori penipuan.

Pada pesan masuk keenam yang memiliki kategori pesan penipuan sebagai kategori sebenarnya. Algoritma penelitian untuk kedua himpunan pesan teruji memprediksi pesan masuk sebagai pesan dengan kategori promo. Tabel bobot kata untuk pesan masuk keenam dapat dilihat pada Tabel 4.18 dan Tabel 4.19.

Tabel 4.18. Tabel Bobot Kata Pesan Keenam Himpunan 300

kata	normal	promo	penipuan
------	--------	-------	----------

no	0	3	33
hp	0	5	4
coin	0	0	0
pulsa	0	22	20
hari	0	0	0
ini	0	0	0
bisa	0	0	0
kamu	0	14	1
ambil	2	2	0
di	0	0	0
bisa	0	0	0
kamu	0	14	1
tukar	0	13	0
pulsa	0	22	20
kuota	0	17	0
segera	0	0	0
hub	0	5	26
dari	0	0	0
hp	0	5	4
ini	0	0	0
utk	0	19	5
ambil	2	2	0
reward	0	0	0
mu	0	1	0

Tabel 4.19. Tabel Bobot Kata Pesan Keenam Himpunan 600

kata	normal	promo	penipuan
no	2	4	81
hp	1	10	10
coin	0	0	0
pulsa	0	40	34
hari	0	0	0
ini	0	0	0
bisa	0	0	0
kamu	3	36	2
ambil	2	3	0

di	0	0	0
bisa	0	0	0
kamu	3	36	2
tukar	0	20	0
pulsa	0	40	34
kuota	0	49	0
segera	0	0	0
hub	0	14	45
dari	0	0	0
hp	1	10	10
ini	0	0	0
utk	0	39	11
ambil	2	3	0
reward	0	0	0
mu	0	3	1

Dari dua tabel, dapat dilihat meskipun bobot kata “no” dan kata “hub” terlihat lebih berat pada himpunan kategori penipuan, kata “pulsa”, “kamu”, “tukar”, “pulsa” dan kata “kuota” melebihi berat bobot kata untuk himpunan kategori pesan promo. Hal ini yang menjadi penyebab algoritma penelitian memprediksi pesan masuk keenam sebagai pesan dengan kategori promo.

Pada pesan ketujuh yang memiliki kategori penipuan sebagai kategori sebenarnya, algoritma penelitian untuk 300 pesan teruji berhasil memprediksi pesan masuk sebagai pesan dengan kategori penipuan, sedangkan himpunan 600 pesan teruji memprediksi pesan masuk sebagai pesan promo. Tabel bobot kata untuk pesan masuk ketujuh dapat dilihat pada Tabel 4.20 dan Tabel 4.21.

Tabel 4.20. Tabel Bobot Kata Pesan Ketujuh Himpunan 300

kata	normal	promo	penipuan
nomor	0	1	8
hp	0	5	4
coin	0	0	0
pulsa	0	22	20

anda	0	0	0
segera	0	0	0
ambil	2	2	0
di	0	0	0
belum	0	0	0
hangus	0	0	0

Tabel 4.21. Tabel Bobot Kata Pesan Ketujuh Himpunan 600

kata	normal	promo	penipuan
nomor	0	9	13
hp	1	10	10
coin	0	0	0
pulsa	0	40	34
anda	0	0	0
segera	0	0	0
ambil	2	3	0
di	0	0	0
belum	0	0	0
hangus	0	0	0

Dari tabel bobot kata himpunan 600 pesan teruji, kata “pulsa” dan kata “ambil” pada himpunan kategori pesan promo melebihi berat bobot pada kategori pesan penipuan. Hal ini yang menjadi penentu algoritma penelitian untuk himpunan 600 pesan teruji memprediksi pesan masuk ketujuh sebagai pesan promo.

Hal yang sama pada pesan ketujuh juga terjadi pada pesan kesembilan. Algoritma penelitian untuk himpunan 300 pesan teruji berhasil memprediksi pesan masuk sebagai pesan penipuan, sedangkan himpunan 600 pesan teruji memprediksi pesan masuk sebagai pesan normal.

Alasan dari prediksi ini memiliki alasan yang sama dengan pesan pertama, kedua, kelima, kedelapan, dan kesepuluh. Meskipun terdapat bobot kata yang lebih berat pada himpunan kategori pesan promo dan pesan penipuan, proses smoothing

melakukan perhitungan yang hasilnya memprediksi pesan masuk sebagai pesan normal.

Setelah melakukan uji coba pada masing-masing kesalahan prediksi pada himpunan pesan masuk. Maka dihitung hasil uji coba penilaian pesan yang berhasil. Hasil uji coba penilaian pesan dapat dilihat pada Tabel 4.22.

Tabel 4. 22. Tabel Penilaian Pesan

300 Pesan	True Positive (TP)	42
	True Negative (TN)	92
	False Positive (FP)	8
	False Negative (FN)	8
600 Pesan	True Positive (TP)	40
	True Negative (TN)	90
	False Positive (FP)	10
	False Negative (FN)	10

Dari perhitungan yang ditunjukkan oleh tabel, terlihat perbedaan antara masing-masing nilai TP, TN, FP, dan FN pada data yang diuji dengan masing-masing perbandingan jumlah data teruji. Data yang diuji pada 600 pesan teruji bernilai berbeda dengan himpunan pada data yang dibandingkan dengan 300 data teruji.

4.3. Pembahasan

4.3.1. Perspektif Informatika

a. Accuracy

Accuracy merupakan rasio prediksi benar, baik itu prediksi benar positif dan prediksi benar negative, dengan keseluruhan data. Pada penelitian, akurasi menjawab

pertanyaan “Berapa persen pesan yang benar diprediksi dan pesan yang tidak dari keseluruhan pesan”.

Selanjutnya akurasi dari masing-masing pengujian pada 2 himpunan data pesan tersebut akan dihitung. Untuk akurasi pada pengujian pesan terhadap himpunan data 300 pesan dapat dilihat pada Persamaan 4.1 dan akurasi pada pengujian pesan terhadap himpunan data 600 pesan dapat dilihat pada Persamaan 4.2.

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

$$Accuracy = \frac{42 + 92}{(42 + 8 + 8 + 92)}$$

$$Accuracy = 0.89$$
(4.1)

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

$$Accuracy = \frac{40 + 90}{(40 + 10 + 10 + 90)}$$

$$Accuracy = 0.86$$
(4.2)

Dari perhitungan diatas, akurasi dari pesan yang diuji pada himpunan 300 pesan teruji bernilai 89% sedangkan akurasi dari pesan yang diuji pada himpunan 600 bernilai 86%.

b. Precision

Presisi merupakan rasio prediksi benar positif dibandingkan dengan keseluruhan hasil yang diprediksi positif. Presisi menjawab pertanyaan “Berapa persen pesan yang benar diprediksi dan bernilai benar dari keseluruhan pesan yang dinilai benar?”

Selanjutnya presisi dari masing-masing pengujian pada 2 himpunan data pesan tersebut akan dihitung. Untuk presisi pada pengujian pesan terhadap himpunan data 300 pesan dapat dilihat pada Persamaan 4.3 dan presisi pada pengujian pesan terhadap himpunan data 600 pesan dapat dilihat pada Persamaan 4.4.

$$\begin{aligned} \text{Precision} &= \frac{TP}{TP + FP} \\ \text{Precision} &= \frac{42}{42 + 8} \\ \text{Precision} &= 0.84 \end{aligned} \tag{4.3}$$

$$\begin{aligned} \text{Precision} &= \frac{TP}{TP + FP} \\ \text{Precision} &= \frac{40}{40 + 10} \\ \text{Precision} &= 0.8 \end{aligned} \tag{4.4}$$

Dari perhitungan diatas, presisi dari pesan yang diuji pada himpunan 300 pesan teruji bernilai 84% sedangkan presisi dari pesan yang diuji pada himpunan 600 bernilai 80%.

c. Recall

Recall (sensitifitas) merupakan rasio prediksi benar positif dibandingkan dengan keseluruhan data yang benar positif. *Recall* menjawab pertanyaan “Berapa persen pesan yang diprediksi benar dibandingkan keseluruhan pesan yang benar diprediksi dan salah tidak diprediksi?”

Selanjutnya sensitifitas dari masing-masing pengujian pada 2 himpunan data pesan tersebut akan dihitung. Untuk sensitifitas pada pengujian pesan terhadap himpunan data 300 pesan dapat dilihat pada Persamaan 4.5 dan sensitifitas pada pengujian pesan terhadap himpunan data 600 pesan dapat dilihat pada Persamaan 4.6.

