

**PENANGANAN KETIDAKSEIMBANGAN DATA MENGGUNAKAN
METODE SMOTE UNTUK MENINGKATKAN KEAKURATAN DALAM
PREDIKSI HUBUNGAN BIDANG STUDI DENGAN PEKERJAAN
PADA LULUSAN UNIVERSITAS SRIWIJAYA**

SKRIPSI

**Sebagai Salah Satu Syarat untuk Memperoleh
Gelar Sarjana Matematika**



OLEH :

**ANNISA KARTIKASARI
NIM. 08011281621045**

**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS SRIWIJAYA
DESEMBER 2019**

LEMBAR PENGESAHAN

**PENANGANAN KETIDAKSEIMBANGAN DATA MENGGUNAKAN
METODE SMOTE UNTUK MENINGKATKAN AKURASI DALAM
PREDIKSI HUBUNGAN BIDANG STUDI DENGAN PEKERJAAN
PADA LULUSAN UNIVERSITAS SRIWIJAYA**

SKRIPSI

**Sebagai Salah Satu Syarat untuk Memperoleh
Gelar Sarjana Matematika**

Oleh

**ANNISA KARTIKASARI
NIM. 08011281621045**

Pembimbing Pembantu




**Drs. Sugandi Yahdin, M.M
NIP. 19580727 198603 1 003**

**Indralaya, Desember 2019
Pembimbing Utama**



**Anita Desiani, S.Si., M.Kom
NIP. 197712112003122002**

**Mengetahui,
Ketua Jurusan Matematika**



**Drs. Sugandi Yahdin, M.M
NIP. 19580727 198603 1 003**

LEMBAR PERSEMBAHAN

Motto

*“Sesungguhnya sesudah kesulitan itu ada kemudahan.”
(QS.Al-Insyirah : 6)*

**“Jika aku tidur maka aku akan bermimpi,
namun jika aku bangun
maka aku dapat meraih mimpi itu”**

Skripsi ini kupersembahkan kepada:

- 1. Allah SWT**
- 2. Kedua Orangtuaku**
- 3. Keluarga Besarku**
- 4. Semua Dosen dan Guruku**
- 5. Teman-temanku**
- 6. Almamaterku**

KATA PENGANTAR

Assalamu'alaikum Wr. Wb.

Segala puji dan syukur kehadirat Allah SWT atas segala kasih sayang, rahmat dan karunia-Nya, penulis dapat menyelesaikan skripsi yang berjudul “**Penanganan Ketidakseimbangan Data Menggunakan Metode SMOTE untuk Meningkatkan Akurasi dalam Prediksi Hubungan Bidang Studi dengan Pekerjaan pada Lulusan Universitas Sriwijaya**” dengan baik. Shalawat serta salam semoga senantiasa tercurahkan kepada Nabi Muhammad SAW beserta keluarga, sahabat, dan seluruh pengikutnya hingga akhir zaman. Skripsi ini disusun untuk memenuhi salah satu syarat guna memperoleh gelar Sarjana Matematika di Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Sriwijaya.

Dengan penuh rasa hormat dan kerendahan hati, penulis mengucapkan terima kasih khusus kepada kedua orang tua, Bapak **Arif Rosyadi (Alm)** dan Ibu **Khomsiah,S.Pd** untuk seluruh kasih sayang, didikan, nasihat, motivasi, perhatian, dan do'a yang tidak pernah putus dipanjatkan kepada penulis. terselesaikannya skripsi ini juga tidak terlepas dari bantuan berbagai pihak, baik secara langsung maupun tidak langsung. Untuk itu penulis menyampaikan ucapan terima kasih yang tak terhingga sekaligus penghargaan kepada :

1. Bapak **Prof. Dr. Ishaq Iskandar, M.Sc** selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Sriwijaya.
2. Bapak **Drs. Sugandi Yahdin, M.M** selaku Ketua Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam sekaligus selaku Dosen Pembimbing

Pembantu yang bersedia meluangkan waktu di tengah kesibukannya untuk membimbing serta memberikan saran untuk pengerjaan skripsi ini.

