

**ANALISIS SENTIMEN TERHADAP KINERJA PELAYANAN
DI PT BANK RAKYAT INDONESIA (PERSERO) TBK.
MENGUNAKAN METODE *SUPPORT VECTOR MACHINE*,
NAIVE BAYES, DAN *K-NEAREST NEIGHBORS***

(Skripsi)

Oleh

DEVI RAMADHIA FITRI

NPM 1917051005



**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG**

2023

ABSTRAK

ANALISIS SENTIMEN TERHADAP KINERJA PELAYANAN DI PT. BANK RAKYAT INDONESIA (PERSERO) TBK. MENGGUNAKAN METODE *SUPPORT VECTOR MACHINE*, *NAÏVE BAYES*, DAN *K-NEAREST NEIGHBORS*

Oleh

DEVI RAMADHIA FITRI

Penelitian ini mengenai analisis sentimen terhadap kinerja pelayanan di PT. Bank Rakyat Indonesia (Persero) Tbk. Dataset yang digunakan sebanyak 1652 data Xs yang dikumpulkan sejak bulan Maret 2022 sampai dengan Februari 2023 yang diklasifikasikan menjadi dua kelas, yaitu sentimen positif dan sentimen negatif. Dalam memberikan pelayanan kepada nasabah antrian yang panjang kerap terjadi PT. Bank Rakyat Indonesia (Persero) Tbk. sehingga dapat menyebabkan komentar-komentar mengenai pelayanan. PT. Bank Rakyat Indonesia (Persero) Tbk. belum terdapat dalam menanggapi komentar nasabah seharusnya dibuatkan secara rutin analisis sentimen terhadap kinerja pelayanan. Analisis sentimen dalam penelitian ini menggunakan metode *Support Vector Machine (SVM)*, *Naïve Bayes*, dan *K-Nearest Neighbors*. Penelitian ini menerapkan penyeimbangan data pada *dataset imbalanced* menggunakan *Synthetic Minority Over-sampling Technique (SMOTE)* untuk mendapatkan performa yang terbaik. Hasil evaluasi menggunakan algoritma *Support Vector Machine (SVM)* pada *dataset balanced* yang menghasilkan kinerja yang lebih baik dibandingkan dengan metode lainnya akurasi sebesar 94,30%, algoritma *Naïve Bayes* sebesar 93,68% pada *dataset balanced*, sementara algoritma *K-Nearest Neighbors (KNN)* sebesar 88,31% pada *dataset balanced*. Perbandingan hasil kinerja berdasarkan ketiga metode yang telah digunakan menunjukkan bahwa algoritma *Support Vector Machine (SVM)* lebih baik dibandingkan *Naïve Bayes*, dan *K-Nearest Neighbors (KNN)* untuk analisis sentimen terhadap kinerja pelayanan di PT. Bank Rakyat Indonesia (Persero) Tbk.

Kata Kunci: Analisis Sentimen, *Support Vector Machine*, *Naïve Bayes*, *K-Nearest Neighbors*, Pelayanan PT. Bank Rakyat Indonesia (Persero) Tbk.

ABSTRACT

SENTIMENT ANALYSIS OF SERVICE PERFORMANCE AT PT. BANK RAKYAT INDONESIA (PERSERO) TBK. USING SUPPORT VECTOR MACHINE, NAÏVE BAYES, AND K-NEAREST NEIGHBORS METHODS

BY

DEVI RAMADHIA FITRI

This research is about sentiment analysis of the service performance at PT. Bank Rakyat Indonesia (Persero) Tbk. The dataset used is 1652 data Xs collected from March 2022 to February 2023, classified into two classes, namely positive and negative sentiment. In providing services to customers, long queues often occur, PT. Bank Rakyat Indonesia (Persero) Tbk. so that it can lead to comments about the service. PT. Bank Rakyat Indonesia (Persero) Tbk. Yet to be available in response to customer comments, we should routinely make sentiment analysis on service performance. Sentiment analysis in this study uses the Support Vector Machine (SVM), Naïve Bayes, and K-Nearest Neighbors methods. This study applies data balancing to *imbalanced* datasets using the Synthetic Minority Over-sampling Technique (SMOTE) for the best performance. The evaluation results use the Support Vector Machine (SVM) algorithm on a balanced dataset, producing better performance than other methods with an accuracy of 94,30%. The Naïve Bayes algorithm is 93,68% on a balanced dataset, while the K-Nearest Neighbors (KNN) algorithm is 88,31%. The performance comparison results based on the three methods used show that the Support Vector Machine (SVM) algorithm is better than the Naïve Bayes and K-Nearest Neighbors (KNN) algorithms for sentiment analysis of service performance at PT. Bank Rakyat Indonesia (Persero) Tbk.

Keywords: Sentiment Analysis, Support Vector Machine, Naïve Bayes, K-Nearest Neighbors, service of PT. Bank Rakyat Indonesia (Persero) Tbk.

**ANALISIS SENTIMEN TERHADAP KINERJA PELAYANAN
DI PT BANK RAKYAT INDONESIA (PERSERO) TBK.
MENGUNAKAN METODE *SUPPORT VECTOR MACHINE*,
NAIVE BAYES, DAN *K-NEAREST NEIGHBORS***

Oleh

DEVI RAMADHIA FITRI

Skripsi

**Sebagai Salah Satu Syarat untuk Mencapai Gelar
SARJANA KOMPUTER**

Pada

**Jurusan Ilmu Komputer
Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Lampung**



**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG**

2023

Judul Skripsi : **ANALISIS SENTIMEN TERHADAP KINERJA PELAYANAN DI PT. BANK RAKYAT INDONESIA (PERSERO) TBK. MENGGUNAKAN METODE *SUPPORT VECTOR MACHINE*, *NAÏVE BAYES*, DAN *K-NEAREST NEIGHBORS***

Nama Mahasiswa : ***Devi Ramadhia Fitri***

Nomor Induk Mahasiswa : **1917051005**

Program Studi : **Ilmu Komputer**

Fakultas : **Matematika dan Ilmu Pengetahuan Alam**

MENYETUJUI

1. **Komisi Pembimbing**


Prof. Adm. Syarif, Ph.D.
NIP. 19670103 199203 1 003

2. **Ketua Jurusan Ilmu Komputer**


Didik Kurniawan, S.Si., M.T.
NIP. 19800419 200501 1 004

MENGESAHKAN

1. Tim Penguji

Ketua : **Prof. Admi Syarif, Ph.D.**



Penguji I : **Dr. rer. nat. Akmal Junaidi, M.Sc.**



Penguji II : **Favorisen R. Lumbanraja, Ph.D.**



2. Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam



Dr. Eng. Heri Satria, S.Si., M.Si.
NIP. 19711001 200501 1 002

Tanggal Lulus Ujian Skripsi: 01 Agustus 2023

PERNYATAAN

Saya yang bertanda tangan di bawah ini:

Nama : Devi Ramadhia Fitri

NPM : 1917051005

Dengan ini menyatakan bahwa skripsi saya yang berjudul “Analisis Sentimen Terhadap Kinerja Pelayanan di PT. Bank Rakyat Indonesia (Persero) Tbk. Menggunakan Metode *Support Vector Machine*, *Naïve Bayes*, dan *K-Nearest Neighbors*.” merupakan karya saya sendiri bukan karya orang lain. Semua tulisan yang tertulis dalam skripsi ini telah mengikuti kaidah penulisan karya ilmiah Universitas Lampung. Jika di kemudian hari terbukti bahwa skripsi saya merupakan hasil penjiplakan karya orang lain, maka saya bersedia menerima sanksi berupa pencabutan gelar yang telah saya peroleh.

Bandar Lampung, 01 Agustus 2023



Devi Ramadhia Fitri

NPM. 1917051005

RIWAYAT HIDUP



Penulis dilahirkan di Tanjung Karang pada tanggal 13 Desember 2001, sebagai anak kedua dari dua bersaudara. Penulis menyelesaikan pendidikan formal di SD Negeri 1 Langkapura, Bandar Lampung dan selesai pada tahun 2013. Kemudian melanjutkan pendidikan menengah pertama di SMP Negeri 14 Bandar Lampung yang diselesaikan pada tahun 2016, lalu melanjutkan ke pendidikan menengah atas di SMA Negeri 3 Bandar Lampung yang diselesaikan pada tahun 2019. Pada tahun 2019 penulis terdaftar sebagai mahasiswa jurusan Ilmu Komputer Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung. Selama menjadi mahasiswa, beberapa kegiatan yang dilakukan penulis antara lain.

1. Menjadi anggota Adapter Himpunan Mahasiswa Jurusan Ilmu Komputer pada periode 2019.
2. Menjadi anggota Bidang External Himpunan Mahasiswa Jurusan Ilmu Komputer pada periode 2020.
3. Menjadi wirausahawan mahasiswa Fakultas Matematika Dan Ilmu Pengetahuan Alam Universitas Lampung periode 2020.
4. Menjadi staff hubungan Dewan Perwakilan Mahasiswa Universitas Lampung periode 2021.
5. Menjadi bendahara Bidang Media Informasi Himpunan Mahasiswa Jurusan Ilmu Komputer pada periode 2021.
6. Menjadi Sekretaris Koordinator Acara Musyawarah Nasional XII Forum Lembaga Legislatif Mahasiswa Indonesia pada periode 2021.
7. Menjadi moderator Kegiatan karya wisata ilmiah XXXII BEM FMIPA Universitas Lampung periode 2021.

8. Mengikuti program Kredensial Mikro Mahasiswa Indonesia (KMMI) di Universitas Amikom Yogyakarta tahun 2021.
9. Melaksanakan kerja praktek periode 2022 di PT. Bank Rakyat Indonesia Tanjung Karang, Bandar Lampung.
10. Melaksanakan KKN 2022 periode 2 di Kelurahan Sumber Agung, Kecamatan Kemiling, Bandar Lampung.
11. Mengikuti ujian sertifikasi dan mendapatkan *sertifikat Microsoft Office Specialist: word Associate* dan *power point associatte* oleh microsoft pada tahun 2022.
12. Mengikuti studi independen kampus merdeka di PT. Maribelajar Cerdas Indonesia pada tahun 2022.

MOTTO

“Yakinlah, ada sesuatu yang menantimu setelah sekian banyak kesabaran yang kau jalani, yang akan membuatmu terpana hingga kau lupa betapa pedihnya rasa sakit”

-Ali Bin Abu Thalib-

“Kegagalan adalah bumbu kehidupan, kegagalan membuat kita bisa menjadi manusia tangguh”

-Bob Sadino-

“Menjadi yang terbaik diantara orang orang baik”

-Anonim-

PERSEMBAHAN

Alhamdulillahillobbilamin

Puji dan syukur tercurahkan kepada Allah Subhanahu Wa Ta'ala atas segala Rahmat dan Karunia-Nya sehingga saya dapat menyelesaikan skripsi ini. Shalawat serta salam selalu tercurahkan kepada Nabi Muhammad SAW.

Kupersembahkan karya ini kepada:

Diri sendiri

Yang telah berjuang dan bertahan hingga saat ini sehingga dapat menyelesaikan perkuliahan.

Ibu tercinta (almh) Nurhasanah

Yang senantiasa memberikan terbaik semasa hidupnya. Serta Ayah yang selalu melantunkan do'a dan memberikan kasih sayang yang tidak ternilai. Kuucapkan terima kasih sebesar-besarnya karena telah mendidik dan membesarkanku, serta pengorbanan yang tidak dapat terbalas.

Seluruh Keluarga Besar Ilmu Komputer 2019

Yang selalu memberikan semangat dan dukungan.

Almamater Tercinta, Universitas Lampung dan Jurusan Ilmu Komputer

Tempat bernaung mengemban semua ilmu untuk menjadi bekal hidup.

SANWACANA

Puji syukur kehadirat Allah *Subhanahu Wa Ta'ala* atas berkat rahmat dan hidayah-Nya, serta petunjuk dan pedoman dari Rasulullah Nabi Muhammad *Shollallahu Alaihi Wasallam* penulis dapat menyelesaikan skripsi yang berjudul “**Analisis Sentimen Terhadap Kinerja Pelayanan PT. Bank Rakyat Indonesia (Persero) Tbk Menggunakan Metode *Support Vector Machine*, *Naïve Bayes*, dan *K-Nearest Neighbors***” dengan baik dan lancar.

Selama proses penulisan skripsi ini tidak terlepas dari dukungan banyak pihak yang telah membimbing, membantu, dan memberi semangat kepada saya, sehingga pada kesempatan ini saya ingin menyampaikan ungkapan terima kasih kepada:

1. Allah *Subhanahu Wa Ta'ala* yang telah memberikan rahmat dan karunia-Nya selama menjalani penelitian hingga laporan dapat diselesaikan dengan baik.
2. (almh)Ibu dan Ayah tercinta, yang selalu memberi dukungan, memotivasi, dan menyemangati saya selama proses perkuliahaan sampai dengan penyusunan skripsi. Semoga Allah *Subhanahu Wa Ta'ala* selalu menyertai, memberkati, dan memberi kesehatan dan kebahagiaan yang berlimpah.
3. Bapak Dr. Eng. Suropto Dwi Yuwono, S.Si., M.T., selaku dekan Fakultas Matematika dan Ilmu Pengetahuan Alam.
4. Bapak Didik Kurniawan, S.Si., M.T., selaku Ketua Jurusan Ilmu Komputer Universitas Lampung.
5. Bapak Prof. Admi Syarif, Ph.D. selaku dosen pembimbing utama dalam penelitian ini yang selalu memberikan bimbingan, ilmu, dan saran.
6. Bapak Dr. rer. nat. Akmal Junaidi, M.Sc. selaku dosen pembahas satu, sekaligus Sekretaris Jurusan Ilmu Komputer yang telah memberikan ilmu dan saran dalam penelitian ini

7. Bapak Favorisen R. Lumbanraja, Ph.D. selaku dosen pembahas dua, yang telah memberikan ilmu dan saran dalam penelitian ini.
8. Seluruh Bapak dan Ibu dosen Jurusan Ilmu Komputer FMIPA Universitas Lampung, atas bimbingan dan pengajarannya selama penulis menjadi mahasiswa FMIPA Universitas Lampung.
9. Seluruh Staf dan karyawan Fakultas MIPA Universitas Lampung: Ibu Ade Nora Maela, Bang Zainuddin, Mas Syam, Mas Ardi Novalia, dan lainnya yang tidak bisa penulis sebutkan satu per satu, yang telah membantu segala urusan administrasi penulis.
10. Teman-teman grup *hangout*, yang telah memberikan banyak dukungan moril, segala bentuk bantuan, dan yang menemani dari awal perkuliahan.
11. Keluarga Ilmu Komputer 2019 serta kakak tingkat dan adik tingkat yang tidak bisa penulis sebutkan satu per satu.

