

SKRIPSI

KLASIFIKASI BERITA DAN ANALISIS SENTIMEN TWITTER TERHADAP KOMPONEN CADANGAN KEMENTERIAN PERTAHANAN RI DENGAN METODE NAÏVE BAYES CLASSIFIER



NURUL FAJRIA

NIM : 185610088

PROGRAM STUDI SISTEM INFORMASI

PROGRAM SARJANA

FAKULTAS TEKNOLOGI INFORMASI

UNIVERSITAS TEKNOLOGI DIGITAL INDONESIA

YOGYAKARTA

2022

SKRIPSI

KLASIFIKASI BERITA DAN ANALISIS SENTIMEN TWITTER TERHADAP KOMPONEN CADANGAN KEMENTERIAN PERTAHANAN RI DENGAN METODE *NAÏVE BAYES CLASSIFIER*



Disusun Oleh

NURUL FAJRIA

NIM : 185610088

**PROGRAM STUDI SISTEM INFORMASI
PROGRAM SARJANA
FAKULTAS TEKNOLOGI INFORMASI
UNIVERSITAS TEKNOLOGI DIGITAL INDONESIA
YOGYAKARTA**

PERNYATAAN KEASLIAN SKRIPSI

Dengan ini saya menyatakan bahwa naskah skripsi ini belum pernah diajukan untuk memperoleh gelar Sarjana Komputer di suatu Perguruan Tinggi, dan sepanjang pengetahuan saya tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara sah diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Yogyakarta,*23 Agustus*.....2022



Nurul Fajria

NIM: 185610088

HALAMAN PERSEMBAHAN

Skripsi ini saya persembahkan untuk:

Untuk Kedua orang tua saya yang tak henti-hentinya mendukung, mendoakan, serta memberikan kasih sayangnya tak terhingga selama ini saya ucapan banyak terimakasih.

Untuk Bapak Ir. Totok Suprawoto, MM, MT selaku dosen pembimbing yang telah memberikan bimbingan pengarahan serta semangat dalam penyusunan skripsi saya.

Kepada Annisa Salsabila dan Nurwahida Besti saya yang selalu mensuport dan selalu memotivasi saya saat saya sudah bosan mengerjakan skripsi saya.

Kepada Yusuf Nur Rahman Wahid yang telah membantu dan memberi pencerahan kepada saya saat saya sudah putus asa bertemu eror.

Serta saya ucapan terima kasih untuk seluruh teman teman yang telah membantu kelancaran pembuatan skripsi ini terima kasih atas bantuan dan kerjasama nya.

Saya ucapan terima kasih kepada Teman teman seperjuangan saya teman angkatan 2018 yang tidak saya bisa sebutkan satu persatu.

Kepada seluruh warga Kwangya saya banyak sekali termotivasi untuk lebih bekerja keras lagi untuk meraih mimpi saya.

HALAMAN MOTTO

**“Maka sesungguhnya bersama kesulitan ada kemudahan, sesungguhnya bersama
kesulitan ada kemudahan”
(QS. Al Insyirah : 5-6)**

**“Do The Things You Want to do, Do The Things You Thought Were Right, Do The
Things You Like”
(Jeong Jaehyun)**

**“I feel like the possibility of all those possibilities being possible is just another
possibility that can possibly happen.”
(Mark Lee)**

KATA PENGANTAR

Assalamu'alaikum warahmatullahi wabarakatuh. Bismillahirrahmanirrohim, segala puji dan syukur kehadiran Allah SWT yang melimpahkan rahmat dan hidayah-Nya sehingga penulis dapat menyelesaikan Skripsi dengan judul "Klasifikasi Bdan Analisis Sentimen Twitter Terhadap Komponen Cadangan Kementerian Pertahanan RI dengan menggunakan metode *naïve bayes classifier*". Laporan Skripsi ini disusun untuk memenuhi salah satu syarat dalam memperoleh gelar Sarjana Komputer Universitas Teknologi Digital Indonesia.

Dengan melakukan penelitian dan penyusunan Tugas Akhir ini penulis telah mendapatkan banyak dukungan dan bantuan dari berbagai pihak. Penulis mengucapkan terima kasih yang tak terhingga kepada:

1. Allah SWT yang telah memberikan ridho, bimbingan, berkah, rahmat, hidayah dan inayah kepada penulis selama melakukan penelitian dan menyusun laporan skripsi ini.
2. Ir. Totok Suprawoto, MM.,MT , selaku Rektor Universitas Teknologi Digital Indonesia sekaligus Dosen Pembimbing saya yang telah berkenan memberikan izin dan membimbing serta memberi arahan dalam pembuatan Skripsi ini.
3. Pulut Suryati, S.Kom., M.Cs. selaku Ketua Program Studi Sistem Informasi Universitas Teknologi Digital Indonesia.
4. Para dosen Program Studi Strata 1 Studi Sistem Informasi Universitas Teknologi Digital Indonesia yang telah memberikan bekal ilmu kepada penulis.
5. Keluarga tercinta, Bapak, Ibu dan saudara yang selalu mendukung dalam menyelesaikan Skripsi ini.
6. Teman-teman organisasi DPM, HMJ, WAMIKA, MENWA dan Falsa yang telah mendukung dan mensuport saya dalam menyusun penelitian ini.
7. Terimakasih kepada sahabat-sahabat penulis yang telah memberi dukungan.