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

$$Recall = \frac{42}{42 + 8}$$

$$Recall = 0.84$$

(4.5)

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

$$Recall = \frac{40}{40 + 10}$$

$$Recall = 0.8$$

(4.6)

Dari perhitungan diatas, sensitifitas dari pesan yang diuji pada himpunan 300 pesan teruji bernilai 84% sedangkan sensitifitas dari pesan yang diuji pada himpunan 600 pesan yang teruji bernilai 80%.

d. Specifity

Spesifitas Merupakan kebenaran memprediksi negatif dibandingkan dengan keseluruhan

data negative. Spesifitas menjawab pertanyaan “Berapa persen pesan yang diprediksi sebagai tidak benar dibandingkan dengan keseluruhan pesan yang tidak benar baik diprediksi maupun tidak diprediksi?”

Selanjutnya spesifitas dari masing-masing pengujian pada 2 himpunan data pesan tersebut akan dihitung. Untuk spesifitas pada pengujian pesan terhadap himpunan data 300 pesan dapat dilihat pada Persamaan 4.7 dan spesifitas pada pengujian pesan terhadap himpunan data 600 pesan teruji dapat dilihat pada Persamaan 4.8.

$$Specifity = \frac{TN}{TN + FP}$$

$$\begin{aligned}
 \text{Specifity} &= \frac{92}{92 + 8} \\
 \text{Specifity} &= 0.92
 \end{aligned}
 \tag{4.7}$$

$$\begin{aligned}
 \text{Specifity} &= \frac{\text{TN}}{\text{TN} + \text{FP}} \\
 \text{Specifity} &= \frac{90}{90 + 10} \\
 \text{Specifity} &= 0.9
 \end{aligned}
 \tag{4.8}$$

Dari perhitungan diatas, spesifitas dari pesan yang diuji pada himpunan 300 pesan teruji bernilai 84% sedangkan spesifitas dari pesan yang diuji pada himpunan 600 pesan teruji bernilai 80%.

e. Skor F1

Skor F1 merupakan perbandingan rata-rata presisi dan sensitifitas yang dibobotkan. Persamaan serta implementasi dari Skor F1 untuk pengujian pesan masuk pada himpunan 300 pesan teruji dapat dilihat pada Persamaan 4.9 sedangkan persamaan serta implementasi dari skor F1 untuk pengujian pesan masuk pada himpunan 600 pesan teruji dapat dilihat pada persamaan 4.10.

$$\begin{aligned}
 \text{Skor F1} &= 2 \cdot \frac{\text{Recall} \cdot \text{Presisi}}{(\text{Recall} + \text{Presisi})} \\
 \text{Skor F1} &= 2 \cdot \frac{0.84 \cdot 0.84}{0.84 + 0.84} \\
 \text{Skor F1} &= 0.84
 \end{aligned}
 \tag{4.9}$$

$$\begin{aligned}
 \text{Skor F1} &= 2 \cdot \frac{\text{Recall} \cdot \text{Presisi}}{(\text{Recall} + \text{Presisi})} \\
 \text{Skor F1} &= 2 \cdot \frac{0.8 \cdot 0.8}{0.8 + 0.8} \\
 \text{Skor F1} &= 0.8
 \end{aligned}$$

(4.10)

4.3.2. Integrasi Islam

Dari pembahasan hasil penelitian baiknya jika setiap pesan yang masuk agar diklarifikasi kebenarannya salah satunya dengan cara memfilter dengan menggunakan metode yang ada pada penelitian. Hal ini senada dengan firman Allah di Q.S Al-Hujurat ayat 6, yaitu

يَا أَيُّهَا الَّذِينَ آمَنُوا إِن جَاءَكُمْ فَاسِقٌ بِنَبَأٍ فَتَبَيَّنُوا أَن تُصِيبُوا قَوْمًا
بِجَهَالَةٍ فَتُصْحَبُوا عَلَى مَا فَعَلْتُمْ نَادِمِينَ

Ayat tersebut memiliki arti, “Wahai orang-orang yang beriman! Jika seseorang yang fasik datang kepadamu membawa suatu berita, maka telitilah kebenarannya, agar kamu tidak mencelakakan suatu kaum karena kebodohan (kecerobohan), yang akhirnya kamu menyesali perbuatanmu itu.”

Al-Hujurat merupakan surat dengan urutan ke 49 dalam Al-Qur'an. Surat Al-Hujurat terdiri dari 18 ayat dan surat ini adalah surat yang tergolong kepada kelompok surat yang turun di kota madinah.

Pada buku tafsir karangan Prof. Teungku Muhammad Hasbi yang berjudul *Tafsir Al-Qur'anul Majid An-Nur*, beliau menjelaskan bahwa surat Al-Hujurat secara garis besar membahas tentang dasar-dasar budi pekerti dan sikap para Muslim terhadap Rasulullah SAW.

Surat ini juga menjelaskan tentang bagaimana cara menerima berita-berita serta kabar-kabar dari sumber yang tidak dapat dipercaya. Ayat yang sangat relevan dengan keadaan dunia saat ini.

Surat Al-Hujurat ayat 6 memiliki asbabun nuzul yang berkaitan dengan kisah al-Walid bin Uqbah. Berdasarkan buku yang ditulis oleh Adian Husaini yang memiliki judul *Penyesatan Opini: Sebuah Rekayasa Mengubah Citra*. Saat Nabi Muhammad SAW mengutus al-Walid bin Uqbah untuk menarik zakat dari Bani Musthaliq yang telah masuk Islam. Walid tidak berhasil melakukannya dan kembali ke Madinah untuk melapor kepada Rasulullah. Dalam laporannya, bahwa kaum Bani Musthaliq telah murtad dan keluar dari Islam.

Kemudian Rasulullah memerintahkan sahabat Khalid bin Walid untuk pergi melakukan klarifikasi atas kabar tersebut. Khalid diutus untuk pergi memantau kaum Bani Musthaliq untuk mencari kebenaran dari kabar yang dibawa oleh al-Walid bin Uqbah dan dia pun memantau kegiatan kaum Bani Musthaliq secara diam-diam bersama dengan pasukannya.

Setelah melakukan pemantauan, ternyata kaum Bani Musthaliq didapati oleh Khalid bin Walid masih memeluk agama Islam dan sahabat pun langsung melaporkan hal tersebut ke Rasulullah SAW. Setelah kejadian itu, maka turunlah surat Al-Hujurat ayat 6 untuk peringatan kepada kaum muslimin bahayanya kabar yang datang dari orang-orang fasik serta memastikan kebenaran kabar-kabar yang datang dari orang-orang fasik tersebut.

BAB V

PENUTUP

5.1. Kesimpulan

Dari hasil implementasi dan uji coba yang telah dilakukan dapat ditarik kesimpulan sebagai berikut:

- a. Algoritma Naïve Bayes pada percobaan kepada himpunan 300 pesan teruji terhadap pesan masuk, akurasi dari algoritma penelitian sebesar 89%, sedangkan untuk percobaan kepada himpunan 600 pesan teruji, akurasi dari algoritma penelitian sebesar 86%. Kedua nilai tersebut cukup tinggi sebagai penentu keabsahan dari pesan masuk.
- b. Beberapa hal yang mempengaruhi besarnya akurasi dari algoritma Naïve Bayes dalam penelitian menentukan klasifikasi pesan spam diantaranya adalah banyaknya bobot kata pada masing-masing himpunan kata-kata pada pesan normal, pesan promosi, serta pesan penipuan serta jumlah data pengujian yang membentuk masing-masing kelas tersebut.

5.2. Saran

Terdapat beberapa kekurangan pada penelitian ini. Beberapa saran agar penelitian ini dapat dikembangkan menjadi lebih baik diantaranya:

- a. Mampu mendeteksi secara khusus atribut-atribut khusus yang terdapat pada pesan dan menjadikan atribut-atribut khusus tersebut masuk ke dalam kalkulasi penelitian. Misalnya, atribut garis miring (/) pada pesan yang memiliki nama website pada kontennya agar tidak dihapus dan masuk ke dalam perhitungan.
- b. Menambah jumlah pesan-pesan yang telah diuji dan telah diklasifikasi sebagai data pembanding.

- c. Memperbaiki perubahan kalimat menjadi kata-kata yang lebih baku dan memperhitungkan kata-kata yang disingkat bukan baku. Seperti “bkn” yang merupakan singkatan dari kata “bukan”.