Disadari masih banyak kekurangan dalam penulisan laporan skripsi ini yang disebabkan terbatasnya kemampuan, pengetahuan, dan pengalaman. Tetapi, semoga skripsi ini memberikan manfaat bagi yang pihak yang membaca.

Bandar Lampung, 01 Agustus 2023

Devi Ramadhia Fitri

NPM. 1917051005

DAFTAR ISI

Halaman

DAFTAR ISI	iii
DAFTAR TABEL	vi
DAFTAR GAMBAR	viii
I. PENDAHULUAN	1
1.1. Latar Belakang dan Masalah	1
1.2. Rumusan Masalah	5
1.3. Tujuan Penelitian.....	5
1.4. Batasan Masalah.....	5
1.5. Manfaat Penelitian.....	6
II. TINJAUAN PUSTAKA	7
2.1. <i>Text Mining</i>	7
2.2. <i>Machine Learning</i>	8
2.3. Analisis Sentimen.....	9
2.4. X.....	10
2.5. <i>Jupyter Notebook</i>	11
2.6. <i>Python</i>	11
2.6.1. Tweepy.....	12
2.6.2. Pandas	12
2.6.3. NumPy.....	12
2.6.4. Matplotlib.....	12
2.6.5. Regular Expression	12
2.6.6. CSV	12
2.6.7. String.....	13
2.6.8. Snsrape	13

2.6.9.	NLTK	13
2.6.10.	Scikit-Learn (Sklearn)	13
2.7.	<i>Support Vector Machine (SVM)</i>	13
2.8.	<i>Naïve Bayes</i>	15
2.9.	<i>K-Nearest Neighbors</i>	17
2.10.	Pembobotan <i>Term Frequency-Inverse Document Frequency</i>	19
2.11.	<i>Cross Validation</i>	19
2.11.1.	<i>K-fold Cross Validation</i>	20
2.11.2.	<i>Leave-One-Out Cross Validation</i>	20
2.11.3.	<i>Hold-out Cross Validation</i>	21
2.12.	<i>Synthetic Minority Over-sampling Technique (SMOTE)</i>	22
2.13.	Pelayanan Bank Rakyat Indonesia	23
2.14.	Penelitian Terdahulu	24
III. METODOLOGI PENELITIAN		30
3.1.	Waktu dan Tempat Penelitian	30
3.2.	Perangkat Penelitian	30
3.2.1.	Perangkat Keras	30
3.2.2.	Perangkat Lunak	30
3.3.	Alur Penelitian	31
3.3.1.	Identifikasi Masalah	32
3.3.2.	Studi literatur	32
3.3.3.	Pengumpulan Data	33
3.3.4.	Pelabelan Data	33
3.3.5.	<i>Preprocessing Data</i>	34
3.3.6.	Pembobotan Kata	36
3.3.7.	Pembagian <i>Data</i>	36
3.3.8.	<i>Synthetic Minority Over-sampling Technique (SMOTE)</i>	37
3.3.9.	Klasifikasi Sentimen	37
3.3.10.	Evaluasi	38
IV. PEMBAHASAN		41
4.1.	Pengumpulan Data	41
4.2.	Pelabelan sentimen	43
4.3.	<i>Preprocessing</i>	44

4.3.1.	<i>Cleaning</i>	44
4.3.2.	<i>Case Folding</i>	45
4.3.3.	<i>Tokenizing</i>	45
4.3.4.	Normalisasi	45
4.3.5.	<i>Stopword Removal</i>	46
4.3.6.	<i>Stemming</i>	46
4.4.	<i>Wordcloud</i>	46
4.5.	Pembobotan TF-IDF.....	50
4.6.	Pembagian Data.....	50
4.7.	<i>Synthetic Minority Over-sampling Technique (SMOTE)</i>	52
4.8.	Klasifikasi Sentimen.....	54
4.8.1.	Klasifikasi <i>Support Vector Machine</i>	54
4.8.2.	Klasifikasi <i>Naïve Bayes</i>	55
4.8.3.	Klasifikasi <i>K-Nearest Neighbors</i>	56
4.9.	Evaluasi Hasil Klasifikasi.....	58
4.10.	Perbandingan Hasil Klasifikasi.....	64
V.	SIMPULAN DAN SARAN	71
5.1.	Simpulan.....	71
5.2.	Saran.....	72
	DAFTAR PUSTAKA	73
	LAMPIRAN	78

DAFTAR TABEL

Tabel	Halaman
1. Jenis <i>kernel Support Vector Machine</i> (SVM).....	15
2. Penelitian Terdahulu	25
3. <i>Confusion Matrix</i>	39
4. Jumlah Data X.....	42
5. Data Pada Penelitian Setiap Bulan.....	42
6. Sentimen Positif.....	43
7. Sentimen Negatif	44
8. Contoh <i>Cleaning</i>	45
9. Contoh <i>Case Folding</i>	45
10. Contoh <i>Tokenizing</i>	45
11. Contoh Normalisasi	46
12. Contoh <i>Stopword Removal</i>	46
13. Contoh <i>Stemming</i>	46
14. Visualisasi <i>Wordcloud</i> dengan Sentimen Negatif Setiap Bulan.....	47
15. Visualisasi <i>Wordcloud</i> dengan Sentimen Positif Setiap Bulan.....	48
16. Perbandingan Perhitungan TF-IDF	50
17. Skema Pembagian Data	51
18. Pembagian Data Latih dan Data Uji Pada <i>Dataset Inbalanced</i>	51
19. Sampel Sintesis Menggunakan Metode SMOTE	53
20. Pembagian <i>Dataset Balanced</i>	53
21. Hasil Pengujian Dengan <i>Kernel SVM</i>	55
22. Hasil Pengujian Nilai Parameter <i>Naïve Bayes</i>	56
23. Hasil Pengujian Parameter KNN	57
24. Kinerja SVM Dengan <i>Dataset Inbalanced</i>	58
25. <i>Confusion Matrix SVM</i> Pada <i>Dataset Inbalanced</i>	59

26. Kinerja SVM Dengan <i>Dataset Balanced</i>	59
27. <i>Confusion Matrix</i> SVM Pada <i>Dataset Balanced</i>	60
28. Kinerja <i>Naïve Bayes</i> Dengan <i>Dataset Inbalanced</i>	60
29. <i>Confusion Matrix Naïve Bayes</i> Pada <i>Dataset Inbalanced</i>	61
30. Kinerja <i>Naïve Bayes</i> Dengan <i>Dataset Balanced</i>	61
31. <i>Confusion Matrix Naïve Bayes</i> Pada <i>Dataset Balanced</i>	62
32. Kinerja KNN Dengan <i>Dataset Inbalanced</i>	62
33. <i>Confusion Matrix</i> KNN Pada <i>Dataset Inbalanced</i>	63
34. Kinerja KNN Dengan <i>Dataset Balanced</i>	63
35. <i>Confusion Matrix</i> KNN Pada <i>Dataset Balanced</i>	64
36. Perbandingan Kinerja Klasifikasi SVM, <i>Naïve Bayes</i> , dan KNN	65
37. Perbandingan Akurasi Penelitian Terdahulu dan Penelitian Ini	69

DAFTAR GAMBAR

Gambar	Halaman
1. SVM Menentukan <i>Hyperplane</i> Terbaik (Pisner & Schnyer, 2019).....	14
2. Ilustrasi <i>K-Nearest Neighbors</i> (Nasution & Hayaty, 2019).....	17
3. Ilustrasi <i>K-Nearest Neighbors</i> K=3 (Nasution & Hayaty, 2019).....	18
4. Ilustrasi <i>K-Nearest Neighbors</i> K=5 (Nasution & Hayaty, 2019).....	18
5. Ilustrasi <i>K-Fold Cross Validation</i> (Ashfaque, 2019).....	20
6. Ilustrasi <i>Leave-One-Out Cross Validation</i> (Chlis, 2013).	21
7. Ilustrasi <i>Hold-out Cross Validation</i> (Chils, 2013).....	22
8. Alur Penelitian.	31
9. Dataset X Terkait Sentimen Positif dan Negatif.....	41
10. Perbandingan Sentimen Pada <i>Dataset</i>	52
11. Sebaran Data	54
12. Perbandingan Akurasi Setiap Metode.....	67
13. Perbandingan Presisi Setiap Metode.....	67
14. Perbandingan <i>Recall</i> Setiap Metode	68
15. Perbandingan <i>F1-score</i> Setiap Metode.....	69

I. PENDAHULUAN

1.1. Latar Belakang dan Masalah

Segala aspek kehidupan telah terpengaruh oleh kemajuan teknologi, terutama dalam hal digitalisasi. Digitalisasi menjadi pendukung bagi sumber daya manusia dalam menangani beragam jenis pekerjaan dengan lebih efisien guna meningkatkan kualitas layanan yang dihasilkan. Salah satu sektor yang terpengaruh adalah industri perbankan. Dalam mencapai kepuasan layanan nasabah, industri perbankan harus memberikan pelayanan yang baik. Kecepatan layanan menjadi hal yang sangat penting bagi bank dalam memberikan pelayanan kepada nasabah di era saat ini (Sari & Irhamah, 2020). Opini masyarakat terbagi menjadi emosi positif atau negatif, berdasarkan perasaan itu dapat digunakan sebagai evaluasi yang membantu perusahaan bank meningkatkan kualitas layanan kepada nasabah.

Industri perbankan adalah institusi keuangan yang berada di Indonesia. Perbankan memiliki tujuan utama yaitu menghimpun dana dari masyarakat dalam bentuk simpanan, kemudian menyediakan kembali kepada masyarakat dengan bentuk pinjaman (Koeswara & Muslimah, 2016). Salah satu lembaga perbankan yang menawarkan fasilitas-fasilitas tersebut ialah PT. Bank Rakyat Indonesia (Persero) Tbk. Dalam upaya memaksimalkan peran serta manfaatnya sebagai entitas keuangan dalam memenuhi kebutuhan masyarakat dan negara, kini jaringan cabang-cabang PT. Bank Rakyat Indonesia (Persero) Tbk. telah tersebar hampir di seluruh wilayah Indonesia. Kepuasan nasabah merupakan respon dari nasabah terhadap perbedaan antara harapan awal (atau standar kinerja lainnya) sebelum melakukan pembelian dan pengalaman nyata menggunakan atau mengonsumsi produk (Maulana & Muhajirin, 2021). Dalam memberikan pelayanan kepada nasabah antrian yang panjang kerap terjadi PT. Bank Rakyat Indonesia (Persero)

Tbk. (BRI) sehingga dapat menyebabkan komentar-komentar mengenai pelayanan. BRI belum terdapat analisis dalam menanggapi komentar nasabah seharusnya dibuatkan secara rutin analisis sentimen terhadap kinerja pelayanan.

X adalah *platform* media sosial yang sebelumnya disebut dengan nama Twitter, banyak pengguna aktif berbagi dan membaca X setiap menit. X tersebut mengandung data berharga yang mencerminkan situasi dan pandangan banyak individu. X telah menjadi tempat yang digunakan untuk berinteraksi, berbagi pendapat, ide, perasaan, dan perspektif mengenai peristiwa dan berbagai hal lainnya. X yang dihasilkan oleh pengguna X ditampung secara alfabetis dalam berbagai isu oleh X (Karo et al., 2022). Penggunaan X juga dapat memberikan wawasan mengenai pendapat nasabah terkait pelayanan PT Bank Rakyat Indonesia (Persero) Tbk.

Analisis sentimen merupakan bidang penelitian yang dirancang untuk mempelajari sikap, emosi, pendapat, evaluasi, dan penilaian orang terhadap suatu topik atau produk. Manfaat dari analisis sentimen adalah memperoleh informasi mengenai apakah suatu teks memiliki nilai positif atau negatif (Romadoni et al., 2020). Dalam konteks pelayanan perbankan, kualitas layanan kepada nasabah menjadi hal yang sangat penting. Oleh karena itu, penelitian ini dilakukan untuk menggali opini masyarakat mengenai pelayanan bank BRI melalui *platform* X." Data yang akan dianalisis adalah *tweet* dengan kata kunci "pelayanan BRI", "CS BRI", "Customer Service BRI", dan "KUR BRI". Pengumpulan data dilakukan dengan menggunakan *API key* dengan menggunakan *library Snsrape* yang tersedia pada X. Data tersebut akan dilakukan pengklasifikasian menjadi sejumlah kelas sentimen. Pada penelitian ini, menggunakan metode *Support Vector Machine* (SVM), *Naïve Bayes*, dan *K-Nearest Neighbors*.

Beberapa penelitian sebelumnya telah dilakukan dalam bidang analisis sentimen, termasuk salah satunya penelitian yang dilakukan oleh (Hakim et al., 2020). Pada penelitian tersebut, metode SMOTE digunakan untuk menjaga keseimbangan data *tweet* antara label positif dan negatif. Hasil penelitian menunjukkan akurasi *Naïve*

Bayes sebesar 66,81%, *Support Vector Machine* sebesar 80,05%, dan *K-Nearest Neighbors* sebesar 51,45%. Dalam penelitian ini, metode *Support Vector Machine* menunjukkan performa yang lebih baik dibandingkan kedua algoritma lainnya.

