PENERAPAN METODE *RANDOM OVER-SAMPLING* UNTUK MENANGANI DATA *IMBALANCED CLASS* PADA DATA KLASIFIKASI EMOSI

SKRIPSI

disusun oleh

Agnes Asthika Setyawinda

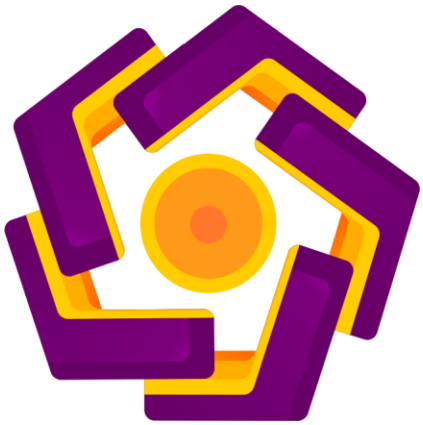
19.21.1343

PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2021

PENERAPAN METODE *RANDOM OVER-SAMPLING* UNTUK MENANGANI DATA *IMBALANCED CLASS* PADA DATA KLASIFIKASI EMOSI

SKRIPSI

untuk memenuhi sebagian persyaratan
mencapai gelar Sarjana
pada Program Studi Informatika



disusun oleh

Agnes Asthika Setyawinda

19.21.1343

PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2021

PERSETUJUAN

SKRIPSI

PENERAPAN METODE *RANDOM OVER-SAMPLING* UNTUK MENANGANI DATA *IMBALANCED CLASS* PADA DATA KLASIFIKASI EMOSI

yang dipersiapkan dan disusun oleh

Agnes Asthika Setyawinda

19.21.1343

telah disetujui oleh Dosen Pembimbing Skripsi

pada tanggal 26 September 2020

Dosen Pembimbing,

Mardhiya Hayaty, S.T., M.Kom.

NIK. 190302108

PENGESAHAN

SKRIPSI

PENERAPAN METODE *RANDOM OVER-SAMPLING* UNTUK MENANGANI DATA *IMBALANCED CLASS* PADA DATA KLASIFIKASI EMOSI

yang dipersiapkan dan disusun oleh

Agnes Asthika Setyawinda

19.21.1343

telah dipertahankan di depan Dewan Penguji
pada tanggal 19 Juli 2021

Susunan Dewan Penguji

Nama Penguji	Tanda Tangan
Windha Mega Pradnya D, M.Kom **NIK. 190302185**	_____
Wiwi Widayani, M.Kom **NIK. 190302272**	_____
Mardhiya Hayaty, S.T., M.Kom **NIK. 190302108**	_____

Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 19 Juli 2021

DEKAN FAKULTAS ILMU KOMPUTER

Hanif Al Fatta, S.Kom., M.Kom.
NIK. 190302096

PERNYATAAN

Saya yang bertandatangan dibawah ini menyatakan bahwa, skripsi ini merupakan karya saya sendiri (ASLI), dan isi dalam skripsi ini tidak terdapat karya yang pernah diajukan oleh orang lain untuk memperoleh gelar akademis di suatu institusi pendidikan tinggi manapun, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis dan/atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Segala sesuatu yang terkait dengan naskah dan karya yang telah dibuat adalah menjadi tanggungjawab saya pribadi.

Yogyakarta, 30 Juli 2021

Agnes Asthika Setyawinda

NIM. 19.21.1343

MOTTO

”Jangan terlalu ambil hati dengan ucapan seseorang, kadang manusia punya mulut tapi belum tentu punya pikiran” – Albert Einstein.

”Tujuan pendidikan itu untuk mempertajam kecerdasan, memperkukuh kemauan serta memperhalus perasaan” – Tan Malaka.

”Takut akan Tuhan adalah permulaan pengetahuan, tetapi orang bodoh menghina hikmat dan didikan” – Amsal 1:7

”Jangan seorang pun menganggap engkau rendah karena engkau muda. Jadilah teladan bagi orang-orang percaya, dalam perkataanmu, dalam tingkah lakumu, dalam kasihmu, dalam kesetiaanmu dan dalam kesucianmu” – 1 Timotius 4:12

”Serahkanlah hidupmu kepada Tuhan dan percayalah kepada-Nya, dan Ia akan bertindak” – Mazmur 37:5

PERSEMBAHAN

Dengan segala kerendahan hati saya ucapkan terima kasih, pada Tuhan Yesus dan setiap pihak yang terkait. Atas terselesaikannya skripsi, yang berjudul ”*Penerapan Metode Random Over-sampling untuk Menangani Data Imbalanced Class pada Data Klasifikasi Emosi*”. Tanpa mengurangi rasa hormat saya persembahkan karya ini untuk :

1. Ibu Windha Mega Pradnya D, M.Kom selaku Ketua Prodi S1 Informatika Universitas Amikom Yogyakarta. Yang telah memberi kesempatan pada saya selaku penulis, dalam menyelesaikan skripsi ini.
2. Kepada Orang Tua saya yang tercinta yaitu Bapak Anung dan Ibu Ita, yang telah merawat, mendidik, mendukung, dan tak hentinya mendoakan saya selama ini.
3. Ibu Mardhiya Hayaty, S.T., M.Kom, yang menjadi dosen pembimbing saya selama membuat skripsi ini. Terimakasih atas kritik dan saran yang membangun, dukungan, nasihat yang berarti serta ilmu dan pengetahuan yang berguna demi terselesaikannya skripsi ini.
4. Bapak/Ibu dosen Prodi S1 Informatika Universitas Amikom Yogyakarta, khususnya Transfer. Terimakasih atas saran, dukungan dan nasihat yang berarti serta ilmu dan pengetahuan yang sangat berguna demi terselesaikannya skripsi ini.
5. Seluruh teman-teman kelas S1 Informatika Transfer angkatan 2019, yang juga selalu memberi saran yang berguna serta dukungan.
6. Orang-orang yang secara tidak langsung telah membantu saya, dalam menyelesaikan skripsi ini.
7. Sahabat dan seluruh teman di kampus tercinta, terimakasih untuk support yang luar biasa, sampai saya bisa menyelesaikan skripsi ini dengan baik.