DAFTAR PUSTAKA

- Adhitya Prayoga Permana, K. A. (2021). Analisis Perbandingan Algoritma Decision Tree, kNN, dan Naive Bayes untuk Prediksi Kesuksesan Start-up. *Jurnal Informatika Sunan Kalijaga*, 178-188.
- Anugroho, P., Winarno, I., & M, N. R. (2015). *Klasifikasi Email Spam dengan Metode Naive Bayes Classifier Menggunakan Java Programming*. Surabaya: Politeknik Elektronika Negeri Surabaya.
- Bustami. (2014). Penerapan Algoritma Naive Bayes untuk Mengklasifikasi Data Nasabah Asuransi. *TECHSI: Jurnal Penelitian Teknik Informatika*, 128-137.
- Ginting, S. L., & Trinanda, R. P. (2013). Teknik Data Mining Menggunakan Metode Bayes Classifier Untuk Optimalisasi Pencarian pada Aplikasi Perpustakaan (Studi Kasus: Perpustakaan Universitas Pasundan - Bandung). *Jurnal Teknologi dan Informasi*, 2-14.
- Husaini, A. (2002). Penyesatan Opini: Sebuah Rekayasa Mengubah Citra. In A. Husaini, *Penyesatan Opini: Sebuah Rekayasa Mengubah Citra* (p. xvi). Jakarta: Gema Insani.
- Irwan Budi Santoso, C. C. (2018). *Optimalisasi Multi Model Naive Bayes Classifier Dalam Mengklasifikasikan Objek Ruang Terbuka Hijau (RTH) Berbasis Citra Google Earth*. Malang: Lembaga Penelitian Dan Pengabdian Kepada Masyarakat (LP2M) Universitas Islam Negeri Maulana Malik Ibrahim Malang.
- Karimi, Z. (2021, October 06). *Confusion Matrix*. Retrieved from researchgate: <https://www.researchgate.net/publication/355096788>
- Larner, A. (2021). *Contingency, Confusion, and the Metrics of Binary Classification*. Cham: Springer.
- Listiowarni, I., & Ramadhani, N. (2019). Implementasi Naïve Bayessian dengan Laplacian Smoothing untuk Peminatan dan Lintas Minat Siswa SMAN 5 Pamekasan. *Jurnal SISFOKOM*, 125-128.
- Putri, A. N. (2017). Penerapan Naive Bayes Untuk Perankingan Kegiatan di Fakultas TIK Universitas Semarang. *Jurnal SIMETRIS*, 604-607.
- Serhad Sarica, J. L. (2020). *Stopwords in Technical Language Processing*. Singapore: ResearchGate.
- Syarli, & Muin, A. A. (2016). Metode Naive Bayes Untuk Prediksi Kelulusan (Studi Kasus: Data Mahasiswa Baru Perguruan Tinggi). *Jurnal FIKOM - UNASMAN*, 23-25.

Wibisono, Y. (2020, July 2). *fpmipa.upi.edu*. Retrieved from Website UPI:
http://fpmipa.upi.edu/staff/yudi/stop_words_list.txt

Yunifa Miftachul Arif, M. W. (2020). Tourism Destinations Popularity Rating In Malang Raya using Naive Bayes Classifier and Selection Sort Based on Twitter World Popularity. *International Journal of Information System & Technology*, 261-267.

LAMPIRAN: TABEL HASIL UJI COBA 300 PESAN TERUJI

No	Pesan	Ground Truth	Prediksi	TP	TN	FP	FN
1	Boleh lah. Datang saja. Sehat kan?	normal	normal	1	2	0	0
2	Berkah dari kantor BPJS Anda mendptkan dana-bantuan Rp.27.500.000 info ketik: BPJS Kirim ke: WA 082291112662	penipuan	penipuan	1	2	0	0
3	Tetap terhubung dengan kerabatmu dengan paket Internet bulanan favoritmu. Beli skrg juga di tsel.me/omg atau *363#	promo	promo	1	2	0	0
4	Kris kmu sempro kemarin presentasi a?	normal	normal	1	2	0	0
5	Anda! m,nerima Rp85.000.000 dr MsGlow PIN Id; 2547799 Kunjungi.. bit.ly/cs-msglow	penipuan	penipuan	1	2	0	0
6	Selamat Anda mendptkn Hadiah dari BANK BTPN KODE : 7874638 C3K Rp.275.000.000 Silahkan dikonfirmasi Ke WhatsApp 082284208647 bit.ly/bantuan-bank-btpn	penipuan	penipuan	1	2	0	0
7	INFO PINJ4M4N D4N4 ONLINE PROSES MUD4H/CEP4TPINJ4MAN MINIM4L 5JT S/D 500JT BUNG4 RENDAH 2% THN MINAT WHATSAPP :0822-5334-6299"	penipuan	penipuan	1	2	0	0
8	ibu Retno WA 085314124447 M3n4w4rk4n PIN74M4N 50-500 Bunga 0,679%/bln T4NP4 J4min4n T4np4 Survey	penipuan	penipuan	1	2	0	0
9	Sudah dibawakan pagi tadi mas	normal	normal	1	2	0	0
10	Pagi aja jam segini gapapa	normal	normal	1	2	0	0
11	Za ganti sabtu minggu aja ya aku mau balik	normal	normal	1	2	0	0

12	INFO PINJ4M4N D4N4 ONLINE PROSES MUD4H/CEP4T PINJ4MAN MINIM4L 5JT S/D 500JT BUNG4 RENDAH 2% THN MINAT WHATSAPP :0822- 9038-2379	penipuan	penipuan	1	2	0	0
13	Aplikasi Uang Terbaru Di Indonesia ! Tebak Skor Salah Sepak Bola Investasi Keuangan yang Paling Bagus ! Buruan Download Dan Hub cs di https://bit.ly/3HkJ002	penipuan	normal	0	1	1	1
14	Ass,Bpk/Ibu Butuh Dana Buat Usaha Dan Tambha Modal Min 5jt S/D 500jt Bunga 2% Minat Silahkan WhatsApp 082316469392	penipuan	penipuan	1	2	0	0
15	Sebentar lagi Liga Champion loh !! Pilih tim favoritmu di AFAPo-Ker cutt.ly/AF4Po-Ker Nonton santai sambil profit !	penipuan	normal	0	1	1	1
16	Datang dan dapatkan penghasilan tambahan Anda, hingga 200RP per hari, hubungi kami dan biarkan kami memberi tahu Anda cara menghasilkan cutt.ly/VUmCTr6	penipuan	normal	0	1	1	1
17	Ekstra pulsa Telkomsel 25rb/1hr(hny utk SMS,tlp sesama Tsel) spesial UNTUKMU! Dengan beli paket Combo hanya di MyTelkomsel tsel.me/combosakti (s/d22Mar).SKB	promo	promo	1	2	0	0
18	Spesial untukmu! Dapatkan Ekstra Kuota 10GB dengan Tukar 1 Poin + Rp10 dan kamu langsung bisa nikmatin hari ini. Klik tsel.me/kejutanpoin . SKB	promo	promo	1	2	0	0

19	Temukan rekomendasi berbagai paket menarik yang KAMU BANGET NIH! Cek lebih lanjut di tsel.me/shop1 atau *888*15# skrg juga!	promo	promo	1	2	0	0
20	ADA KABAR GEMBIRA UNTUKMU! Sambut MyTelkomsel Anniversary yang ke-6, Yuk Dapatkan Cashback 60% hanya di MyTelkomsel Cek di tsel.me/annivmytsel	promo	promo	1	2	0	0
21	Mas nanti klo udah landing pean tlp y.. q yg jemput d bandara	normal	normal	1	2	0	0
22	ntar tinggal minta kunci kalo ke sini	normal	normal	1	2	0	0
23	Enggak sih za, coba kamu chat alam dia kan relawan siapa tau dia faham tempat2nya	normal	normal	1	2	0	0
24	Jadi nanti jadi satu aja kalau ke penguji	normal	normal	1	2	0	0
25	Paket di kursi kayu depan rumah	normal	normal	1	2	0	0
26	Nggak itu cuma untuk kompre kuliah. Pak imam kan kompre agama	normal	normal	1	2	0	0
27	Lumayan 10 menit dr kontrakan yg dulu	normal	normal	1	2	0	0
28	Nah gak tau lek itu, coba tanyain langsung ke admin aja biar lebih jelasnya	normal	normal	1	2	0	0
29	express kalau gk salah sediain request client juga kan?	normal	normal	1	2	0	0
30	nabilah maulida agar hub sayaYes, nah sebelum itu nanti ada gang sebelah kanan jalan	normal	penipuan	0	1	1	1
31	Tetap terhubung dengan kerabatmu dengan paket Internet bulanan favoritmu. Beli skrg juga di tsel.me/omg atau *363#.	promo	promo	1	2	0	0
32	MURAH BANGET! Combo Sakti 20GB + UL Kuota Apps +Telp&SMS 30hr Hrg mulai 89Rb! PAKET MURAH lainnya	promo	promo	1	2	0	0

