KLASIFIKASI SHORT TEXT MESSAGE SPAM MENGGUNAKAN ALGORITMA NAÏVE BAYES CLASSIFIER

SKRIPSI

Disusun Oleh:

JANNIO ALVARES

18.11.2373

PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2022

KLASIFIKASI SHORT TEXT MESSAGE SPAM MENGGUNAKAN ALGORITMA NAÏVE BAYES CLASSIFIER

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana

Program Studi Informatika

diajukan oleh

JANNIO ALVARES

18.11.2373

Kepada

PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA

2022

HALAMAN PERSETUJUAN

SKRIPSI

KLASIFIKASI SHORT TEXT MESSAGE SPAM MENGGUNAKAN ALGORITMA NAÏVE BAYES CLASSIFIER

yang disusun dan diajukan oleh

Jannio Alvares

18.11.2373

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 29 September 2021

Dosen Pembimbing,

Uyock Anggoro Saputro, M.Kom
NIK. 190302419

HALAMAN PENGESAHAN

SKRIPSI

KLASIFIKASI SHORT TEXT MESSAGE SPAM MENGGUNAKAN ALGORITMA NAÏVE BAYES CLASSIFIER

yang disusun dan diajukan oleh

Jannio Alvares

18.11.2373

Telah dipertahankan di depan Dewan Penguji
pada tanggal 18 Agustus 2022

Susunan Dewan Penguji

Nama Penguji	Tanda Tangan
Wahyu Sukestyastama Putra, S.T., M.Eng **NIK. 190302328**	
Yoga Pristyanto, S.Kom, M.Eng **NIK. 190302412**	
Uyock Anggoro Saputro, M.Kom **NIK. 190302419**	

Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 25 Agustus 2022

DEKAN FAKULTAS ILMU KOMPUTER

Hanif Al Fatta,S.Kom., M.Kom.
NIK. 190302096

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Jannio Alvares
NIM : 18.11.2373

Menyatakan bahwa Skripsi dengan judul berikut:

KLASIFIKASI SHORT TEXT MESSAGE SPAM MENGGUNAKAN ALGORITMA NAÏVE BAYES CLASSIFIER

Dosen Pembimbing : Uyock Anggoro Saputro, M.Kom

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 18 Agustus 2022

Yang Menyatakan,

Jannio Alvares

KATA PENGANTAR

Puji dan syukur kepada Tuhan Yang Maha Esa atas rahmat dan karunia-Nya, sehingga penulis dapat menyelesaikan skripsi dengan judul "Klasifikasi Short Text Message Spam Menggunakan Algoritma Naïve Bayes Classifier" sebagai syarat untuk memperoleh gelar sarjana komputer dari program studi Informatika Universitas Amikom Yogyakarta.

Selama proses pengerjaan penelitian serta penulisan tugas akhir ini, penulis mendapatkan dukungan, bantuan, bimbingan dan nasehat dari berbagai pihak sehingga pada kesempatan ini penulis menyampaikan terima kasih kepada:

1. Bapak Uyock Anggoro Saputro, M.Kom selaku dosen pembimbing skripsi yang telah memberikan bimbingan dan masukan kepada penulis selama menyelesaikan skripsi.
2. Seluruh dosen Informatika Universitas Amikom Yogyakarta yang telah mendidik dan memberikan ilmu kepada penulis.
3. Orang tua secara khusus ibu dan nenek serta seluruh keluarga yang ada di Buntok yang telah memberikan dukungan dan doa selama pengerjaan skripsi.
4. Sahabat dan teman dekat yang selalu memberikan saran serta membantu dalam proses pengerjaan skrips.
5. Semua pihak yang tidak bisa disebutkan satu per satu yang telah membantu penulis dalam mengerjakan skripsi ini

Penulis menyadari bahwa penulisan tugas akhir ini masih memiliki banyak kekurangan. Unutk itu penulis sangat membutuhkan kritik dan saran untuk perbaikan dimasa yang akan datang. Semoga penulisan tugas akhir ini dapat bermanfaat bagi semua pihak.

Yogyakarta, 18 Agustus 2022

Penulis

Jannio Alvares

DAFTAR ISI

DAFTAR GAMBAR

DAFTAR TABEL

INTISARI

Short Message Service atau yang lebih dikenal dengan SMS merupakan salah satu komunikasi dengan media teks melalui perangkat telepon seluler atau smartphone. Saat ini perkembangan teknologi informasi memungkinkan pengaksesan data lebih praktis, cepat, dan efisien, oleh karena itu dapat dimanfaatkan sebagian orang untuk melakukan spam untuk melakukan promosi terhadap suatu toko atau jasa. Spam adalah singkatan dari Sending and Posting Advertisement in Mass, yang artinya mengirim pesan secara massal. Selain untuk promosi ada juga yang menyalahgunakannya untuk melakukan tindakan kejahatan seperti penipuan berhadiah. Sebagian orang mungkin dapat mengenali apakah SMS tersebut berupa penipuan atau tidak, namun bagi orang awam yang baru saja mendapat SMS seperti itu tentu kebingunan dan mungkin akan tergiur. Oleh karena itu, tujuan penelitian ini adalah untuk mengklasifikasi jenis-jenis SMS apakah termasuk pesan normal, penipuan, atau promo. Metode yang digunakan dalam penelitian ini adalah Naïve Bayes Classifier. Sebelum data digunakan dilakukan proses preprocessing terlebih dahulu, kemudian dilakukan proses feature extraction menggunakan TF-IDF. Data yang sudah diproses akan dibagi menjadi dua bagian dan akan dilakukan proses cross validation untuk menghasilkan kinerja yang baik. Berdasarkan hasil yang diperoleh, Multinomial Naïve Bayes yang menggunakan proses cross validation mendapat akurasi mencapai 97,7%.

Kata Kunci: Klasifikasi, SMS, Spam, Naive Bayes, TF-IDF

ABSTRACT

Short Message Service or better known as SMS is a form of communication with text media via mobile phones or smartphones. Currently the development of information technology allows access to data more practical, fast, and efficient, therefore some people can use it to do spam to promote a store or service. Spam stands for Sending and Posting Advertisement in Mass, which means sending messages in bulk. In addition to promotion, there are also those who misuse it to commit crimes such as fraud with prizes. Some people may be able to recognize whether the SMS is a fraud or not, but for ordinary people who have just received such an SMS, they are certainly confused and may be tempted. Therefore, the purpose of this study is to classify the types of SMS whether they are normal, fraudulent, or promo messages. The method used in this research is the Naïve Bayes Classifier. Before the data is used, the preprocessing process is carried out first, then the feature extraction process is carried out using TF-IDF. The processed data will be divided into two parts and a cross validation process will be carried out to produce good performance. Based on the results obtained, Multinomial Nave Bayes which uses the cross validation process has an accuracy of 97,7%.

Keyword: Classification, SMS, Spam, Naive Bayes, TF-IDF