Penulis menyadari sepenuhnya bahwa laporan Skripsi ini masih sangat jauh dari sempurna. Untuk itu, semua jenis saran, kritik dan masukkan yang bersifat membangun sangat penulis harapkan. Akhir kata, semoga tulisan ini dapat memberikan manfaat dan memberikan wawasan tambahan bagi pembaca dan khususnya bagi penulis sendiri.

Terimakasih,

Wassalamualaikum warahmatullahi wabarakatuh.

Yogyakarta, 2022

Penulis

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERSETUJUAN	ii
HALAMAN PENGESAHAN.....	iii
PERNYATAAN KEASLIAN SKRIPSI.....	iv
HALAMAN PERSEMBAHAN	v
HALAMAN MOTTO	vi
KATA PENGANTAR	vii
DAFTAR ISI.....	ix
DAFTAR GAMBAR	xii
DAFTAR TABEL.....	xiii
INTISARI.....	xiv
ABSTRACT	xv
BAB I	1
PENDAHULUAN	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah	3
1.3 Ruang Lingkup	3
1.4 Tujuan Penelitian.....	3
1.5 Manfaat Penelitian.....	4
1.6 Sistematika Penelitian	4
BAB II.....	5
TINJAUAN PUSTAKA DAN DASAR TEORI	5
2.2.1 Machine Learning.....	10
2.2.2 Data Science	12

Gambar 2. 1Ruang Lingkup Data Science.....	13
2.2.3 Sentiment Analysis.....	13
2.2.4 Twitter	14
2.2.5 Crawling Data Twitter.....	15
2.2.6 Text Mining	15
2.2.8 Naive Bayes Classifier.....	18
2.2.9 Python.....	19
2.2.10 Jupyter Notebook.....	20
BAB III.....	21
METODE PENELITIAN.....	21
3.1 ANALISIS SITEM.....	21
3.1.1 KEBUTUHAN DATA.....	21
3.1.2 Kebutuhan Proses	21
3.1.3 Kebutuhan Perangkat Lunak	22
3.1.4 Kebutuhan Perangkat Keras.....	22
3.2 Prosedur dan Pengambilan Data	23
3.2.1 Metode Pengumpulan Data	23
3.2.2 Prosedur Pengumpulan Data	23
3.3 Analisis dan Perancangan Sistem	23
3.3.1 Alur Kerja Sistem	23
BAB IV	28
IMPLEMENTASI DAN PEMBAHASAN	28
4.1 Implementasi dan Pembahasan Sistem	28
4.1.1 Scraping Data Twitter	28
4.1.2 Preprocessing	29
4.1.3 Implementasi Pemodelan Sistem	37
4.1.4 Implementasi Naïve Bayes Classifier.....	38

4.1.5 Pengujian Akurasi Perhitungan Naïve Bayes Classifier	43
4.1.6 Visualisai.....	44
4.2 Pembahasan Output	45
BAB V.....	48
PENUTUP	48
5.1 Kesimpulan.....	48
5.2 Saran.....	49
DAFTAR PUSTAKA.....	50

LAMPIRAN

DAFTAR GAMBAR

GAMBAR 2. 1RUANG LINGKUP DATA SCIENCE	13
GAMBAR 3. 1 DIAGRAM BLOK	24
GAMBAR 3. 2 CRAWLING DATA	24
GAMBAR 3. 3 TAHAP PRE-PROCESSING DATA	25
GAMBAR 3. 4 PREDICTIVE MODEL.....	27
GAMBAR 4. 1 POTONGAN PROGRAM SCRAPING DATA.....	28
GAMBAR 4. 2 POTONGAN HASIL CRAWLING DATA	29
GAMBAR 4. 3 POTONGAN PROGRAM CASE FOLDING.....	30
GAMBAR 4. 4 POTONGAN PROGRAM TOKENIZE	33
GAMBAR 4. 5 POTONGAN PROGRAM FILTERING-STOPWORDS	34
GAMBAR 4. 6 HASIL FILTERING-STOPWORDS.....	34
GAMBAR 4. 7 POTONGAN PROGRAM STEMMING	36
GAMBAR 4. 8 HASIL STEMMING.....	36
GAMBAR 4. 9 POTONGAN GAMBAR PEMODELAN SISTEM	37
GAMBAR 4. 10 POTONGAN PROGRAM PENERAPAN MODELING.....	38
GAMBAR 4. 11 POTONGAN PROGRAM NAIVE BAYES.....	39
GAMBAR 4. 12 POTONGAN PROGRAM TESTING	39
GAMBAR 4. 13 HASIL TESTING 1	40
GAMBAR 4. 14 POTONGAN PROGRAM UJI KLASIFIKASI LABEL.....	41
GAMBAR 4. 15 HASIL KLASIFIKASI LABEL	42
GAMBAR 4. 16 CONTOH KLASIFIKASI YANG SALAH.....	43
GAMBAR 4. 17 POTONGAN PROGRAM MENGHITUNG AKURASI.....	44
GAMBAR 4. 18 POTONGAN PROGRAM VISUALISASI	45
GAMBAR 4. 19 HASIL VISUALISASI DATA	45
GAMBAR 4. 20 HASIL VISUALISAI PENGUJIAN KE-2.....	46
GAMBAR 4. 21 HASIL KESELURUHAN.....	47

