PENGARUH SMOTE (*SYNTHETIC MINORITY OVERSAMPLING TECHNIQUE*) UNTUK MENGATASI *IMBALANCE DATA* PADA ANALISIS SENTIMEN MENGGUNAKAN ALGORITMA *K-NEAREST NEIGHBORS*

Diajukan Sebagai Syarat Untuk Menyelesaikan
Pendidikan Program Strata-1 Pada
Jurusan Teknik Informatika

Oleh:

RAISHA FATIYA
NIM: 09021381823128

Jurusan Teknik Informatika

FAKULTAS ILMU KOMPUTER UNIVERSITAS SRIWIJAYA

2021

LEMBAR PENGESAHAN SKRIPSI

PENGARUH SMOTE (*SYNTHETIC MINORITY OVERSAMPLING TECHNIQUE*) UNTUK MENGATASI *IMBALACE DATA* PADA ANALISIS SENTIMEN MENGGUNAKAN ALGORITMA *K-NEAREST NEIGHBORS*

Oleh :

RAISHA FATIYA
NIM: 09021381823128

Palembang, 28 Desember 2021

Pembimbing I,

Novi Yusliani, M.T.
NIP. 198211082012122001

Pembimbing II,

Mastura Diana Marieska, S.T., M.T.
NIP. 198603212018032001

Mengetahui,
Ketua Jurusan Teknik Informatika

Alvi Syahrini Utami, M.Kom.
NIP. 197812222006042003

TANDA LULUS UJIAN SIDANG SKRIPSI

Pada hari Selasa tanggal 28 Desember 2021 telah dilaksanakan ujian sidang skripsi oleh Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

Nama : Raisha Fatiya
NIM : 09021381823128
Judul : Pengaruh SMOTE (*Synthetic Minority Oversampling Technique*) untuk Mengatasi *Imbalance Data* pada Analisis Sentimen Menggunakan Algoritma *K-Nearest Neighbors*

1. Pembimbing I

   Novi Yusliani, M.T. .....
   NIP. 198211082012122001

2. Pembimbing II

   Mastura Diana Marieska, S.T., M.T. .....
   NIP. 198603212018032001

3. Penguji I

   Alvi Syahrini Utami, M.Kom .....
   NIP. 197812222006042003

4. Penguji II

   Desty Rodiah, M.T. .....
   NIP. 198912212020122011

Mengetahui,
Ketua Jurusan Teknik Informatika

Alvi Syahrini Utami, M.Kom.
NIP. 197812222006042003

HALAMAN PERNYATAAN

Yang bertanda tangan dibawah ini:

Nama	:	Raisha Fatiya
NIM	:	09021381823128
Program Studi	:	Teknik Informatika
Judul	:	Pengaruh SMOTE (*Synthetic Minority Oversampling Technique*) untuk Mengatasi *Imbalance Data* pada Analisis Sentimen Menggunakan Algoritma *K-Nearest Neighbors*

Hasil Pengecekan *Software iThenticate/Turnitin* : 8%

Menyatakan bahwa laporan tugas akhir saya merupakan hasil karya sendiri dan bukan hasil penjiplakan/plagiat. Apabila ditemukan unsur penjiplakan atau plagiat dalam laporan tugas akhir ini, maka saya bersedia menerima sanksi akademik dari Universitas Sriwijaya.

Demikian, pernyataan ini saya buat dengan sebenarnya dan tidak ada paksaan dari siapapun.

Palembang, 28 Desember 2021

Raisha Fatiya
NIM. 09021381823128

ABSTRACT

The problem of imbalanced data is one of the most often problems that appears in machine learning field. A data is said to be imbalanced if the dataset is divided into a majority class and a minority class. The majority class has far more data than the minority class so that the classification results will be biased towards the majority class. Synthetic Minority Oversampling Technique (SMOTE) can be used to overcome the problem of imbalanced data that occurs. SMOTE will overcome this problem by forming synthetic data on the minority class so that the number of minority class data is balanced with the majority class. This research will carry out the process of classifying sentiment analysis using the K-Nearest Neighbors algorithm. The results of the evaluation in this study resulted in an increase in the average values of accuracy, precision, recall, and f-measure of about 8%, 4%, 10%, and 10% respectively on KNN+SMOTE. This research shows that SMOTE can be used to overcome the problem of imbalanced data and can improve the performance results of the classification model.