3. Ibu **Anita Desiani, S.Si., M.Kom** selaku Dosen Pembimbing Utama yang telah bersedia meluangkan waktu, tenaga, pikiran, motivasi, nasehat, saran serta kesabaran dalam memberi bimbingan terbaik kepada penulis selama menyelesaikan skripsi ini.
4. Ibu **Des Alwine Zayanti, M.Si** selaku Sekretaris Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Sriwijaya.
5. Ibu **Endang Sri K, M.Si** selaku dosen Pembimbing Akademik yang telah sangat baik membimbing dan mengarahkan urusan akademik kepada penulis di setiap semester selama belajar di Jurusan Matematika FMIPA Universitas Sriwijaya.
6. Ibu **Des Alwine Zayanti, M.Si, Novi Rustiana Dewi, M.Si,** dan Ibu **Oki Dwipurwani, M.Si.** sebagai Dosen Pembahas skripsi yang telah memberikan tanggapan dan saran yang bermanfaat dalam pengerjaan skripsi ini.
7. **Seluruh Dosen** di Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Sriwijaya yang telah memberikan ilmu yang bermanfaat kepada penulis selama menempuh pendidikan.
8. Kakak-kakakku **Kak Fito, Kak Arie** dan **Kak Dhika** serta **Yuk Ayu** dan **Mbak Tika** atas kasih sayang, semangat, motivasi, nasihat, dan do'anya.
9. **Keluarga Besarku** terima kasih untuk segala dukungan dan semangat yang telah diberikan kepada penulis.

10. Teman-teman di bangku perkuliahan, **Dwian, Gina, Nana, Muti, Naurah, Indah, Rima, Giskha, Anisa, Sisca, Sandra, Hariani, Anita, Rachma**, Terima kasih untuk semuanya, untuk bantuannya, semangat dan kebersamaan selama kuliah.
11. Teman, Kakak, dan Adik Asisten Laboratorium Komputasi Matematika, **Gina, Fero, Ilham, Fathur, Filda, Ayu, Yogi, Gibran, Yudha, Kak Muthia, Kak Wili, Kak Ayu, Kak Daus, dan Bang Arden** serta kakak pembimbing, **Kak Elen dan Kak Nisa**, Terima kasih untuk semuanya, untuk bantuannya, semangat dan kebersamaan selama kuliah dan bimbingan mengerjakan skripsi ini.
12. Teman-teman satu Angkatan **2016**, kakak-kakak tingkat Angkatan **2012, 2013, 2014, 2015**. dan adik-adik tingkat Angkatan **2017, 2018**.
13. Teman-teman UKM Teater Gabi 91 Universitas Sriwijaya, Terimakasih atas bantuannya, semangat, dan kebersamaan dalam mengasah softskill di bidang teater, manajerial dan penulisan naskah.
14. Pak **Irwan** dan Ibu **Hamidah** yang telah banyak membantu dalam proses administrasi.
15. Semua pihak yang tidak dapat penulis sebutkan satu persatu yang telah memberikan bantuan dalam menyelesaikan skripsi ini. Semoga segala kebaikan yang diberikan mendapatkan balasan dari Allah SWT.

Penulis menyadari penulisan skripsi ini masih jauh dari kata sempurna. Oleh karena itu, saran yang bermanfaat untuk penulisan skripsi ini sangat diharapkan. Semoga skripsi ini dapat menambah pengetahuan dan bermanfaat bagi mahasiswa/mahasiswi Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Sriwijaya dan semua pihak yang memerlukan.

Wassalamu'alaikum Wr. Wb

Indralaya, Desember 2019

Penulis

**HANDLING IMBALANCE DATA USING SMOTE
TO INCREASE THE ACCURATION IN THE PREDICTION OF
THE RELEVANCE OF THE EDUCATION BACKGROUND WITH
GRADUATES EMPLOYMENT FOR UNIVERSITAS SRIWIJAYA**

By:

**Annisa Kartikasari
08011281621045**

ABSTRACT

The tracer study dataset of Universitas Sriwijaya is the result of a questionnaire completed by graduates of Universitas Sriwijaya from Career Development Center Unsri (CDC-Unsri). The tracer study dataset has 2934 data of graduates from academic years 2014-2016. There are 12 attributes in the data set such as a level of learning in lectures, a research project experience, an internship experience, a fieldwork experience, a level competence of education background, general education skill, an english skill, an internet user skill, a computer user skill, a research skill, a communication skill and a relevance of the education background with the graduates employment. The attribute label in this research is attribute a relevance of the education background with the graduates employment but there is an imbalance data for this attribute because the distribution of data in some classes is very small compared to other classes. This attribute consist of 5 classes namely the very closely label has 214 data, the closely label has 304 data, the quite closely label has 475 data, the less closely label has 463 data , and the not closely label has 787 data. The amount of data in each class that is not balanced can affect the value of accuration in the prediction such as accuracy, precision and recall so it can be hard for find the information for each class. In this study the SMOTE method will be used to make synthetic data in the minority class so that the amount of data becomes balanced. Based on the test classification with KNN the value of accuracy, precision, and recall increased while in the ANN and SVM classification the accuracy value decreased but the value of precision and recall increased so it was concluded the SMOTE method did not always increase accuracy but could increase the value of precision and recall so that it could obtain information on each label .

Keyword : Imbalance Data, SMOTE, Accuration , Prediction, The Relevance of The Education Background with Graduates Employment.

**PENANGANAN KETIDAKSEIMBANGAN DATA MENGGUNAKAN
METODE SMOTE UNTUK MENINGKATKAN KEAKURATAN DALAM
PREDIKSI HUBUNGAN BIDANG STUDI DENGAN PEKERJAAN
PADA LULUSAN UNIVERSITAS SRIWIJAYA**

Oleh :

**Annisa Kartikasari
08011281621045**

ABSTRAK

Dataset tracer study Universitas Sriwijaya merupakan hasil kuisisioner yang diisi oleh lulusan Universitas Sriwijaya yang disediakan oleh *Career Development Center* (CDC-Unsri). *Dataset tracer study* untuk lulusan tahun 2014, 2015, dan 2016 memiliki 2934 data. Ada 12 atribut yang digunakan pada penelitian ini seperti tingkat pembelajaran pada perkuliahan, proyek riset, magang, kerja lapangan, tingkat kompetensi bidang ilmu, pendidikan umum, kemampuan bahasa inggris, kemampuan penggunaan internet, kemampuan penggunaan komputer, kemampuan dalam riset, kemampuan berkomunikasi, dan hubungan bidang studi dengan pekerjaan. Pada *dataset* ini terjadi ketidakseimbangan data pada atribut hubungan bidang studi dengan pekerjaan karena sebaran data di beberapa kelas sangat kecil daripada kelas lain. Atribut ini memiliki 5 kelas yaitu kelas tidak erat yang terdiri dari 787 data, kelas kurang erat yang terdiri dari 463 data, kelas cukup erat yang terdiri dari 475 data, kelas erat yang terdiri dari 304 data, dan kelas sangat erat yang terdiri dari 214 data. Jumlah data pada tiap kelas yang tidak seimbang dapat memengaruhi nilai keakuratan pada prediksi yaitu nilai akurasi, presisi dan *recall* sehingga sulit untuk menemukan informasi pada tiap label. Dalam penelitian ini akan digunakan metode SMOTE untuk membuat data buatan pada kelas minoritas sehingga jumlah data menjadi seimbang. Berdasarkan klasifikasi uji dengan KNN nilai akurasi, presisi, dan *recall* meningkat sedangkan pada klasifikasi ANN dan SVM nilai akurasi menurun tetapi nilai presisi dan *recall* meningkat sehingga dapat disimpulkan metode SMOTE tidak selalu meningkatkan akurasi namun dapat meningkatkan nilai presisi dan *recall* sehingga dapat memperoleh informasi pada tiap label.

Kata kunci : Ketidakseimbangan Data, SMOTE, Keakuratan, Prediksi, Hubungan Bidang Studi dengan Pekerjaan.

DAFTAR ISI

	Halaman
HALAMAN JUDUL	i
HALAMAN PENGESAHAN	ii
HALAMAN PERSEMBAHAN	iii
KATA PENGANTAR	iv
ABSTRACT	viii
ABSTRAK	ix
DAFTAR ISI	x
DAFTAR TABEL	xii
DAFTAR GAMBAR	xiv
BAB I PENDAHULUAN	
1.1. Latar Belakang	1
1.2. Perumusan Masalah	3
1.3. Pembatasan Masalah	3
1.4. Tujuan	4
1.5. Manfaat	4
BAB II TINJAUAN PUSTAKA	
2.1. Tracer Study	5
2.2. Data Tidak Seimbang	5
2.3. Metode SMOTE	6
2.4. Evaluasi Metode	8