Penelitian terdahulu lainnya dalam analisis sentimen dilakukan oleh (Risnantoyo et al., 2020). Penelitian ini menggunakan metode *Support Vector Machine* (SVM), *K-Nearest Neighbors* (KNN), dan *Naïve Bayes* dengan menggunakan pembobotan TF-IDF. Selain itu, penelitian ini juga melakukan pengujian data dengan metode *K-fold Cross Validation*. Hasil penelitian menunjukkan bahwa akurasi pada metode SVM, KNN, dan *Naïve Bayes* adalah 76,21%, 71,89%, dan 75,81%. Algoritma SVM menunjukkan hasil yang lebih baik dibandingkan dengan algoritma KNN dan *Naïve Bayes*.

Penelitian ini menggunakan algoritma *Support Vector Machine* (SVM) untuk melakukan analisis sentimen terhadap data pelayanan di PT. Bank Rakyat Indonesia (Persero) Tbk. yang diperoleh melalui observasi di *platform X*. SVM merupakan metode klasifikasi yang efektif dalam membangun model prediktif dengan data yang terbatas. Metode ini memiliki keunggulan dalam mencari *hyperplane* terbaik untuk memisahkan dua kelas dalam ruang fitur, serta menerapkan strategi *Structural Risk Minimization* (SRM) untuk hasil yang lebih optimal (Sari & Irhamah, 2020).

Dalam penelitian ini, menggunakan algoritma *Naïve Bayes* sebagai *classifier* lain. *Naïve Bayes* adalah metode yang menggunakan pendekatan probabilitas untuk mengklasifikasikan data dengan memperhatikan hubungan statistik antara peristiwa yang terjadi. Metode ini berdasarkan aturan *Bayes* yang sederhana, yang mempertimbangkan kondisi dan peluang dari setiap kondisi. Keunggulan algoritma *Naïve Bayes* adalah kemampuannya dalam menyesuaikan proses klasifikasi dengan sifat dan kebutuhan data. Selain itu, metode ini membutuhkan jumlah data pelatihan yang relatif kecil untuk mengestimasi parameter yang diperlukan dalam proses pengklasifikasian (Sadewo et al., 2019).

Dalam penelitian ini juga, menerapkan algoritma *K-Nearest Neighbors* (KNN) yang banyak digunakan dalam klasifikasi teks dan data. KNN adalah metode klasifikasi yang mengelompokkan data baru berdasarkan atribut dan data latih yang ada. Prediksi dari sampel uji baru dalam KNN didasarkan pada mayoritas kategori tetangga terdekatnya. Data latih dibangun dengan memperhatikan keseimbangan dokumen satu sama lain (Sodik & Kharisudin, 2021). Keunggulan KNN terletak pada prinsip kerjanya yang sederhana, yaitu dengan menghitung jarak terpendek antara sampel data latih. Penelitian sebelumnya telah banyak dilakukan dalam analisis sentimen menggunakan KNN, salah satunya oleh (Ogi et al., 2021). Penelitian tersebut menjelaskan akurasi pada *dataset* OVO dengan algoritma *Naïve Bayes*, *Support Vector Machine*, dan *K-Nearest Neighbors* yaitu 81,00%, 82,00%, dan 76,00%. Akurasi pada dataset DANA dengan algoritma *Naïve Bayes*, *Support Vector Machine*, *K-Nearest Neighbors* yaitu 86,00%, 87,00%, dan 86,00%. Berdasarkan penelitian yang telah dilakukan bahwa *Support Vector Machine* lebih baik dibandingkan *Naïve Bayes*, dan *K-Nearest Neighbors*.

Dalam penelitian ini, akan menganalisis dan mengklasifikasikan (positif atau negatif) data hasil observasi pada X terkait kinerja pelayanan PT. Bank Rakyat Indonesia (Persero) Tbk. menggunakan tiga metode yaitu *Support Vector Machine* (SVM), *Naïve Bayes*, dan *K-Nearest Neighbors* (KNN) dalam analisis tersebut. Selanjutnya, akan membandingkan tingkat kinerja ketiga metode tersebut dalam penelitian ini. Berdasarkan konteks yang telah dijelaskan, judul penelitian ini adalah "Analisis Sentimen Terhadap Kinerja Pelayanan PT. Bank Rakyat Indonesia (Persero) Tbk. Menggunakan Metode *Support Vector Machine* (SVM), *Naïve Bayes*, dan *K-Nearest Neighbors* (KNN)". Penulis berharap hasil penelitian ini dapat membantu PT. Bank Rakyat Indonesia (Persero) Tbk. dalam meningkatkan dan memperbaiki pelayanannya.

1.2. Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, rumusan masalah pada penelitian ini adalah sebagai berikut.

- a. Bagaimana sentimen nasabah terhadap kinerja pelayanan PT. Bank Rakyat Indonesia (Persero) Tbk. pada X menggunakan algoritma *Support Vector Machine* (SVM), *Naïve Bayes*, dan *K-Nearest Neighbors* (KNN)?
- b. Bagaimana perbandingan kinerja Algoritma *Support Vector Machine* (SVM), *Naïve Bayes*, dan *K-Nearest Neighbors* (KNN) dalam melakukan analisis sentimen terhadap kinerja pelayanan PT. Bank Rakyat Indonesia (Persero) Tbk di X?

1.3. Tujuan Penelitian

Tujuan dari penelitian ini adalah sebagai berikut:

- a. Menerapkan algoritma *Support Vector Machine* (SVM), *Naïve Bayes*, dan *K-Nearest Neighbors* (KNN) dalam analisis sentimen terhadap kinerja pelayanan PT Bank Rakyat Indonesia (Persero) Tbk.
- b. Melakukan perbandingan dan evaluasi hasil akurasi, presisi, *recall*, dan *F1-score* menggunakan algoritma *Support Vector Machine* (SVM), *Naïve Bayes*, dan *K-Nearest Neighbors* (KNN) terhadap analisis sentimen.

1.4. Batasan Masalah

Penelitian ini memiliki beberapa batasan yang telah ditetapkan sebagai berikut:

- a. Penelitian ini berfokus pada studi kasus pelayanan PT. Bank Rakyat Indonesia (Persero) Tbk. Data *tweet* yang digunakan terbatas pada bahasa Indonesia.

- b. Pengambilan data *tweet* dilakukan menggunakan X API sebagai *library* untuk mengumpulkan data.
- c. Pengolahan data dilakukan menggunakan bahasa pemrograman *Python*.
- d. Algoritma yang digunakan dalam penelitian ini meliputi *Support Vector Machine* (SVM), *Naïve Bayes*, dan *K-Nearest Neighbors* (KNN).

1.5. Manfaat Penelitian

Manfaat penelitian ini adalah sebagai berikut:

- a. Dapat mengetahui performa algoritme *Support Vector Machine* (SVM), *Naïve Bayes*, dan *K-Nearest Neighbors* (KNN) dalam analisis sentimen yang mengklasifikasikan data mengenai opini positif, dan negatif pada X terhadap kinerja pelayanan PT. Bank Rakyat Indonesia (Persero) Tbk.
- b. Dapat membantu PT. Bank Rakyat Indonesia (Persero) Tbk. untuk meningkatkan dan memperbaiki dalam memberikan layanan kepada nasabah.

II. TINJAUAN PUSTAKA

2.1. *Text Mining*

Text mining adalah proses penggalian informasi di mana pengguna berinteraksi dengan sejumlah dokumen menggunakan alat analisis yang merupakan komponen dalam *data mining*. Tujuannya adalah untuk menemukan informasi atau tren baru yang sebelumnya tidak terungkap dengan menganalisis dan memproses data dalam jumlah besar. Dalam menganalisis teks yang tidak terstruktur, *text mining* mencoba untuk menghubungkan bagian teks satu dengan yang lain berdasarkan aturan tertentu (Ariyanti & Iswardani, 2020). Teknik *text mining* digunakan untuk menggali data guna pemenuhan kebutuhan informasi dengan menggunakan metode *data mining*, *machine learning*, *natural language processing*, pencarian informasi, dan manajemen pengetahuan. *Text mining* melibatkan proses awal dokumen seperti kategorisasi teks, ekstraksi informasi, dan ekstraksi kata. Pendekatan ini diimplementasikan untuk mengambil informasi dari berbagai sumber data dengan mengidentifikasi dan mengeksplorasi pola-pola menarik yang ada (Firdaus & Firdaus, 2021).

Text mining merupakan teknologi yang diterapkan untuk menganalisis informasi yang tidak terstruktur dalam bentuk teks melibatkan pendekatan khusus. Proses analisis melalui *text mining* memiliki dua tahap inti. Pertama, tahap pra-pemrosesan dan penggabungan berbagai data yang tak terstruktur. Kedua, dilanjutkan dengan analisis statistik atas data yang telah melewati tahap pra-pemrosesan tersebut. Tujuannya adalah untuk mengekstraksi esensi dari teks yang ada. *Text mining* merubah informasi yang tidak teratur menjadi informasi yang teratur dengan mengonversi teks menjadi representasi numerik (Pakpahan & Widyastuti, 2014). *Text mining* merupakan bidang khusus dalam *data mining* yang fokus pada

ekstraksi informasi dari teks. Informasi diambil dari dokumen sebagai sumber data, dengan tujuan mengidentifikasi frasa-frasa yang mencerminkan substansi dokumen. Tindakan ini bertujuan untuk memungkinkan analisis keterkaitan antara dokumen-dokumen tersebut. biasanya berupa dokumen, dan tujuannya adalah mencari kata-kata yang mewakili isi dokumen sehingga dapat dilakukan analisis hubungan antara dokumen-dokumen tersebut. Teknik *text mining* dapat menganalisis dokumen, mengelompokkan dokumen berdasarkan kata-kata yang terkandung di dalamnya, serta menentukan kesamaan antara dokumen untuk memahami hubungannya dengan variabel lain. Contoh penerapan *text mining* yang umum meliputi penyaringan spam, analisis sentimen, pengukuran preferensi pelanggan, ringkasan dokumen, pengelompokan topik penelitian, dan lain sebagainya (Fathonah & Herliana, 2021).

Proses *text mining* melibatkan ekstraksi dan analisis data tidak terstruktur dalam jumlah besar menggunakan perangkat lunak. Metode ini mencakup identifikasi konsep, pola, tema, kata kunci, dan atribut lain dalam data. Analisis teks merupakan salah satu teknik dalam *text mining* yang digunakan untuk menyortir kumpulan data. *Text mining* berhubungan dengan bidang ilmu seperti *information retrieval*, *data mining*, *machine learning*, statistika, dan linguistik komputasi. *Text mining* digunakan untuk mencari pola-pola dalam teks berbahasa alami yang tidak terstruktur, seperti buku, surel, artikel, situs web, dan lain-lain. Mengingat ketidakteraturan struktur data teks, proses *text mining* membutuhkan tahap awal yang melibatkan persiapan agar teks dapat diubah menjadi lebih terstruktur (Sari & Irhamah, 2020). Berdasarkan definisi-definisi di atas, berdasarkan informasi yang ada, dapat diambil kesimpulan bahwa *text mining* adalah langkah pencarian teks yang bertujuan mendapatkan wawasan berharga dari data.

2.2. Machine Learning

Machine learning merupakan salah satu aspek dalam kecerdasan buatan (*artificial intelligence*) yang memungkinkan komputer untuk belajar dan berperilaku secara

cerdas seperti manusia. Tujuannya adalah memungkinkan komputer untuk meningkatkan pemahamannya melalui pengalaman dan dapat membuat keputusan tanpa perlu diprogram secara eksplisit oleh manusia. Hal ini memungkinkan mesin untuk beradaptasi dengan perubahan yang terjadi. *Machine learning* bekerja dengan menganalisis data yang tersedia sebagai *input*, terutama pada kumpulan data besar (*big data*), untuk menemukan pola-pola tertentu. Data digunakan sebagai *input* untuk melatih algoritma dalam *machine learning* sehingga mesin dapat menghasilkan analisis yang akurat. Dalam *machine learning*, terdapat istilah "*data training*" dan "*data testing*". *Data training* digunakan untuk melatih algoritma, sedangkan *data testing* digunakan untuk menguji performa algoritma ketika dihadapkan pada data baru yang belum pernah diberikan pada tahap pelatihan (Retnoningsih & Pramudita, 2020). Teknik *machine learning* berdasarkan teknik pembelajarannya, terdapat beberapa tipe *machine learning*, antara lain *supervised learning*, *unsupervised learning*, *semi-supervised learning*, dan *reinforcement learning*. *Supervised learning* menggunakan *dataset* yang sudah memiliki label (*labeled data*) untuk melatih mesin sehingga mesin dapat mengidentifikasi label *input* berdasarkan fitur-fitur yang ada dan melakukan prediksi atau klasifikasi. Di sisi lain, *unsupervised learning* adalah teknik yang menganalisis *dataset* tanpa adanya *label response*, dan berusaha untuk menarik kesimpulan atau menemukan struktur tersembunyi dalam data (Retnoningsih & Pramudita, 2020).

2.3. Analisis Sentimen

Analisis sentimen merupakan metode komputasi yang berfokus pada pendapat, sentimen, emosi, dan penilaian dari berbagai individu terhadap entitas seperti layanan, produk, organisasi, masalah, peristiwa, dan topik lainnya (Anesca et al., 2021). Terdapat dua pendekatan dalam analisis sentimen, yaitu pelabelan otomatis yang menggunakan kamus kata-kata positif dan negatif untuk mencocokkan dengan kalimat yang akan dianalisis, dan pelabelan kalimat secara manual (Habibi et al., 2016). Analisis sentimen bertujuan untuk menemukan informasi berharga dari data yang tidak terstruktur. Pendekatan ini melibatkan Penerapan pemrosesan bahasa alami, evaluasi teks, studi linguistik komputasional, dan teknologi biometrik guna

sistematisasi pengenalan, pengestrakan, pengukuran, serta penelitian terhadap aspek afektif dan informasi subjektif (Giovani et al., 2020).