Key Word : Sentiment Analysis, Imbalanced Data, Natural Language Processing, Synthetic Minority Oversampling Technique, K-Nearest Neighbors

ABSTRAK

Permasalahan data tidak seimbang adalah salah satu permasalahan yang sering muncul pada penelitian di bidang *machine learning*. Suatu data dikatakan tidak seimbang apabila *dataset* terbagi menjadi kelas mayoritas dan kelas minoritas. Kelas mayoritas memiliki data yang jauh lebih banyak dari kelas minoritas sehingga hasil klasifikasi akan menjadi bias terhadap kelas mayoritas. *Synthetic Minority Oversampling Technique* (SMOTE) dapat digunakan untuk mengatasi permasalahan data tidak seimbang yang terjadi. SMOTE akan mengatasi permasalahan tersebut dengan melakukan pembentukan data *synthetic* pada kelas minoritas agar jumlah data kelas minoritas seimbang dengan kelas mayoritas. Penelitian ini akan melakukan proses pengklasifikasian analisis sentimen dengan menggunakan algoritma *K-Nearest Neighbors.* Hasil evaluasi pada penelitian ini menghasilkan peningkatan nilai rata-rata *accuracy, precision, recall,* dan *f-measure* masing-masing bernilai sekitar 8%, 4%, 10%, dan 10% pada KNN+SMOTE. Penelitian ini menunjukkan bahwa SMOTE dapat digunakan untuk mengatasi permasalahan data tidak seimbang dan dapat meningkatkan hasil kinerja model klasifikasi.

Kata Kunci : Analisis Sentimen, Data Tidak Seimbang, *Natural Language Processing, Synthetic Minority Oversampling Technique, K-Nearest Neighbors*

KATA PENGANTAR

Puji dan syukur atas kehadirat Allah Subhanahu Wa Ta'ala, atas segala karunia dan rahmat-Nya sehingga penulis dapat menyelesaikan Skripsi dengan judul "**Pengaruh SMOTE (*Synthetic Minority Oversampling Technique*) untuk Mengatasi *Imbalance Data* pada Analisis Sentimen Menggunakan Algoritma *K-Nearest Neighbors***".

Tujuan dari penulisan Skripsi ini adalah untuk melengkapi salah satu syarat dalam memperoleh gelar Sarjana Komputer di Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya. Adapun sebagai bahan penulisan, penulis mengambil berdasarkan hasil penelitian serta observasi dari berbagai sumber literatur yang mendukung dalam penulisan Skripsi ini.

Atas selesainya Skripsi ini, penulis mengucapkan rasa syukur kepada Allah SWT. Dan penulis menyampaikan rasa terima kasih kepada yang terhormat :

1. Kedua Orang Tua serta keluarga penulis tercinta, yang telah memberikan doa dan restu serta dukungan yang sangat besar selama mengikuti dan melaksanakan perkuliahan di Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya hingga penulis dapat menyelesaikan Skripsi ini.
2. Bapak Jaidan Jauhari, S.Pd., M.T., selaku Dekan Fakultas Ilmu Komputer Universitas Sriwijaya.
3. Ibu Alvi Syahrini Utami, M.Kom, selaku Ketua Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.
4. Ibu Yunita, S.SI., M.CS., selaku Dosen Pembimbing Akademik di Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.
5. Ibu Novi Yusliani, M.T. dan Ibu Mastura Diana Marieska, S.T., M.T. selaku Dosen Pembimbing Skripsi di Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.
6. Mba Wiwin selaku admin Jurusan Teknik Informatika yang telah membantu mengurus seluruh berkas.
7. Seluruh dosen dan staff Fakultas Ilmu Komputer Universitas Sriwijaya.