BAB III METODOLOGI PENELITIAN

3.1. Tempat	11
3.2. Waktu	11
3.3. Alat	11
3.4. Metode Penelitian	11

BAB IV HASIL DAN PEMBAHASAN

4.1. Deskripsi Data	13
4.2. Seleksi Data	15
4.3. Penerapan Metode SMOTE	17
4.3.1. Ilustrasi Perhitungan Manual Metode SMOTE pada Data Sederhana	17
4.3.2. Penerapan Metode SMOTE pada Data Penelitian	28
4.4. Pengujian Hasil SMOTE	31
4.4.1 KNN	31
4.4.2 ANN	36
4.4.3 SVM	39

BAB V KESIMPULAN DAN SARAN

5.1. Kesimpulan	53
5.2. Saran	54

DAFTAR PUSTAKA	55
-----------------------------	-----------

DAFTAR TABEL

	Halaman
Tabel 2.1. Confusion Matrix untuk Dua Kelas	9
Tabel 4.1. Atribut dan Keteranganannya	14
Tabel 4.2. Dataset Tracer Study.....	15
Tabel 4.3. Data Sederhana	17
Tabel 4.4. Hasil Normalisasi pada Data Sederhana	18
Tabel 4.5. Jarak Antar Data Minoritas pada Label SE	19
Tabel 4.6. Jarak Antar Data Minoritas pada Label CE	20
Tabel 4.7. Data Buatan pada Data Sederhana	27
Tabel 4.8. Data Setelah SMOTE	28
Tabel 4.9. Data Penelitian	29
Tabel 4.10. Confusion Matrix pada Data Penelitian Tanpa SMOTE Menggunakan Klasifikasi KNN	32
Tabel 4.11. Nilai Presisi dan Recall pada Data Penelitian Tanpa SMOTE Menggunakan Klasifikasi KNN	34
Tabel 4.12. Confusion Matrix pada Data Penelitian Dengan SMOTE Menggunakan Klasifikasi KNN	34
Tabel 4.13. Nilai Presisi dan Recall pada Data Penelitian Dengan SMOTE Menggunakan Klasifikasi KNN	35
Tabel 4.14. Confusion Matrix pada Data Penelitian Tanpa SMOTE Menggunakan Klasifikasi ANN	36
Tabel 4.15. Nilai Presisi dan Recall pada Data Penelitian Tanpa SMOTE Menggunakan Klasifikasi ANN	37

Tabel 4.16. Confusion Matrix pada Data Penelitian Dengan SMOTE	
Menggunakan Klasifikasi ANN	38
Tabel 4.17. Nilai Presisi dan Recall pada Data Penelitian Dengan SMOTE	
Menggunakan Klasifikasi ANN	38
Tabel 4.18. Confusion Matrix pada Data Penelitian Tanpa SMOTE	
Menggunakan Klasifikasi SVM	39
Tabel 4.19. Nilai Presisi dan Recall pada Data Penelitian Tanpa SMOTE	
Menggunakan Klasifikasi SVM	40
Tabel 4.20. Confusion Matrix pada Data Penelitian Dengan SMOTE	
Menggunakan Klasifikasi SVM	41
Tabel 4.21. Nilai Presisi dan Recall pada Data Penelitian Dengan SMOTE	
Menggunakan Klasifikasi SVM	41
Tabel 4.22. Hasil Akurasi pada Data Penelitian	42
Tabel 4.23. Hasil Presisi dan <i>Recall</i> Pada Data Penelitian	44
Tabel 4.24. Rata-rata Nilai Presisi	50
Tabel 4.25 Rata-rata Nilai <i>Recall</i>	50