Proses analisis sentimen melibatkan klasifikasi polaritas teks pada tingkat dokumen, kalimat, atau fitur, untuk menentukan apakah opini yang diungkapkan dalam dokumen, kalimat, atau fitur tersebut bersifat positif atau negatif. Jenis analisis sentimen yang umum meliputi analisis sentimen berbasis aspek, analisis sentimen penilaian (positif dan negatif), serta analisis sentimen multibahasa dan deteksi emosi (Romadoni et al., 2020). Dalam penelitian ini, digunakan teknik berbasis pengetahuan (*knowledge-based techniques*) dan metode statistik (*statistical methods*) untuk membangun model sentimen, karena kedua teknik ini mampu melakukan kategorisasi sentimen dengan mempertimbangkan kemunculan kata-kata dan menganalisis struktur bahasa. Alur proses analisis sentimen dimulai dengan memasukkan dokumen berupa *tweet*, kemudian dilakukan analisis sentimen dengan menggunakan teknik *sentiment analysis* untuk menghasilkan keluaran sentimen positif dan negatif. Analisis sentimen, yang juga dikenal sebagai *opinion mining*, digunakan secara luas dalam berbagai konteks seperti ulasan dan tanggapan survei, media online dan sosial, serta dalam bidang perawatan kesehatan, termasuk pemasaran, layanan pelanggan, dan kedokteran klinis (Giovani et al., 2020). Tugas dasar dalam analisis sentimen adalah mengklasifikasikan polaritas teks pada tingkat dokumen, kalimat, atau fitur, untuk menentukan apakah opini yang diungkapkan dalam dokumen, kalimat, atau fitur tersebut bersifat positif atau negatif (Giovani et al., 2020).

2.4. X

X adalah sebuah *platform* media sosial yang sebelumnya bernama Twitter. X setelah diakuisisi oleh X Corp memiliki tujuan memastikan kebebasan berbicara dan akselerator. Twitter resmi berganti nama menjadi X pada tanggal 25 Juli 2023. X adalah sosial yang menggunakan model *microblogging* (Novantirani et al., 2015). X adalah salah satu *platform* media sosial yang populer dan berada di peringkat lima teratas. Xs ini mengandung informasi berharga yang mencerminkan situasi

yang melibatkan banyak orang. X juga telah menjadi tempat di mana orang dapat bertukar pendapat, ide, perasaan, dan perspektif tentang peristiwa dan topik yang beragam (Karo et al., 2022). Terdapat beberapa karakteristik unik pada X. Pertama, X memiliki batasan panjang karakter maksimum sebesar 140 karakter untuk setiap *tweet*. Pengguna X menggunakan berbagai model bahasa dan media untuk memposting pesan. Frekuensi kesalahan pengejaan, penggunaan bahasa gaul, dan penggunaan singkatan juga lebih tinggi di X dibandingkan dengan media sosial lainnya (Habibi et al., 2016). Dalam penelitian ini, X digunakan sebagai sumber informasi atau data. Pengambilan informasi dilakukan melalui X API, yang merupakan perpustakaan pemrograman yang digunakan untuk mengakses dan mengambil informasi dari X.

2.5. *Jupyter Notebook*

Jupyter Notebook adalah sebuah aplikasi pembuatan dokumen interaktif yang terdiri dari kode komputer dan elemen rich text seperti paragraf, persamaan matematika, Gambar, dan tautan. Dokumen yang dihasilkan oleh *Jupyter Notebook App* memiliki ekstensi file *ipynb*. *Jupyter Notebook* adalah sebuah aplikasi *open source* yang digunakan untuk membuat dan berbagi dokumen yang menggunakan bahasa pemrograman *Python* (Retnoningsih & Pramudita, 2020). Dalam penelitian ini, *Jupyter Notebook* digunakan sebagai alat untuk membuat *script program*.

2.6. *Python*

Python adalah bahasa pemrograman yang digunakan oleh organisasi dan pengembang untuk mengembangkan aplikasi berbasis desktop, web, dan mobile. Bahasa ini diciptakan oleh Guido Van Rossum pada tahun 1990 di Belanda. *Python* mendukung implementasi perintah multi-fungsi secara langsung dengan menggunakan paradigma pemrograman berorientasi objek dan memiliki semantik dinamis yang memungkinkan tingkat keterbacaan yang tinggi dalam penulisan sintaks (Muhammad Romzi & Kurniawan, 2020). Bahasa pemrograman *Python*

dapat dijalankan di berbagai sistem operasi seperti *Linux*, *Microsoft Windows*, *Mac OS*, *Android*, *Symbian OS*, *Amiga*, dan *Palm* (Muhammad Romzi & Kurniawan, 2020). Dalam penelitian ini, penulis menggunakan *Python* versi 3.10 serta menggunakan berbagai pustaka (*library*) sebagai berikut:

2.6.1. Tweepy

Tweepy adalah sebuah pustaka (*library*) *Python* yang digunakan untuk mengambil data dari X menggunakan X API.

2.6.2. Pandas

Pandas adalah sebuah pustaka (*library*) *Python* yang digunakan untuk analisis data. Salah satu fitur yang sering digunakan dalam *Pandas* adalah data wrangling, yaitu proses mengubah data mentah menjadi format yang lebih terstruktur. Data yang telah diproses dalam bentuk CSV/TSV atau SQL dapat diubah menjadi objek *Python* dengan baris dan kolom yang disebut *DataFrame*.

2.6.3. NumPy

NumPy adalah sebuah pustaka (*library*) pada *Python* yang digunakan untuk melakukan komputasi saintifik, termasuk operasi aritmatika.

2.6.4. Matplotlib

Matplotlib adalah sebuah pustaka (*library*) *Python* yang digunakan untuk mempresentasikan hasil analisis dalam bentuk grafik dan chart.

2.6.5. Regular Expression

Regular Expression adalah sebuah pustaka (*library*) *Python* yang digunakan untuk manipulasi *string*, sehingga memungkinkan untuk membersihkan data.

2.6.6. CSV

CSV (Comma Separated Value) adalah sebuah pustaka (*library*) *Python* yang digunakan untuk menyimpan data dalam format ".csv".

2.6.7. String

String adalah sebuah pustaka (*library*) yang digunakan dalam proses pembersihan (*cleaning*) dan transformasi *string*.

2.6.8. Snsrape

Snsrape adalah sebuah program yang digunakan untuk menyaring data dari layanan media sosial, seperti mencari profil pengguna, tagar, atau melakukan pencarian tertentu.

2.6.9. NLTK

NLTK (*Natural Language Tool Kit*) adalah sebuah pustaka (*library*) yang digunakan dalam pemrosesan teks, termasuk pembersihan data, tokenisasi, penghapusan kata yang tidak penting (*stopword removal*), dan lain-lain.

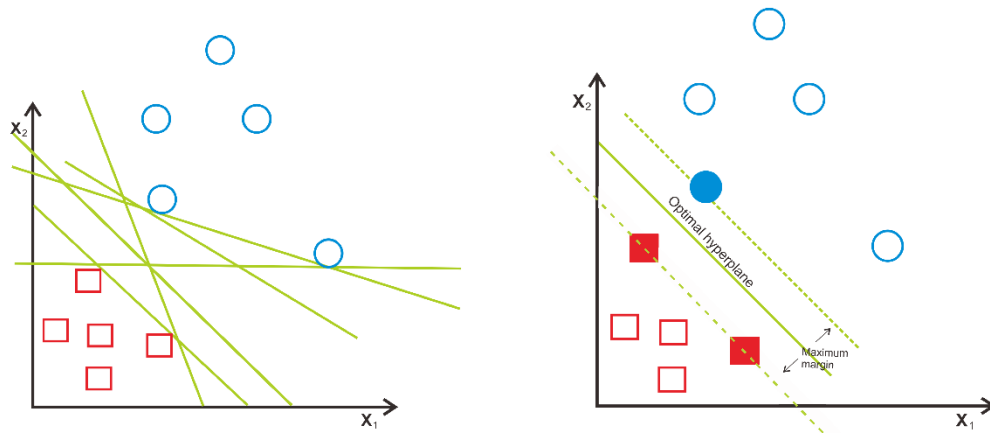
2.6.10. Scikit-Learn (Sklearn)

Scikit-Learn (Sklearn) adalah sebuah pustaka (*library*) yang membantu dalam penerapan model *machine learning*.

2.7. *Support Vector Machine (SVM)*

Support Vector Machine (SVM) untuk pertama kalinya pada tahun 1992 Vanik memperkenalkan sebagai suatu kumpulan rancangan yang unggul dalam bidang pengenalan pola. Dasar teori SVM salah satunya adalah konsep *margin hyperplane* (Duda & Hart tahun 1973, Cover tahun 1965, Vapnik tahun 1964, dan lain-lain) serta pengenalan *kernel* oleh Aronszajn pada tahun 1950. SVM merupakan algoritma supervised *machine learning* yang sering digunakan untuk prediksi dan klasifikasi (Pisner & Schnyer, 2019). *Support Vector Machine (SVM)* memiliki dasar teori yang mencari fungsi pemisah (*hyperplane*) optimal di antara fungsi yang ada untuk memisahkan dua kelas. Fungsi *hyperplane* dianggap baik jika memiliki *margin* terbesar, yaitu dua kali jarak antara *hyperplane* dan *support vector*, yang merujuk pada titik terdekat dengan *hyperplane* (Sari & Irhamah, 2020). Gambar 1 menjelaskan ilustrasi *Support Vector Machine* Mencari *hyperplane* terbaik dengan

cara memisahkan dua buah *class* pada ruang fitur. Apabila data input memiliki kemampuan untuk dipisahkan secara linear, pemisahan mengenai *hyperplane*.



Gambar 1. SVM Menentukan *Hyperplane* Terbaik (Pisner & Schnyer, 2019).

Pencarian *hyperplane* terbaik sebagai pemisah antara kedua kelas melibatkan alternatif garis pemisah (*discrimination boundaries*) yang dapat diukur menggunakan *margin hyperplane* dan mencari titik maksimumnya. *Margin* adalah selisih jarak antara *hyperplane* dan pola terdekat dari setiap kelas, yang dikenali sebagai *support vector*. Gambar 1, garis kontinu menggambarkan *hyperplane* teroptimal yang terletak ditengah kedua kelas, sementara objek persegi berwarna merah, dan bulat berwarna biru adalah contoh *support vector*. Proses pembelajaran pada *Support Vector Machine* (SVM) melibatkan upaya mencari lokasi *hyperplane* ini (Pisner & Schnyer, 2019).

Dalam pemodelan menggunakan metode *Support Vector Machine* (SVM) untuk mencapai akurasi terbaik, digunakan fungsi klasifikasi *Support Vector Machine* (SVM) dengan *impor Support Vector Classification* (SVC) dalam *Python*. SVM adalah teknik *supervised learning* yang dimanfaatkan untuk klasifikasi, regresi, dan deteksi outlier. Penerapan SVM didasari oleh akurasi yang tinggi dalam menganalisis teks. *Kernel* adalah kumpulan fungsi matematika yang diterapkan dalam SVM, berperan dalam mentransformasi data *input* ke dalam dimensi yang lebih tinggi (*kernel space*). Beberapa *kernel* yang populer dalam SVM adalah *Linear*, *Polynomial*, *Sigmoid*, dan *Radial Basis Function* (RBF) (Pisner & Schnyer, 2019). Sebuah fungsi dianggap sebagai fungsi *kernel* jika memenuhi Teorema

Mercer yang menyatakan bahwa matriks *kernel* yang dihasilkan harus positif *semi-definite* (Muis et al., 2015). Jenis-jenis *kernel Support Vector Machine* (SVM) dapat dilihat pada Tabel 1:

Tabel 1. Jenis *kernel Support Vector Machine* (SVM)

Jenis <i>kernel</i>	Definisi
<i>Linear</i>	$K(\vec{x}_i, \vec{x}_j) = \vec{x}_i^t \vec{x}_j$
<i>Polynomial</i>	$K(\vec{x}_i, \vec{x}_j) = (\vec{x}_i, \vec{x}_j + 1)^p$
RBF	$K(\vec{x}_i, \vec{x}_j) = \exp - \frac{\ \vec{x}_i - \vec{x}_j\ ^2}{2\sigma^2}$
<i>Sigmoid</i>	$K(\vec{x}_i, \vec{x}_j) = \tanh(a\vec{x}_i, \vec{x}_j + \beta)$

Data yang tidak dapat dikelompokkan secara *linear* dapat diatasi dengan mentransformasikan data ke dalam ruang fitur (*feature space*). Biasanya, *feature space* memiliki dimensi yang lebih tinggi daripada vektor *input* (*input space*). Namun, hal ini dapat menyebabkan komputasi yang rumit karena jumlah fitur yang besar dan sulitnya mengetahui fungsi transformasi yang tepat. Untuk mengatasi masalah ini, *Support Vector Machine* (SVM) menggunakan *kernel trick* (Muis et al., 2015). Dalam penelitian ini, metode *Support Vector Machine* (SVM) digunakan dengan *kernel Radial Basis Function* (RBF) untuk pengolahan data.

