

**PENGARUH *SYNTHETIC MINORITY OVERSAMPLING
TECHNIQUE* (SMOTE) PADA SENTIMEN ANALISIS
MENGGUNAKAN ALGORITMA *NAIVE BAYES CLASSIFIER***

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



Oleh:

Nurmasita Anawula MP

NIM: 09021981621146

**Jurusun Teknik Informatika
FAKULTAS ILMU KOMPUTER UNIVERSITAS SRIWIJAYA
2020**

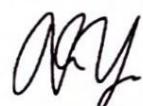
LEMBAR PENGESAHAN TUGAS AKHIR

PENGARUH SYNTHETIC MINORITY OVER SAMPLING TECHNIQUE
(SMOTE) PADA SENTIMENT ANALYSIS MENGGUNAKAN
ALGORITMA NAÏVE BAYES CLASSIFIER

Oleh :

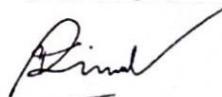
NURMASITA ANAWULA MP
NIM : 09021981621146

Pembimbing I



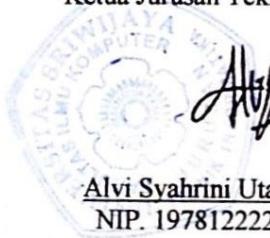
Novi Yusliani, M.T.
NIP. 198211082012122001

Palembang, Desember 2020
Pembimbing II,



Mastura Diana Marieska, S.T., M.T.
NIP. 1986032112018032001

Mengetahui,
Ketua Jurusan Teknik Informatika



Alvi Syahrini Utami, M.Kom.
NIP. 197812222006042003

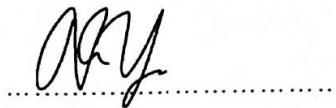
TANDA LULUS UJIAN SIDANG TUGAS AKHIR

Pada hari Rabu tanggal 18 Desember 2020 telah dilaksanakan ujian sidang tugas akhir oleh Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

Nama : Nurmasita Anawula MP
NIM : 09021981621146
Judul : Pengaruh *Synthetic Minority Over Sampling Technique* (SMOTE) pada Sentimen Analisis Menggunakan Algoritma *Naïve Bayes Classifier*.

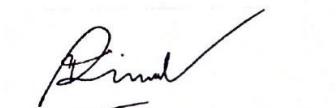
1. Pembimbing I

Novi Yusliani, M.T.
NIP. 198211082012122001



2. Pembimbing II

Mastura Diana Marieska, S.T., M.T.
NIP. 1986032112018032001



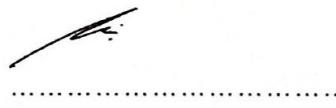
3. Penguji I

Alvi Syahrini Utami, M.Kom.
NIP. 197812222006042003

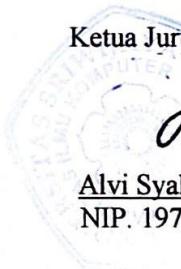


4. Penguji II

Rizki Kurniati, S.Kom, MT
NIP. 199107122019032016



Mengetahui,
Ketua Jurusan Teknik Informatika



Alvi Syahrini Utami, M.Kom.
NIP. 197812222006042003

HALAMAN PERNYATAAN

Yang bertanda tangan dibawah ini:

Nama :Nurmasita Anawula MP

Nim :09021981621146

Program studi :Teknik Informatika

Judul : Pengaruh *Synthetic Minority Oversampling Technique* (SMOTE) pada sentimen analisis menggunakan algoritma *Naïve Bayes Classifier*.

Hasil pengecekan software *iThenticate/turnitin* : 5%

Menyatakan bahwa laporan projek saya merupakan hasil karya sendiri dan bukan hasil penjiplakan/plagiat. Apabila ditemukan unsur penjiplakan/plagiat dalam laporan projek ini, maka saya bersedia menerima sanksi akademik dari universitas sriwijaya sesuai dengan ketentuan yang berlaku.

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



Palembang, Desember 2020



Nurmasita Anawula MP
NIM. 09021981621146

MOTTO DAN PERSEMBAHAN

Motto

“Allah tidak membebani seseorang melainkan sesuai dengan
kesanggupannya.
[Al-Baqoroh (128)]”

Jadi percaya diri aja pada setiap tantangan yang datang dalam kehidupan.

“*Let's do what we love and do a lot of it*”
[Mark Jacobs]

“*Love self much, always be grateful, never comparing, struggling and
Praying*”
[Nurmasita Anawula MP]

Kupersembahkan kepada:

Allah SWT

Orangtuaku Muslan dan Trinovitasari

Saudaraku Jingga dan Anaway

Kawan Kawan Seperjuangan

Almamater

THE EFFECT OF SYNTHETIC MINORITY OVER SAMPLING
TECHNIQUE (SMOTE) ON SENTIMENT ANALYSIS USING
NAÏVE BAYES CLASSIFIER ALGORITHM

by:

Nurmasita Anawula MP
09021981621146

ABSTRACT

Sentiment analysis is a multidisciplinary field of study that is used to analyze people's sentiments towards an entity. Sentiment analysis can be conducted using Twitter data from social media Twitter. However, the amount of tweet data is sometimes unbalanced, where the amount of data is more in one class compared to other classes or what is known as imbalanced data. Naïve Bayes is an algorithm that can be used to analyze a sentiment. However, Naïve Bayes itself is not equipped to solve the problem of imbalanced data. To solve this problem, one approach can be done by applying SMOTE (Synthetic Minority Oversampling Technique). From the test results using four different datasets, it shows that the application of SMOTE by synthesizing data in the minor class in sentiment analysis using the Naïve Bayes algorithm has an effect with an average increase of accuracy is 5%.

Key Word: Sentiment Analysis, Imbalanced Data, Naïve Bayes, SMOTE (Synthetic Minority Oversampling Technique).

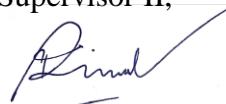
Palembang, December 2020

Supervisor I,



Novi Yusliani, M.T.
NIP. 198211082012122001

Supervisor II,



Mastura Diana Marieska, S.T., M.T.
NIP. 1986032112018032001

Approve,
Head of the Informatics Engineering Departement,



Alvi Syahrini Utami, M.Kom.
NIP. 19781222200604200

**PENGARUH SYNTHETIC MINORITY OVER SAMPLING TECHNIQUE
(SMOTE) PADA SENTIMEN ANALISIS MENGGUNAKAN
ALGORITMA NAÏVE BAYES CLASSIFIER**

Oleh:

Nurmasita Anawula MP
09021981621146

ABSTRAK

Sentiment analysis merupakan bidang studi multidisiplin yang digunakan untuk menganalisis sentimen orang-orang terhadap suatu entitas. *Sentiment analysis* dapat dilakukan dengan menggunakan data kicauan yang berasal dari sosial media *Twitter*. Akan tetapi jumlah data kicauan terkadang tidak seimbang dimana jumlah data yang lebih banyak pada suatu kelas terhadap kelas lainnya atau yang dikenal dengan *imbalanced data*. *Naïve Bayes* adalah salah satu algoritma yang dapat digunakan untuk menganalisis sebuah sentimen. Namun *Naïve Bayes* sendiri tidak dilengkapi kemampuan untuk mengatasi permasalahan *imbalanced data*. Untuk mengatasi permasalahan tersebut salah satu pendekatan yang bisa dilakukan dengan menerapkan SMOTE (*Synthetic Minority Oversampling Technique*). Metode SMOTE akan membentuk data sintetis yang baru pada kelas minor sehingga jumlah data pada masing-masing kelas dapat terhindar dari *imbalanced data*. Dari hasil pengujian dengan menggunakan empat *dataset* yang berbeda menunjukkan jika penerapan SMOTE dengan mensistesis data di kelas minor pada analisis sentimen menggunakan algoritma *Naïve Bayes* memberikan pengaruh dengan adanya rata-rata peningkatan akurasi sebesar 5%.