	sesuai kebutuhanmu jg ada, Beli di *363# atau tsel.me/combo1						
33	Mau kuota Youtube, Instagram, Facebook yg murah meriah? Beli aja kuota ketengan! Dijamin gk mahal loh. Aktifin pakatnya di tsel.me/ketengan & *363*847# SEKARANG	promo	promo	1	2	0	0
34	120GB GRATIS SELAMA SETAHUN HANYA HARI INI di aplikasi ROLi Telkomsel! Download aplikasi ROLi Telkomsel sekarang tsel.me/ROLi_MD S&K	promo	promo	1	2	0	0
35	Langganan Netflix skrg bisa pake pulsa! Aktifkan paket berlangganan Netflix mulai dr Rp 62rb dpt kuota 6GB + BONUS kuota hingga 20GB di tsel.me/netflixpack. SKB	promo	promo	1	2	0	0
36	1/2 PROMO NSP Rp 0/7hr bls SMS ketik YA utk langganan NSP Cinta Sampai Mati-Kangen Band. Perpanjangan Rp.9990/30hari atau sesuai saldo prabayar/tagihan	promo	promo	1	2	0	0
37	MODEM TANPA KABEL! Praktis dibawa kemana aja. Harga paket mulai 20K. Dapatkan ORBIT di konter GT RELOAD LOWOKWARU. Info: bit.ly/TELKOMSELORBIT	promo	normal	0	1	1	1
38	Ada Extra Kuota hingga 10GB di myTELKOMSEL! Mau? AYO Segera beli paket data bulanan mu & lakukan Daily Cek In melalui apps myTELKOMSEL. Info Klik : tsel.me/dci	promo	promo	1	2	0	0

39	Aktifkan ulang Paket Data Combo Sakti/Internet Sakti atau OMG anda melalui aps myTELKOMSEL. Dgn Daily Check-In rutin, bisa dapat EXTRA Kuota hingga 7,5GB #SKB	promo	promo	1	2	0	0
40	Baru! Buat pengguna Shopee. Beli Paket Combo/Internet sakti bisa pakai shopeepay. Klik aja langsung di tsel.me/shopee-app . Pilih paket lalu bayar. SK	promo	promo	1	2	0	0
41	No HP 081217299196, COIN PULSA Hari Ini: +1000. Bisa Kamu Ambil di *500*220#. Bisa Kamu Tukar PULSA+KUOTA. SEGERA Hub *500*220# dari HP ini utk Ambil REWARD mu!	penipuan	promo	0	1	1	1
42	Nomor HP 081217299196, COIN PULSA Anda: +2000. SEGERA Ambil di *500*75# Sebelum Hangus!	penipuan	penipuan	1	2	0	0
43	SELAMAT Anda mndptkn Hadiah berkah msglow Rp.157.000.000 PIN:MSG-130299 1NF0 HUB bit.ly/give-away-ms-glow Whatsapp 082297299219	penipuan	penipuan	1	2	0	0
44	Proses Depo & WD secepat kilat di T063LMANDIRI dik.si/Ob1NG Main Games b4ccar4t,P@K3R,Ce'Me,To_Gel japan, sydney,Bo'la sbo & Slot'Games online.	penipuan	normal	0	1	1	1
45	kami dr P-E-G-A-D-A-I-A-N Galeri24 mnwrkan barang yang tidak ditebus pemilik seperti 'Emas,iphone,laptop,dll.klik https://tinyurl.com/434csppa	penipuan	penipuan	1	2	0	0

46	J01N BR0T@63L & Main CardGames Do'mino B4cc4r4t DinG'DonG hkb,Slots,T*Gel Sydney,sbo C@S1N0 & Bo'la. cutt.ly/BR0899 Raih Bonusnya skrg jg	penipuan	penipuan	1	2	0	0
47	PINJAMAN DANA Buat Nmbah Modal Dan DLL Siap Bantu Min 5Jt S/D 500jt Bunga 2% Pertahun WhatsApp:085822577247	penipuan	penipuan	1	2	0	0
48	YTH.. Bpk/Ibu Sekluarga K@mi T4w@rkan 4nda P1nj4m@n Proses Cepat Inf0 WhatsApp +6282288799147	penipuan	penipuan	1	2	0	0
49	Permisi Bpk/Ibu Butuh Dana INGIN KEMBANGKAN USAHA & Kebutuhan Lain Min 5Jt S/D 500Jt Bunga 2% Info Lanjut WA : +6282213927356	penipuan	penipuan	1	2	0	0
50	Raih Jackpot dgn main Slots online di TEXASP0 KERCC s.id/TPK999 Ada jg B4cc4r4t,P@K3R & DinG'DonG,To_Gel Sydney,Bo'la & C@51N0 Sbo. Mainkan skrg!	penipuan	normal	0	1	1	1
Jumlah				42	92	8	8

LAMPIRAN: TABEL HASIL UJI COBA 600 PESAN TERUJI

No	Pesan	Ground Truth	Prediksi	TP	TN	FP	FN
1	Boleh lah. Datang saja. Sehat kan?	normal	normal	1	2	0	0
2	Berkah dari kantor BPJS Anda mendptkan dana-bantuan Rp.27.500.000 infO ketik: BPJS Kirim ke: WA 082291112662	penipuan	penipuan	1	2	0	0
3	Tetap terhubung dengan kerabatmu dengan paket Internet bulanan favoritmu. Beli skrg juga di tsel.me/omg atau *363#	promo	promo	1	2	0	0
4	Kris kmu sempro kemarin presentasi a?	normal	normal	1	2	0	0
5	Anda! m,nerima Rp85.000.000 dr MsGlow PIN Id; 2547799 Kunjungi.. bit.ly/cs-msglow	penipuan	penipuan	1	2	0	0
6	Selamat Anda mendptkn Hadiah dari BANK BTPN KODE : 7874638 C3K Rp.275.000.000 Silahkan dikonfirmasi Ke WhatsApp 082284208647 bit.ly/bantuan-bank-btpn	penipuan	penipuan	1	2	0	0
7	INFO PINJ4M4N D4N4 ONLINE PROSES MUD4H/CEP4TPINJ4MAN MINIM4L 5JT S/D 500JT BUNG4 RENDAH 2% THN MINAT WHATSAPP :0822-5334-6299"	penipuan	penipuan	1	2	0	0
8	ibu Retno WA 085314124447 M3n4w4rk4n PIN74M4N 50-500 Bunga 0,679%/bln T4NP4 J4min4n T4np4 Survey	penipuan	penipuan	1	2	0	0
9	Sudah dibawakan pagi tadi mas	normal	normal	1	2	0	0
10	Pagi aja jam segini gapapa	normal	normal	1	2	0	0
11	Za ganti sabtu minggu aja ya aku mau balik	normal	normal	1	2	0	0

12	INFO PINJ4M4N D4N4 ONLINE PROSES MUD4H/CEP4T PINJ4MAN MINIM4L 5JT S/D 500JT BUNG4 RENDAH 2% THN MINAT WHATSAPP :0822- 9038-2379	penipuan	penipuan	1	2	0	0
13	Aplikasi Uang Terbaru Di Indonesia ! Tebak Skor Salah Sepak Bola Investasi Keuangan yang Paling Bagus ! Buruan Download Dan Hub cs di https://bit.ly/3HkJ002	penipuan	normal	0	1	1	1
14	Ass,Bpk/Ibu Butuh Dana Buat Usaha Dan Tambha Modal Min 5jt S/D 500jt Bunga 2% Minat Silahkan WhatsApp 082316469392	penipuan	penipuan	1	2	0	0
15	Sebentar lagi Liga Champion loh !! Pilih tim favoritmu di AFAPo-Ker cutt.ly/AF4Po-Ker Nonton santai sambil profit !	penipuan	normal	0	1	1	1
16	Datang dan dapatkan penghasilan tambahan Anda, hingga 200RP per hari, hubungi kami dan biarkan kami memberi tahu Anda cara menghasilkan cutt.ly/VUmCTr6	penipuan	normal	0	1	1	1
17	Ekstra pulsa Telkomsel 25rb/1hr(hny utk SMS,tlp sesama Tsel) spesial UNTUKMU! Dengan beli paket Combo hanya di MyTelkomsel tsel.me/combosakti (s/d22Mar).SKB	promo	promo	1	2	0	0
18	Spesial untukmu! Dapatkan Ekstra Kuota 10GB dengan Tukar 1 Poin + Rp10 dan kamu langsung bisa nikmatin hari ini. Klik tsel.me/kejutanpoin . SKB	promo	promo	1	2	0	0

