PERBANDINGAN KOMBINASI MODEL ALGORITMA *NAIVE BAYES* DENGAN TEKNIK PEMBOBOTAN KATA DALAM ANALISIS SENTIMEN

SKRIPSI

disusun oleh

Muttafi'ah

17.11.1236

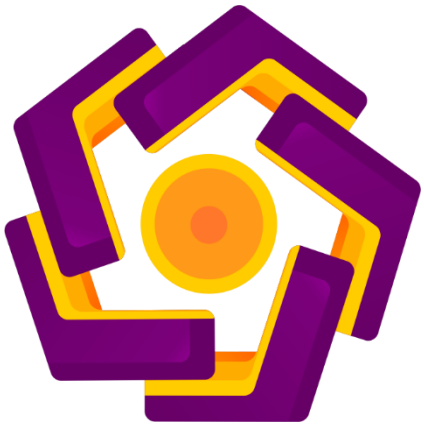
PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA

2021

PERBANDINGAN KOMBINASI MODEL ALGORITMA *NAÏVE BAYES* DENGAN TEKNIK PEMBOBOTAN KATA DALAM ANALISIS SENTIMEN

SKRIPSI

untuk memenuhi sebagian persyaratan
mencapai gelar Sarjana
pada Program Studi Informatika



disusun oleh

Muttafi'ah

17.11.1236

PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
2021

PERSETUJUAN

SKRIPSI

PERBANDINGAN KOMBINASI MODEL ALGORITMA *NAÏVE BAYES* DENGAN TEKNIK PEMBOBOTAN KATA DALAM ANALISIS SENTIMEN

yang dipersiapkan dan disusun oleh

Muttafi'ah
17.11.1236

telah disetujuai oleh Dosen Pembimbing Skripsi
pada tanggal 21 November 2020

Dosen Pembimbing,

Rizqi Sukma Kharisma, M.kom
NIK. 190302215

PENGESAHAN

SKRIPSI

PERBANDINGAN KOMBINASI MODEL ALGORITMA *NAÏVE BAYES* DENGAN TEKNIK PEMBOBOTAN KATA DALAM ANALISIS SENTIMEN

yang dipersiapkan dan disusun oleh

Muttafi'ah

17.11.1236

telah dipertahankan di depan Dewan Penguji
pada tanggal 17 Juni 2021

Susunan Dewan Penguji

Nama Penguji	Tanda Tangan
Wiwi Widayani, M.Kom **NIK. 190302272**	_____
Muhammad Tofa Nurcholis, M.Kom **NIK. 190302281**	_____
Rizqi Sukma Kharisma, M.Kom **NIK. 190302215**	_____

Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 18 Juni 2021

DEKAN FAKULTAS ILMU KOMPUTER

Hanif Al Fatta, M.Kom
NIK. 190302096

PERNYATAAN

Saya yang bertandatangan dibawah ini menyatakan bahwa, skripsi ini merupakan karya saya sendiri (ASLI), dan isi dalam skripsi ini tidak terdapat karya yang pernah diajukan oleh orang lain untuk memperoleh gelar akademis di suatu institusi pendidikan tinggi manapun, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis dan/atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Segala sesuatu yang terkait dengan naskah dan karya yang telah dibuat adalah menjadi tanggungjawab saya pribadi.

Yogyakarta, 26 Juni 2021

Muttafi'ah

NIM. 17.11.1236

MOTTO

“Hargai waktumu, karna 5 menit dalam hidupmu dapat mengubah duniamu”

KATA PENGANTAR

Puji dan syukur penulis persembahkan atas kehadirat Allah SWT yang telah melimpahkan rahmat, taufik, dan hidayah sehingga penulis dapat menyelesaikan skripsi ini. Tak lupa pula shalawat serta salam penulis haturkan kepada junjungan Nabi Muhammad SAW, yang telah menjadi teladan bagi seluruh umatnya.