Kata Kunci: Analisis sentimen, *imbalanced data*, *Naïve Bayes*, SMOTE (*Synthetic Minority Oversampling Technique*).

Pembimbing I,

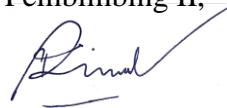


Novi Yusliani, M.T.

NIP. 198211082012122001

Palembang, Desember 2020

Pembimbing II,



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

NIP. 1986032112018032001

Mengetahui,
Ketua Jurusan Teknik Informatika,



Alvi Syahrini Utami, M.Kom.

NIP. 197812222006042003

KATA PENGANTAR

Puji syukur kepada Allah SWT atas rahmat dan karunia-Nya sehingga saya dapat menyelesaikan penelitian dan skripsi yang berjudul pengaruh SMOTE (*Synthetic Minority Oversampling Technique*) pada sentimen analisis menggunakan algoritma *Naïve Bayes Classifier*. Skripsi ini disusun untuk memenuhi salah satu syarat kelulusan dalam meraih derajat sarjana Komputer program Strata Satu (S-1) Fakultas Ilmu Komputer Universitas Sriwijaya.

Selama penelitian dan penyusunan skripsi, penulis tidak luput dari kendala. Kendala tersebut dapat diatasi berkat doa, bantuan, dan dukungan dari berbagai pihak. Penulis ingin menyampaikan rasa terima kasih kepada:

1. **Jaidan Jauhari, M.T.**, selaku Dekan Fakultas Ilmu Komputer Universitas Sriwijaya.
2. **Rifkie Primartha, M.T.**, selaku Ketua Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.
3. **Novi Yusliani, M.T.**, selaku Pembimbing Skripsi Pertama yang banyak sekali membantu saya dalam pembuatan skripsi ini.
4. **Mastura Diana Marieska, S.T., M.T.**, selaku Pembimbing Skripsi Kedua yang membimbing saya dari awal masuk kuliah, kerja praktik, hingga lulus.
5. **Alvi Syahrini Utami, M.Kom.** dan **Rizki Kurniati, S.Kom, MT** selaku Pengaji Skripsi yang banyak membantu koreksi dan memberi saran dalam pembuatan skripsi.

6. **Winda Kurnia Sari**, dan **Ricy Firnando**, selaku Admin Teknik Informatika yang membantu administrasi saya dari awal kuliah hingga lulus.
7. **Fathoni, S.T., M.MSI.** selaku wakil dekan III yang sangat membantu dalam hal dukungan untuk mengikuti kompetisi sehingga penulis mendapatkan kesempatan ikut kompetisi nasional dan internasional.
8. **Dosen-dosen Fakultas Ilmu Komputer** yang tidak bisa saya sebutkan satu persatu, terima kasih telah membagi ilmunya kepada kami.
9. **Orangtuaku, Muslan Polingay** dan **Trinovitasari Polingay**, terima kasih telah memberikan banyak doa dan dukungan sampai saat ini.
10. **Untuk kedua adikku Jingga dan Anaway**, terimakasih untuk doa doa yang kalian panjatkan untuk kakakmu ini.
11. **Suci Inayah** sahabat dekat saya yang seperjuangan untuk menjadi pribadi yang lebih baik, yang selalu mengingatkan banyak hal dan selalu mendukung penulis serta tempat berkeluh kesah haha. Terimakasih sudah bersedia menjadi sahabat ana.
12. **Indah Rosita** sahabat lomba dan jalan-jalan. Terimakasih banyak sudah banyak ngajarin ana banyak hal. Suka dan duka penulis dengan sahabat ini sangat banyak. Terimakasih sudah senantiasa hadir dalam cerita penulis. Rasanya mau nangis kalau ingat perjuangan-perjuangan kita.
13. **Muhammad Irfan Triananto Putra** yang selalu ada kalau penulis susah dalam menghadapi skripsi dan persoalan kuliah. Terimakasih sudah banyak membantu dan suka mengirimkan makanan di kosan.

14. **Ade Fajri**, sahabat organisasi yang memberikan dukungan kepada penulis sehingga penulis lebih berani mengikuti kompetisi dan sangat memotivasi penulis untuk menjadi pribadi yang lebih baik.
15. **M. Rafli Hakim dan kak Julianti**, sahabat *hanging out* yang selalu gas mau kemanapun terimakasih sudah sering mengajak ana untuk bermain. Rafli dan kak Juli telah menambah warna baru dalam kehidupan ana di kota Palembang.
16. **Sahabat Organisasi di WIFI dan FASCO** Dina Eliyatun Nasuha, Ratih Dewi Sari, kak Yeni Anggaraini, kak Aziz, mba Mey terimakasih sudah membantu penulis menemukan jalan penulis selama kuliah. Berkat dukungan kalian penulis lebih semangat dalam mengeksplorasi kehidupan kampus.
17. **Teman-teman INTEL**, terimakasih banyak atas wadah dan dukungannya organisasi inilah awal mula penulis menitik karir dibidang kompetisi.
18. **Rifdah Yumna Farha Maitsa, Dwi Novitasari, Cikita Merly Febiola, Dita Ayu Savitri, Riska Savitri, dan sahabat HIMAGIBAH** lainnya, terimakasih banyak atas bantuannya selama perkuliahan. Berkat bantuan dan motivasi dari kalian ana bisa menyelesaikan *milestone* perkuliahan yang jika tanpa kalian mungkin ana tidak semangat menghadapi semuanya.
19. **Friska, Fadli, Dhiya dan Abdi**, Terimakasih tim *United Production* yang udah ngasih pengalaman baru berkat kalian ana lebih banyak belajar untuk lebih kreatif.
20. **Irsyad Masyhudin, Sultan Alif Utama dan Farid Landriandani**, Terimakasih sahabat-sahabat *project* penulis berkat kalian penulis jadi ada portofolio dan pastinya Alhamdulillah bisa menghasilkan pundi-pundi uang.

