IMPLEMENTASI ALGORITMA *NAÏVE BAYES* TERHADAP ANALISIS SENTIMEN PADA TIMNAS BULU TANGKIS INDONESIA DI JEJARING SOSIAL TWITTER

SKRIPSI

ALPIN ANUGERAH

1810511118

UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAKARTA

FAKULTAS ILMU KOMPUTER

PROGRAM STUDI INFORMATIKA

2022

IMPLEMENTASI ALGORITMA *NAÏVE BAYES TERHADAP* ANALISIS SENTIMEN PADA TIMNAS BULU TANGKIS INDONESIA DI JEJARING SOSIAL TWITTER

SKRIPSI

Diajukan Sebagai Salah Satu Syarat untuk Memperoleh Gelar Sarjana Komputer

ALPIN ANUGERAH

1810511118

UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAKARTA

FAKULTAS ILMU KOMPUTER

PROGRAM STUDI INFORMATIKA

2022

PERNYATAAN ORISINALITAS

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

Nama : Alpin Anugerah

NIM : 1810511118

Tanggal : 19 Januari 2023

Bilamana di kemudian hari ditemukan ketidaksesuaian dengan pernyataan saya ini, maka saya bersedia dituntut dan diproses sesuai dengan ketentuan yang berlaku.

Jakarta, 19 Januari 2023

Yang Menyatakan,

Alpin Anugerah

PERNYATAAN PERSETUJUAN PUBLIKSASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS

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

Nama : Alpin Anugerah

NIM : 1810511118

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:

Implementasi Algoritma Naive Bayes Terhadap Analisis Sentimen Pada Timnas Bulu Tangkis Indonesia Di Jejaring Sosial Twitter

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

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Jakarta

Pada tanggal : 10 Januari 2023

Yang menyatakan,

Alpin Anugerah

LEMBAR PENGESAHAN

Tugas Akhir ini diajukan oleh:

Nama	: Alpin Anugerah
NIM	: 1810511118
Program Studi	: S1 Informatika
Judul Tugas Akhir	: Implementasi Algoritma Naïve Bayes Terhadap Analisis Sentimen Pada Timnas Bulu tangkis Indonesia Di Jejaring Sosial Twitter

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

Dr. Ermatita, M. Kom.
Penguji 1

Bambang Tri W., S.Kom., MSi.
Penguji 2

Bayu Hananto, S.Kom., M.Kom.
Pembimbing 1

Helena Nurramdhani Irmanda, S.Pd., M.Kom.
Pembimbing 2

Dr. Ermatita, M. Kom.
Dekan

Dr. Widya Cholil, M.I.T
Kepala Program Studi

Ditetapkan di : Jakarta
Tanggal Ujian : 09 Januari 2023

IMPLEMENTASI ALGORTIMA *NAÏVE BAYES* TERHADAP ANALISIS SENTIMEN PADA TIMNAS BULUTANGKIS INDONESIA PADA JEJARING SOSIAL TWITTER

Alpin Anugerah

ABSTRAK

Jejaring sosial bukan hanya berguna sebagai media untuk berkomuniksi yang tergolong optimal, tetapi jejaring sosial ini juga bisa berperan sebagai tempat dalam menampungkan pendapat ataupun opini dari masyarakat secara meluas. Salah satu jejaring sosial yang kerap dipergunakan dalam menampung seluruh opini masyarakat ini ialah *Twitter*. Pihak yang mempergunakan *Twitter* di tanah air kerap kali menyampaikan opininya terkait dengan seluruh bidang, tanpa terkecuali bidang olahraga, terkhusus berkaitan dengan Timnas Bulutangkis Indonesia. Ditemukan banyak opini yang disampaikan masyarakat terkait dengan Timnas Bulutangkis tanah air, dimana hal ini bisa ditemukan pada jejaring sosial *Twitter*. Dari keadaan tersebut, maka dilaksanakannya penelitian yang berhubungan dengan opini masyarakat perihal Timnas Bulutangkis Indonesia. Salah satu metode yang diterapkan yakni melalui pelaksanaan penganalisisan terhadap sentimen pada Timnas Bulutangkis Indonesia dalam jejaring sosial *Twitter* dengan memanfaatkan metode klasifikasi serta algoritma *Naïve Bayes* guna memperolehkan klasifikasi terhadap tweet yang bernilai positif ataupun negatif dari masyarakat terhadap Timnas Bulutangkis Indonesia. Jumlah data yang di crawling sebanyak 3277 data lalu data melalui proses *data duplicate removal* sehingga memperolehkan hasil data sebanyak 272 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 akan diberi bobot pada setiap 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 yaitu dilaksanakannya pembagian data yang besarannya yakni 80% 20% dan diklasifikasikan dengan metode Naive Bayes. Maka hasil yang diperolehkan dari pelaksanaan penelitiannya tersebut ialah didapat bahwa data uji memperolehkan accuracy yang besaran persentasenya yakni 96%, precision dengan besaran persentasenya 100%, recall dengan besaran persentasenya 92%.

