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

SKRIPSI

ALESSANDRO SAMUEL TAMADA

1810511070

PROGRAM STUDI S1 INFORMATIKA

FAKULTAS ILMU KOMPUTER

UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAKARTA

2024

PERNYATAAN ORISINALITAS

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

Nama : Alesandro Samuel Tamada

NIM : 1810511070

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

Jakarta, 10 Juli 2024

Yang Menyatakan,

Alesandro Samuel Tamada

PERNYATAAN PERSETUJUAN PUBLIKASI

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

Nama	: Alesandro Samuel Tamada
NIM	: 1810511070
Fakultas	: Ilmu Komputer
Program Studi	: S1 Informatika
Jenis Karya	: Skripsi

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 & *K-Nearest Neighbor* Pada Ulasan Pengguna Aplikasi Maxim

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/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat Di : Jakarta
Pada tanggal : 19 Juli 2024
Yang Menyatakan,

Alesandro Samuel Tamada

LEMBAR PENGESAHAN SKRIPSI

Dengan ini dinyatakan bahwa Tugas Akhir/Skripsi berikut :

Nama	: Alessandro Samuel Tamada
NIM	: 1810511070
Program Studi	: S1 Informatika
Judul Tugas Akhir	: Klasifikasi Sentimen Data Tidak Seimbang Menggunakan Algoritma *SMOTE & K-NEAREST NEIGHBOR* Pada Ulasan Pengguna Aplikasi MAXIM

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

(Dr. Widya Cholil, M.I.T)	**(Nindy Irzavika, S.Sl., M.T.)**
Penguji I	Penguji II
(Dr. Ermatita, M.Kom)	**(Neny Rosmawarni, M.Kom)**
Pembimbing I	Pembimbing II
(Prof. Dr. Ir. Supriyanto, ST., M.Sc., IPM)	**(Dr. Widya Cholil, M.I.T)**
Dekan Fakultas Ilmu Komputer	Ketua Program Studi

Disahkan di : Jakarta

Tangal : 30 Juli 2024

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

Alessandro Samuel Tamada

Abstrak

Maxim dalah aplikasi ojek *online* yang dipakai banyak masyarakat Indonesia, sebagai pendatang baru namun mampu bersaing dengan ojek *online* lainnya yang terlebih dahulu, dengan memberi banyak diskon dan harga yang lebih murah, pada awalnya banyak kontroversi mengenai aplikasi ini karena pelayanan nya yang kurang baik, namun dari penelitian ini di dapatkan bahwa banyak penilaian yang baik dari pada yang buruk. Penelitian ini mengambil data dari ulsan aplikasi Maxim dan di olah sehingga dapat diklsifikasi dengan algoritma KNN. Namun dari data yang di dapatkan juga terdapat selisih yang cukup jauh antara data berlabel negatif dan data berlabel positif, karena itu harus menggunakan algoritma SMOTE agar mendapatkan label data yang seimbang. Hasil dari penelitian ini menggunakan data yang tidak seimbang dengan data positif sebesar 842 dan data negatif 158, didapatkan nilai tertinggi pada nilai K = 2, yaitu akurasi sebesar 0.875, presisi 0.877, *f1-score* 0.932, spesifisitas 0.111, dan sensitivitas 0.994. sedangkan penelitian kedua dilakukan *oversampling* dengan teknik SMOTE agar data dapat seimbang yang kemudian data positif menjadi 842 dan data negatif juga 842, hasilnya tertinggi didapatkan pada nilai K = 1 dan 2 dengan nilai yang sama yaitu akurasi 0.985, presisi 1, *f1-score* 0.986, spesifisitas 1 dan sensitivitas 0.973.Penelitian dengan data yang sudah melalui proses SMOTE ini mendapatkan hasil yang sangat baik dapat dilihat dari nilai spesifisitas dan sensitivitas yang memiliki selisih tipis, menunjukkan bahwa data penelitian yang digunakan seimbang. Sehingga dapat disimpulkan bahwa data yang baik untuk penelitian adalah data yang seimbang.

Kata kunci : Maxim, KNN, SMOTE.

UNBALANCED DATA SENTIMENT CLASSIFICATION USING SMOTE & K-NEAREST NEIGHBOR ALGORITHM ON USER REVIEWS OF THE MAXIM APPLICATION

Alessandro Samuel Tamada

Abstract

Maxim is an online motorbike taxi application that is used by many Indonesian people, as a newcomer but able to compete with other online motorbike taxis that were earlier, by providing lots of discounts and cheaper prices, at first there was a lot of controversy regarding this application because the service was not good, however From this research it was found that there were more good assessments than bad. This research takes data from reviews of the Maxim application and processes it so that it can be classified using the KNN algorithm. However, from the data obtained there is also quite a large difference between data labeled as negative and data labeled as positive, therefore you have to use the SMOTE algorithm to get balanced data labels. The results of this study used unbalanced data with positive data of 842 and negative data of 158. The highest value was obtained at K = 2, namely accuracy of 0.875, precision of 0.877, f1-score of 0.932, specificity of 0.111, and sensitivity of 0.994. while the second research carried out oversampling using the SMOTE technique so that the data could be balanced, then positive data became 842 and negative data also 842, the highest results were obtained at values K = 1 and 2 with the same values, namely accuracy 0.985, precision 1, f1-score 0.986, specificity 1 and sensitivity 0.973. Research with data that has gone through the SMOTE process has obtained very good results which can be seen from the specificity and sensitivity values which have a slight difference, indicating that the research data used is balanced. So it can be concluded that good data for research is balanced data.

Keywords : *Maxim, KNN, SMOTE.*

KATA PENGANTAR

Puji dan syukur penulis panjatkan kehadirat Tuhan Yang Maha Esa atas segala anugerah-Nya, sehingga skripsi 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 Aplikasi Maxim" 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 Prof. Dr. Ir. Supriyanto, ST., M.Sc., IPM. selaku dekan Fakultas Ilmu Komputer.
3. Ibu Dr. Widya Cholil, M.I.T. selaku Kaprodi Informatika yang telah memberikan informasi mengenai tugas akhir.
4. Bapak Henki Bayu Seta, S.Kom, MTI. selaku dosen pembimbing akademik.
5. Ibu Dr. Ermatita, M.Kom . selaku dosen pembimbing yang telah memberikan saran yang bermanfaat selama proses pembuatan Proposal hingga menyelesaikan skripsi.
6. Ibu Neny Rosmawarni, M.Kom. selaku dosen pembimbing yang telah memberikan saran yang bermanfaat selama proses pembuatan Proposal hingga menyelesaikan skripsi.
7. 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, sehingga dapat terselesaikan dengan baik.

Jakarta, 30 Juli 2024

Penulis

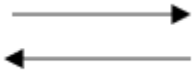
DAFTAR ISI

\

DAFTAR TABEL

DAFTAR GAMBAR

DAFTAR SIMBOL

No	Simbol	Nama	Keterangan
1.		Terminal (*start, end)*	Simbol ini menggambarkan proses dimulai atau proses berakhir
2.		*Process*	Simbol ini menggambarkan penjelasan dari suatu proses yang akan dijalankan
3.		*Flow Direction*	Simbol ini menggambarkan hubungan antar simbol yang menyatakan berjalannya suatu sistem
4.		*Data*	Simbol ini menggambarkan suatu proses yang dilakukan memiliki data masukan dan keluaran

DAFTAR LAMPIRAN