21. **Edu, Alif, Acap, Rama, Zikri, Haikal, Reyhan, Rani, Atan sahabat INFORGEN** yang telah membantu penulis selama perkuliahan.
22. **AkuJuara dan Sahabat Gensri**, terimakasih sudah menjadi organisasi penulis pasca tidak ada mata kuliah yang harus kekampus lagi.
23. **Untuk Semua Yang Mendoakan**, terimakasih berkat doa doa kalian penulis yang merantau ini jauh dari SULAWESI bisa menyelesaikan studi ini di SUMATERA SELATAN. Penulis sangat bersyukur telah dikasih kesempatan untuk berkuliah disini karena tanpa ketentuan tersebut penulis tidak akan bertemu dengan orang-orang luarbiasa yang penulis telah sebutkan diatas.

Palembang, 06 Januari 2021



Nurmasita Anawula MP

DAFTAR ISI

HALAMAN JUDUL.....	i
LEMBAR PENGESAHAN TUGAS AKHIR	ii
TANDA LULUS UJIAN SIDANG TUGAS AKHIR	iii
HALAMAN PERNYATAAN	iv
MOTTO DAN PERSEMPAHAN	v
ABSTRACT	vi
ABSTRAK	vii
KATA PENGANTAR	viii
DAFTAR ISI.....	xii
DAFTAR TABEL.....	xvi
DAFTAR GAMBAR	xviii
DAFTAR LAMPIRAN.....	xx
 BAB I PENDAHULUAN	I-1
1.1 Pendahuluan	I-1
1.2 Latar Belakang Masalah	I-1
1.3 Rumusan Masalah	I-4
1.4 Tujuan Penelitian.....	I-5
1.5 Manfaat Penelitian.....	I-6
1.6 Batasan Masalah.....	I-6
1.7 Sistematika Penulisan.....	I-7
1.8 Kesimpulan.....	I-8
 BAB II KAJIAN LITERATUR	II-1
2.1 Pendahuluan	II-1
2.2 Analisis Sentimen.....	II-1
2.3 <i>Text Preprocessing</i>	II-3
2.3.1 <i>Case folding</i>	II-3
2.3.2 Normalisasi	II-3
2.3.3 <i>Remove Punctuation</i>	II-4
2.3.4 <i>Remove Number</i>	II-4
2.3.5 <i>Stop Words Removal</i>	II-5
2.3.6 <i>Stemming</i>	II-5
2.3.7 <i>Tokenizing</i>	II-6

2.4	<i>Twitter</i>	II-7
2.5	<i>Multinomial Naïve Bayes</i>	II-8
2.6	<i>Imbalanced Dataset</i> (Ketidakseimbangan Dataset)	II-9
2.7	<i>SMOTE (Synthetic Minority Oversampling Technique)</i>	II-10
2.8	Term Frequency/Inverse Document Frequency (TF-IDF)	II-13
2.9	<i>Confusion Matrix</i>	II-14
2.10	RUP (Rational Unified Process).....	II-16
2.11	Penelitian Lain yang Relevan.....	II-17
2.12	Kesimpulan.....	II-20
	BAB III METODOLOGI PENELITIAN.....	III-1
3.1	Pendahuluan	III-1
3.2	Pengumpulan Data.....	III-1
3.2.1	Jenis dan Sumber Data.....	III-1
3.2.2	Metode Pengumpulan Data	III-2
3.3	Tahapan Penelitian	III-3
3.3.1	Menetapkan Kerangka Kerja / Framework	III-3
3.3.2	Menetapkan Kriteria Pengujian.....	III-7
3.3.3	Menetapkan Format Data Pengujian	III-8
3.3.4	Menentukan Alat yang Digunakan dalam Pelaksanaan Penelitian	III-9
3.3.5	Melakukan Pengujian Penelitian.....	III-9
3.3.6	Menetapkan Analisis Hasil Pengujian dan Kesimpulan Penelitian	III-10
3.4	Metode Pengembangan Perangkat Lunak	III-11
3.4.1	Fase Insepsi	III-11
3.4.2	Fase Elaborasi	III-12
3.4.3	Fase Konstruksi.....	III-12
3.4.4	Fase Transisi.....	III-13
3.5	Manajemen Proyek Penelitian.....	III-13
3.6	Kesimpulan.....	III-14
	BAB IV PENGEMBANGAN PERANGKAT LUNAK.....	IV-1
4.1	Pendahuluan	IV-1
4.2	Fase Insepsi	IV-1

4.2.1	Pemodelan Bisnis	IV-1
4.2.2	Kebutuhan Sistem	IV-3
4.2.3	Analisis dan Desain.....	IV-6
4.3	Fase Elaborasi.....	IV-55
4.3.1	Pemodelan Bisnis	IV-55
4.3.2	Perancangan Data.....	IV-55
4.3.3	Perancangan Antarmuka	IV-57
4.3.4	Kebutuhan Sistem	IV-59
4.3.5	Diagram Aktivitas	IV-60
4.3.6	Diagram Alur	IV-63
4.4	Fase Konstruksi	IV-66
4.4.1	Kebutuhan Sistem	IV-66
4.4.2	Diagram Kelas.....	IV-66
4.4.3	Implementasi.....	IV-68
4.5	Fase Transisi	IV-71
4.5.1	Pemodelan Bisnis	IV-71
4.5.2	Kebutuhan Sistem	IV-71
4.5.3	Rencana Pengujian.....	IV-72
4.6	Kesimpulan.....	IV-77
BAB V HASIL DAN ANALISIS PENELITIAN		V-1
5.1	Pendahuluan	V-1
5.2	Data Hasil Penelitian	V-1
5.2.1	Konfigurasi Percobaan.....	V-1
5.2.2	Hasil Pengujian Pada <i>General Sentiment Dataset</i>	V-3
5.2.3	Hasil Pengujian Pada <i>Dataset Covid</i>	V-8
5.2.4	Hasil Pengujian Pada <i>Dataset PILKADA1</i>	V-13
5.2.5	Hasil Pengujian Pada <i>Dataset PILKADA2</i>	V-18
5.3	Analisis Hasil Pengujian Secara Keseluruhan.....	V-23
5.4	Kesimpulan.....	V-31
BAB VI KESIMPULAN DAN SARAN		VI-1
6.1	Pendahuluan	VI-1
6.2	Kesimpulan.....	VI-1