DAFTAR GAMBAR

	Halaman
Gambar 4.1 Grafik Banyaknya Data pada Setiap Label	16
Gambar 4.2 Banyaknya Data Setelah Proses SMOTE pada Data Berlabel SE	29
Gambar 4.3 Banyaknya Data Setelah Proses SMOTE pada Data Berlabel E	30
Gambar 4.4 Banyaknya Data Setelah Proses SMOTE pada Data Berlabel CE	30
Gambar 4.5 Banyaknya Data Setelah Proses SMOTE pada Data Berlabel KE	30
Gambar 4.6 Hasil Akurasi Prediksi dengan 3 Klasifikasi Uji	43
Gambar 4.7 Perubahan Nilai Presisi pada Hasil Klasifikasi KNN	46
Gambar 4.8 Perubahan Nilai Presisi pada Hasil Klasifikasi ANN	56
Gambar 4.9 Perubahan Nilai Presisi pada Hasil Klasifikasi SVM	47
Gambar 4.10 Perubahan Nilai Recall pada Hasil Klasifikasi KNN	48
Gambar 4.11 Perubahan Nilai Presisi pada Hasil Klasifikasi ANN	48
Gambar 4.12 Perubahan Nilai Presisi pada Hasil Klasifikasi SVM	49
Gambar 4.13 Perbedaan Rata-rata Nilai Presisi pada Semua Klasifikasi	51
Gambar 4.14 Perbedaan Rata-rata Nilai <i>Recall</i> pada Semua Klasifikasi	52

BAB I

PENDAHULUAN

1.1. Latar Belakang

Tracer study merupakan salah satu bentuk studi empiris yang dapat memberikan informasi berharga untuk mengevaluasi hasil pendidikan pada perguruan tinggi yang digunakan sebagai pengembangan lebih lanjut dalam menjamin kualitas dan dapat dijadikan indikator keberhasilan pendidikan yang diselenggarakan perguruan tinggi itu sendiri. (Cassar, 2016; Schomburg, 2003).

Universitas Sriwijaya memiliki kumpulan data (*dataset*) *tracer study* pada setiap tahunnya yang dilakukan oleh pusat pengembangan karakter dan karir mahasiswa Universitas Sriwijaya atau yang dikenal dengan CDC-Unsri. Hasil *tracer study* ini diharapkan dapat menjadi dasar pengembangan kualitas proses belajar mengajar agar lulusan Unsri terserap di pasar kerja dengan maksimal. Alumni Universitas Sriwijaya yang telah lulus minimal 2 tahun dapat mengisi *tracer study* ini pada website cdc.unsri.ac.id dimana data ini merupakan kuisioner online yang terdiri dari 17 pertanyaan yang mengacu pada standar DIKTI.

Pada *dataset tracer study* Universitas Sriwijaya untuk lulusan tahun 2014-2016 tepatnya pada atribut hubungan bidang studi dengan pekerjaan terdapat 5 pilihan jawaban yaitu sangat erat, erat, cukup erat, kurang erat dan tidak erat sama sekali. Hasilnya diperoleh 787 data pada kelas tidak erat sama sekali, 463 data pada kelas kurang erat, 475 data pada kelas cukup erat serta kelas erat dan sangat erat masing-masing sebanyak 304 dan 214 data. Penyebaran data pada atribut ini tidak seimbang

karena terdapat kelas dengan label tidak erat sama sekali yang mendominasi daripada kelas lainnya dengan jangkauan data sebesar 573 data.

Menurut Chawla *et al.*, (2002) Penyebaran data yang tidak seimbang ini dapat mempengaruhi tingkat keakuratan dalam prediksi yaitu akurasi, presisi, dan *recall*. Akurasi merupakan tingkat keakuratan prediksi benar terhadap semua prediksi (Ali *et al.*, 2013) namun perhitungan akurasi kurang baik untuk data tidak seimbang karena perhitungan akurasi tidak memperhitungkan keakuratan pada tiap kelas sehingga perlu dilakukan perhitungan *recall* atau tingkat keberhasilan mesin dalam menemukan kembali sebuah informasi dan presisi yang merupakan tingkat keakuratan antara permintaan dengan jawaban prediksi (Beckmann *et al.*, 2015; Ha *et al.*, 2012; Octaviani *et al.*, 2019). Ada banyak macam perhitungan keakuratan dalam prediksi seperti akurasi, presisi, *recall*, kappa, *specificity*, *sensitivity* dan sebagainya, namun untuk kasus data tidak seimbang perhitungan keakuratan prediksi direkomendasikan untuk menggunakan akurasi, presisi, dan *recall* (Ha *et al.*, 2012).