2.8. *Naïve Bayes*

Metode *Naïve Bayes* adalah suatu pendekatan berbasis statistik dan probabilitas yang diperkenalkan oleh Thomas Bayes. Dalam *Naïve Bayes*, prediksi atas data masa depan ditentukan oleh pengalaman dari masa lalu. Ciri utama yang mencolok dari pendekatan ini adalah asumsi yang sangat sederhana (naif) terkait independensi antara setiap peristiwa. Dalam konteks teknik klasifikasi, metode *Naïve Bayes* umumnya diterapkan pada pemrosesan data X, dengan variasi-variasi seperti *Unigram Naïve Bayes*, *Multinomial Naïve Bayes*, *Bernoulli Naïve Bayes*, *Complement Naïve Bayes*, serta *Maximum Entropy Classification*. Kelebihan yang ditemukan pada metode *Naïve Bayes* adalah kemampuannya menghasilkan hipotesis yang signifikan dari segala kondisi atau peristiwa. (Verawardina et al.,

2021). Perhitungan probabilitas dalam *Naïve Bayes* dilakukan pendekatan algoritma *Bayes* menggunakan Persamaan (1):

$$P(Y|X) = \frac{P(X|Y)P(Y)}{P(X)} \quad (1)$$

Persamaan (1) menunjukkan variabel Y yang mewakili kelas tertentu, X menunjukkan data pada dengan kelas yang belum diidentifikasi, $P(Y|X)$ adalah probabilitas hipotesis Y berdasarkan kondisi X , sementara $P(Y)$ adalah hipotesis Y dan $P(X|Y)$ merupakan probabilitas X berdasarkan kondisi hipotesis Y , dan $P(X)$ merupakan probabilitas X .

Dalam klasifikasi *Naïve Bayes*, kemudian dilakukan pengembangan sehingga menjadi Persamaan (2).

$$\begin{aligned} P(Y + X_1, X_2, \dots, X_n) &= \frac{P(X_1, X_2, \dots, X_n)P(Y)}{P(X_1, X_2, \dots, X_n)} \\ &= \frac{P(X_1|Y)P(X_2|Y) \dots P(X_n|Y)P(Y)}{P(X_1, X_2, \dots, X_n)} \end{aligned} \quad (2)$$

Dimana $(Y|X_1, X_2, \dots, X_n)$ merupakan hasil perhitungan dari seluruh probabilitas posterior terhadap nilai X untuk seluruh nilai di Y . Sehingga, *Naïve Bayes* akan menghasilkan prediksi berdasarkan probabilitas maksimum dari probabilitas *posterior* dengan memanfaatkan metode *Laplace*, di mana c merupakan jumlah nilai dalam Y , seperti yang ditunjukkan dalam persamaan tersebut.

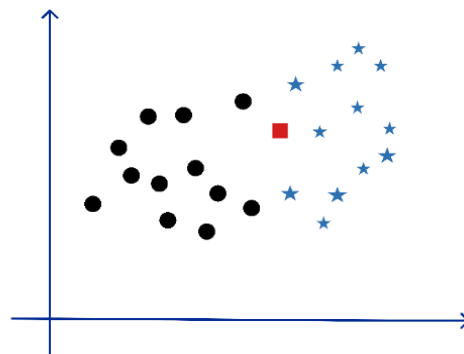
$$P(X_i|Y) = \frac{N_{ic} + 1}{N_c + c} \quad (3)$$

Pada metode teknik *opinion mining* menggunakan data dari X , metode klasifikasi *Naïve Bayes* digunakan. Pada tahap ini, model dibangun dan digunakan untuk mengklasifikasikan sentimen dengan akurasi yang tinggi (Verawardina et al., 2021). Data yang digunakan dalam penelitian ini merupakan sejumlah pesan yang diambil dari platform X , kemudian disimpan dalam format CSV. Terdapat dua

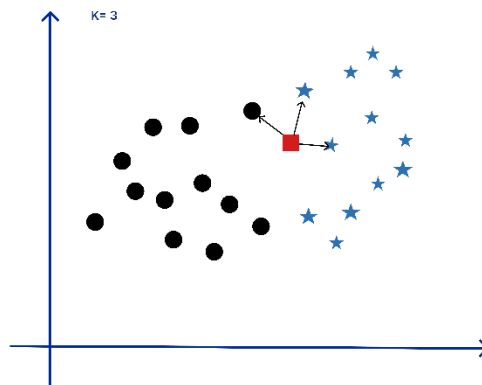
dataset yaitu data pelatihan dan data pengujian. Pemberian label pada data yaitu sentimen positif dan negatif. Metode *Naive Bayes* digunakan dalam tahap klasifikasi sentimen dan interpretasi hasil analisis sentimen (Verawardina et al., 2021).

2.9. *K-Nearest Neighbors*

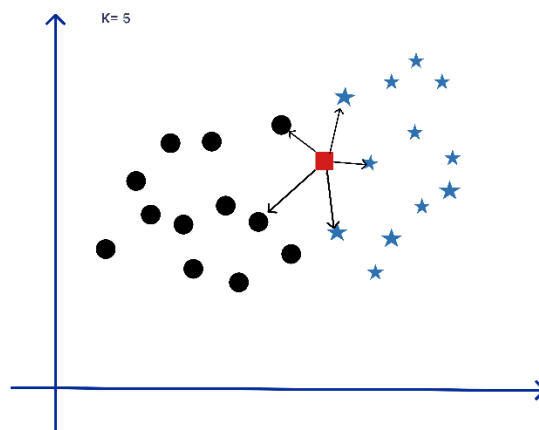
K-Nearest Neighbors adalah algoritma yang banyak digunakan untuk melakukan klasifikasi teks dan data. Cara kerja algoritma *K-Nearest Neighbors* dengan menentukan nilai K , kemudian hitung jarak data baru dengan *data training* untuk menghitung jarak, 3 metrik jarak yang biasa digunakan adalah *Euclidean Distance*, *Manhattan Distance*, dan *Minkowsky Distance*. Setelah menghitung jarak, selanjutnya mencari *K-Neighbors* yang dekat dengan data baru tersebut, dalam memprediksi label, pilih kelas data latih yang paling dekat dengan data baru dan memiliki jumlah terbanyak. Prinsip sederhana metode ini adalah “Sebuah data yang akan diprediksi apakah tergolong kelas positif atau negatif” (Nasution & Hayaty, 2019). Gambar 2 mengilustrasikan metode klasifikasi *K-Nearest Neighbors*.



Gambar 2. Ilustrasi *K-Nearest Neighbors* (Nasution & Hayaty, 2019).



Gambar 3. Ilustrasi *K-Nearest Neighbors* $K=3$ (Nasution & Hayaty, 2019).



Gambar 4. Ilustrasi *K-Nearest Neighbors* $K=5$ (Nasution & Hayaty, 2019).

Gambar 3 dan 4 menggambarkan metode klasifikasi. Dalam gambar tersebut, kotak berwarna merah mewakili data yang akan diprediksi ke dalam salah satu kelas. Kelas berwarna hitam mengindikasikan kelas negatif, sementara kelas berwarna biru mengindikasikan kelas positif. Dalam kasus $k=3$, kemungkinan objek akan masuk ke dalam kelas biru (kelas positif) karena perhitungan menunjukkan adanya perbedaan yang lebih besar, yaitu $2-1$, dibandingkan dengan kelas hitam. Pada kasus $k=5$, objek juga akan diprediksi masuk ke dalam kelas biru karena perhitungan menunjukkan perbedaan yang lebih besar, yaitu $3-2$, dibandingkan dengan kelas hitam (Nasution & Hayaty, 2019).

Dalam penelitian ini, data yang digunakan memiliki label, sehingga memudahkan dalam proses pengelompokan ke dalam kelas yang paling sesuai. Keunggulan metode ini adalah kemampuannya untuk mengklasifikasikan data berdasarkan

tetangga. Selain itu, hasil dan akurasi prediksi dapat dengan mudah diinterpretasikan, dengan memperhatikan nilai k terdekat secara akurat sebelumnya. *K-Nearest Neighbors* akan dihitung memakai jarak *Euclidean* dengan Persamaan (4).

$$\begin{aligned}
 d(x, y) &= \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 \dots + (x_n - y_n)^2} \\
 &= \sqrt{\sum_{i=1}^k (x_i - y_i)^2} \tag{4}
 \end{aligned}$$

Dengan $d(x, y)$ sebagai jarak antar variabel x_1 dan y_2 , x sebagai variabel dan k melambangkan total jumlah variabel (Furqan et al., 2022).

2.10. Pembobotan *Term Frequency-Inverse Document Frequency*

TF-IDF adalah sebuah algoritma pembobotan yang terdiri dari dua nilai dalam dua algoritma yang berbeda, yaitu *Term Frequency* (TF) dan *Inverse Document Frequency* (IDF). Algoritma ini berfungsi dengan menghitung jumlah kemunculan sebuah term dalam sebuah dokumen (TF) dan *Inverse Document Frequency* dari term tersebut (IDF) (Risnantoyo et al., 2020). Metode TF-IDF digunakan untuk memberikan bobot pada hubungan antara kata (term) dan dokumen. Setelah *dataset* melalui tahap *preprocessing*, dilakukan proses pembobotan. Tujuan dari tahap ini adalah untuk mendapatkan nilai dari kata dasar yang telah diekstraksi. Kata-kata dasar tersebut akan diubah menjadi vektor yang merepresentasikan kata tersebut sehingga dapat dikenali oleh sistem (Nurjannah & Astuti, 2013). Penelitian ini, menerapkan metode *Term Frequency-Inverse Document Frequency* (TF-IDF).

2.11. *Cross Validation*

Cross-validation adalah suatu metode untuk mengevaluasi model dan menguji generalisasi dari kumpulan data independen. Metode ini membagi data menjadi dua bagian, yaitu data pelatihan dan data validasi, yang digunakan untuk menguji model (Watanabe, 2009). Terdapat beberapa jenis *cross-validation*, antara lain:

2.11.1. *K-fold Cross Validation*

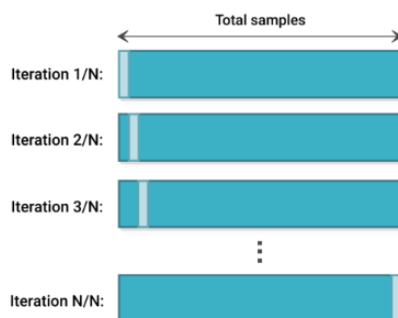
K-fold Cross Validation adalah metode yang digunakan untuk memperkirakan tingkat kesalahan dalam klasifikasi. Data dibagi menjadi k bagian dengan ukuran yang sama, kemudian dilakukan iterasi di mana setiap iterasi menggunakan bagian yang berbeda sebagai data validasi, sedangkan sisa bagian digunakan sebagai data pelatihan. Tujuan dari *K-fold Cross Validation* adalah untuk mengurangi bias dalam memilih data secara acak (Sari & Irhamah, 2020). Gambar 5 mengilustrasikan *k-fold cross validation* tersebut dengan memperhitungkan kinerja klasifikasi.



Gambar 5. Ilustrasi *K-Fold Cross Validation* (Ashfaque, 2019).

2.11.2. *Leave-One-Out Cross Validation*

Leave-One-Out Cross Validation adalah kasus khusus dari *K-fold Cross Validation* di mana k memiliki nilai yang sama dengan jumlah instance dalam data. Dalam setiap iterasi, hampir semua data kecuali satu observasi digunakan sebagai data pelatihan, dan model yang dibangun diuji menggunakan observasi tunggal tersebut. Metode ini masih sering digunakan ketika data yang tersedia sangat terbatas, terutama dalam bidang bioinformatika (Watanabe, 2009). Gambar 6 mengilustrasikan *Leave-One-Out Cross Validation*.



Gambar 6. Ilustrasi *Leave-One-Out Cross Validation* (Chlis, 2013).

2.11.3. *Hold-out Cross Validation*

Hold-out Cross Validation adalah pembagian data dibagi secara acak menjadi dua bagian: data pelatihan dan data pengujian, dengan rasio pembagian tipikal seperti 60%:40% atau 80%:20%. Metode ini digunakan untuk melatih model dengan data pelatihan dan menguji model dengan data pengujian. Evaluasi model tergantung pada jenis masalah yang diteliti, namun jika data pelatihan atau pengujian tidak mewakili data secara lengkap, hasilnya mungkin tidak akurat (Nugraha, 2013). Ketika data yang digunakan untuk pelatihan dan pengujian terbatas, diperlukan metode untuk mengukur tingkat akurasi metode dalam *machine learning*. Salah satu metode validasi yang digunakan adalah metode *holdout*, di mana sebagian data digunakan sebagai data pengujian dan sisanya sebagai data pelatihan (Nugraha, 2013). Namun, saat proses pengacakan data untuk pembagian menjadi data pelatihan dan data pengujian, kemungkinan besar salah satu klasifikasi akan menjadi lebih dominan daripada klasifikasi lainnya, sehingga data pelatihan dan pengujian yang dihasilkan tidak mewakili secara proporsional. Oleh karena itu, diperlukan prosedur stratifikasi *holdout* yang memastikan setiap klasifikasi terwakili secara proporsional dalam data pelatihan dan pengujian. Proses ini dapat diulang beberapa kali dengan data pelatihan dan pengujian yang diacak, dan nilai rata-ratanya diambil. Prosedur ini dikenal sebagai *repeated holdout*. Gambar 7 mengilustrasikan *Hold-out Cross Validation*.



Gambar 7. Ilustrasi *Hold-out Cross Validation* (Chils, 2013).

2.12. *Synthetic Minority Over-sampling Technique (SMOTE)*

Data yang tidak seimbang terjadi ketika distribusi kelas dalam *dataset* tidak proporsional, dengan salah satu kelas memiliki jumlah yang lebih sedikit atau lebih banyak dibandingkan kelas lainnya. Kelas yang memiliki jumlah lebih sedikit disebut sebagai kelas minoritas, sementara kelas yang lainnya disebut sebagai kelas mayoritas. Untuk menyeimbangkan distribusi data tersebut, salah satu metode yang digunakan adalah *oversampling*, di mana jumlah data pada kelas minoritas ditingkatkan. Salah satu teknik *oversampling* yang populer adalah *Synthetic Minority Over-sampling Technique (SMOTE)*. SMOTE diperkenalkan oleh Nitesh V. Chawla dan bekerja dengan membuat replikasi data pada kelas minoritas. Data replikasi yang dihasilkan disebut sebagai data sintetis (Siringoringo, 2018).