6.3 Saran	VI-3
DAFTAR PUSTAKA	xxi
LAMPIRAN	L-1

DAFTAR TABEL

Tabel III- 1. Rancangan Tabel <i>Confusion Matrix</i> Hasil klasifikasi.....	III-8
Tabel III- 2. Rancangan Tabel Hasil Analisis Klasifikasi	III-11
Tabel III- 3. Penjadwalan Penelitian dalam Bentuk WBS.....	III-15
Tabel IV- 1. Kebutuhan Fungsional.....	IV-5
Tabel IV- 2. Kebutuhan Non-Fungsional.....	IV-6
Tabel IV- 3. Contoh Data Kicauan.....	IV-10
Tabel IV- 4. Hasil <i>Case Folding</i>	IV-12
Tabel IV- 5. Hasil Tahapan Normalisasi.....	IV-13
Tabel IV- 6. Hasil Tahapan <i>Remove Punctuation</i>	IV-15
Tabel IV- 7. Hasil Tahapan <i>Remove Number</i>	IV-16
Tabel IV- 8. Hasil Tahapan <i>Stop Words Removal</i>	IV-17
Tabel IV- 9. Hasil Tahapan <i>Stemming</i>	IV-19
Tabel IV- 10. Hasil Tahapan <i>Tokenizing</i>	IV-20
Tabel IV- 11. Hasil TF-IDF Dari Contoh Data Kicauan.....	IV-22
Tabel IV- 12. Hasil Pembobotan Kata Dari Contoh Data Kicauan	IV-24
Tabel IV- 13. Hasil Pembentukan KNN Kelas Minor Positif Dari Contoh Data Kicauan	IV-28
Tabel IV- 14. Hasil Pembentukan KNN Kelas Minor Netral Dari Contoh Data Kicauan	IV-31
Tabel IV- 15. Hasil Pembentukan Data Sintetis Kelas Minor Positif	IV-35
Tabel IV- 16. Hasil Pembentukan Data Sintetis Kelas Minor Netral	IV-37
Tabel IV- 17. Frekuensi Kemunculan Kata Pada Data Sintetis	IV-41
Tabel IV- 18. Hasil <i>Conditional Probability</i> Pada Contoh Data Kicauan.....	IV-45
Tabel IV- 19. Klasifikasi Pada Data <i>Testing</i>	IV-47
Tabel IV- 20. Contoh <i>Confusion Matrix</i>	IV-48
Tabel IV- 21. Definisi Aktor	IV-50
Tabel IV- 22. Definisi <i>Use Case</i>	IV-51
Tabel IV- 23. Skenario Memilih <i>Dateset</i>	IV-51
Tabel IV- 24. Skenario Menguji Sistem Analisis sentimen	IV-53
Tabel IV- 25. Rancangan Data	IV-56
Tabel IV- 26. Rencana Pengujian <i>Use Case</i> Memilih <i>Dataset</i>	IV-73
Tabel IV- 27. Rencana Pengujian <i>Use Case</i> Pengujian Sistem Analisis Sentimen	IV-74
Tabel IV- 28. Pengujian <i>Use Case</i> Memilih <i>Dataset</i>	IV-75
Tabel IV- 29. Hasil Pengujian Pada <i>Use Case</i> Sistem Analisis sentimen Menggunakan <i>Naïve Bayes</i> Tanpa SMOTE dan <i>Naïve Bayes</i> Dengan SMOTE.	IV-76
Tabel V- 1. Hasil <i>Confusion Matrix Naive Bayes</i> Tanpa SMOTE Pada <i>General Sentiment Dataset</i>	V-3

Tabel V- 2. Hasil <i>Confusion Matrix Naïve Bayes Dengan SMOTE Pada General Sentiment Dataset</i>	V-4
Tabel V- 3. Hasil Evaluasi <i>Naïve Bayes Tanpa SMOTE Pada General Sentiment Dataset</i>	V-5
Tabel V- 4. Hasil Evaluasi <i>Naïve Bayes Dengan SMOTE Pada General Sentiment Dataset</i>	V-5
Tabel V- 5. Hasil <i>Confusion Matrix Naïve Bayes Tanpa SMOTE Pada Dataset Covid</i>	V-9
Tabel V- 6. Hasil <i>Confusion Matrix Pada Klasifikasi Naïve Bayes Dengan SMOTE Pada Dataset Covid</i>	V-9
Tabel V- 7. Hasil Evaluasi <i>Naïve Bayes Tanpa SMOTE Pada Dataset Covid</i>	V-10
Tabel V- 8. Hasil Evaluasi <i>Naïve Bayes Dengan SMOTE Pada Dataset Covid</i> ..	V-10
Tabel V- 9. Hasil <i>Confusion Matrix Naïve Bayes Tanpa SMOTE Pada Dataset PILKADA1</i>	V-14
Tabel V- 10. Hasil <i>Confusion Matrix Naïve Bayes Dengan SMOTE Pada Dataset PILKADA1</i>	V-14
Tabel V- 11. Hasil Evaluasi <i>Naïve Bayes Tanpa SMOTE Pada Dataset PILKADA1</i>	V-15
Tabel V- 12. Hasil Evaluasi <i>Naïve Bayes Dengan SMOTE Pada Dataset PILKADA1</i>	V-15
Tabel V- 13. Hasil <i>Confusion Matrix Naïve Bayes Tanpa SMOTE Pada Dataset PILKADA2</i>	V-19
Tabel V- 14. Hasil <i>Confusion Matrix Naïve Bayes Dengan SMOTE Pada Dataset PILKADA2</i>	V-19
Tabel V- 15. Hasil Evaluasi <i>Naïve Bayes Tanpa SMOTE Pada Dataset PILKADA2</i>	V-20
Tabel V- 16. Hasil Evaluasi <i>Naïve Bayes Dengan SMOTE Pada Dataset PILKADA2</i>	V-20

DAFTAR GAMBAR

Gambar II- 1. Contoh kerangka sistem analisis sentimen (Rozi et al., 2019) ...	II-2
Gambar II- 2. Contoh tahap normalisasi (Septian et al., 2019).	II-4
Gambar II- 3. Contoh tahapan <i>stopword removal</i> (Septian et al., 2019).	II-5
Gambar II- 4. Contoh tahap <i>stemming</i> (Septian et al., 2019).....	II-6
Gambar II- 5. Contoh tahap <i>tokenizing</i> (Septian et al., 2019).	II-7
Gambar II- 6. <i>Pseudocode</i> algoritma SMOTE.....	II-11
Gambar II- 7. <i>Whale Chart RUP</i> ((Hughes, 2016)	II-16
Gambar III-1. Diagram Tahapan Proses Perangkat Lunak.....	III-4
Gambar III-2. Diagram Tahapan <i>Text Preprocessing</i>	III-5
Gambar III-3. Diagram Tahapan Pengujian Penelitian.....	III-10
Gambar III-4. Penjadwalan Pada Tahap Menentukan Ruang Lingkup Dan Unit Penelitian.....	III-21
Gambar III-5. Penjadwalan untuk menentukan dasar teori dan kriteria pengujian.....	III-22
Gambar III-6. Penjadwalan untuk menentukan alat pada penelitian fase insepsi.....	III-22
Gambar III-7. Penjadwalan untuk menentukan alat pada penelitian fase Elaborasi.....	III-23
Gambar III-8. Penjadwalan untuk menentukan alat pada penelitian fase Kontruksi.....	III-24
Gambar III-9. Penjadwalan Untuk Menentukan Pada Penelitian Fase Transisi.....	III-25
Gambar III-10. Penjadwalan Untuk Tahap Pengujian Dan Analisa Hasil Pengujian.....	III-26
Gambar IV- 1. Contoh Perhitungan <i>Prior Probability</i>	IV-44
Gambar IV- 2. Contoh Perhitungan <i>Prior Probability</i>	IV-44
Gambar IV- 3. Hasil Perhitungan Probabilitas Tertinggi Pada Data Testing	IV-47
Gambar IV- 4. Perhitungan Akurasi, <i>Precision</i> , <i>Recall</i> dan <i>F-Measure</i>	IV-49
Gambar IV- 5. Diagram <i>Use Case</i>	IV-50
Gambar IV- 6. Rancangan Antar Muka Pilih <i>Dataset</i>	IV-58
Gambar IV- 7. Rancangan Antar Muka Informasi Parameter SMOTE Dan Detail <i>Dataset</i>	IV-58
Gambar IV- 8. Rancangan Antar Muka Hasil Pengujian Sistem	IV-59
Gambar IV- 9. Diagram Aktivitas Memilih <i>Dataset</i>	IV-61
Gambar IV- 10. Diagram Aktivitas Pengujian Sistem	IV-62
Gambar IV- 11. Diagram Alur Memilih Dataset.	IV-64
Gambar IV- 12. Diagram Alur Menguji Sistem Analisis sentimen	IV-65
Gambar IV- 13. Diagram Kelas.	IV-67
Gambar IV- 14. Implementasi Antarmuka Pilih <i>Dataset</i>	IV-70
Gambar IV- 15. Implementasi Antarmuka Informasi Paramter SMOTE dan Detail Dataset.....	IV-70
Gambar IV- 16. Implementasi Antarmuka Pengujian Sistem.....	IV-71