19	Temukan rekomendasi berbagai paket menarik yang KAMU BANGET NIH! Cek lebih lanjut di tsel.me/shop1 atau *888*15# skrg juga!	promo	promo	1	2	0	0
20	ADA KABAR GEMBIRA UNTUKMU! Sambut MyTelkomsel Anniversary yang ke-6, Yuk Dapatkan Cashback 60% hanya di MyTelkomsel Cek di tsel.me/annivmytsel	promo	promo	1	2	0	0
21	Mas nanti klo udah landing pean tlp y.. q yg jemput d bandara	normal	normal	1	2	0	0
22	ntar tinggal minta kunci kalo ke sini	normal	normal	1	2	0	0
23	Enggak sih za, coba kamu chat alam dia kan relawan siapa tau dia faham tempat2nya	normal	normal	1	2	0	0
24	Jadi nanti jadi satu aja kalau ke penguji	normal	normal	1	2	0	0
25	Paket di kursi kayu depan rumah	normal	normal	1	2	0	0
26	Nggak itu cuma untuk kompre kuliah. Pak imam kan kompre agama	normal	normal	1	2	0	0
27	Lumayan 10 menit dr kontrakan yg dulu	normal	normal	1	2	0	0
28	Nah gak tau lek itu, coba tanyain langsung ke admin aja biar lebih jelasnya	normal	normal	1	2	0	0
29	express kalau gk salah sediain request client juga kan?	normal	normal	1	2	0	0
30	nabilah maulida agar hub saya	normal	penipuan	0	1	1	1
31	Tetap terhubung dengan kerabatmu dengan paket Internet bulanan favoritmu. Beli skrg juga di tsel.me/omg atau *363#.	promo	promo	1	2	0	0
32	MURAH BANGET! Combo Sakti 20GB + UL Kuota Apps +Telp&SMS 30hr Hrg mulai 89Rb! PAKET MURAH lainnya sesuai kebutuhanmu jg ada, Beli di *363# atau tsel.me/combo1	promo	promo	1	2	0	0

33	Mau kuota Youtube, Instagram, Facebook yg murah meriah? Beli aja kuota ketengan! Dijamin gk mahal loh. Aktifin paketnya di tsel.me/ketengan & *363*847# SEKARANG	promo	promo	1	2	0	0
34	120GB GRATIS SELAMA SETAHUN HANYA HARI INI di aplikasi ROLi Telkomsel! Download aplikasi ROLi Telkomsel sekarang tsel.me/ROLi_MD S&K	promo	promo	1	2	0	0
35	Langganan Netflix skrg bisa pake pulsa! Aktifkan paket berlangganan Netflix mulai dr Rp 62rb dpt kuota 6GB + BONUS kuota hingga 20GB di tsel.me/netflixpack. SKB	promo	promo	1	2	0	0
36	1/2 PROMO NSP Rp 0/7hr bls SMS ketik YA utk langganan NSP Cinta Sampai Mati-Kangen Band. Perpanjangan Rp.9990/30hari atau sesuai saldo prabayar/tagihan	promo	promo	1	2	0	0
37	MODEM TANPA KABEL! Praktis dibawa kemana aja. Harga paket mulai 20K. Dapatkan ORBIT di konter GT RELOAD LOWOKWARU. Info: bit.ly/TELKOMSELORBIT	promo	normal	0	1	1	1
38	Ada Extra Kuota hingga 10GB di myTELKOMSEL! Mau? AYO Segera beli paket data bulanan mu & lakukan Daily Cek In melalui apps myTELKOMSEL. Info Klik : tsel.me/dci	promo	promo	1	2	0	0
39	Aktifkan ulang Paket Data Combo Sakti/Internet Sakti atau OMG anda melalui aps myTELKOMSEL. Dgn Daily	promo	promo	1	2	0	0

	Check-In rutin, bisa dapat EXTRA Kuota hingga 7,5GB #SKB						
40	Baru! Buat pengguna Shopee. Beli Paket Combo/Internet sakti bisa pakai shopeepay. Klik aja langsung di tsel.me/shopee-app . Pilih paket lalu bayar. SK	promo	promo	1	2	0	0
41	No HP 081217299196, COIN PULSA Hari Ini: +1000. Bisa Kamu Ambil di *500*220#. Bisa Kamu Tukar PULSA+KUOTA. SEGERA Hub *500*220# dari HP ini utk Ambil REWARD mu!	penipuan	promo	0	1	1	1
42	Nomor HP 081217299196, COIN PULSA Anda: +2000. SEGERA Ambil di *500*75# Sebelum Hangus!	penipuan	promo	0	1	1	1
43	SELAMAT Anda mndptkn Hadiah berkah msglow Rp.157.000.000 PIN:MSG-130299 1NF0 HUB bit.ly/give-away-ms-glow Whatsapp 082297299219	penipuan	penipuan	1	2	0	0
44	Proses Depo & WD secepat kilat di T063LMANDIRI dik.si/Ob1NG Main Games b4ccar4t,P@K3R,Ce'Me,To_Gel japan, sydney,Bo'la sbo & Slot'Games online.	penipuan	normal	0	1	1	1
45	kami dr P-E-G-A-D-A-I-A-N Galeri24 mnwrkan barang yang tidak ditebus pemilik seperti 'Emas,iphone,laptop,dll.klik https://tinyurl.com/434csppa	penipuan	penipuan	1	2	0	0
46	J01N BR0T@63L & Main CardGames Do'mino B4cc4r4t DinG'DonG hkb,Slots,T*Gel Sydney,sbo C@S1N0 & Bo'la. cutt.ly/BR0899 Raih Bonusnya skrg jg	penipuan	normal	0	1	1	1

47	PINJAMAN DANA Buat Nmbah Modal Dan DLL Siap Bantu Min 5Jt S/D 500jt Bunga 2% Pertahun WhatsApp:085822577247	penipuan	penipuan	1	2	0	0
48	YTH.. Bpk/Ibu Sekluarga K@mi T4w@rkan 4nda P1nj4m@n Proses Cepat Inf0 WhatsApp +6282288799147	penipuan	penipuan	1	2	0	0
49	Permisi Bpk/Ibu Butuh Dana INGIN KEMBANGKAN USAHA & Kebutuhan Lain Min 5Jt S/D 500Jt Bunga 2% Info Lanjut WA : +6282213927356	penipuan	penipuan	1	2	0	0
50	Raih Jackpot dgn main Slots online di TEXASP0 KERCC s.id/TPK999 Ada jg B4cc4r4t,P@K3R & DinG'DonG,To_Gel Sydney,Bo'la & C@51N0 Sbo. Mainkan skrg!	penipuan	normal	0	1	1	1
Jumlah				40	90	10	10

LAMPIRAN: TABEL KATA STOP WORD

No	Word
1	yang
2	di
3	dan
4	itu
5	dengan
6	untuk
7	tidak
8	ini
9	dari
10	dalam
11	akan
12	pada
13	juga
14	saya
15	ke
16	karena
17	tersebut
18	bisa
19	ada
20	mereka
21	lebih
22	kata
23	tahun
24	sudah
25	atau
26	saat
27	oleh
28	menjadi
29	orang
30	ia
31	telah
32	adalah
33	seperti
34	sebagai
35	bahwa
36	dapat

37	para
38	harus
39	namun
40	kita
41	dua
42	satu
43	masih
44	hari
45	hanya
46	mengatakan
47	kepada
48	kami
49	setelah
50	melakukan
51	lalu
52	belum
53	lain
54	dia
55	kalau
56	terjadi
57	banyak
58	menurut
59	anda
60	hingga
61	tak
62	baru
63	beberapa
64	ketika
65	saja
66	jalan
67	sekitar
68	secara
69	dilakukan
70	sementara
71	tapi
72	sangat
73	hal

74	sehingga
75	seorang
76	bagi
77	besar
78	lagi
79	selama
80	antara
81	waktu
82	sebuah
83	jika
84	sampai
85	jadi
86	terhadap
87	tiga
88	serta
89	pun
90	salah
91	merupakan
92	atas
93	sejak
94	membuat
95	baik
96	memiliki
97	kembali
98	selain
99	tetapi
100	pertama
101	kedua
102	memang
103	pernah
104	apa
105	mulai
106	sama
107	tentang
108	bukan
109	agar
110	semua

111	sedang
112	kali
113	kemudian
114	hasil
115	sejumlah
116	juta
117	persen
118	sendiri
119	katanya
120	demikian
121	masalah
122	mungkin
123	umum
124	setiap
125	bulan
126	bagian
127	bila
128	lainnya
129	terus
130	luar
131	cukup
132	termasuk
133	sebelumnya
134	bahkan
135	wib
136	tempat
137	perlu
138	menggunakan
139	memberikan
140	rabu
141	sedangkan
142	kamis
143	langsung
144	apakah
145	pihak
146	melalui
147	diri
148	mencapai
149	minggu