Skripsi ini disusun untuk menyelesaikan studi jenjang Strata Satu (S1) pada program studi Informatika fakultas Ilmu Komputer Universitas Amikom Yogyakarta. Selain itu juga sebagai bukti bahwa mahasiswa telah memenuhi salah satu syarat untuk memperoleh gelar Sarjana Komputer.

Dalam penyusunan skripsi ini, penulis menyadari bahwa masih banyak kekurangan dan jauh dari kata sempurna. Selain itu dengan terselesaikannya skripsi ini, penulis ingin menyampaikan rasa terima kasih kepada:

1. Allah SWT atas segala berkah dan karunia serta ridho-Nya sehingga penulis dapat menyelesaikan skripsi ini dengan baik.
2. Nabi Muhammad SAW yang selalu menjadi panutan dan suri tauladan.
3. Kedua orang tua dan keluarga penulis yang selalu memberikan dukungan dalam bentuk apapun.
4. Bapak Prof. Dr. M. Suyanto, M.M., selaku Rektor Universitas Amikom Yogyakarta.
5. Bapak Rizqi Sukma Kharisma, M.kom, selaku dosen pembimbing yang telah memberikan bimbingan, saran dan arahan, serta masukan kepada penulis.

6. Seluruh staf pengajar dan karyawan Universitas Amikom Yogyakarta yang telah memberikan ilmu yang bermanfaat.
7. Dia yang selalu menemani, memberi semangat dan bantuan kepada saya.
8. Teman-teman 17 IF 05, dan gimang yang menemani dan memberikan semangat.
9. Serta semua pihak yang telah membantu penulis baik dukungan moral maupun materi, pikiran dan tenaga dalam menyelesaikan skripsi ini.

Semoga Allah SWT memberikan balasan lebih kepada semua yang telah ikut membantu penulis hingga terselesainya skripsi ini. Semoga skripsi ini dapat bermanfaat bagi penulis maupun pembaca.

Yogyakarta, 14 Juni 2021

Muttafi'ah

PERSEMBAHAN

Dengan penuh rasa syukur, segala puji Allah SWT atas rahmat dan karunia yang telah diberikan. Skripsi ini saya persembahkan kepada semua pihak yang terlibat secara langsung maupun tidak langsung dalam proses pembuatan skripsi.

1. Kedua orang tua dan kakak – kakak saya yang selalu mendoakan, memberi motivasi dan dukungan kepada saya.
2. Bapak Rizqi Sukma Kharisma, M.kom selaku dosen pembimbing yang telah membimbing saya dari awal hingga akhir pembuatan skripsi ini.
3. Dosen-dosen Universitas Amikom Yogyakarta yang telah memberikan banyak ilmu pengetahuan kepada saya selama perkuliahan.
4. Teman seperjuangan saya, Benedicta Kristi yang rela meluangkan waktu menemani saya dan membantu memberikan solusi, semangat, serta motivasi selama pembuatan skripsi.
5. Teman – teman Gimang, Destri Herliana Irianti, Akbar Hari Mukti, Muhammad Gufron Hawaly, Herlandro Tribiakto, Jordan Kurnia, dan Rhaka Noviansyah D. yang telah menjadi sahabat dan keluarga baru, dari awal perkuliahan hingga saat ini.
6. Teman-teman kelas 17 IF 05 yang selalu menemani perkuliahan.
7. Martha Presina yang selalu menemani, memberikan semangat, dorongan, motivasi dan menjadi tempat berkeluh kesah bagi saya.

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

INTISARI

Text mining merupakan konsep dasar analisis sentimen dan disiplin ilmu yang menggabungkan ilmu bahasa dan ilmu komputer dengan teknik pembelajaran mesin (*machine leaning*). *Text mining* digunakan untuk mengubah teks menjadi lebih terstruktur. Sedangkan pembelajaran mesin berfokus untuk mencari dan mengembangkan algoritma untuk membangun sebuah sistem yang dapat mensimulasikan atau meniru sebuah pola dari kumpulan data. Dalam penelitian ini menggunakan teknik supervised learning yang merupakan teknik dasar pembelajaran mesin dengan mengkomparasikan model algoritma *Naive Bayes Classifier* yaitu *Multinomial Naive Bayes* dan *Bernoulli Naive Bayes* dengan objek sentimen dari Twitter. Pada penelitian ini juga menggunakan teknik pembobotan kata yaitu TF-IDF dan TF- RF pada masing – masing model. Penelitian ini dilakukan untuk mengetahui kombinasi terbaik dari setiap model dengan pembobotan kata.