8. Kak Nurmasita dan Kak Zikry yang telah membantu dalam penulisan Skripsi ini.
9. Vepi Puspitasari, Neta Fransisca, Clara Putri Herlin, Cindy Monica, Fiyna Rahmawati, Kgs. M. Rusdiansyah Muharrom, Rafliandi Ardana, Pratama Yanuarta, M. Zufar Alkautsar, Hafizh Safwan Rafa, Denta Mustofa, Syechky Al Qodrin Aruda dan teman-teman IFBILA serta seluruh teman-teman seperjuangan angkatan 2018 Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.
10. Isnaini Ramadhanti, Nyayu Adila, Della Pradita, dan Dwi Wulan yang telah memberikan *support* dalam mengerjakan Skripsi ini.
11. Kak Amel, Kak Gina, Kak Dina, Kak Wiwik, Pretty Fujianti, Virani Amanda, Uswatun Khasanah, Nurul Izzah dan teman-teman BPH HMIF.

Penulis menyadari bahwa masih banyak kekurangan dalam penulisan Skripsi ini. Oleh karena itu, segala saran dan kritik sangatlah penting bagi penulis. Akhir kata, semoga Skripsi ini dapat bermanfaat dan berguna bagi khalayak.

Palembang, 28 Desember 2021

Penulis

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

DAFTAR LAMPIRAN

BAB I
PENDAHULUAN

1.1 Pendahuluan

Bab pendahuluan akan membahas latar belakang masalah, rumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah penelitian, dan sistematika penulisan. Bab ini juga memuat penjelasan mengenai gambaran umum dari keseluruhan kegiatan penelitian yang dilakukan.

Pendahuluan dimulai dengan membahas mengenai Analisis Sentimen serta penelitian yang berkaitan dengan *K-Nearest Neighbors* dan SMOTE (*Synthetic Minority Oversampling Technique*).

1.2 Latar Belakang

Perkembangan teknologi informasi dan komunikasi pada saat ini sangat berpengaruh pada kehidupan sehari-hari. Teknologi berkembang dengan sangat pesat khususnya pada media sosial. Melalui media sosial semua orang dapat menerima maupun membagikan informasi dengan cepat dan mudah. Salah satu media sosial yang sangat terkenal dalam penyebaran informasi adalah *twitter*.

Informasi atau komentar yang tersebar dalam *twitter* terdiri dari berbagai macam jenis yaitu komentar positif, negatif, dan netral. Saat ini telah banyak dilakukan penelitian di bidang *Natural Language Processing* (NLP) khususnya mengenai analisis sentimen. Analisis sentimen dapat digunakan untuk menganalisis komentar yang diberikan apakah termasuk dalam komentar positif, negatif, ataupun netral. Dalam survei yang terdiri dari sekitar 2000 orang di Amerika dapat diketahui bahwa komentar atau *review* mampu mempengaruhi peningkatan penjualan sekitar

73% - 87% dan pelanggan bersedia untuk membeli barang yang harganya lebih mahal sekitar 20% - 99% kepada penjual yang mendapatkan *review* yang baik (Pang & Lee, 2018). Hal ini menunjukkan bahwa analisis sentimen dapat memberikan pengaruh yang sangat besar.

Algoritma *machine learning* dapat diterapkan dalam penelitian di bidang analisis sentimen. Ada banyak algoritma yang dapat digunakan karena menghasilkan akurasi yang cukup baik seperti algoritma *Support Vector Machine*, *Naïve Bayes*, dan *Logistic Regression* (Satriaji & Kusumaningrum, 2018). Selain ketiga jenis algoritma tersebut, algoritma *K-Nearest Neighbors* (KNN) juga dapat digunakan karena dapat mengolah data dalam jumlah yang besar serta dapat diterapkan dengan perhitungan yang sederhana (Sudira et al., 2019).

Seiring berjalannya waktu terdapat permasalahan yang muncul dalam penelitian analisis sentimen yaitu terdapat data yang tidak seimbang (*imbalanced data*) pada kelas baik jumlah data lebih banyak ke kelas positif ataupun negatif (Satriaji & Kusumaningrum, 2018). Permasalahan tersebut dapat diatasi dengan melakukan *resample dataset*. Salah satu metode *resample dataset* yang sering digunakan dalam penelitian adalah *Synthetic Minority Oversampling Technique* (SMOTE).