SMOTE (*Synthetic Minority Over-sampling Technique*) adalah metode statistika yang digunakan untuk menyeimbangkan jumlah data antara kelas minoritas dan mayoritas. Berbeda dengan metode *random oversampler*, SMOTE tidak hanya menduplikasi data minoritas, tetapi juga menggunakan algoritma KNN (*K-Nearest Neighbors*) untuk menciptakan data baru berdasarkan data yang ada. Caranya adalah dengan memilih secara acak satu data dari kelas minoritas, kemudian menggunakan algoritma KNN untuk mencari data tetangga dan menghubungkan kedua data tersebut dengan garis. Data sintetis kemudian dihasilkan melalui kombinasi *convex* antara kedua data yang telah dipilih (Fahlapi et al., 2022). Dengan menggunakan SMOTE, dapat meningkatkan jumlah data pada kelas minoritas secara sintetis, yang pada gilirannya dapat membantu meningkatkan kinerja model dalam mengenali kelas minoritas yang jarang terjadi. Hal ini berguna dalam situasi di mana data tidak seimbang dan ingin mencegah model cenderung memprediksi hanya pada kelas mayoritas.

2.13. Pelayanan Bank Rakyat Indonesia

Pelayanan merupakan suatu kegiatan yang dilakukan untuk memenuhi kebutuhan dan keinginan orang lain. Pelayanan yang baik sangat penting bagi perusahaan, karena dapat mempengaruhi keputusan pelanggan dalam membeli produk yang ditawarkan. Kinerja pelayanan adalah ukuran dari kualitas yang diharapkan untuk memenuhi keinginan konsumen (Wibowati, 2021). Menurut Philip Kotler, jasa adalah setiap kegiatan atau manfaat yang diberikan oleh satu pihak kepada pihak lain. Jasa bersifat tidak berwujud dan tidak mempengaruhi kepemilikan sesuatu, meskipun produksinya mungkin berhubungan dengan produk fisik. Pelayanan, dalam konteks ini, merujuk pada aktivitas penyampaian jasa dari satu pihak ke pihak lain. Pelayanan yang baik adalah pelayanan yang teruji dalam hal ramah, adil, cepat, tepat, dan beretika untuk memenuhi kepentingan dan kepuasan penerima layanan (Hariyono, 2007).

Bank adalah lembaga keuangan yang memberikan jasa keuangan seperti kredit, tabungan, pembayaran jasa, dan berbagai fungsi keuangan lainnya secara profesional. Keberhasilan suatu bank ditentukan oleh kemampuan untuk mengidentifikasi permintaan masyarakat terhadap jasa keuangan, memberikan pelayanan secara efisien, dan menjualnya dengan harga yang bersaing (Regaer et al., 2016). Menurut definisi dalam UU No. 7 tahun 1992 tentang Perbankan Indonesia, bank merupakan badan usaha yang menghimpun dana dari masyarakat dalam bentuk simpanan dan menyalurkannya kepada masyarakat dalam bentuk kredit dan bentuk lainnya untuk meningkatkan taraf hidup banyak orang.

Salah satu jenis pelayanan di bank adalah pelayanan *frontliner*, yang berkaitan dengan pelayanan langsung kepada nasabah di bagian depan bank. Pelayanan ini memiliki peran penting dalam memberikan gambaran evaluasi baik dan buruk terhadap pelayanan bank. Pelayanan yang dilakukan dapat mempengaruhi persepsi nasabah terhadap bank tersebut. *Frontline service* mencakup pelayanan nasabah, ATM, dan keamanan (Vidyandari et al., 2021). Kepuasan nasabah adalah

tanggapan nasabah terhadap perbedaan antara harapan awal atau standar kinerja sebelum melakukan pembelian dengan kinerja aktual produk setelah digunakan atau dikonsumsi. Kepuasan nasabah bersifat relatif dan bergantung pada apa yang diharapkan oleh nasabah (Maulana & Muhajirin, 2021). Pada penelitian ini, data yang berkaitan dengan kinerja pelayanan PT. Bank Rakyat Indonesia (Persero) Tbk. di X akan digunakan untuk melakukan klasifikasi analisis sentimen.

2.14. Penelitian Terdahulu

Penelitian berkaitan pada *Text mining* untuk studi kasus analisis sentimen sudah banyak dilakukan terlebih dahulu. Penelitian terlebih dahulu dengan studi kasus analisis sentimen pada Tabel 2.

Tabel 2. Penelitian Terdahulu

Perbandingan teori	Peneliti (tahun penelitian)	Tujuan	Langkah-langkah	metode	output
<i>Sentiment Analysis on Corona Virus Pandemic Using Machine Learning Algorithm</i>	(Risnantoyo et al., 2020)	Mencari performansi terbaik dari algoritma <i>machine learning</i> menggunakan beberapa metode, termasuk <i>Naïve Bayes</i> , SVM, dan KNN.	<ol style="list-style-type: none"> 1. pengambilan data dengan Twitter API 2. Melakukan <i>text Preprocessing</i> 3. Pembobotan dengan TF-IDF 4. Pembagian <i>Data training</i> dan <i>testing 10-fold Cross Validation</i> 5. Melakukan analisis menggunakan <i>Naïve Bayes</i>, SVM, dan KNN 	Metode <i>Naïve Bayes</i> , <i>Support Vector Machine</i> (SVM), dan <i>K-Nearest Neighbors</i> (KNN)	Penelitian ini menggunakan <i>dataset</i> 833 data <i>tweets</i> . Hasil akurasi algoritma <i>Naïve Bayes</i> , SVM <i>kernel Linear</i> , dan KNN K=5 adalah 75,81%, 76,21%, dan 71,89%. Hasil presisi adalah 74,33%, 77,59%, dan 66,95%. Hasil <i>recall</i> adalah 70,83%, 71,42%, dan 66,03%. perbedaan dalam nilai akurasi, presisi, dan <i>recall</i> ketiga algoritma ini, tidak signifikan (<5%). ketiga algoritma tersebut memberikan klasifikasi yang baik dalam analisis sentimen dengan data X.

Perbandingan teori	Peneliti (tahun penelitian)	Tujuan	Langkah-langkah	metode	output
Sentimen Analisis <i>Stay Home</i> menggunakan metode klasifikasi <i>Naïve Bayes</i> , <i>Support Vector Machine</i> , dan <i>K-Nearest Neighbors</i>	(Hakim et al., 2020)	Penelitian ini bertujuan untuk mengamati dampak kebijakan <i>Stay Home</i> terhadap kehidupan masyarakat di negara-negara yang terkena pandemi yang menerapkan kebijakan tersebut. Selain itu, penelitian ini juga membandingkan performa tiga algoritma, yaitu <i>Naïve Bayes</i> , SVM, dan KNN.	<ol style="list-style-type: none"> 1. Pengumpulan data dengan <i>Snsrapper</i> 2. Tahap <i>preprocessing</i> (<i>regex Removal</i>, <i>Remove URL</i>, <i>Remove Annotation</i>, <i>Remove Duplicate tweets</i>, <i>extract sentiment</i>, <i>transform case</i>) 3. Penerapan SMOTE pada data dan melakukan analisis dengan algoritma <i>Naïve Bayes</i>, SVM, dan KNN 	<i>Naïve Bayes</i> , <i>Support Vector Machine</i> , dan <i>K-Nearest Neighbors</i>	Penelitian ini menggunakan <i>dataset</i> sebanyak 1652 <i>tweets</i> . Hasil penelitian dari ketiga algoritma, setelah dilakukan proses pengujian data didapatkan hasil akurasi sebagai berikut <i>Naïve Bayes</i> 66,81%, <i>Support Vector Machine</i> 80,05% dan <i>K-Nearest Neighbors</i> 51,45%. Hasil terbaik yaitu pada Algoritma <i>Support Vector Machine</i>

Perbandingan Teori	Peneliti (Tahun Penelitian)	Tujuan	Langkah-langkah	Metode	Output
Analisis Sentimen Kualitas Layanan Pembayaran Elektronik pada Twitter	(Ogi et al., 2021)	Melakukan analisis sentimen pengguna terhadap layanan OVO dan DANA di Twitter berdasarkan cuitan pengguna menggunakan tiga model klasifikasi, yaitu <i>Naïve Bayes</i> , SVM, dan KNN.	<ol style="list-style-type: none"> 1. Pengumpulan Data menggunakan API Twitter 2. <i>Preprocessing</i> 3. Pembobotan menggunakan TF-IDF 4. Klasifikasi Teks 5. Validasi menggunakan <i>k-fold</i> ke 10 dan evaluasi performa klasifikasi 	Metode yang digunakan yaitu <i>Support Vector Machine</i> , <i>Naïve Bayes</i> , dan <i>K-Nearest Neighbors</i>	Penelitian ini menggunakan data tweets ovo 6387 dan dana 6476. Algoritma SVM mendapatkan hasil akurasi paling tinggi pada data OVO dan DANA yaitu 81,33% dan 86,23%, <i>Naïve Bayes</i> yaitu 80,63% dan 86,19% dan KNN yaitu 75,83% dan 86,19.

Perbandingan Teori	Peneliti (Tahun Penelitian)	Tujuan	Langkah-langkah	Metode	Output
Analisis Sentimen Masyarakat Indonesia Terhadap Pindahan Ibu Kota Negera Indonesia Pada Twitter	(Lestari et al., 2022)	dilakukan analisis sentimen masyarakat Indonesia terkait pemindahan ibu kota negara Indonesia. Dan dilakukan perbandingan kinerja tiga algoritma klasifikasi yaitu SVM, <i>Naïve Bayes</i> , dan KNN	<ol style="list-style-type: none"> 1. Pengumpulan data dengan API Twitter 2. <i>Preprocessing</i> (<i>case folding</i>, <i>cleaning</i>, filter <i>stopword</i>, <i>tokenize</i>) 3. Pelabelan data 4. Analisis dengan Algoritma <i>Support Vector Machine</i>, <i>Naïve Bayes</i>, dan <i>K-Nearest Neighbors</i> dengan pembagian data latih dan data uji 80% dan 20%. Dan melakukan evaluasi <i>Cross Validation</i> dengan <i>10-Fold Cross Validation</i> 	SVM, <i>Naïve Bayes</i> , dan <i>K-Nearest Neighbors</i>	Penelitian ini menggunakan sebanyak 1732 data <i>tweets</i> . Dari proses ini, Hasil Algoritma SVM untuk <i>accuracy</i> 85,71%, <i>precision</i> 72,56%, dan <i>recall</i> 85,88%. Hasil Algoritma <i>Naïve Bayes</i> untuk <i>accuracy</i> 76,70%, <i>precision</i> 79,41%, dan <i>recall</i> 48,31%. Sedangkan hasil Algoritma KNN untuk <i>accuracy</i> 52,74%, <i>precision</i> 43,20%, dan <i>recall</i> 96,02%.

Perbandingan Teori	Peneliti (Tahun Penelitian)	Tujuan	Langkah-langkah	Metode	Output
Analisis Sentimen dengan SVM, <i>Naïve Bayes</i> , dan KNN untuk Studi Tanggapan Masyarakat Indonesia Terhadap Pandemi Covid-19 pada Media Sosial Twitter	(Sodik & Kharisudin, 2021)	Penelitian ini bertujuan untuk membandingkan metode yang paling baik untuk mengklasifikasikan data tanggapan.	<ol style="list-style-type: none"> 1. Pengumpulan data. 2. <i>Pre-processing</i> data. 3. Klasifikasi menggunakan metode <i>Support Vector Machine</i>, <i>K-Nearest Neighbors</i>, dan <i>Naïve Bayes</i>. 4. Evaluasi Model. 	Metode <i>Support Vector Machine</i> , <i>Naïve Bayes</i> , dan <i>K-Nearest Neighbors</i> .	Penelitian ini menggunakan 10000 data, Algoritma <i>Support Vector Machine</i> dengan kernel <i>linear</i> , <i>Naïve Bayes</i> dengan <i>laplace correction</i> adalah 1, dan <i>K-Nearest Neighbors</i> dengan $K=20$, Serta evaluasi model <i>10-Fold Cross Validation</i> yaitu akurasi sebesar 90.1%, 79.2%, dan 62.1%.

III. METODOLOGI PENELITIAN

3.1. Waktu dan Tempat Penelitian

Penelitian dilakukan di Jurusan Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung yang terletak di Jalan Soemantri Brojonegoro No. 1 Gedung Meneng, Bandar Lampung. Pelaksanaan penelitian mulai semester ganjil tahun ajaran 2022/2023 hingga bulan Juni 2023.

3.2. Perangkat Penelitian

Perangkat untuk penelitian yang digunakan adalah sebagai berikut.

