KLASIFIKASI SENTIMEN DATA TIDAK SEIMBANG MENGGUNAKAN ALGORITMA SMOTE DAN *K-NEAREST NEIGHBOR* PADA ULASAN PENGGUNA APLIKASI PEDULILINDUNGI

SKRIPSI

SHEILA GABRIELA BARUS

1810511072

UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN" JAKARTA

FAKULTAS ILMU KOMPUTER

PROGRAM STUDI S1 INFORMATIKA

2022

KLASIFIKASI SENTIMEN DATA TIDAK SEIMBANG MENGGUNAKAN ALGORITMA SMOTE DAN *K-NEAREST NEIGHBOR* PADA ULASAN PENGGUNA APLIKASI PEDULILINDUNGI

SKRIPSI

Diajukan Sebagai Salah Satu Syarat Untuk Memperoleh

Gelar Sarjana Komputer

SHEILA GABRIELA BARUS

1810511072

UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN" JAKARTA

FAKULTAS ILMU KOMPUTER

PROGRAM STUDI S1 INFORMATIKA

2022

PERNYATAAN ORISINALITAS

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

Nama : Sheila Gabriela Barus

NIM : 1810511072

Tanggal : 24 Juni 2022

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

Jakarta, 24 Juni 2022

Yang Menyatakan,

(Sheila Gabriela Barus)

PERNYATAAN PERSETUJUAN PUBLIKASI

Sebagai civitas akademik Universitas Pembangunan Nasional Veteran Jakarta, saya yang bertandatangan di bawah ini :

Nama : Sheila Gabriela Barus

NIM : 1810511072

Fakultas : Ilmu Komputer

Program Studi : S1 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 SENTIMEN DATA TIDAK SEIMBANG MENGGUNAKAN ALGORITMA SMOTE DAN *K-NEAREST NEIGHBOR* PADA ULASAN PENGGUNA APLIKASI PEDULILINDUNGI

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 Skripsi saya selama tetap mencantumkan nama saya sebagai penulis dan sebagai pemilik Hak Cipta. Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat Di : Jakarta

Pada tanggal : 24 Juni 2022

Yang Menyatakan,

(Sheila Gabriela Barus)

LEMBAR PENGESAHAN

Dengan ini menyatakan bahwa Proposal Skripsi berikut :

Nama	: Sheila Gabriela Barus
NIM	: 1810511072
Program Studi	: S1 Informatika
Judul	: Klasifikasi Sentimen Data Tidak Seimbang Menggunakan Algoritma SMOTE dan *K-Nearest Neighbor* pada Ulasan Pengguna Aplikasi PeduliLindungi

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

Henki Bayu Seta, S.Kom, MTI.	**Desta Sandya Prasvita, S.Kom., M.Kom.**
Penguji I	Penguji II
Dr. Didit Widiyanto, S.Kom, M.Si	**Mayanda Mega Santoni, S.Kom., M.Kom.**
Pembimbing I	Pembimbing II
Dr. Ermatita, M.Kom.	**Desta Sandya Prasvita, S.Kom., M.Kom.**
Dekan	Ketua Program Studi

Ditetapkan : Jakarta

Tanggal Persetujuan : 24 Juni 2022

KLASIFIKASI SENTIMEN DATA TIDAK SEIMBANG MENGGUNAKAN ALGORITMA SMOTE DAN *K-NEAREST NEIGHBOR* PADA ULASAN PENGGUNA APLIKASI PEDULILINDUNGI