Gambar V- 1. Grafik Perbandingan Akurasi NB tanpa SMOTE dan NB dengan SMOTE Pada <i>General Sentiment Dataset</i>	V-6
Gambar V- 2. Grafik Perbandingan <i>Precision</i> NB tanpa SMOTE dan NB dengan SMOTE Pada <i>General Sentiment Dataset</i>	V-7
Gambar V- 3. Grafik Perbandingan <i>Recall</i> NB tanpa SMOTE dan NB dengan SMOTE Pada <i>General Sentiment Dataset</i>	V-7
Gambar V- 4. Grafik Perbandingan <i>F-measure</i> NB tanpa SMOTE dan NB dengan SMOTE Pada <i>General Sentiment Dataset</i>	V-8
Gambar V- 5. Grafik Perbandingan Akurasi NB tanpa SMOTE dan NB dengan SMOTE Pada <i>Dataset Covid</i>	V-11
Gambar V- 6. Grafik Perbandingan <i>Precision</i> NB tanpa SMOTE dan NB dengan SMOTE Pada <i>Dataset Covid</i>	V-12
Gambar V- 7. Grafik Perbandingan <i>Recall</i> NB tanpa SMOTE dan NB dengan SMOTE Pada <i>Dataset Covid</i>	V-12
Gambar V- 8. Grafik Perbandingan <i>F-Measure</i> NB tanpa SMOTE dan NB dengan SMOTE Pada <i>Dataset Covid</i>	V-13
Gambar V- 9. Grafik Perbandingan Akurasi NB tanpa SMOTE dan NB dengan SMOTE Pada <i>Dataset PILKADA1</i>	V-16
Gambar V- 10. Grafik Perbandingan <i>Precision</i> NB tanpa SMOTE dan NB dengan SMOTE Pada <i>Dataset PILKADA1</i>	V-17
Gambar V- 11. Grafik Perbandingan <i>Recall</i> NB tanpa SMOTE dan NB dengan SMOTE Pada <i>Dataset PILKADA1</i>	V-17
Gambar V- 12. Grafik Perbandingan <i>F-measure</i> NB tanpa SMOTE dan NB dengan SMOTE Pada <i>Dataset PILKADA1</i>	V-18
Gambar V- 13. Grafik Perbandingan Akurasi NB tanpa SMOTE dan NB dengan SMOTE Pada <i>Dataset PILKADA2</i>	V-21
Gambar V- 14. Grafik Perbandingan <i>Precision</i> NB tanpa SMOTE dan NB dengan SMOTE Pada <i>Dataset PILKADA2</i>	V-22
Gambar V- 15. Grafik Perbandingan <i>Recall</i> NB tanpa SMOTE dan NB dengan SMOTE Pada <i>Dataset PILKADA2</i>	V-22
Gambar V- 16. Grafik Perbandingan <i>F-measure</i> NB tanpa SMOTE dan NB dengan SMOTE Pada <i>Dataset PILKADA2</i>	V-23
Gambar V- 17. Grafik Rata-rata Akurasi Pada Seluruh <i>Dataset</i>	V-24
Gambar V- 18. Grafik Rata-rata <i>Precision</i> Pada Seluruh <i>Dataset</i>	V-24
Gambar V- 19. Grafik Rata-rata <i>Recall</i> Pada Seluruh <i>Dataset</i>	V-25
Gambar V- 20. Grafik Rata-rata <i>F-measure</i> Pada Seluruh <i>Dataset</i>	V-26

DAFTAR LAMPIRAN

Halaman

1. Dokumentasi *Source Code*.....L-1

BAB I

PENDAHULUAN

1.1 Pendahuluan

Bab pendahuluan akan memberikan penjelasan umum mengenai latar belakang masalah, rumusan masalah, tujuan penelitian, manfaat penelitian dan batasan masalah serta bagaimana sistematika penulisan dalam penelitian ini.

Bab ini dimulai dengan pembahasan kebutuhan dan tantangan terhadap klasifikasi analisis sentimen serta penelitian yang berkaitan dengan *Synthetic Minority Oversampling Technique* (SMOTE).

1.2 Latar Belakang Masalah

Salah satu *platform* media sosial di Indonesia yang mengalami pertumbuhan sangat pesat adalah *twitter* yang saat ini mampu menjadi sarana komunikasi, promosi, ataupun kampanye politik. Media sosial *twitter* memiliki tingkat popularitas yang cukup tinggi dikarenakan *twitter* menjadi saluran pilihan masyarakat untuk menunjukkan perasaan dan pikiran mereka melalui ulasan, *posting* dan *tweet*. (Gupta & Gupta, 2019).

Pengguna *twitter* dapat melakukan *update* dan *post* berupa opini, kritik maupun saran melalui *twitter* berupa teks pesan dengan batas maksimal 140 karakter yang disebut dengan *tweet*. *Tweet* ini memberikan keanekaragaman pesan dengan penggunaan bahasa Indonesia yang kurang bahkan tidak baku. Dengan keanekaragaman *tweet* ini maka diperlukan *sentiment analysis* atau *opinion mining*

untuk mengetahui kategori opini yang diberikan oleh pengguna. *Sentiment analysis* adalah teknik untuk menilai suatu bahasa tertulis maupun lisan apakah bernilai positif, negatif atau netral (Alsaedi & Khan, 2019). *Sentiment analysis* memungkinkan kita untuk mendapatkan pengetahuan tentang opini yang diberikan oleh pengguna melalui *tweet* dengan jumlah besar apakah positif, negatif, atau netral yang sulit dilakukan secara manual oleh manusia melalui *sentiment classification*. Pada dasarnya teknik *sentiment classification* terdiri dari pendekatan *machine learning*, *lexicon based*, dan *hybrid* (Sharma & Singh, 2018). Dalam penelitian ini akan menggunakan pendekatan *machine learning* dalam proses *sentiment classification*.