DAFTAR TABEL

TABEL 2. 1 : TINJAUAN PUSTAKA 1	6
TABEL 4. 1 TABEL PERBANDING SEBELUM DAN SESUDAH CASE FOLDING.....	30

INTISARI

Twitter merupakan layanan media sosial yang biasa digunakan penggunanya untuk mengutarakan pendapat mengenai kegiatan atau kejadian kejadian yang dialami melalui tweet pada akun Twitter mereka. Twitter biasanya digunakan untuk penyampaian opini ataupun kritik mengenai suatu topik berita yang sedang ramai dibicarakan, oleh sebab itu twitter dapat dijadikan sarana yang baik untuk menampung suara atau pendapat masyarakat. Selain itu, sifat Twitter yang selalu up-to-date juga memudahkan dalam penerapan metode analisis dalam segi data. Machine Learning adalah ilmu yang mempelajari tentang algoritma komputer yang bisa mengenali pola-pola didalam data, dengan tujuan mengubah beragam macam data menjadi suatu tindakan yang nyata dengan sesedikit mungkin campur tangan manusia. Dengan *Machine Learning*, kita dapat menciptakan mesin (komputer) yang “ belajar ” dari data yang ada, selanjutnya dia bisa membuat keputusan secara mandiri tanpa perlu diprogram lagi. Secara umum, *Machine Learning* berada dibawah payung *Artifical Intelligence/AI*, (Kecerdasan Buatan).

Penelitian ini menggunakan metode *Naïve Bayes Classifier* dengan 3 tahapan proses yaitu pre-processing yang terdiri dari cleansing, case folding, tokenizing, filtering-stopword, dan stemming. Tahap kedua yaitu pemodelan sistem dengan menentukan *feature* dan *label* pada training dataset. Tahap terakhir adalah klasifikasi sentimen yang menghasilkan 2 kategori yaitu sentimen positif dan negatif.

Data yang digunakan berupa opini masyarakat terhadap Komponen Cadangan Kementerian Pertahanan dan Keamanan Negara Republik Indonesia dengan jumlah total 152 data tweet yang dibagi menjadi 70% data latih dan 30% data uji, dan diperoleh dari hasil scrapping menggunakan snscraper. Hasil klasifikasi sentimen yang dilakukan oleh sistem diperoleh sebanyak 68,72% sentimen positif dan 27,8% sentimen negatif. Dari hasil pengujian pada sistem yang pertama diperoleh akurasi sebesar 85% kemudian peneliti melakukan pengujian ulang pada sistem dengan data uji baru, diperoleh akurasi sebesar 90% dan untuk keseluruhan hasil akurasi sebesar 90%. Hasil tersebut menunjukkan bahwa dari data yang di proses masyarakat lebih banyak berkomentar positif dari pada berkomentar negatif.

Kata Kunci : *Analisis Sentiment, Twitter, Machine Learning, Naïve Bayes Classifier, komcad.*

ABSTRACT

Twitter is a social media service that is usually used by users to express opinions about activities or events experienced through tweets on their Twitter account. Twitter is usually used to convey opinions or criticisms about a news topic that is being discussed, therefore Twitter can be used as a good means to accommodate people's voices or opinions. In addition, Twitter's up-to-date nature also makes it easier to apply analytical methods in terms of data. Machine Learning is the study of computer algorithms that can recognize patterns in data, with the aim of turning various kinds of data into real actions with as little human intervention as possible. With *Machine Learning*, we can create machines (computers) that "learn" from existing data, then they can make decisions independently without needing to be programmed again. In general, *Machine Learning* is under the umbrella of ***Intelligence / AI***, (Artificial Intelligence).

This study uses the *Naïve Bayes Classifier* with 3 stages of process, namely pre-processing which consists of cleansing, case folding, tokenizing, filtering-stopword, and stemming. The second stage is system modeling by determining the *features* and *labels* on the training dataset. The last stage is sentiment classification which produces 2 categories, namely positive and negative sentiments.

The data used is in the form of public opinion on the Reserve Components of the Ministry of Defense and State Security of the Republic of Indonesia with a total of 152 tweet data which is divided into 70% training data and 30% test data, and obtained from scrapping using snscreper. The results of the sentiment classification carried out by the system obtained 68.72% positive sentiment and 27.8% negative sentiment. From the results of testing on the first system, an accuracy of 85% was obtained, then the researchers retested the system with new test data, obtained an accuracy of 90% and for the overall accuracy of 90%. These results indicate that from the data processed, the community made more positive comments than negative comments.

Keywords: *Sentiment Analysis, Twitter, Machine Learning, Naïve Bayes Classifier, komcad.*