150	aku
151	berada
152	tinggi
153	ingin
154	sebelum
155	tengah
156	kini
157	the
158	tahu
159	bersama
160	depan
161	selasa
162	begitu
163	merasa
164	berbagai
165	mengenai
166	maka
167	jumlah
168	masuk
169	mengalami
170	sering
171	ujar
172	kondisi
173	akibat
174	hubungan
175	empat
176	paling
177	mendapatkan
178	selalu
179	lima
180	meminta
181	melihat
182	sekarang
183	mengaku
184	mau
185	kerja
186	acara
187	menyatakan
188	masa

189	proses
190	tanpa
191	selatan
192	sempat
193	adanya
194	hidup
195	datang
196	senin
197	rasa
198	maupun
199	seluruh
200	mantan
201	lama
202	jenis
203	segera
204	misalnya
205	mendapat
206	bawah
207	jangan
208	meski
209	terlihat
210	akhirnya
211	jumat
212	punya
213	yakni
214	terakhir
215	kecil
216	panjang
217	badan
218	juni
219	of
220	jelas
221	jauh
222	tentu
223	semakin
224	tinggal
225	kurang
226	mampu
227	posisi

228	asal
229	sekali
230	sesuai
231	sebesar
232	berat
233	dirinya
234	memberi
235	pagi
236	sabtu
237	ternyata
238	mencari
239	sumber
240	ruang
241	menunjukkan
242	biasanya
243	nama
244	sebanyak
245	utara
246	berlangsung
247	barat
248	kemungkinan
249	yaitu
250	berdasarkan
251	sebenarnya
252	cara
253	utama
254	pekan
255	terlalu
256	membawa
257	kebutuhan
258	suatu
259	menerima
260	penting
261	tanggal

262	bagaimana
263	terutama
264	tingkat
265	awal
266	sedikit
267	nanti
268	pasti
269	muncul
270	dekat
271	lanjut
272	ketiga
273	biasa
274	dulu
275	kesempatan
276	ribu
277	akhir
278	membantu
279	terkait
280	sebab
281	menyebabkan
282	khusus
283	bentuk
284	ditemukan
285	diduga
286	mana
287	ya
288	kegiatan
289	sebagian
290	tampil
291	hampir
292	bertemu
293	usai
294	berarti
295	keluar

296	pula
297	digunakan
298	justru
299	padahal
300	menyebutkan
301	gedung
302	apalagi
303	program
304	milik
305	teman
306	menjalani
307	keputusan
308	a
309	upaya
310	mengetahui
311	mempunyai
312	berjalan
313	menjelaskan
314	b
315	mengambil
316	benar
317	lewat
318	belakang
319	ikut
320	barang
321	meningkatkan
322	kejadian
323	kehidupan
324	keterangan
325	penggunaan
	masing-
326	masing
327	menghadapi

**LAMPIRAN: TABEL HASIL NAÏVE BAYES UJI COBA PADA HIMPUNAN
300 PESAN**

No	Pesan	Ground Truth	Nilai Normal	Nilai Promo	Nilai Penipuan
1	Boleh lah. Datang saja. Sehat kan?	normal	1.32E-17	1.85E-20	1.87E-20
2	Berkah dari kantor BPJS Anda mendptkan dana-bantuan Rp.27.500.000 info ketik: BPJS Kirim ke: WA 082291112662	penipuan	1.95E-50	9.47E-50	2.01E-48
3	Tetap terhubung dengan kerabatmu dengan paket Internet bulanan favoritmu. Beli skrg juga di tsel.me/omg atau *363#	promo	2.52E-50	4.16E-48	1.86E-54
4	Kris kmu sempro kemarin presentasi a?	normal	6.03E-19	1.85E-20	1.12E-19
5	Anda! m,nerima Rp85.000.000 dr MsGlow PIN Id; 2547799 Kunjungi.. bit.ly/cs-msglow	penipuan	4.50E-39	4.25E-39	8.96E-37
6	Selamat Anda mendptkn Hadiah dari BANK BTPN KODE : 7874638 C3K Rp.275.000.000 Silahkan dikonfirmasi Ke WhatsApp 082284208647 bit.ly/bantuan-bank-btpn	penipuan	1.09E-59	3.63E-59	1.11E-57
7	INFO PINJ4M4N D4N4 ONLINE PROSES MUD4H/CEP4TPINJ4MAN MINIM4L 5JT S/D 500JT BUNG4 RENDAH 2% THN MINAT WHATSAPP :0822-5334-6299"	penipuan	2.29E-69	4.22E-71	1.83E-63
8	ibu Retno WA 085314124447 M3n4w4rk4n PIN74M4N 50-500 Bunga 0,679%/bln T4NP4 J4min4n T4np4 Survey	penipuan	1.43E-61	8.07E-64	2.08E-55
9	Sudah dibawakan pagi tadi mas	normal	5.61E-15	3.09E-17	9.36E-17
10	Pagi aja jam segini gapapa	normal	1.30E-13	1.54E-15	4.89E-16

11	Za ganti sabtu minggu aja ya aku mau balik	normal	1.30E-26	6.00E-29	1.62E-29
12	INFO PINJ4M4N D4N4 ONLINE PROSES MUD4H/CEP4T PINJ4MAN MINIM4L 5JT S/D 500JT BUNG4 RENDAH 2% THN MINAT WHATSAPP :0822-9038-2379	penipuan	2.52E-72	1.21E-73	5.29E-66
13	Aplikasi Uang Terbaru Di Indonesia ! Tebak Skor Salah Sepak Bola Investasi Keuangan yang Paling Bagus ! Buruan Download Dan Hub cs di https://bit.ly/3HkJ002	penipuan	8.96E-63	1.08E-65	8.15E-65
14	Ass,Bpk/Ibu Butuh Dana Buat Usaha Dan Tambha Modal Min 5jt S/D 500jt Bunga 2% Minat Silahkan WhatsApp 082316469392	penipuan	3.03E-53	2.60E-57	1.44E-50
15	Sebentar lagi Liga Champion loh !! Pilih tim favoritmu di AFAPo-Ker cutt.ly/AF4Po-Ker Nonton santai sambil profit !	penipuan	2.72E-60	7.60E-64	9.00E-64
16	Datang dan dapatkan penghasilan tambahan Anda, hingga 200RP per hari, hubungi kami dan biarkan kami memberi tahu Anda cara menghasilkan cutt.ly/VUmCTr6	penipuan	1.08E-68	8.23E-73	1.15E-72
17	Ekstra pulsa Telkomsel 25rb/1hr(hny utk SMS,tlp sesama Tsel) spesial UNTUKMU! Dengan beli paket Combo hanya di MyTelkomsel tsel.me/combosakti (s/d22Mar).SKB	promo	1.83E-78	1.07E-68	6.60E-80
18	Spesial untukmu! Dapatkan Ekstra Kuota 10GB dengan Tukar 1 Poin + Rp10 dan kamu langsung bisa nikmatin hari ini. Klik tsel.me/kejutanpoin . SKB	promo	1.64E-66	4.84E-60	3.09E-68

19	Temukan rekomendasi berbagai paket menarik yang KAMU BANGET NIH! Cek lebih lanjut di tsel.me/shop1 atau *888*15# skrg juga!	promo	5.92E-56	4.25E-55	2.41E-59
20	ADA KABAR GEMBIRA UNTUKMU! Sambut MyTelkomsel Anniversary yang ke-6, Yuk Dapatkan Cashback 60% hanya di MyTelkomsel Cek di tsel.me/annivmytsel	promo	2.72E-60	3.86E-60	4.50E-64
21	Mas nanti klo udah landing pean tlp y.. q yg jemput d bandara	normal	5.33E-37	1.47E-40	3.45E-39
22	ntar tinggal minta kunci kalo ke sini	normal	5.94E-20	1.11E-23	2.25E-23
23	Enggak sih za, coba kamu chat alam dia kan relawan siapa tau dia faham tempat2nya	normal	6.82E-46	6.35E-50	4.53E-52
24	Jadi nanti jadi satu aja kalau ke penguji	normal	8.32E-23	3.33E-26	1.35E-26
25	Paket di kursi kayu depan rumah	normal	5.13E-18	5.15E-19	1.09E-19
26	Nggak itu cuma untuk kompre kuliah. Pak imam kan kompre agama	normal	3.63E-31	9.84E-36	8.74E-36
27	Lumayan 10 menit dr kontrakan yg dulu	normal	3.92E-17	1.48E-18	4.77E-18
28	Nah gak tau lek itu, coba tanyain langsung ke admin aja biar lebih jelasnya	normal	1.83E-40	6.01E-45	6.29E-46
29	express kalau gk salah sediain request client juga kan?	normal	3.04E-27	4.00E-30	4.05E-30
30	nabilah maulida agar hub saya	normal	5.02E-16	1.85E-16	8.32E-16
31	Tetap terhubung dengan kerabatmu dengan paket Internet bulanan favoritmu. Beli skrg juga di tsel.me/omg atau *363#.	promo	2.52E-50	4.16E-48	1.86E-54
32	MURAH BANGET! Combo Sakti 20GB + UL Kuota Apps +Telp&SMS 30hr Hrg mulai 89Rb! PAKET MURAH lainnya	promo	2.15E-78	5.64E-73	9.41E-82