Kata Kunci: Analisis Sentimen, Klasifikasi, Bulu tangkis, Naïve Bayes

IMPLEMENTATION *NAÏVE BAYES* ALGORTIMA ON SENTIMENT ANALYSIS OF THE INDONESIAN BADMINTON NATIONAL TEAM ON TWITTER SOCIAL NETWORKS

Alpin Anugerah

ABSTRACT

Social networks are not only useful as an effective medium of communication, but social networks can also be a place to accommodate opinions or opinions among the general public. One social network that is widely used to accommodate these opinions is Twitter.users Twitter in Indonesia often express their opinions on all fields including sports, especially the Indonesian Badminton National Team. There are various kinds of opinions about the Indonesian Badminton National Team, which can be seen on Twitter. Based on these conditions, research was carried out on public opinion about the Indonesian Badminton National Team. One way to do this is to carry out a sentiment analysis of the Indonesian Badminton National Team on the Twitter using the classification method and the Naïve Bayes to classify positive or negative tweets that people express about the Indonesian Badminton National Team. The amount of data that was crawled was 3,277 data and then the data went through a data duplicate removal so that the results obtained were 272 data tweets and had not been labeled. Prior to the process of classifying the data obtained, the data must be labeled on the data as well as cleaning the data first before entering the text processing, then the data will be weighted for each word with the Term Frequency– Inverse Document Frequency (TF-IDF) which will be said in the future. will be used as a feature. Then, because the data labeled positive and negative have a much different amount, the Synthetic Minority Oversampling Technique (SMOTE) method is used to carry out balancing of the data. The next stage is the implementation of the distribution of data whose magnitude is 80% 20% and is classified using the Naive Bayes method. Then the results obtained from the implementation of this research are that the test data obtained accuracy with a percentage of 96%, precision with a percentage of 100%, recall with a percentage of 92%.

Keywords: *Sentiment Analysis, Classification, Badminton, 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. Bayu Hananto, S.Kom., M.Kom. selaku dosen pembimbing 1 dari pihak jurusan.
4. Helena Nurramdhani Irmanda, S.Pd., 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, 20 November 2022

Penulis

DAFTAR ISI

PERNYATAAN ORISINALITAS ii
PERNYATAAN PERSETUJUAN PUBLIKSASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS iii
LEMBAR PENGESAHAN iv
ABSTRAK v
ABSTRACT vi
KATA PENGANTAR vii
DAFTAR ISI viii
DAFTAR TABEL x
DAFTAR GAMBAR xi
DAFTAR SIMBOL xii
BAB I PENDAHULUAN 1
1.1. Latar Belakang 1
1.2. Rumusan Masalah 2
1.3. Ruang Lingkup 2
1.4. Tujuan 3
1.5. Manfaat 3
1.6. Luaran yang Diharapkan 3
1.7. Sistematika Penulisan 3
BAB II TINJAUAN PUSTAKA 5
2.1. Twitter 5
2.2. Analisis Sentimen 6
2.3. Fleiss Kappa 7
2.4. Text Analysis 8
2.4.1 Text Mining 8
2.4.2 Pra Proses Data 8
2.5. Pembobotan Kata 10
2.5.1 TF (*Term Frequency*) 10
2.5.2 IDF (*Inverse Document Frequency*) 10
2.6. Metode Klasifikasi 11
2.6.1 Algoritma Naïve Bayes 11
2.7. Python 12
2.8. Penelitian Terdahulu 12