Pada penelitian Amaral (2014) tentang prediksi tanda alarm pasien yang terkena serangan jantung menunjukkan nilai akurasi tinggi namun nilai presisi dan *recall* rendah sehingga tidak dapat mengenali pasien serangan jantung karena jumlah pasien serangan jantung sangat sedikit. Jika nilai akurasi tinggi namun presisi atau *recall*-nya rendah mesin hanya dapat mengenali sebuah kejadian yang sering terjadi tetapi tidak mengenali kejadian yang jarang terjadi. Hal ini akan berakibat fatal apabila seorang pasien terkena serangan jantung namun alarm mendeteksi tidak terkena serangan jantung. Penelitian ini menunjukkan bahwa perlu metode untuk meningkatkan keakuratan prediksi baik akurasi, presisi ataupun *recall* untuk kasus data tidak seimbang.

Salah satu metode yang digunakan untuk mengatasi ketidakseimbangan data adalah *synthetic minority oversampling technique* (SMOTE). Chawla *et al.*, (2002) menyatakan bahwa metode SMOTE bekerja dengan cara memperbanyak jumlah data pada kelas minor agar setara dengan kelas mayor dengan cara membangkitkan data buatan yang dibuat berdasarkan *k-nearest neighbor* atau KNN. Metode SMOTE telah banyak digunakan untuk penelitian seperti penelitian Ramentol *et al.*, (2011), Guo *et al.*, (2018) dan Luengo *et al.*, (2011) yang menunjukkan bahwa SMOTE dapat digunakan dalam proses *pre-processing* untuk kasus data tidak seimbang. Penelitian Jeatrakul *et al.*, (2010) menunjukkan bahwa metode SMOTE berkinerja baik untuk menyeimbangkan data pada beberapa kasus uji seperti ANN, SVM dan KNN.

Pada penelitian ini akan membahas penanganan ketidakseimbangan data pada dataset *tracer study* Universitas Sriwijaya dengan menggunakan metode SMOTE untuk meningkatkan keakuratan prediksi hubungan bidang studi dengan pekerjaan.

1.2. Perumusan Masalah

Bagaimana penanganan ketidakseimbangan *dataset tracer study* menggunakan metode SMOTE untuk meningkatkan keakuratan pada prediksi hubungan bidang studi dengan pekerjaan.

1.3. Pembatasan Masalah

Data alumni Universitas Sriwijaya berdasarkan kuisisioner yang disediakan oleh CDC-Unsri untuk lulusan tahun 2014, 2015 dan 2016 yang terisi lengkap serta nilai keakuratan prediksi hanya menggunakan perhitungan akurasi, presisi dan *recall*.

1.4. Tujuan

Menangani ketidakseimbangan data pada *dataset tracer study* menggunakan metode SMOTE untuk meningkatkan keakuratan dalam prediksi hubungan bidang studi dengan pekerjaan berdasarkan nilai akurasi, presisi, dan *recall*.

1.5. Manfaat

Dapat melihat bagaimana kemampuan dari metode SMOTE dalam penanganan ketidakseimbangan data pada *dataset tracer study* terhadap keakuratan hasil survei.

Daftar Pustaka

- Ali, A., Shamsuddin, S. M., & Ralescu, A. L. (2013). Classification with class imbalance problem: A review. *International Journal of Advances in Soft Computing and Its Applications*, 5(3), 176–204.
- Ali, H., Salleh, M. N. M., Saedudin, R., Hussain, K., & Mushtaq, M. F. (2019). Imbalance class problems in data mining: A review. *Indonesian Journal of Electrical Engineering and Computer Science*, 14(3), 1552–1563. <https://doi.org/10.11591/ijeecs.v14.i3.pp1552-1563>
- Amaral, A. C. K. . (2014). The Art of Making Predictions : Statistics versus Bedside Evaluation. *American Journal of Respiratory and Critical Care Medicine*, 190(6).
- Anonymous. (2012). *Tertiary Education Commission Graduate Tracer Study*.
- Bain, S., & Knight, M. (n.d.). *The successful graduate student : a review of the factors for success*. 1–9.
- Bala, R., & Kumar, D. (2017). *Classification Using ANN : A Review*. 13(7), 1811–1820.
- Beckmann, M., Ebecken, N. F. F., & Pires de Lima, B. S. L. (2015). A KNN Undersampling Approach for Data Balancing. *Journal of Intelligent Learning Systems and Applications*, 07(04), 104–116. <https://doi.org/10.4236/jilsa.2015.74010>
- Cassar, E. (2016). *Graduate tracer study*. Malta: NCFHE.
- Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research*, (Sept. 28), 321–357. <https://doi.org/10.1613/jair.953>
- Chawla, N. V., Japkowicz, N., & Kotcz, A. (2004). Editorial: Special Issue on Learning from Imbalanced Data Sets. *ACM SIGKDD Explorations Newsletter*, 6(1), 1. <https://doi.org/10.1145/1007730.1007733>
- Goel, A., & Mahajan, S. (2017). *Comparison : KNN & SVM Algorithm*. 5(Xii), 165–168.
- Guo, H., Zhou, J., & Wu, C. A. (2018). Imbalanced learning based on data-partition and SMOTE. *Information (Switzerland)*, 9(9). <https://doi.org/10.3390/info9090238>