Dalam penelitian sebelumnya, hasil penelitian yang menggunakan data tidak seimbang dengan menerapkan teknik SMOTE akurasinya cenderung meningkat dibandingkan tanpa menggunakan SMOTE. Metode *Naïve Bayes* + SMOTE menghasilkan akurasi lebih tinggi dibandingkan dengan metode *Naïve bayes* tanpa SMOTE dengan selisih 0.855% (Sulistiyowati & Jajuli, 2020). Metode

Support Vector Machine juga cenderung mengalami peningkatan akurasi saat menerapkan SMOTE. Metode SVM + SMOTE menghasilkan akurasi sebesar 83.16% sedangkan SVM tanpa SMOTE menghasilkan akurasi sebesar 80.97% dengan menggunakan 70:30 split (Flores et al., 2018).

Berdasarkan hal tersebut, maka penelitian ini akan melakukan pengujian terhadap pengaruh *Synthetic Minority Oversampling Technique* (SMOTE) pada analisis sentimen dengan menggunakan algoritma *K-Nearest Neighbors* (KNN).

1.3 Rumusan Masalah

Berdasarkan penjelasan latar belakang sebelumnya, rumusan masalah dari penelitian ini adalah bagaimana pengaruh *Synthetic Minority Oversampling Technique* (SMOTE) pada analisis sentimen dengan menggunakan algoritma *K-Nearest Neighbors* (KNN). Terdapat beberapa *research question* (RQ) dalam penelitian ini yaitu sebagai berikut :

1. Bagaimana cara mengembangkan analisis sentimen menggunakan metode *K-Nearest Neighbors*?
2. Bagaimana cara mengembangkan analisis sentimen menggunakan metode *K-Nearest Neighbors* yang dikombinasikan dengan SMOTE?
3. Bagaimana kinerja metode *K-Nearest Neighbors* pada analisis sentimen?
4. Bagaimana kinerja metode *K-Nearest Neighbors* yang dikombinasikan dengan SMOTE pada analisis sentimen?

1.4 Tujuan Penelitian

Adapun tujuan dari penelitian ini adalah sebagai berikut :

1. Menghasilkan sistem analisis sentimen menggunakan metode *K-Nearest Neighbors*.
2. Menghasilkan sistem analisis sentimen menggunakan metode *K-Nearest Neighbors* yang dikombinasikan dengan SMOTE.
3. Mengetahui kinerja metode *K-Nearest Neighbors* pada analisis sentimen.
4. Mengetahui kinerja metode *K-Nearest Neighbors* yang dikombinasikan dengan SMOTE pada analisis sentimen.

1.5 Manfaat Penelitian

Adapun manfaat dari penelitian ini adalah sebagai berikut :

1. Memahami mekanisme klasifikasi *K-Nearest Neighbors* pada analisis sentimen dan SMOTE untuk mengatasi *imbalance dataset*.
2. Hasil penelitian dapat digunakan untuk menangani masalah *imbalance dataset* pada analisis sentimen.
3. Menjadi referensi pada penelitian terkait.

1.6 Batasan Masalah

Adapun batasan masalah pada penelitian ini adalah sebagai berikut :

1. Data yang digunakan merupakan komentar atau opini yang berasal dari media sosial *twitter* dalam Bahasa Indonesia.
2. Sentimen atau komentar yang digunakan tidak mengandung emoji atau *emoticon*.
3. Data yang digunakan terdiri dari dua kelas yaitu kelas positif dan kelas

negatif.

4. Data yang digunakan pada penelitian ini disimpan di dalam file berekstensi txt.
5. Pengujian dilakukan dengan menggunakan *confusion matrix*.
6. Metode pembobotan yang digunakan adalah *Term Frequency-Inverse Document Frequency* (TF-IDF).

1.7 Sistematika Penulisan

Adapun sistematika penulisan pada penelitian ini adalah sebagai berikut:

BAB I. PENDAHULUAN

Bab ini menguraikan mengenai latar belakang, rumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah, dan sistematika penulisan pada penelitian ini.

