KLASIFIKASI DAN ANALISIS SENTIMEN PADA DATA TWITTER MENGGUNAKAN ALGORTIMA NAÏVE BAYES (STUDI KASUS: PEKAN OLAHRAGA NASIONAL XX 2021)

SKRIPSI

KRISNA JONATHAN SITORUS

1810511090

UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAKARTA

FAKULTAS ILMU KOMPUTER

PROGRAM STUDI INFORMATIKA

2022

KLASIFIKASI DAN ANALISIS SENTIMEN PADA DATA TWITTER MENGGUNAKAN ALGORTIMA NAÏVE BAYES (STUDI KASUS: PEKAN OLAHRAGA NASIONAL XX 2021)

SKRIPSI

Diajukan Sebagai Salah Satu Syarat untuk Memperoleh Gelar Sarjana Komputer

KRISNA JONATHAN SITORUS

1810511090

UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAKARTA

FAKULTAS ILMU KOMPUTER

PROGRAM STUDI INFORMATIKA

2022

PERNYATAAN ORISINALITAS

Skripsi ini adalah hasil karya sendiri, dan sumber yang dikutip maupun yang dirujuk telah saya nyatakan dengan benar.

Nama : Krisna Jonathan Sitorus

NIM : 1810511090

Tanggal : 13 Juni 2022

Bilamana dikemudian hari ditemukan ketidaksamaan dengan pernyataan ini, maka saya bersedia dituntut dan diproses sesuai dengan ketentuan yang berlaku.

Jakarta, 13 Juni 2022

Krisna Jonathan Sitorus

PERNYATAAN PERSETUJUAN PUBLIKSASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS

Sebagai civitas akademik Universitas Pembangunan Nasional Veteran Jakarta, saya yang bertanda tangan dibawah ini:

Nama : Krisna Jonathan Sitorus

NIM : 1810511090

Fakultas : Ilmu Komputer

Program Studi : Informatika

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Pembangunan Nasional Veteran Jakarta Hak Bebas Royalti Non Eksklusif (*Non-exclusive Royalty Free Right*) atas karya ilmiah saya yang berjudul:

Klasifikasi dan Analisis Sentimen Pada Data Twitter Menggunakan Algoritma Naïve Bayes. (Studi Kasus: Pekan Olahraga Nasional XX 2021)

Beserta perangkata yang ada (jika diperlukan). Dengan Hak Bebas Royalti ini Universitas Pembangunan Nasional Veteran Jakarta berhak menyimpan, mengalih-meida/formatkan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan mempublikasikan Tugas akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Jakarta

Pada tanggal : 22 Juli 2022

Yang menyatakan,

(Krisna Jonathan Sitorus)

LEMBAR PENGESAHAN

Dengan ini dinyatakan bahwa Skripsi berikut:

Nama : Krisna Jonathan Sitorus

NIM : 1810511090

Program Studi : S1 Informatika

Judul : Klasifikasi dan Analisis Sentimen Pada Data Twitter Menggunakan Algoritma Naïve Bayes. (Studi Kasus: Pekan Olahraga Nasional XX 2021)

Telah berhasil dipertahankan di hadapan Tim penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana Komputer pada Program Studi S1 Informatika, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional Veteran Jakarta.

Henki Bayu Seta, S.Kom, MTI.
Penguji 1

Mayanda Mega Santoni, S.Kom, M.Kom.
Penguji 2

Anita Muliawati, S.Kom., MTI.
Dosen Pembimbing 1

Sarika, S.Kom., M.Kom.
Dosen Pembimbing 2

Dr. Ermatita, M.Kom.
Dekan

Desta Sandya Prasvita, S.Kom., M.Kom.
Ketua Program Studi

Ditetapkan di : Jakarta

Tanggal Pengesahan : 20 Juli 2022

KLASIFIKASI DAN ANALISIS SENTIMEN PADA DATA TWITTER MENGGUNAKAN ALGORITMA NAÏVE BAYES. (STUDI KASUS: PEKAN OLAHRAGA NASIONAL XX 2021)