- Ha, I., Oh, K.-J., Hong, M.-D., & Jo, G.-S. (2012). *Social Filtering Using Social Relationship for Movie Recommendation*. 7653, 395–404. https://doi.org/10.1007/978-3-642-34630-9_41
- Han, J., Kamber, M., & Pei, J. (2011). Data mining: Data mining concepts and techniques. In *Proceedings - 2013 International Conference on Machine Intelligence Research and Advancement, ICMIRA 2013*. <https://doi.org/10.1109/ICMIRA.2013.45>
- I.H. Witten, E. Frank, and M.A Hall. (2011). “Data Mining Practical Machine Learning Tools and Techniques”, Third Edition, Elsevier Publisher, USA.
- Jeatrakul, P., Wong, K. W., & Fung, C. C. (2010). Classification of imbalanced data by combining the complementary neural network and SMOTE algorithm. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 6444 LNCS(PART 2), 152–159. https://doi.org/10.1007/978-3-642-17534-3_19
- Kalra, K., Kumar Goswami, A., & Gupta, R. (2013). a Comparative Study of Supervised Image Classification Algorithms for Satellite Images. *International Journal of Electrical*, 1(10), 2320–2084. Retrieved from http://www.iraj.in/journal/journal_file/journal_pdf/1-10-139036756510-16.pdf
- Khamis, H. S., Cheruiyot, K. W., & Kimani, S. (2014). Application of k- Nearest Neighbour Classification in Medical Data Mining. *International Journal of Information and Communication Technology Research*, 4(4), 8. Retrieved from <http://www.esjournals.org>
- Luengo, J., Fernández, A., García, S., & Herrera, F. (2011). Addressing data complexity for imbalanced data sets: Analysis of SMOTE-based oversampling and evolutionary undersampling. *Soft Computing*, 15(10), 1909–1936. <https://doi.org/10.1007/s00500-010-0625-8>
- Moosavian, A., Ahmadi, H., Tabatabaeefar, A., & Khazaei, M. (2013). *Comparison of two classifiers; K-nearest neighbor and artificial neural network, for fault diagnosis on a main engine*. 20, 263–272. <https://doi.org/10.3233/SAV-2012-00742>
- Octaviani, T. L., Rustam, Z., & Siswantining, T. (2019). Ovarian Cancer Classification using Bayesian Logistic Regression. *IOP Conference Series: Materials Science and Engineering*, 546(5). <https://doi.org/10.1088/1757-899X/546/5/052049>
- Patro, S. G. K., & Kumar, K. (2015). *Normalization : A Preprocessing Stage*. (April). <https://doi.org/10.17148/IARJSET.2015.2305>
- Phoungphol, P. (2013). A Classification Framework for Imbalanced Data. Georgia State University.

- Ramentol, E., Caballero, Y., Bello, R., & Herrera, F. (2011). SMOTE-RSB *: A hybrid preprocessing approach based on oversampling and undersampling for high imbalanced data-sets using SMOTE and rough sets theory. *Knowledge and Information Systems*, 33(2), 245–265. <https://doi.org/10.1007/s10115-011-0465-6>
- Schomburg, H. (2003). Handbook for Graduate Tracer Studies. In *University of Kassel, Germany: Centre for Research on Higher Education and Work*. <https://doi.org/10.1017/CBO9781107415324.004>
- Zacharski, R. (2015). Evaluating algorithms and kNN. In *A Programmer's Guide to Data Mining*.