3.2.1. Perangkat Keras

Perangkat keras yang dibutuhkan dengan spesifikasi laptop sebagai berikut:

- Prosesor : Intel® Core™ i3-7020U CPU @2.30GHz
- RAM : 12.0 GB
- *Operating System* : Windows 10 Ultimate 64-bit Operating System

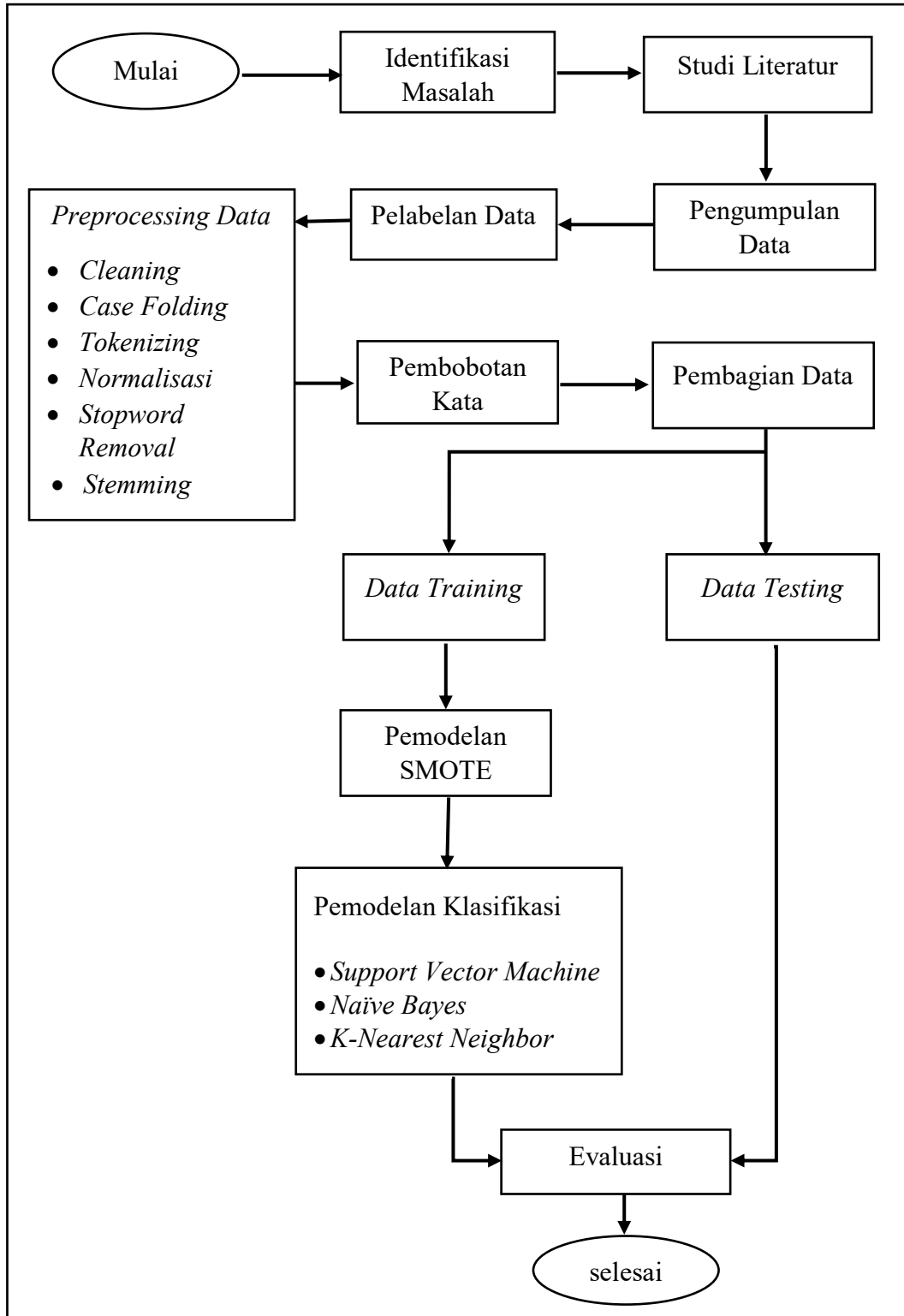
3.2.2. Perangkat Lunak

Perangkat lunak yang dibutuhkan dalam penelitian sebagai berikut:

- *Python* 3.10.5
- Anaconda Navigator 3
- *Jupyter Notebook*
- Microsoft Excel 2019
- Web Browser Google Chrome

3.3. Alur Penelitian

Alur penelitian ini digambarkan sebagai berikut:



Gambar 8. Alur Penelitian.

Gambar 8 menjelaskan tahapan-tahapan yang dilalui dalam penelitian. Penjelasan lebih lanjut mengenai tahapan penelitian sebagai berikut:

3.3.1. Identifikasi Masalah

PT Bank Rakyat Indonesia (Persero) Tbk. adalah salah satu bank milik pemerintah yang dipercaya oleh masyarakat untuk menghimpun dana dan memberikan pinjaman. Dapat dilihat di PT Bank Rakyat Indonesia (Persero) Tbk kerap terjadi antrian yang cukup panjang untuk mendapatkan pelayanan PT. Bank Rakyat Indonesia (Persero) Tbk. Dari permasalahan tersebut muncul komentar mengenai kinerja pelayanan positif dan negatif.

Berdasarkan analisis permasalahan yang ada, pada tahapan ini topik penelitian yang akan dibahas adalah analisis sentimen untuk klasifikasi sentimen negatif dan sentimen positif terhadap data yang didapatkan dari X dengan menggunakan algoritma *Support Vector Machine (SVM)*, *Naïve Bayes*, dan *K-Nearest Neighbors* tentang kinerja pelayanan di PT. Bank Rakyat Indonesia (Persero) Tbk. Hasil dari penelitian tersebut diharapkan dapat membantu PT. Bank Rakyat Indonesia (Persero) Tbk. untuk melakukan evaluasi dan memperbaiki kinerja pelayanan agar lebih baik terhadap nasabah PT Bank Rakyat Indonesia (Persero) Tbk. pada penelitian ini akan dilakukan perbandingan antara algoritma *Support Vector Machine (SVM)*, *Naïve Bayes*, dan *K-Nearest Neighbors (KNN)*.

3.3.2. Studi literatur

Tahapan studi literatur merupakan langkah awal dalam penelitian ini, yang melibatkan pengumpulan literatur yang relevan dengan analisis permasalahan yang akan diteliti. Tujuan dari tahap ini adalah untuk memperoleh pemahaman yang mendalam mengenai penelitian terdahulu yang telah dilakukan sebelumnya, baik terkait dengan topik penelitian maupun metode penelitian yang digunakan. Teori dan penelitian terdahulu tersebut menjadi acuan yang penting untuk memahami dan

melakukan analisis dengan benar, sesuai dengan kerangka berpikir ilmiah yang telah ditetapkan. Dalam konteks penelitian ini, telah dilakukan banyak penelitian mengenai analisis sentimen dengan topik yang berbeda-beda. Informasi mengenai penelitian-penelitian terdahulu ini akan disajikan dalam Tabel 2 sebagai bagian dari studi literatur yang relevan dengan penelitian ini, sehingga dapat memberikan pemahaman yang lebih lengkap dan komprehensif terhadap topik yang sedang diteliti.

3.3.3. Pengumpulan Data

Pada tahap pengumpulan data untuk analisis sentimen terhadap kinerja pelayanan PT. Bank Rakyat Indonesia (Persero) Tbk. dilakukan menggunakan *platform X*. Data yang dikumpulkan berupa teks dalam bahasa Indonesia, dengan menggunakan bahasa pemrograman *Python 3* sebagai alat pengembangan. Untuk mengakses data *tweet* di media sosial X, digunakan API key. API key ini berperan penting dalam pencarian *tweet* menggunakan *library Snsrape* pada *Python*. Kata kunci yang digunakan dalam pengumpulan data adalah "Pelayanan BRI", "KUR BRI", "CS BRI", "Customer Service BRI", dan "Brimo". Data *tweet* yang terkumpul kemudian disimpan dalam bentuk tabel dengan atribut tanggal dan waktu, ID, *username*, dan isi X. Proses pengumpulan data dilakukan mulai bulan Maret 2022 hingga Februari 2023, dan total terkumpul sebanyak 1652 data *tweet*. Hasil data tersebut disimpan dalam format JSON, kemudian dikonversi ke dalam format *Comma Separated Values (CSV)*.

3.3.4. Pelabelan Data

Pelabelan pada *dataset* yang telah dikumpulkan dilakukan secara manual oleh tiga orang pelabel. Dalam proses pelabelan, terdapat dua pelabel utama yang bertanggung jawab dalam menentukan nilai sentimen setiap *tweet* yang terdapat dalam *dataset*. Jika kedua pelabel utama memiliki penilaian sentimen yang berbeda, maka pelabel pendukung akan ikut menentukan penilaian akhir. Dalam penelitian ini, polaris digunakan sebagai ukuran sentimen. Polaritas positif digunakan untuk

data yang mengandung sentimen positif, sementara polaritas negatif digunakan untuk data yang mengandung sentimen negatif. Proses pelabelan secara manual untuk memastikan konsistensi dan akurasi penilaian sentimen pada setiap *tweet*. Dengan melibatkan tiga pelabel, perbedaan penilaian dapat diselesaikan dengan menggunakan pelabel pendukung sebagai penentu akhir. Hal ini membantu mengurangi potensi bias dan meningkatkan hasil analisis sentimen yang dilakukan pada *dataset* yang telah dilabeli secara manual.

3.3.5. Preprocessing Data

Tahap *preprocessing* data dilakukan setelah tahap pengumpulan data untuk membersihkan dan mempersiapkan data sebelum dilakukan analisis lebih lanjut. Pada tahap ini, terdapat beberapa proses yang dilakukan, yaitu *cleaning*, *case folding*, *tokenizing*, normalisasi, *stopword removal*, dan *stemming*. Penjelasan lebih detail mengenai fungsi-fungsi pada pengolahan *dataset* sebagai berikut.

3.3.5.1. Cleaning

Cleaning adalah proses untuk menghilangkan *noise* atau gangguan pada data. Pada tahap ini, seluruh data teks penelitian akan diolah dengan beberapa langkah, seperti penghapusan *username*, URL, tanda hashtag (#), tanda mention (@), angka, tanda baca, karakter HTML, dan karakter simbol lainnya. Hal ini bertujuan untuk menghilangkan elemen-elemen yang tidak relevan atau mengganggu dalam teks dan mempertahankan hanya informasi teks yang penting. Atribut-atribut yang tidak berpengaruh juga akan dihilangkan dari dokumen.

3.3.5.2. Case Folding

Case folding adalah proses mengubah semua huruf dalam teks menjadi huruf kecil atau huruf besar. Hal ini dilakukan agar tidak ada perbedaan dalam pengolahan teks berdasarkan kapitalisasi huruf. Misalnya, kata "Data" dan "data" akan dianggap sama setelah proses *case folding*.

3.3.5.3. Tokenizing

Tokenizing adalah proses memecah teks menjadi unit-unit kecil yang disebut token. Token dapat berupa kata-kata, frasa, atau simbol-simbol tertentu. Proses *tokenizing* memudahkan dalam analisis dan pengolahan lebih lanjut, karena memisahkan teks menjadi bagian-bagian yang lebih terstruktur.

3.3.5.4. Normalisasi

Tahap normalisasi adalah tahap proses perubahan kata-kata yang tidak sesuai dengan Ejaan yang Disempurnakan (EYD) atau standar bahasa yang digunakan. Pada tahap ini, dilakukan perubahan kata-kata yang berupa singkatan menjadi kata baku, serta kata-kata asing yang diubah menjadi bentuk kata normal sesuai dengan aturan bahasa yang berlaku. Proses normalisasi memiliki tujuan untuk memastikan bahwa kata-kata dalam teks mengikuti standar ejaan yang telah ditetapkan. Misalnya, jika terdapat kata singkatan seperti "bkn" yang merupakan kependekan dari "bukan", maka pada tahap normalisasi kata tersebut akan diubah menjadi "bukan". Selain itu, jika terdapat kata-kata asing yang tidak umum dalam bahasa Indonesia, seperti "restoran" yang diubah menjadi "resto", maka pada tahap normalisasi kata tersebut akan dikembalikan menjadi "restoran".

3.3.5.5. Stopword Removal

Stopwords adalah kumpulan kata-kata yang sering muncul dalam teks, namun jika dihapus tidak akan mengubah makna dari teks tersebut. Pada tahap *stopwords removal*, kata-kata *stopwords* diidentifikasi dan dihapus dari teks. Tujuan dari tahap ini adalah untuk menghilangkan kata-kata yang tidak memiliki arti atau tidak relevan dalam analisis sentimen.

3.3.5.6. Stemming

Stemming adalah proses penghilangan afiks atau imbuhan pada kata dalam suatu kalimat sehingga kata tersebut direduksi menjadi bentuk kata dasar dalam dokumen. Tujuan dari stemming adalah untuk mengurangi variasi kata yang

memiliki akar kata yang sama, sehingga memperbaiki kualitas data dan informasi yang ditemukan dalam pencarian informasi. Dalam pencarian informasi, stemming digunakan untuk memperoleh keterkaitan antara kata-kata yang memiliki akar kata yang sama. Dengan mereduksi kata-kata tersebut menjadi bentuk kata dasar, sehingga dapat mengidentifikasi kata-kata yang memiliki makna serupa atau berkaitan secara lebih efisien.

3.3.6. Pembobotan Kata

Setelah melalui tahap *preprocessing*, *dataset* akan dilakukan pembobotan kata. Tahap ini bertujuan untuk memberikan nilai bobot pada setiap kata yang terdapat dalam *dataset*. Dalam penelitian ini, pembobotan dilakukan menggunakan metode TF-IDF (*Term Frequency-Inverse Document Frequency*). Metode TF-IDF digunakan untuk menentukan tingkat pentingnya suatu kata atau kumpulan kata dalam *dataset*. Proses pembobotan ini melibatkan dua komponen utama: *Term Frequency* (TF) dan *Inverse Document Frequency* (IDF). *Term Frequency* (TF) mengukur seberapa sering kata tertentu muncul dalam sebuah dokumen. Semakin sering kata tersebut muncul, semakin tinggi bobotnya. *Inverse Document Frequency* (IDF) mengukur tingkat keunikan kata tersebut dalam seluruh *dataset*. Dengan menggunakan metode TF-IDF, kata-kata yang sering muncul dalam sebuah dokumen tetapi jarang muncul dalam dokumen lain akan memiliki bobot yang tinggi. Sebaliknya, kata-kata yang sering muncul dalam seluruh *dataset* akan memiliki bobot yang rendah karena kurangnya keunikan kata tersebut. Penelitian ini dapat mengevaluasi tingkat pentingnya setiap kata dalam *dataset* dan mengoptimalkan penggunaan algoritma pembelajaran mesin yang telah ditentukan.