BAB II. KAJIAN LITERATUR

Bab ini membahas mengenai dasar-dasar teori mengenai analisis sentimen, *preprocessing*, algoritma *K-Nearest Neighbors*, dan SMOTE. Bab ini juga menguraikan penelitian-penelitian terdahulu yang terkait dengan penelitian ini.

BAB III. METODOLOGI PENELITIAN

Bab ini berisi pembahasan mengenai metodologi dan tahapan perancangan penelitian seperti pengumpulan data, metode pengembangan perangkat lunak, dan manajemen proyek penelitian.

BAB IV. PENGEMBANGAN PERANGKAT LUNAK

Bab ini berisi pembahasan mengenai setiap tahapan pengembangan perangkat lunak yang dilakukan. Pengembangan sistem analisis sentimen dilakukan dengan menggunakan algoritma KNN dengan SMOTE dan KNN tanpa SMOTE. Proses pengembangan perangkat lunak dibuat berdasarkan metode RUP (*Rational Unified Process).*

BAB V. HASIL DAN ANALISIS PENELITIAN

Bab ini berisi hasil pengujian pada perangkat lunak yang telah dikembangkan dan bab ini juga akan memaparkan pembahasan mengenai analisis dari hasil pengujian yang telah dilakukan.

BAB VI. KESIMPULAN DAN SARAN

Bab ini berisi kesimpulan dari hasil penelitian yang telah dilakukan serta saran yang dapat digunakan untuk penelitian selanjutnya.

1.8 Kesimpulan

Berdasarkan latar belakang penelitian yang telah diuraikan maka penelitian ini akan menguji pengaruh *Synthetic Minority Oversampling Technique* (SMOTE) pada analisis sentimen menggunakan algoritma *K-Nearest Neighbors*.

DAFTAR PUSTAKA

Bunkhumpornpat, C., Sinapiromsaran, K., & Lursinsap, C. (2009). Safe-level-SMOTE: Safe-level-synthetic minority over-sampling technique for handling the class imbalanced problem. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 5476 LNAI*, 475–482. https://doi.org/10.1007/978-3-642-01307-2_43

Chawla, N. V, Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE : Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research*, *16*, 321–357.

Deviyanto, A., & Wahyudi, M. D. R. (2018). Penerapan Analisis Sentimen Pada Pengguna Twitter Menggunakan Metode K-Nearest Neighbor. *JISKA (Jurnal Informatika Sunan Kalijaga)*, *3*(1), 1. https://doi.org/10.14421/jiska.2018.31-01

Flores, A. C., Icoy, R. I., Peña, C. F., & Gorro, K. D. (2018). *An Evaluation of SVM and Naive Bayes with SMOTE on Sentiment Analysis Data Set*. 1–4.

Ghosh, K., Banerjee, A., Chatterjee, S., & Sen, S. (2019). Imbalanced Twitter Sentiment Analysis using Minority Oversampling. *2019 IEEE 10th International Conference on Awareness Science and Technology, ICAST 2019 - Proceedings*, 1–5. https://doi.org/10.1109/ICAwST.2019.8923218

Hakim, A. A., Erwin, A., Eng, K. I., Galinium, M., & Muliady, W. (2014). Automated document classification for news article in Bahasa Indonesia based

on term frequency inverse document frequency (TF-IDF) approach. *Proceedings - 2014 6th International Conference on Information Technology and Electrical Engineering: Leveraging Research and Technology Through University-Industry Collaboration, ICITEE 2014*, 0–3. https://doi.org/10.1109/ICITEED.2014.7007894

Istia, S. S., & Purnomo, H. D. (2018). Sentiment analysis of law enforcement performance using support vector machine and K-nearest neighbor. *Proceedings - 2018 3rd International Conference on Information Technology, Information Systems and Electrical Engineering, ICITISEE 2018*, 84–89. https://doi.org/10.1109/ICITISEE.2018.8720969

Kasanah, A. N., Muladi, M., & Pujianto, U. (2019). Penerapan Teknik SMOTE untuk Mengatasi Imbalance Class dalam Klasifikasi Objektivitas Berita Online Menggunakan Algoritma KNN. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, *3*(2), 196–201. https://doi.org/10.29207/resti.v3i2.945

Kaur, S., Sikka, G., & Awasthi, L. K. (2018). Sentiment Analysis Approach Based on N-gram and KNN Classifier. *ICSCCC 2018 - 1st International Conference on Secure Cyber Computing and Communications*, 13–16. https://doi.org/10.1109/ICSCCC.2018.8703350

Kruchten, P. (2014). The Rational Unified Process -- An Introduction. *Rational Software*.