Tahap awal dalam penelitian ini adalah *crawling data* menggunakan API Twitter, kemudian data tersebut di*labeling*. Setelah data diberi label, data tersebut akan masuk ke tahap penting dalam penelitian, yaitu *preprocessing* dan pembobotan kata. Data yang telah dilabeli dibersihkan dan diubah menjadi data yang terstruktrur sehingga data siap untuk dianalisis. Data hasil *prepocessing* diberi bobot dengan teknik TF-IDF dan TF-RF, kemudian diklasifikasi satu-persatu menggunakan 2 model NBC sehingga dalam penelitian ini terdapat 4 skema model, yaitu Multinomial dan TF-IDF, Bernoulli dan TF-IDF, Multinomial dan TF-RF serta Bernoulli dan TF-RF. Tahap terakhir dari penelitian ini adalah pengujian menggunakan *Confusion Matrix* kemudian divalidasi dengan *K-Fold Cross Validation*, pengujian dilakukan untuk melihat performa terbaik dari 4 skema tersebut.

Dari 4 skema yang dilakukan, skema TF-IDF dan TF-RF dengan *Bernoulli Naive Bayes* dari hasil pengujian *Confusion Matrix* mengahasilkan akurasi terbaik yaitu 61%, dan rata - rata nilai akurasi dari validasi 5-*fold* sebesar 60%. Dan yang memiliki nilai akurasi terendah terletak pada model *Multinomial Naive Bayes* dan TF-IDF yaitu 58% dari *Confusion Matrix*, dengan rata-rata nilai akurasi 59% dari validasi 5-*fold*.

Kata-kunci: Multinomial Naïve Bayes, Bernoulli Naïve Bayes, TF-IDF, TF-RF, Analisis sentimen

ABSTRACT

Text mining is a basic concept of sentiment analysis and a discipline that combines linguistics and computer science with machine learning techniques. Text mining is used to change the text to be more structured. While machine learning focuses on finding and developing algorithms to build a system that can simulate or imitate a pattern from a dataset. In this study, supervised learning is used which is a basic machine learning technique with comparing the Naive Bayes Classifier algorithm model, namely Multinomial Nave Bayes and Bernoulli Naive Bayes with sentiment objects from Twitter. This study also uses Term Weighting techniques, namely TF-IDF and TF-RF in each model. This study was conducted to determine the best combination of each model with Term Weighting.

The first step in this research is crawling the data using the Twitter API, then the data is labeled. After the data is labeled, the data will enter an important step in the research, namely preprocessing and term weighting. The data that has been labeled is cleaned and converted into structured data so that the data is ready for analysis. The preprocessing data are weighted using the TF-IDF and TF-RF techniques, then classified one by one using 2 NBC models, so in this study there are 4 model schemes, namely Multinomial and TF-IDF, Bernoulli and TF-IDF, Multinomial and TF-RF as well as Bernoulli and TF-RF. The last stage of this research is testing using Confusion Matrix, and then validated with K-Fold Cross Validation, testing is carried out to see the best performance of the 4 schemes.

Of the 4 schemes, TF-IDF and TF-RF with Bernoulli Naive Bayes schemes from the results of the Confusion Matrix test produce the best accuracy 61%, and the average accuracy value of the 5-fold validation is 60%. And the one with the lowest accuracy value lies in the Multinomial Naive Bayes model and TF-IDF which is 58% from Confusion Matrix, with an average value of 59% from the 5-fold validation.

***Keywords**: Multinomial Nave Bayes, Bernoulli Nave Bayes, TF-IDF, TF-RF, Sentiment analysis*