Salah satu pendekatan *machine learning* dalam klasifikasi *Sentiment analysis* dengan menggunakan *Naïve Bayes*. *Naïve Bayes* merupakan algoritma yang mampu melakukan pengklasifikasian probablistik sederhana dengan menghitung kumpulan probabilitas melalui penjumlahan frekuensi dan kombinasi nilai dari *dataset* yang diberikan (Saputra *et al.*, 2018). *Naïve Bayes* juga disebut *Bayesian Classification* yang terbukti memiliki akurasi serta kecepatan yang cukup tinggi saat di implementasikan dalam dataset yang cukup besar (Muhammad *et al.*, 2017).

Dalam penelitian Lorosae, Prakoso, Saifudin dan kusrini (2018) pada analisis sentimen jasa pengiriman barang menggunakan algoritma *Naïve Bayes* sebagai metode klasifikasi sistem mampu mengklasifikasikan sentimen pada opini *tweet* kedalam tiga label sentiment positif, negatif dan netral dengan mendapatkan hasil akurasi sebesar 84%. *Naïve Bayes* juga pernah diterapkan dalam analisis

sentimen *mobile reviews* dengan akurasi sebesar 79.66% (Mathapati *et al.*, 2017).

Dari hasil penelitian tersebut terlihat *Naïve Bayes* cukup baik untuk diterapkan dalam klasifikasi *Sentiment analysis*.

Permasalahan yang kemudian timbul dalam *Sentiment analysis* adalah data opini yang cenderung tidak seimbang dalam jumlah masing-masing kelas atau condong pada salah satu kelas. Hal ini akan menyebabkan ketidakseimbangan dalam *dataset* (*imbalanced dataset*). Dalam pembelajaran mesin dan *data mining* klasifikasi pada *dataset* yang tidak seimbang telah menjadi masalah utama (Zheng *et al.*, 2015). Ketika terdapat ketidakseimbangan dalam data latih algoritma klasifikasi akan cenderung mengklasifikasikan kedalam kelas mayoritas karena peningkatan probabilitas sebelumnya sehingga mengakibatkan kelas minoritas lebih cenderung salah diklasifikasikan dibanding dengan kelas mayoritas (Johnson & Khoshgoftaar, 2019).

Dalam penelitian ini menggunakan algoritma naïve bayes classifier dimana hampir semua algoritma klasifikasi termasuk *Naive Bayes*, *Decision Tree*, *K-Nearest Neighbor* dan yang lainnya memperlihatkan performa yang sangat buruk ketika melakukan proses klasifikasi pada data dengan kelas yang tidak seimbang karena algoritma tersebut tidak lengkap dengan kemampuan untuk menangani permasalahan ketidak seimbangan kelas dalam dataset (Siringoringo, 2018). Oleh karena itu sangat penting untuk menangani permasalahan *imbalanced dataset* pada analisis sentimen untuk mendapatkan kinerja klasifikasi yang lebih baik.

Salah satu pendekatan dalam menangani permasalahan *imbalanced data* adalah SMOTE (Synthetic Minority Oversampling Technique). SMOTE adalah

pendekatan yang menggunakan teknik *over-sampling* dengan mensintesis data pada kelas minor sehingga data menjadi seimbang antara data pada kelas mayoritas dan minoritas (Chawla *et al.*, 2002). Beberapa penelitian yang menerapkan SMOTE diantaranya Putri dan Wahono (2015) mengkombinasikan SMOTE dengan *Naïve Bayes* serta *information gain* pada kasus cacat *software* dari hasil penelitian menunjukan jika model yang diusulkan mencapai hasil akurasi yang lebih tinggi dengan rata-rata AUC *Naïve Bayes SMOTE+IG* dengan 0.798 sedangkan jika hanya menggunakan *Naïve Bayes* adalah 0.753. SMOTE juga pernah dikombinasikan dengan algoritma *random forest classifier* untuk mengklasifikasikan dataset penyakit Parkinson. Hasil penelitian ini terbukti kombinasi antara SMOTE dan *random forest classifier* mencapai akurasi klasifikasi hingga 94.89% dibanding dengan hanya menggunakan *random forest classifier* yang akurasinya 87.037% (Polat, 2019).

Dari hasil uraian diatas, maka penelitian ini akan melakukan pengujian terhadap pengaruh SMOTE pada *sentiment analysis* dengan menggunakan algoritma *Naïve Bayes*.

1.3 Rumusan Masalah

Berdasarkan latar belakang masalah yang telah dijelaskan sebelumnya rumusan masalah pada penelitian ini adalah bagaimana pengaruh SMOTE pada *sentiment analysis* menggunakan *Naïve Bayes classifier*. Dalam menyelesaikan permasalahan tersebut penelitian ini dibagi menjadi beberapa *research question* yakni sebagai berikut :

1. Bagaimana implementasi dari klasifikasi sentimen analisis menggunakan *Multinomial Naïve Bayes* ?
2. Bagaimana implementasi klasifikasi sentimen analisis menggunakan algortima *Multinomial Naïve Bayes* yang dikombinasikan dengan SMOTE ?
3. Bagaimana nilai akurasi, *precision*, *recall* dan F-measure pada *sentiment analysis* menggunakan algoritma *Naïve Bayes Classifier* ?
4. Bagaimana nilai akurasi, *precision*, *recall* dan F-measure pada *sentiment analysis* menggunakan algoritma *Naïve Bayes Classifier* dan SMOTE ?

1.4 Tujuan Penelitian

Berikut tujuan penelitian ini :

1. Mengetahui mekanisme analisis sentimen menggunakan *Naïve Bayes*.
2. Mengetahui mekanisme analisis sentimen menggunakan algortima SMOTE dan *Naïve Bayes Classifier* .
3. Mengetahui nilai akurasi, *precision*, *recall* dan F-measure menggunakan *confusion matrix* paada analisis sentimen menggunakan algoritma *Naïve Bayes Classifier*.
4. Mengetahui nilai akurasi, *precision*, *recall* dan F-measure menggunakan *confusion matrix* pada analisis sentimen menggunakan algoritma *Naïve Bayes Classifier* dan SMOTE.

5. Membangun perangkat lunak yang dapat melakukan klasifikasi analisis sentimen menggunakan algoritma *Naïve Bayes* dan algoritma *Naïve Bayes* yang dikombinasikan dengan SMOTE.