	sesuai kebutuhanmu jg ada, Beli di *363# atau tsel.me/combo1				
33	Mau kuota Youtube, Instagram, Facebook yg murah meriah? Beli aja kuota ketengan! Dijamin gk mahal loh. Aktifin pakatnya di tsel.me/ketengan & *363*847# SEKARANG	promo	9.84E-66	2.48E-65	7.46E-72
34	120GB GRATIS SELAMA SETAHUN HANYA HARI INI di aplikasi ROLi Telkomsel! Download aplikasi ROLi Telkomsel sekarang tsel.me/ROLi_MD S&K	promo	1.97E-65	4.10E-62	6.39E-69
35	Langganan Netflix skrg bisa pake pulsa! Aktifkan paket berlangganan Netflix mulai dr Rp 62rb dpt kuota 6GB + BONUS kuota hingga 20GB di tsel.me/netflixpack. SKB	promo	4.43E-75	2.44E-64	5.34E-76
36	1/2 PROMO NSP Rp 0/7hr bls SMS ketik YA utk langganan NSP Cinta Sampai Mati-Kangen Band. Perpanjangan Rp.9990/30hari atau sesuai saldo prabayar/tagihan	promo	1.19E-71	5.69E-67	4.43E-71
37	MODEM TANPA KABEL! Praktis dibawa kemana aja. Harga paket mulai 20K. Dapatkan ORBIT di konter GT RELOAD LOWOKWARU. Info: bit.ly/TELKOMSELORBIT	promo	1.38E-62	6.94E-67	9.98E-69
38	Ada Extra Kuota hingga 10GB di myTELKOMSEL! Mau? AYO Segera beli paket data bulanan mu & lakukan Daily Cek In melalui apps myTELKOMSEL. Info Klik : tsel.me/dci	promo	1.12E-78	1.45E-71	4.94E-83

39	Aktifkan ulang Paket Data Combo Sakti/Internet Sakti atau OMG anda melalui aps myTELKOMSEL. Dgn Daily Check-In rutin, bisa dapat EXTRA Kuota hingga 7,5GB #SKB	promo	2.10E-73	5.12E-71	4.54E-79
40	Baru! Buat pengguna Shopee. Beli Paket Combo/Internet sakti bisa pakai shopeepay. Klik aja langsung di tsel.me/shopee-app . Pilih paket lalu bayar. SK	promo	8.13E-72	8.93E-67	4.46E-79
41	No HP 081217299196, COIN PULSA Hari Ini: +1000. Bisa Kamu Ambil di *500*220#. Bisa Kamu Tukar PULSA+KUOTA. SEGERA Hub *500*220# dari HP ini utk Ambil REWARD mu!	penipuan	1.79E-71	1.07E-65	2.84E-70
42	Nomor HP 081217299196, COIN PULSA Anda: +2000. SEGERA Ambil di *500*75# Sebelum Hangus!	penipuan	1.62E-30	1.96E-30	2.12E-30
43	SELAMAT Anda mndptkn Hadiah berkah msglow Rp.157.000.000 PIN:MSG-130299 1NF0 HUB bit.ly/give-away-ms-glow Whatsapp 082297299219	penipuan	3.37E-54	1.98E-55	1.80E-51
44	Proses Depo & WD secepat kilat di T063LMANDIRI dik.si/Ob1NG Main Games b4ccar4t,P@K3R,Ce'Me,To_Gel japan, sydney,Bo'la sbo & Slot'Games online.	penipuan	1.97E-65	8.11E-70	2.60E-69
45	kami dr P-E-G-A-D-A-I-A-N Galeri24 mnwrkan barang yang tidak ditebus pemilik seperti 'Emas,iphone,laptop,dll.klik https://tinyurl.com/434csppa	penipuan	3.18E-71	4.02E-73	2.88E-70
46	J01N BR0T@63L & Main CardGames Do'mino B4cc4r4t DinG'DonG hkb,Slots,T*Gel	penipuan	7.31E-80	3.63E-83	3.87E-82

	Sydney,sbo C@S1N0 & Bo'la. cutt.ly/BR0899 Raih Bonusnya skrg jg				
47	PINJAMAN DANA Buat Nmbah Modal Dan DLL Siap Bantu Min 5Jt S/D 500jt Bunga 2% Pertahun WhatsApp:085822577247	penipuan	3.45E- 48	1.13E- 50	1.81E-47
48	YTH.. Bpk/Ibu Sekluarga K@mi T4w@rkan 4nda P1nj4m@n Proses Cepat Inf0 WhatsApp +6282288799147	penipuan	5.82E- 27	7.20E- 29	7.20E-27
49	Permisi Bpk/Ibu Butuh Dana INGIN KEMBANGKAN USAHA & Kebutuhan Lain Min 5Jt S/D 500Jt Bunga 2% Info Lanjut WA : +6282213927356	penipuan	1.14E- 53	1.18E- 56	2.78E-50
50	Raih Jackpot dgn main Slots online di TEXASP0 KERCC s.id/TPK999 Ada jg B4cc4r4t,P@K3R & DinG'DonG,To_Gel Sydney,Bo'la & C@51N0 Sbo. Mainkan skrg!	penipuan	3.34E- 70	4.21E- 73	3.85E-74

**LAMPIRAN: TABEL HASIL NAÏVE BAYES UJI COBA PADA HIMPUNAN
600 PESAN**

No	Pesan	Ground Truth	Nilai Normal	Nilai Promo	Nilai Penipuan
1	Boleh lah. Datang saja. Sehat kan?	normal	2.53E-18	2.85E-22	3.03E-22
2	Berkah dari kantor BPJS Anda mendptkan dana-bantuan Rp.27.500.000 info ketik: BPJS Kirim ke: WA 082291112662	penipuan	2.37E-54	3.21E-53	2.71E-51
3	Tetap terhubung dengan kerabatmu dengan paket Internet bulanan favoritmu. Beli skrg juga di tsel.me/omg atau *363#	promo	1.18E-54	8.5E-51	5.02E-58
4	Kris kmu sempro kemarin presentasi a?	normal	9.06E-21	2.85E-22	1.82E-21
5	Anda! m,nerima Rp85.000.000 dr MsGlow PIN Id; 2547799 Kunjungi.. bit.ly/cs-msglow	penipuan	4.69E-42	8.31E-42	5.98E-38
6	Selamat Anda mendptkn Hadiah dari BANK BTPN KODE : 7874638 C3K Rp.275.000.000 Silahkan dikonfirmasi Ke WhatsApp 082284208647 bit.ly/bantuan-bank-btpn	penipuan	5.68E-64	5.86E-62	9.85E-60
7	INFO PINJ4M4N D4N4 ONLINE PROSES MUD4H/CEP4TPINJ4MAN MINIM4L 5JT S/D 500JT BUNG4 RENDAH 2% THN MINAT WHATSAPP :0822-5334-6299"	penipuan	2.39E-75	1.66E-76	5.88E-68
8	ibu Retno WA 085314124447 M3n4w4rk4n PIN74M4N 50-500 Bunga 0,679%/bln T4NP4 J4min4n T4np4 Survey	penipuan	7.67E-67	7.47E-68	5.35E-59
9	Sudah dibawakan pagi tadi mas	normal	2.52E-16	2.85E-18	6.06E-18
10	Pagi aja jam segini gapapa	normal	4.61E-14	2.44E-16	4.37E-17
11	Za ganti sabtu minggu aja ya aku mau balik	normal	1E-28	5.23E-31	1.23E-31