Krisna Jonathan Sitorus

ABSTRAK

Dunia teknologi sekarang ini sedang mengalami perkembangan yang cukup signifikan, masyarakat kini bisa dengan bebas mengakses dan memanfaatkan media sosial sebagai contoh yakni Twitter. Banyak sekali pengguna Twitter mencurahkan pendapatkan melalui tweet-tweet yang mereka kirimkan pada media sosial tersebut, khususnya mengenai Pekan Olahraga Nasional (PON) XX 2021 silam. Banyak sekali tweet yang bersifat mendukung, tetapi tidak jarang juga ada tweet yang bersifat keluhan mengenai penyelenggaraan PON XX tersebut. Dari masalah tersebut, dilaksanakanlah penelitian mengenai analisis sentimen pada data twitter yang berkaitan dengan PON XX dan mempergunakan metodeNaïve Bayes. Jumlah data yang di crawling sebanyak 1000 data lalu data melalui proses *dataduplicate removal* sehingga memperolehkan hasil data sebanyak 218 data tweet dan belum terlabelkan. Sebelum proses mengklasifikasikan data yang diperoleh, data harus dilaksanakan pemberian label pada datanya serta pembersihan data terlebih dulu sebelum masuk pada tahapan *text processing*, selanjutnya data diberi bobot pada tiap kata dengan *Term Frequency– Inverse Document Frequency* (TF-IDF) yang akan kedepannya kata tersebut akan dijadikan sebagai fitur. Lalu, karena data berlabel positif serta negatif mempunyai jumlah yang jauh berbeda, maka dimanfaatkan metode *Synthetic Minority Oversampling Technique* (SMOTE) guna melaksanakan penyeimbangan terhadap datanya. Tahapan selanjutnya dilaksanakan pembagian data yang besarannya yakni 80% 20% dan diklasifikasikan dengan metode Naive Bayes. Hasil yang diperolehkan dari pelaksanaan penelitiannya tersebut ialah diperoleh bahwa data uji memperolehkan accuracy yang besaran persentasenya yakni 99%, precision dengan besaran persentasenya 100%, recall dengan besaran persentasenya 98%.

Kata Kunci: Analisis Sentimen, Klasifikasi, PON XX, Naïve Bayes

CLASSIFICATION AND SENTIMENT ANALYSIS ON TWITTER DATA USING NAÏVE BAYES. (CASE STUDY: PEKAN OLAHRAGA NASIONAL XX 2021)

Krisna Jonathan Sitorus

ABSTRACT

The world of technology is currently developing very rapidly, people can now freely accessand express themselves using social media as an example, namely Twitter. Many Twitter users express their opinions through the tweets they send on social media, especiallyregarding the XX 2021 National Sports Week (PON) ago. Several types of tweets that are appreciative of the event are often seen, but not infrequently there are also tweets that are complaining about the implementation of the XX PON. From this problem, a research was conducted on sentiment analysis on twitter data regarding PON XX using the Naïve Bayes method. The amount of data that is crawled is 1000 data and then the data goes through a duplicate data removal process so that it produces 218 tweets and has not been labeled. Before the process of classifying the data obtained, the data must be labeled and cleaned firstbefore entering the text processing stage, then the data is given a weight for each word witha Term Frequency–Inverse Document Frequency (TF-IDF) which will be used in the future.as a feature. Then, because the data with positive and negative labels have significantly different amounts, the Synthetic Minority Oversampling Technique (SMOTE) method was used to balance the data. The next stage is the distribution of data by 80% 20% and classifiedby the Naive Bayes method. The results obtained from this study are obtained that the test data get 99% accuracy, 100% precision, 98% recall.

Keyword: Sentiment Analysis, Classification, PON XX, Naïve Bayes

KATA PENGANTAR

Segala puji dan syukur penulis panjatkan kehadirat Tuhan Yang Mahakuasa atas limpahan rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan Skripsi (Tugas Akhir). Skripsi ini disusun sebagai syarat kelulusan Studi Informatika Fakultas Ilmu Komputer Universitas Pembangunan Nasional Veteran Jakarta.

Dalam penyelesaian karya tulis ilmiah ini tidak lepas dari bantuan banyak pihak yang telah memberikan masukan kepada penulis. Untuk itu penulis mengucapkan terima kasih kepada:

1. Dr. Ermatita, M.Kom., selaku dekan Fakultas Ilmu Komputer
2. Desta Sandya Prasvita, S.Kom, M.Kom., selaku Ketua Program Studi Sarjana Jurusan S1 Informatika.
3. Anita Muliawati, S.Kom., MTI. selaku dosen pembimbing 1 dari pihak jurusan.
4. Sarika, S.Kom., M.Kom. selaku dosen pembimbing 2 dari pihak jurusan.
5. Orang tua yang telah memberikan dukungan baik secara moril maupun materil.
6. Seluruh pihak yang terlibat dalam kelancaran pembuatan makalah karya ilmiah ini dan yang belum disebutkan di atas, penulis ucapkan terima kasih.

Penulis menyadari bahwa masih banyak kekurangan dari laporan ini, baik dari materi maupun teknik penyajiannya. Oleh karena itu, kritik dan saran yang membangun penulis harapkan.

Jakarta, 8 Januari 2021

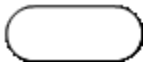
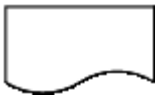
Krisna Jonathan Sitorus

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

DAFTAR SIMBOL

Simbol Flowchart			
NO	**GAMBAR**	**NAMA**	**KETERANGAN**
1		*Terminal (start, end)*	MengGambarkan bagaimana kegiatan dimulai atau kegiatan berakhir.
2		*Flow Direction*	MengGambarkan hubungan antar simbol yang menyatakan suatu jalannya proses dalam sistem.
3		*Process*	MengGambarkan deskripsi dari proses yang dijalankan.
4		*Document*	MengGambarkan bahwa masukkan (*input*) berasal dari sebuah data dokumen yang dapat berupa kertas atau keluaran (*output*) yang dicetak ke kertas.
5		*Predefine process*	MengGambarkan pelaksanaan dari sebuah proses atau disebut sebagai subprogram.