1.5 Manfaat Penelitian

Manfaat dari penelitian ini adalah :

1. Memahami mekanisme SMOTE dalam menyeimbangkan dataset dan *Naïve Bayes* dalam proses klasifikasi *sentiment analysis*.
2. Hasil penelitian dapat berkontribusi dalam penelitian terkait penanganan ketidakseimbangan *dataset* analisis sentimen dan mampu digunakan sebagai referensi dipenelitian terkait.

1.6 Batasan Masalah

Adapun batasan masalah pada penelitian ini adalah sebagai berikut :

1. Data *sentiment analysis* merupakan opini yang berasal dari *tweet* pengguna pada media sosial *twitter*.
2. Data *sentiment analysis* yang digunakan berbahasa Indonesia.
3. Algoritma *Naïve Bayes* yang diterapkan untuk klasifikasi ini merupakan *Multinomial Naïve Bayes*.
4. Pengujian dilakukan dengan menggunakan *confusion matrix*
5. Metode pembobotan yang di pakai yakni TF-IDF (*Term Frequency- Inverse Document Frequency*)

6. Pengujian menggunakan empat *dataset* untuk mengetahui pengaruh SMOTE dengan *Naïve Bayes* pada data tersebut.

1.7 Sistematika Penulisan

Berikut alur sistematika penulisan skripsi ini:

BAB I. PENDAHULUAN.

Bab ini memberikan penjelasan latar belakang, perumusan masalah, tujuan dan manfaat penelitian, batasan masalah, metodologi penelitian, dan sistematika penulisan.

BAB II. KAJIAN LITERATUR

Bab ini membahas dasar-dasar teori yang digunakan seperti *sentiment analysis*, *preprocessing*, *Naïve Bayes Classifier*, *SMOTE* (*Synthetic Minority Oversampling Technique*). Bab ini akan menjelaskan penelitian-penelitian sebelumnya yang terkait dengan topik dalam penelitian ini.

BAB III. METODOLOGI PENELITIAN

Bab ini memberikan penjelasan tahapan yang dilaksanakan pada penelitian ini. Perencanaan tahapan penelitian akan dideskripsikan secara rinci dengan mengacu pada suatu kerangka kerja.

BAB IV. PENGEMBANGAN PERANGKAT LUNAK

Bab ini menguraikan tahapan dalam proses pengembangan perangkat lunak sentimen analisis menggunakan *Naïve Bayes* dengan SMOTE dan *Naïve Bayes* tanpa SMOTE. Tahapannya disesuaikan dengan metode RUP (*Rational Unified Porcess*).

BAB V. HASIL DAN ANALISIS PENELITIAN

Bab ini menguraikan hasil pengujian pada perangkat lunak yang telah dibangun serta bab ini akan memberikan analisis dari hasil pengujian tersebut.

BAB VI. KESIMPULAN DAN SARAN

Bab memberikan kesimpulan dari hasil penelitian yang telah dilakukan serta saran oleh penulis untuk penelitian kedepannya.

1.8 Kesimpulan

Bab ini telah membahas latar belakang masalah dari penelitian yang akan dilakukan. Penelitian ini akan melakukan pengujian pengaruh SMOTE (*Synthetic Minority Oversampling Technique*) terhadap klasifikasi *sentiment analysis* dengan menggunakan *Naïve Bayes classifier* dengan klasifikasi *Naïve Bayes classifier* tanpa menggunakan SMOTE.

DAFTAR PUSTAKA

- Alsaedi, A. & Khan, M.Z. 2019. A Study on Sentiment Analysis Techniques of Twitter Data. (February).
- Barro,Rossi; Sulvianti, D.A.F. 2013. Penerapan Synthetic Minority Oversampling Technique (Smote) Terhadap Data Tidak Seimbang Pada Pembuatan Model Komposisi Jamu. Xplore: Journal of Statistics, 1(1).
- Cahyono, Y. 2017. Analisis Sentiment pada Sosial Media Twitter Menggunakan Naïve Bayes Classifier dengan Feature Selection Particle Swarm Optimization dan Term Frequency. Jurnal Informatika Universitas Pamulang, 2(1): 14.
- Chawla, N. V, Bowyer, K.W. & Hall, L.O. 2002. SMOTE : Synthetic Minority Over-sampling Technique. 16: 321–357.
- Douzas, G., Bacao, F. & Last, F. 2018. Improving imbalanced learning through a heuristic oversampling method based on k-means and SMOTE. Information Sciences, 465: 1–20. (<https://doi.org/10.1016/j.ins.2018.06.056>).
- Ferdiana, R., Jatmiko, F., Purwanti, D.D., Sekar, A., Ayu, T., Dicka, W.F., Sentimen, A.A. & Indonesia, B. 2019. Dataset Indonesia untuk Analisis Sentimen. 8(4): 334–339.
- Flores, A.C. & Gorro, K.D. 2018. on Sentiment Analysis Data Set. 1–4.
- Fuadin, D.N., Pembimbing, D., Magister, P., Elektro, D.T. & Elektro, F.T. 2017. Deteksi Botnet Menggunakan Naïve Bayes. Thesis Fuadin, Didin Nizarul.
- Gupta, G. & Gupta, P. 2019. Twitter mining for sentiment analysis in tourism industry. Proceedings of the 3rd World Conference on Smart Trends in

- Systems, Security and Sustainability, WorldS4 2019, 302–306.
- Haddi, E., Liu, X. & Shi, Y. 2013. The role of text pre-processing in sentiment analysis. *Procedia Computer Science*, 17(December): 26–32. (<http://dx.doi.org/10.1016/j.procs.2013.05.005>).
- Hanssen, G.K., Westerheim, H. & Bjørnson, F.O. 2005. Using rational unified process in an SME - A case study. *Lecture Notes in Computer Science* (including subseries *Lecture Notes in Artificial Intelligence* and *Lecture Notes in Bioinformatics*), 3792 LNCS(2): 142–150.
- Hughes, R. 2016. Introduction to Alternative Iterative Methods. *Agile Data Warehousing for the Enterprise*, 31–54.
- Johnson, J.M. & Khoshgoftaar, T.M. 2019. Survey on deep learning with class imbalance. *Journal of Big Data*, 6(1). (<https://doi.org/10.1186/s40537-019-0192-5>).
- Kamal, A., Abulaish, M. & Jahiruddin 2016. Sentiment Analysis and Ontology Engineering. *Studies in Computational Intelligence*, 639(August 2016): 399–423. (<http://www.scopus.com/inward/record.url?eid=2-s2.0-84961589899&partnerID=tZOTx3y1>).
- Karyadi, S. & Yasin, H. 2016. 1 , 2 , 3. 5: 763–770.
- KURNIAWAN, R. & APRILIANI, A. 2020. Analisis Sentimen Masyarakat Terhadap Virus Corona Berdasarkan Opini Dari Twitter Berbasis Web Scraper. *Jurnal INSTEK (Informatika Sains dan Teknologi)*, .
- Li, H. & Sun, J. 2012. Forecasting business failure: The use of nearest-neighbour support vectors and correcting imbalanced samples - Evidence from the