SHEILA GABRIELA BARUS

ABSTRAK

Salah satu penanganan pemerintah dalam mengatasi penyebaran Covid-19 yang terjadi di Indonesia yaitu dengan membuat sebuah aplikasi yaitu aplikasi PeduliLindungi. Aplikasi ini berfungsi dalam melacak dan memantau penyebaran Covid-19, oleh karena itu banyak masyarakat Indonesia yang harus mempunyai aplikasi ini. Banyak juga ulasan yang diberikan pada aplikasi ini, dari komentar yang positif hingga komentar negatif. Ulasan tersebut yang menjadi data dalam penelitian ini untuk mengetahui hasil sentimen masyarakat dan menguji klasifikasi algoritma *K-Nearest Neighbor.* Pengumpulan data dilakukan dengan *scraping* di google play menggunakan bahasa pemrograman *Python*, dimana data yang diperoleh mendapatkan 750 label negatif dan 250 label positif. Sehingga data yang tidak seimbang ini harus diseimbangkan dengan teknik *undersampling* dan *oversampling* SMOTE. Oleh karena itu, penelitian ini dilakukan tiga data yang berbeda jumlah yaitu dari data yang tidak seimbang, data yang sudah di *undersampling* dan data yang sudah di *oversampling* dengan SMOTE. Hasil dari ketiga percobaan tersebut diperoleh nilai terbaik menggunakan teknik SMOTE pada K = 1 dengan nilai akurasi sebesar 0.9766, nilai presisi sebesar 0.9691, nilai *F1 score* 0.9781, nilai spesifisitas sebesar 0.9645, dan nilai sensitivitas sebesar 0.9874.

Kata Kunci : Sentimen, *K-Nearest Neighbor,* SMOTE, PeduliLindungi

CLASSIFICATION OF UNBALANCED DATA SENTIMENTS USING SMOTE ALGORITHM AND K-NEAREST NEIGHBOR ON USER REVIEWS OF CARE APPLICATIONS

SHEILA GABRIELA BARUS

ABSTRACT

One of the government's methods in dealing with the spread of Covid-19 that occurred in Indonesia is to create an application, namely the PeduliLindungi application. This application functions in tracking and monitoring the spread of Covid-19, therefore many Indonesian people must have this application. Many reviews are also given on this application, from positive comments to negative comments. These reviews are used as data in this study to determine the results of community sentiment and to test the classification of the K-Nearest Neighbor algorithm. Data collection was done by scraping on google play using the Python programming language, where the data obtained got 750 negative labels and 250 positive labels. So this unbalanced data must be balanced with SMOTE undersampling and oversampling techniques. Therefore, this study carried out three experiments, namely from unbalanced data, data that had been undersampled and data that had been oversampled with SMOTE. The results of the three experiments obtained the best value using the SMOTE technique at K = 1 with an accuracy value of 0.9766, a precision value of 0.9691, an F1 score of 0.9781, a specificity value of 0.9645, and a sensitivity value of 0.9874.

***Keywords** : Sentiment, K-Nearest Neighbor, SMOTE, PeduliLindungi*

KATA PENGANTAR

Puji dan syukur penulis panjatkan kehadirat Tuhan Yang Maha Esa atas segala anugerah-Nya sehingga proposal ini dapat terselesaikan. Judul dari penelitian ini yang dilaksanakan sejak bulan November 2021 ini adalah "Klasifikasi Sentimen Data Tidak Seimbang Menggunakan Algoritma SMOTE dan *K-Nearest Neighbor* Pada Ulasan Pengguna Aplikai PeduliLindungi" berhasil diselesaikan. Tak lupa penulis ingin mengucapkan banyak terima kasih kepada:

1. Orang tua, keluarga yang selalu memberikan dukungan kepada penulis sehingga dapat menyelesaikan skripsi ini.
2. Bapak Dr. Didit Widiyanto, S.Kom, M.Si dan Ibu Mayanda Mega Santoni, S.Kom., M.Kom. selaku dosen pembimbing yang telah memberikan saran yang bermanfaat selama proses pembuatan Proposal hingga menyelesaikan skripsi.
3. Bapak Henki Bayu Seta, S.Kom, MTI. selaku dosen pembimbing akademik.
4. Bapak Desta Sandya Prasvita, M.Kom. selaku Kaprodi Informatika yang telah memberikan informasi mengenai tugas akhir.
5. Ibu Dr. Ermatita, M.Kom. selaku dekan Fakultas Ilmu Komputer.
6. Teman-teman Informatika 2018 yang telah berjuang bersama dalam setiap proses perkuliahan serta saling memberikan semangat untuk dapat menyelesaikan Skripsi.

Dan semua pihak yang telah membantu penulisan dalam menyelesaikan skripsi ini.

Jakarta, 24 Juni 2022

Sheila Gabriela Barus

DAFTAR ISI

DAFTAR GAMBAR

DAFTAR TABEL

DAFTAR LAMPIRAN