12	INFO PINJ4M4N D4N4 ONLINE PROSES MUD4H/CEP4T PINJ4MAN MINIM4L 5JT S/D 500JT BUNG4 RENDAH 2% THN MINAT WHATSAPP :0822- 9038-2379	penipuan	1.43E- 78	3.91E- 79	7.91E-71
13	Aplikasi Uang Terbaru Di Indonesia ! Tebak Skor Salah Sepak Bola Investasi Keuangan yang Paling Bagus ! Buruan Download Dan Hub cs di https://bit.ly/3HkJ002	penipuan	7.67E- 68	1.24E- 69	6E-69
14	Ass,Bpk/Ibu Butuh Dana Buat Usaha Dan Tambha Modal Min 5jt S/D 500jt Bunga 2% Minat Silahkan WhatsApp 082316469392	penipuan	5.2E- 57	8.48E- 62	1.21E-53
15	Sebentar lagi Liga Champion loh !! Pilih tim favoritmu di AFAPo-Ker cutt.ly/AF4Po-Ker Nonton santai sambil profit !	penipuan	1.42E- 65	3.27E- 69	3.32E-69
16	Datang dan dapatkan penghasilan tambahan Anda, hingga 200RP per hari, hubungi kami dan biarkan kami memberi tahu Anda cara menghasilkan cutt.ly/VUmCTr6	penipuan	2.3E- 74	6.86E- 78	8.41E-79
17	Ekstra pulsa Telkomsel 25rb/1hr(hny utk SMS,tlp sesama Tsel) spesial UNTUKMU! Dengan beli paket Combo hanya di MyTelkomsel tsel.me/combosakti (s/d22Mar).SKB	promo	1.2E- 84	1.3E- 70	2.6E-85
18	Spesial untukmu! Dapatkan Ekstra Kuota 10GB dengan Tukar 1 Poin + Rp10 dan kamu langsung bisa nikmatin hari ini. Klik tsel.me/kejutanpoin . SKB	promo	1.02E- 71	4.65E- 63	1.62E-73
19	Temukan rekomendasi berbagai paket menarik yang KAMU BANGET NIH! Cek lebih lanjut di tsel.me/shop1 atau *888*15# skrg juga!	promo	2.04E- 59	6.07E- 58	3.62E-63

20	ADA KABAR GEMBIRA UNTUKMU! Sambut MyTelkomsel Anniversary yang ke-6, Yuk Dapatkan Cashback 60% hanya di MyTelkomsel Cek di tsel.me/annivmytsel	promo	5.68E- 65	1.81E- 64	3.01E-69
21	Mas nanti klo udah landing pean tlp y.. q yg jemput d bandara	normal	7.03E- 40	1.91E- 43	4.8E-41
22	ntar tinggal minta kunci kalo ke sini	normal	9.58E- 21	1.71E- 25	8.18E-25
23	Enggak sih za, coba kamu chat alam dia kan relawan siapa tau dia faham tempat2nya	normal	4.15E- 47	1.07E- 53	6.08E-56
24	Jadi nanti jadi satu aja kalau ke penguji	normal	6.84E- 25	2.9E- 28	1.36E-28
25	Paket di kursi kayu depan rumah	normal	1.2E- 19	2.21E- 20	6.97E-21
26	Nggak itu cuma untuk kompre kuliah. Pak imam kan kompre agama	normal	2.45E- 33	1.36E- 38	1.47E-38
27	Lumayan 10 menit dr kontrakan yg dulu	normal	8.79E- 18	3.85E- 20	7.66E-19
28	Nah gak tau lek itu, coba tanyain langsung ke admin aja biar lebih jelasnya	normal	7.09E- 42	1.9E- 48	9.27E-50
29	express kalau gk salah sediain request client juga kan?	normal	2.02E- 29	7.68E- 33	8.18E-33
30	nabilah maulida agar hub saya	normal	1.51E- 17	1.49E- 17	4.51E-17
31	Tetap terhubung dengan kerabatmu dengan paket Internet bulanan favoritmu. Beli skrg juga di tsel.me/omg atau *363#.	promo	1.18E- 54	8.5E- 51	5.02E-58
32	MURAH BANGET! Combo Sakti 20GB + UL Kuota Apps +Telp&SMS 30hr Hrg mulai 89Rb! PAKET MURAH lainnya sesuai kebutuhanmu jg ada, Beli di *363# atau tsel.me/combo1	promo	4.81E- 84	1.1E- 74	3.55E-88
33	Mau kuota Youtube, Instagram, Facebook yg murah meriah? Beli aja kuota ketengan! Dijamin gk	promo	8.83E- 71	1.44E- 68	4.24E-77

	<p>mahal loh. Aktifin pakatnya di tsel.me/ketengan & *363*847# SEKARANG</p>				
34	<p>120GB GRATIS SELAMA SETAHUN HANYA HARI INI di aplikasi ROLi Telkomsel! Download aplikasi ROLi Telkomsel sekarang tsel.me/ROLi_MD S&K</p>	promo	6.13E-71	2.39E-64	1.19E-73
35	<p>Langganan Netflix skrg bisa pake pulsa! Aktifkan paket berlangganan Netflix mulai dr Rp 62rb dpt kuota 6GB + BONUS kuota hingga 20GB di tsel.me/netflixpack. SKB</p>	promo	9.02E-81	2.72E-66	4.71E-80
36	<p>1/2 PROMO NSP Rp 0/7hr bls SMS ketik YA utk langganan NSP Cinta Sampai Mati-Kangen Band. Perpanjangan Rp.9990/30hari atau sesuai saldo Prabayar/tagihan</p>	promo	6.63E-77	8.37E-71	1.1E-75
37	<p>MODEM TANPA KABEL! Praktis dibawa kemana aja. Harga paket mulai 20K. Dapatkan ORBIT di konter GT RELOAD LOWOKWARU. Info: bit.ly/TELKOMSELOBIT</p>	promo	1.11E-67	1E-70	6.9E-73
38	<p>Ada Extra Kuota hingga 10GB di myTELKOMSEL! Mau? AYO Segera beli paket data bulanan mu & lakukan Daily Cek In melalui apps myTELKOMSEL. Info Klik : tsel.me/dci</p>	promo	6.01E-85	1.82E-73	4.99E-88
39	<p>Aktifkan ulang Paket Data Combo Sakti/Internet Sakti atau OMG anda melalui aps myTELKOMSEL. Dgn Daily Check-In rutin, bisa dapat EXTRA Kuota hingga 7,5GB #SKB</p>	promo	2.32E-79	1.1E-73	6.61E-85
40	<p>Baru! Buat pengguna Shopee. Beli Paket Combo/Internet sakti bisa pakai shopeepay. Klik aja langsung</p>	promo	5.96E-77	2.53E-70	5.47E-83

	di tsel.me/shopee-app . Pilih paket lalu bayar. SK				
41	No HP 081217299196, COIN PULSA Hari Ini: +1000. Bisa Kamu Ambil di *500*220#. Bisa Kamu Tukar PULSA+KUOTA. SEGERA Hub *500*220# dari HP ini utk Ambil REWARD mu!	penipuan	1.59E-75	7.52E-69	1.04E-74
42	Nomor HP 081217299196, COIN PULSA Anda: +2000. SEGERA Ambil di *500*75# Sebelum Hangus!	penipuan	7.05E-33	4.57E-32	1.22E-32
43	SELAMAT Anda mndptkn Hadiah berkah msglow Rp.157.000.000 PIN:MSG-130299 1NF0 HUB bit.ly/give-away-ms-glow Whatsapp 082297299219	penipuan	5.92E-59	2.37E-58	5.25E-55
44	Proses Depo & WD secepat kilat di T063LMANDIRI dik.si/Ob1NG Main Games b4ccar4t,P@K3R,Ce'Me,To_Gel japan, sydney,Bo'la sbo & Slot'Games online.	penipuan	4.6E-71	1.23E-74	3.79E-75
45	kami dr P-E-G-A-D-A-I-A-N Galeri24 mnwrkan barang yang tidak ditebus pemilik seperti 'Emas,iphone,laptop,dll.klik https://tinyurl.com/434csppa	penipuan	1.33E-76	1.72E-78	3.69E-75
46	J01N BR0T@63L & Main CardGames Do'mino B4cc4r4t DinG'DonG hkb,Slots,T*Gel Sydney,sbo C@S1N0 & Bo'la. cutt.ly/BR0899 Raih Bonusnya skrg jg	penipuan	3.22E-86	5.05E-90	6.22E-89
47	PINJAMAN DANA Buat Nmbah Modal Dan DLL Siap Bantu Min 5Jt S/D 500jt Bunga 2% Pertahun WhatsApp:085822577247	penipuan	1.44E-51	1.41E-54	2.8E-50
48	YTH.. Bpk/Ibu Sekluarga K@mi T4w@rkan 4nda P1nj4m@n	penipuan	1.76E-29	7.3E-31	3.11E-28

	Proses Cepat Inf0 WhatsApp +6282288799147				
49	Permisi Bpk/Ibu Butuh Dana INGIN KEMBANGKAN USAHA & Kebutuhan Lain Min 5Jt S/D 500Jt Bunga 2% Info Lanjut WA : +6282213927356	penipuan	6.13E- 58	3.79E- 61	9.03E-54
50	Raih Jackpot dgn main Slots online di TEXASP0 KERCC s.id/TPK999 Ada jg B4cc4r4t,P@K3R & DinG'DonG,To_Gel Sydney,Bo'la & C@51N0 Sbo. Mainkan skrg!	penipuan	8.94E- 76	5.39E- 78	5.05E-80