- Chinese hotel industry. *Tourism Management*, 33(3): 622–634. (<http://dx.doi.org/10.1016/j.tourman.2011.07.004>).
- Liu, B. 2015. Sentiment analysis: Mining opinions, sentiments, and emotions. *Sentiment Analysis: Mining Opinions, Sentiments, and Emotions*, (May): 1–367.
- Lohmann, S., Burch, M., Schmauder, H. & Weiskopf, D. 2012. Visual analysis of microblog content using time-varying co-occurrence highlighting in tag clouds. *Proceedings of the Workshop on Advanced Visual Interfaces AVI*, (May): 753–756.
- Mathapati, P.M., Shahapurkar, A.S. & Hanabaratti, K.D. 2017. Sentiment Analysis using Naïve bayes Algorithm. *International Journal of Computer Sciences and Engineering*, 5(7): 75–77.
- Mathew, J., Pang, C.K., Luo, M. & Leong, W.H. 2018. Classification of Imbalanced Data by Oversampling in Kernel Space of Support Vector Machines. *IEEE Transactions on Neural Networks and Learning Systems*, 29(9): 4065–4076.
- Mohasseb, A., Bader-El-Den, M., Cocea, M. & Liu, H. 2018. Improving Imbalanced Question Classification Using Structured Smote Based Approach. *Proceedings - International Conference on Machine Learning and Cybernetics*, 2: 593–597.
- Muhamad, H., Prasojo, C.A., Sugianto, N.A., Surtiningsih, L. & Cholissodin, I. 2017. Optimasi Naïve Bayes Classifier Dengan Menggunakan Particle Swarm Optimization Pada Data Iris. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 4(3): 180.

- Mujilahwati, S. 2016. Pre-Processing Text Mining Pada Data Twitter. Seminar Nasional Teknologi Informasi dan Komunikasi, 2016(Sentika): 2089–9815.
- Murnawan, M. 2017. Pemanfaatan Analisis Sentimen Untuk Pemeringkatan Popularitas Tujuan Wisata. Jurnal Penelitian Pos dan Informatika, 7(2): 109.
- Mustaqim, M., Warsito, B. & Surarso, B. 2019. Kombinasi Synthetic Minority Oversampling Technique (SMOTE) dan Neural Network Backpropagation untuk menangani data tidak seimbang pada prediksi pemakaian alat kontrasepsi implan Combination of Synthetic Minority Oversampling Technique (SMOTE) and Bac. 5(34): 116–127.
- Pak, A. & Paroubek, P. 2010. Twitter as a corpus for sentiment analysis and opinion mining. Proceedings of the 7th International Conference on Language Resources and Evaluation, LREC 2010, 1320–1326.
- Polat, K. 2019. A Hybrid Approach to Parkinson Disease Classification using speech signal : The combination of SMOTE and Random Forests. 2019 Scientific Meeting on Electrical-Electronics & Biomedical Engineering and Computer Science (EBBT), 1–3.
- Prusty, M.R., Jayanthi, T. & Velusamy, K. 2017. Weighted-SMOTE: A modification to SMOTE for event classification in sodium cooled fast reactors. Progress in Nuclear Energy, 100(May 2018): 355–364. (<http://dx.doi.org/10.1016/j.pnucene.2017.07.015>).
- Pujianto, U., Hidayat, M.F. & Rosyid, H.A. 2019. Text Difficulty Classification Based on Lexile Levels Using K-Means Clustering and Multinomial Naive Bayes. Proceedings - 2019 International Seminar on Application for

- Technology of Information and Communication: Industry 4.0: Retrospect, Prospect, and Challenges, iSemantic 2019, 163–170.
- Putra, R.S. 2017. Analisis sentimen twitter dengan klasifikasi naïve bayes menggunakan seleksi fitur mutual information dan inverse document frequency riky sutriadi putra.
- Rambocas, M. & Gama, J. 2013. Marketing Research : The Role of Sentiment Analysis. Working Papers (FEP) - Universidade Do Porto, (489): 1–24.
- Ratnawati, F. 2018. Implementasi Algoritma Naive Bayes Terhadap Analisis Sentimen Opini Film Pada Twitter.
- Rossi, A., Lestari, T., Setya Perdana, R. & Fauzi, M.A. 2017. Analisis Sentimen Tentang Opini Pilkada DKI 2017 Pada Dokumen Twitter Berbahasa Indonesia Menggunakan Näive Bayes dan Pembobotan Emojo. Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer, 1(12): 1718–1724. (<http://j-ptiik.ub.ac.id>).
- Rozi, I.F., Hamdana, E.N., Balya, M. & Alfahmi, I. n.d. Pengembangan Aplikasi Analisis Sentimen Twitter (Studi Kasus SAMSAT Kota Malang). 149–154.
- Sanguanmak, Y. & Hanskunatai, A. 2016. DBSM: The combination of DBSCAN and SMOTE for imbalanced data classification. 2016 13th International Joint Conference on Computer Science and Software Engineering, JCSSE 2016.
- Saputra, R.A., Taufik, A.R., Ramdhani, L.S., Oktapian, R. & Marsusanti, E. 2018. Sistem Pendukung Keputusan Dalam Menentukan Metode Kontrasepsi Menggunakan Algoritma Naive Bayes. Seminar Nasional Inovasi dan Tren, 106–111.

- Satriaji, W. & Kusumaningrum, R. 2019. 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.
- Septian, J.A., Fahrudin, T.M. & Nugroho, A. 2019. Analisis Sentimen Pengguna Twitter Terhadap Polemik Persepakbolaan Indonesia Menggunakan Pembobotan TF - IDF dan K - Nearest Neighbor. (September).
- Sharma, S. & Singh, D. 2018. Study of Sentiment Classification Techniques. International Journal of Computer Sciences and Engineering, 6(5): 479–783.
- Siringoringo, R. n.d. Klasifikasi Data Tidak Seimbang Menggunakan. 3(1): 44–49.
- Song, J., Kim, K.T., Lee, B., Kim, S. & Youn, H.Y. 2017. A novel classification approach based on Naïve Bayes for Twitter sentiment analysis. KSII Transactions on Internet and Information Systems, 11(6): 2996–3011.
- Tallo, T.E. & Musdholifah, A. 2018. The Implementation of Genetic Algorithm in Smote (Synthetic Minority Oversampling Technique) for Handling Imbalanced Dataset Problem. Proceedings - 2018 4th International Conference on Science and Technology, ICST 2018, 1: 1–4.
- Tsai, M.F. & Yu, S.S. 2016. Distance Metric Based Oversampling Method for Bioinformatics and Performance Evaluation. Journal of Medical Systems, 40(7). (<http://dx.doi.org/10.1007/s10916-016-0516-3>).
- Zheng, Z., Cai, Y. & Li, Y. 2015. Oversampling method for imbalanced classification. Computing and Informatics, 34(5): 1017–1037.