KATA PENGANTAR

Puji dan syukur kehadirat Tuhan Yang Maha Esa atas berkat rahmat serta kasih-Nya sehingga penulis dapat menyelesaikan skripsi ini yang mengambil judul ”*Penerapan Metode Random Over-sampling untuk Menangani Data Imbalanced Class pada Data Klasifikasi Emosi*”.

Tujuan penulisan skripsi ini untuk memenuhi sebagian syarat memperoleh gelar Sarjana Komputer (S.Kom) bagi mahasiswa program S-1 di program studi Informatika Universitas Amikom Yogyakarta. Penulis menyadari bahwa skripsi ini masih jauh dari kesempurnaan, oleh sebab itu penulis mengharapkan kritik dan saran yang membangun dari semua pihak demi kesempurnaan skripsi ini.

Terselesaikannya skripsi ini tidak lepas dari bantuan banyak pihak, sehingga pada kesempatan ini dengan segala kerendahan hati dan penuh rasa hormat penulis menghaturkan terima kasih yang sebesar-besarnya bagi semua pihak yang telah memberikan bantuan moril maupun materil baik langsung maupun tidak langsung dalam penyusunan skripsi ini hingga selesai, terutama kepada yang saya hormati :

1. Bapak Prof. Dr. M. Suyanto, M.M selaku Rektor Universitas Amikom Yogyakarta.
2. Bapak Hanif Al Fatta, S.Kom., M.Kom selaku Dekan Fakultas Ilmu Komputer Universitas Amikom Yogyakarta.
3. Ibu Windha Mega Pradnya D, M.Kom selaku Ketua Prodi S1 Informatika Universitas Amikom Yogyakarta.
4. Ibu Mardhiya Hayaty, S.T., M.Kom selaku dosen pembimbing skripsi saya yang telah memberikan kritik dan saran bimbingan maupun arahan yang sangat berguna dalam penyusunan skripsi ini.
5. Bapak/Ibu dosen dan staff di lingkungan Fakultas Ilmu Komputer Universitas Amikom Yogyakarta, khususnya Program Studi S1 Informatika yang telah banyak membantu kami untuk dapat melaksanakan penulisan dalam studi.
6. Teristimewa kepada Orang Tua penulis Ign. Anung Irianto dan Bernadetta Yunita K.U yang selalu mendoakan, memberikan motivasi

dan pengorbanannya baik dari segi moril, materi kepada penulis sehingga penulis dapat menyelesaikan skripsi ini. Buat sahabat dan juga teman-teman saya terimakasih atas dukungan dan doanya.

7. Terimakasih juga kepada semua pihak yang telah membantu dalam penyelesaian skripsi ini yang tidak dapat disebutkan satu per satu.

Akhir kata penulis mengucapkan terimakasih kepada semua pihak yang telah membantu dan penulis berharap semoga skripsi ini dapat bermanfaat bagi kita semua dan menjadi bahan masukkan dalam dunia pendidikan.

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

INTISARI

Ketidak seimbangan data merupakan masalah yang sering terjadi dengan dicirikan jumlah salah satu kelas jauh lebih banyak dibandingkan jumlah kelas lainnya. Hal ini sangat berpengaruh terhadap hasil akurasi yang akan didapatkan yang akan digunakan untuk mengambil keputusan sebuah system.

Pada penelitian ini digunakan 40000 *text* yang sudah diklasifikasikan menjadi beberapa emosi. Oleh karena itu untuk menangani ketidak seimbangan data tersebut maka diusulkan teknik pengambilan *sampling* yaitu *over-sampling* data dengan metode *Random Oversampling* dan algoritma KNN.

Setelah dilakukan dua kali percobaan dengan menggunakan metode *oover-sampling* dan tidak menggunakan metode *over-sampling* hasil akurasi menunjukkan bahwa metode *Random Over-sampling* + KNN dapat menyeimbangkan data dengan akurasi sebesar 65% dan metode SMOTE *Over-sampling* sebesar 67%.

Kata Kunci: Ketidak seimbangan data, *Oversampling*, KNN, *Random Over-sampling*, SMOTE *Oversampling.*

ABSTRACT

Data imbalance is a problem that often occurs, characterized by the number of one class being far more than the number of other classes. This is very influential on the accuracy results that will be obtained which will be used to make decisions on a system.

In this study, 40000 texts that have been classified into several emotions were used. Therefore, to deal with the data imbalance, a sampling technique is proposed, namely over-sampling data using the Random Oversampling method and the KNN algorithm.

After two experiments using the over-sampling method and not using the over-sampling method, the accuracy results show that the Random Over-sampling + KNN method can balance the data with an accuracy of 65% and the SMOTE Over-sampling method of 67%.

Keyword: *Data imbalance, Oversampling, KNN, Random Over-sampling, SMOTE Oversampling.*