Kurniawan, R., & Apriliani, A. (2020). Analisis Sentimen Masyarakat Terhadap Virus Corona Berdasarkan Opini Dari Twitter Berbasis Web Scraper. In *Jurnal INSTEK (Informatika Sains dan Teknologi)* (Vol. 5, Issue 1, p. 67).

https://doi.org/10.24252/instek.v5i1.13686

Lestari, A. R. T., Perdana, R. S., & Fauzi, M. A. (2017). Analisis Sentimen Tentang Opini Pilkada DKI 2017 Pada Dokumen Twitter Berbahasa Indonesia Menggunakan Näive Bayes dan Pembobotan Emoji. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, *1*(12), 1718–1724.

Nurjannah, M., & Fitri Astuti, I. (2013). PENERAPAN ALGORITMA TERM FREQUENCY-INVERSE DOCUMENT FREQUENCY (TF-IDF) UNTUK TEXT MINING Mahasiswa S1 Program Studi Ilmu Komputer FMIPA Universitas Mulawarman Dosen Program Studi Ilmu Komputer FMIPA Universitas Mulawarman. *Jurnal Informatika Mulawarman*, *8*(3), 110–113.

Pang, B., & Lee, L. (2018). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, *2*, 1–135. https://doi.org/10.3748/wjg.v22.i45.9898

Peterson, M. R., Doom, T. E., & Raymer, M. L. (2005). GA-facilitated classifier optimization with varying similarity measures. *GECCO 2005 - Genetic and Evolutionary Computation Conference*, 1549–1550. https://doi.org/10.1145/1068009.1068253

Satriaji, W., & Kusumaningrum, R. (2018). Effect of Synthetic Minority Oversampling Technique (SMOTE), Feature Representation, and Classification Algorithm on Imbalanced Sentiment Analysis. *2018 2nd International Conference on Informatics and Computational Sciences, ICICoS 2018*, 99–103. https://doi.org/10.1109/ICICOS.2018.8621648

Soyusiawaty, D., & Zakaria, Y. (2018). Book data content similarity detector with

cosine similarity (case study on digilib.uad.ac.id). *Proceeding of 2018 12th International Conference on Telecommunication Systems, Services, and Applications, TSSA 2018*, 1–6. https://doi.org/10.1109/TSSA.2018.8708758

Sudira, H., Diar, A. L., & Ruldeviyani, Y. (2019). Instagram Sentiment Analysis with Naive Bayes and KNN: Exploring Customer Satisfaction of Digital Payment Services in Indonesia. *2019 International Workshop on Big Data and Information Security, IWBIS 2019*, 21–26. https://doi.org/10.1109/IWBIS.2019.8935700

Sulistiyowati, N., & Jajuli, M. (2020). Integrasi Naive Bayes Dengan Teknik Sampling Smote Untuk Menangani Data Tidak Seimbang. *Nuansa Informatika*, *14*(1), 34. https://doi.org/10.25134/nuansa.v14i1.2411

Supriadi, F., & Hardian, R. (2019). Penerapan Metode Rational Unified Process Pada Perancangan Sistem Pengolah Data Arisankita. *Infotekmesin*, *10*(2), 22–27. https://doi.org/10.35970/infotekmesin.v10i2.45

Wibawa, D. W., Nasrun, M., & Setianingsih, C. (2018). Sentiment Analysis on User Satisfaction Level of Cellular Data Service Using the K-Nearest Neighbor (K-NN) Algorithm. *Proceedings - 2018 International Conference on Control, Electronics, Renewable Energy and Communications, ICCEREC 2018*, 235–241. https://doi.org/10.1109/ICCEREC.2018.8711992

Zheng, Z., Cai, Y., & Li, Y. (2015). Oversampling method for imbalanced classification. *Computing and Informatics*, *34*(5), 1017–1037.