BAB III METODOLOGI PENELITIAN 14
3.1. Kerangka Pikir 14
3.1.1. Identifikasi masalah 15
3.1.2. Studi Literatur 15
3.1.3. Akuisisi Data 15
3.1.4. Pelabelan Data 16
3.1.5. Praproses Data 16
3.1.6. Pembobotan TF IDF 18
3.1.7. Synthetic Minority Oversampling Technique (SMOTE) 18
3.1.8. Klasifikasi 19
3.1.9. Hasil Klasifikasi 19
3.1.10. Evaluasi 19
3.1.11. Analisis Hasil Klasifikasi 21
3.2. Perangkat Penelitian 21
BAB IV HASIL DAN PEMBAHASAN 22
4.1 Pengumpulan Data 22
4.2 Pelabelan Data 22
4.3 Praproses 26
4.3.1 Case Folding 26
4.3.2 Data Cleaning 27
4.3.3 Normalization 28
4.3.4 Stemming 29
4.3.5 Stopwords Removal 30
4.3.6 Tokenizing 30
4.4 Pembobotan TF-IDF 31
4.5 SMOTE (Synthetic Minority Oversampling Technique) 32
4.6 Klasifikasi Naïve Bayes 33
4.6.1 Data Latih 33
4.6.2 Data Uji 36
4.7 Evaluasi 40
4.8 Analisis Hasil Kegiatan 42
4.8.1 *Wordcloud* Sentimen Positif terhadap Data Tweet Timnas Bulu Tangkis Indonesia 42
4.8.2 *Wordcloud* Sentimen Negatif terhadap Data Tweet Timnas Bulu Tangkis Indonesia 43
BAB V PENUTUP 45
5.1 Kesimpulan 45
5.2 Saran 45
DAFTAR PUSTAKA 46
RIWAYAT HIDUP 48
LAMPIRAN 49

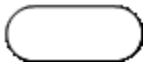
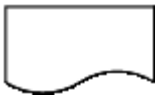
DAFTAR TABEL

Tabel 3. 1 Tabel Cleaning 17
Tabel 3. 2 Stopwords Removal 18
Tabel 3. 3 Confusion Matrix 20
Tabel 4. 1 Tweet dan Hasil Pelabelan 23
Tabel 4. 2 Hasil Penilaian Anotator 24
Tabel 4. 3 Tabel Hasil Perhitungan Nilai Pi 24
Tabel 4. 4 Case Folding 27
Tabel 4. 5 Data Cleaning 28
Tabel 4. 6 Normalization 28
Tabel 4. 7 Stemming 29
Tabel 4. 8 Stopwords Removal 30
Tabel 4. 9 Tokenization 31
Tabel 4. 10 Contoh Data TF-IDF 31
Tabel 4. 11 Perhitungan TF-IDF 32
Tabel 4. 12 Jumlah Data Sebelum dan Sesudah SMOTE 33
Tabel 4. 13 Contoh Data Latih 33
Tabel 4. 14 Nilai Probabilitas Pada Data Latih 36
Tabel 4. 15 Contoh Data Uji 36
Tabel 4. 16 Data Uji yang Setelah Praproses 37
Tabel 4. 17 Probabilitas Sampel Data Uji 37
Tabel 4. 18 Nilai Probabilitas Kelas Positif 38
Tabel 4. 19 Nilai Probabilitas Kelas Negatif 39
Tabel 4. 20 Confusion Matrix Model 41

DAFTAR GAMBAR

Gambar 3. 1 Tahapan Penelitian 14
Gambar 4. 1 Hasil Crawling Data Twitter 22
Gambar 4. 2 Wordcloud Sentimen Positif 43
Gambar 4. 3 Wordcloud Sentimen Negatif 44

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.