3.3.7. Pembagian *Data*

Pada penelitian ini, *dataset* akan dibagi menjadi dua bagian: *data training* dan *data testing*. Tujuan dari pembagian ini adalah untuk melakukan pelatihan (*training*) dan pengujian (*testing*) model klasifikasi yang akan dikembangkan. *Data training* adalah subset dari *dataset* yang akan digunakan untuk melatih model klasifikasi.

Proses pelatihan dilakukan dengan memasukkan data training ke dalam model klasifikasi, sehingga model dapat belajar pola atau hubungan antara fitur (*features*) yang ada dalam data dengan label klasifikasi yang terkait. Dalam tahap ini, model akan mengoptimalkan parameter dan memperbaiki kemampuannya dalam memprediksi label klasifikasi. Setelah model selesai dilatih menggunakan data training, langkah selanjutnya adalah menguji performanya menggunakan data testing. *Data testing* digunakan untuk menguji sejauh mana model yang telah dilatih dapat menggeneralisasi dan memprediksi label klasifikasi dengan benar untuk data yang belum pernah dilihat sebelumnya.

3.3.8. Synthetic Minority Over-sampling Technique (SMOTE)

Dalam mengatasi masalah data yang tidak seimbang (*inbalanced data*), pada *data training* akan dilakukan penyeimbangan. Penyeimbangan data dilakukan menggunakan metode SMOTE (*Synthetic Minority Over-sampling Technique*). Metode SMOTE bekerja dengan mensintesis sampel baru dari kelas minoritas untuk menciptakan keseimbangan pada *dataset*. Dengan cara melakukan sampling ulang pada sampel kelas minoritas, metode ini membantu meningkatkan representasi dan akurasi klasifikasi pada kelas minoritas. Penerapan metode SMOTE pada data training bertujuan untuk menciptakan distribusi yang seimbang antara kelas sentimen yang ada. Hal ini dilakukan agar model klasifikasi tidak cenderung bias terhadap kelas mayoritas dan dapat memberikan hasil yang lebih akurat dalam mengklasifikasikan sentimen pada data yang belum pernah dilihat sebelumnya. Dengan melakukan pembagian *dataset* menjadi *data training* dan *data testing*, serta menerapkan metode SMOTE untuk penyeimbangan data training, penelitian ini dapat mengoptimalkan proses pelatihan dan evaluasi model klasifikasi terhadap sentimen pada dataset yang digunakan.

3.3.9. Klasifikasi Sentimen

Tahap klasifikasi sentimen dalam penelitian ini menggunakan tiga algoritma yaitu *Support Vector Machine (SVM)*, *Naïve Bayes*, dan *K-Nearest Neighbors*. Data yang

digunakan untuk klasifikasi sentimen adalah data *tweet* yang telah diproses pada tahap pengolahan *dataset* sebelumnya. Pada tahap awal, *dataset* dibagi menjadi dua bagian, yaitu data latih (*training data*) dan data uji (*testing data*). Data latih digunakan untuk melatih algoritma klasifikasi, sedangkan *data testing* digunakan untuk mengevaluasi performa algoritma yang telah dilatih. Pembagian data ini penting untuk menguji kemampuan algoritma dalam mengklasifikasikan sentimen pada data yang belum pernah dilihat sebelumnya.

Setelah dilakukan pembagian data, dilakukan pelatihan algoritma menggunakan data latih yang telah dibagi sebelumnya, data tersebut digunakan sebagai *input* untuk melatih algoritma SVM, *Naïve Bayes*, dan *K-Nearest Neighbors*. Proses pelatihan ini bertujuan untuk menghasilkan model klasifikasi yang dapat digunakan untuk mengklasifikasikan sentimen pada data yang belum diketahui labelnya. Setelah algoritma dilatih, dilakukan klasifikasi pada data uji menggunakan algoritma SVM, *Naïve Bayes*, dan *K-Nearest Neighbors* yang telah terlatih sebelumnya. Hasil klasifikasi berupa prediksi sentimen positif dan negatif dievaluasi menggunakan metrik evaluasi seperti *Confusion Matrix*, akurasi, presisi, *recall*, dan *F1-score*. Matrik-matrik ini memberikan informasi mengenai performa algoritma dalam mengklasifikasikan sentimen, seperti sejauh mana algoritma mampu mengenali sentimen positif dan negatif, serta tingkat kesalahan yang terjadi dalam klasifikasi.

3.3.10. Evaluasi

Evaluasi yang dilakukan dalam penelitian ini mencakup perbandingan analisis sentimen menggunakan tiga algoritma, yaitu *Support Vector Machine* (SVM), *Naïve Bayes*, dan *K-Nearest Neighbors* (KNN) dengan *dataset imbalanced* dan *dataset balanced*. Pengujian pada setiap *dataset* bertujuan untuk mengetahui perubahan yang terjadi pada nilai akurasi, presisi, *recall*, dan *F1-score*. Kemudian, dilakukan perbandingan antara hasil klasifikasi dari ketiga algoritma tersebut. Evaluasi ini menghasilkan metrik evaluasi seperti *Confusion Matrix* dan

Classification Report yang meliputi akurasi, presisi, *recall*, dan *F1-score*. *Confusion Matrix* memberikan informasi tentang hasil prediksi klasifikasi dari sistem yang dibandingkan dengan data aktual yang telah dilabeli secara manual. *Confusion Matrix* ini terdiri dari empat bagian yaitu *True Positive*, *False Positive*, *True Negative*, dan *False Negative*, yang menggambarkan sejauh mana algoritma mampu mengenali sentimen positif dan negatif dengan benar. Kinerja sistem klasifikasi dihitung menggunakan data dalam *Confusion Matrix* dijelaskan pada Tabel 3.

Tabel 3. *Confusion Matrix*

Fakta	Prediksi	
	Negatif	Positif
Negatif	TN (<i>True Negative</i>)	FP (<i>False Positive</i>)
positif	FN (<i>False Negative</i>)	TP (<i>True Positive</i>)

Nilai *True Negative* (TN) merupakan jumlah data negatif yang terdeteksi dengan benar, sedangkan *False Negative* (FN) merupakan data positif yang terdeteksi negatif. Nilai *True Positive* (TP) merupakan data positif yang terdeteksi benar, sedangkan *False Positive* (FP) merupakan data negatif namun terdeteksi positif. Tabel *Confusion Matrix* digunakan untuk mengukur kinerja suatu metode klasifikasi dengan menghitung nilai akurasi, presisi, *recall*, dan *F1-score* (Muludi et al., 2021).

a. Akurasi

Akurasi (*accuracy*) merupakan perbandingan prediksi benar positif dan negatif dari keseluruhan data. Rumus akurasi menggunakan Persamaan (5).

$$\text{Akurasi (accuracy)} = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \quad (5)$$

b. Presisi

Presisi (*precision*) merupakan ukuran keakuratan hasil suatu model. Presisi adalah perbandingan prediksi benar positif dengan keseluruhan hasil yang diprediksi positif. Rumus presisi menggunakan Persamaan (6).

$$\text{Presisi (precision)} = \frac{TP}{TP + FP} \times 100\% \quad (6)$$

c. Recall

Recall merupakan ukuran kelengkapan suatu model. *Recall* adalah perbandingan prediksi benar positif dibandingkan dengan data yang memiliki label positif. Rumus *recall* menggunakan Persamaan (7).

$$\text{Recall} = \frac{TP}{TP + FN} \times 100\% \quad (7)$$

d. F1-score

F1-score adalah perbandingan rata-rata nilai yang dihasilkan dari presisi dan *recall*. Rumus F1-score menggunakan Persamaan (8).

$$\text{F1 Score} = 2 \times \frac{(\text{recall} \times \text{precision})}{(\text{recall} + \text{precision})} \quad (8)$$

V. SIMPULAN DAN SARAN

5.1. Simpulan

Berdasarkan dari hasil penelitian yang telah dilakukan maka dapat disimpulkan sebagai berikut.

1. Jumlah data yang digunakan pada penelitian ini yaitu sebanyak 1652 data yang diperoleh dari X. Opini masyarakat mengenai kinerja pelayanan PT Bank Rakyat Indonesia (Persero) Tbk didominasi oleh sentimen positif sebanyak 903 bersentimen positif atau 54,7% dan 749 bersentimen negatif atau 45,3%. Pada penelitian ini dilakukan penyeimbangan data menggunakan *Synthetic Minority Over-sampling Technique* (SMOTE). Penyeimbangan data dilakukan pada data latih sehingga data positif dan data negatif sama. Penelitian ini berhasil mengimplementasikan algoritma *Support Vector Machine*, *Naïve Bayes*, dan *K-Nearest Neighbors*. Penambahan *Synthetic Minority Over-sampling Technique* (SMOTE) untuk *dataset balanced* memberikan hasil akurasi yang lebih baik dengan rata-rata akurasi pada algoritma *Support Vector Machine* sebesar 93.55%.
2. Perbandingan hasil kinerja klasifikasi Algoritma *Support Vector Machine*, *Naïve Bayes*, dan *K-Nearest Neighbors* bahwa klasifikasi terbaik pada metode *Support Vector Machine* pada *dataset balanced*. Hasil klasifikasi dengan menerapkan *Synthetic Minority Over-sampling Technique* (SMOTE), Klasifikasi *Support Vector Machine* menghasilkan akurasi tertinggi pada skema ke-4 dengan *dataset* 65% data latih - 35% data uji sebesar 94,30%, presisi sebesar 94,92%, *recall* sebesar 94,62%, dan *F1-score* sebesar 94,77%. Klasifikasi *Naïve Bayes* menghasilkan akurasi tertinggi pada skema ke-2 dengan *dataset* 55% data latih - 45% data uji sebesar 93,82%, presisi sebesar 95,69%, *recall* sebesar 92,63%, dan *F1-score* sebesar 94,13%. Klasifikasi *K-Nearest Neighbors* menghasilkan

akurasi tertinggi pada skema ke-5 dengan *dataset* 70% data latih - 30% data uji sebesar 88,31%, presisi sebesar 84,12%, *recall* sebesar 94,16%, dan F1-*score* sebesar 88,85%.

5.2. Saran

Setelah dilakukan penelitian ini terdapat saran untuk penelitian selanjutnya adalah sebagai berikut:

1. Pengembangan mengenai analisis sentimen selanjutnya, dapat menambahkan optimasi jenis klasifikasi lainnya seperti *Deep Learning*, sehingga mampu meningkatkan performa klasifikasi.
2. Pada penelitian selanjutnya dapat melakukan filterisasi data *tweet* untuk mendeteksi *tweet* yang berulang dengan menggunakan *bot* otomatis pendeteksi *tweet* duplikat untuk mengurangi *noise*.
3. Pengembangan penyeimbangan data menggunakan metode lainnya seperti SMOTEENN, sehingga mendapatkan hasil *resampling* yang lebih baik.

DAFTAR PUSTAKA

- Anesca, N. P. A., Muludi, K., & Shofiana, D. A. (2021). Sentiment Analysis Protokol Kesehatan Virus Corona Dari Tweet Menggunakan Word2Vec Model Dan Recurrent Neural Network Learning. *Jurnal Pepadun*, 2(3), 432–439. <https://doi.org/10.23960/pepadun.v2i3.86>
- Ariyanti, D., & Iswardani, K. (2020). Teks Mining untuk Klasifikasi Keluhan Masyarakat Pada Pemkot Probolinggo Menggunakan Algoritma *Naïve Bayes*. *Jurnal IKRA-ITH Informatika*, 4(3), 125–132.
- Fahlapi, R., Asra, T., Kuntoro, A. Y., Ocanitra, R., Effendi, L., Syukmana, F., Bina, U., Informatika, S., Pusat, J., Mandiri, U. N., & Pusat, J. (2022). Analisa sentimen vaksinasi covid-19 dengan metode support vector machine dan naïve bayes berbasis teknik smote. *Jurnal Informatika Kaputama*, 6(1), 57–64.
- Fathonah, F., & Herliana, A. (2021). Penerapan Text Mining Analisis Sentimen Mengenai Vaksin Covid - 19 Menggunakan Metode *Naïve Bayes*. *Jurnal Sains Dan Informatika*, 7(2), 155–164. <https://doi.org/10.34128/jsi.v7i2.331>
- Firdaus, A., & Firdaus, W. I. (2021). Text Mining Dan Pola Algoritma Dalam Penyelesaian Masalah Informasi : (Sebuah Ulasan). *Jurnal JUPITER*, 13(1), 66.
- Giovani, A. P., Ardiansyah, A., Haryanti, T., Kurniawati, L., & Gata, W. (2020). Analisis Sentimen Aplikasi Ruang Guru Di Twitter Menggunakan Algoritma Klasifikasi. *Jurnal Teknoinfo*, 14(2), 115. <https://doi.org/10.33365/jti.v14i2.679>
- Habibi, R., Setyohadi, D. B., & Wati, E. (2016). Analisis Sentimen Pada Twitter Mahasiswa Menggunakan Metode Backpropagation. *Jurnal Informatika*, 12(1), 103–109. <https://doi.org/10.21460/inf.2016.121.462>

- Hakim, I., Nugroho, A., Sukmana, S. H., & Gata, W. (2020). *Sentimen Analisis Stay Home menggunakan metode klasifikasi Naive Bayes , Support Vector Machine , dan k-Nearest Neighbor*. 22(2), 169–174.
- Karo, I. M. K., Fudzee, M. F. M., Kasim, S., & Ramli, A. A. (2022). Sentiment Analysis in Karonese Tweet using Machine Learning. *Indonesian Journal of Electrical Engineering and Informatics*, 10(1), 219–231. <https://doi.org/10.52549/ijeei.v10i1.3565>
- Koeswara, S., & Muslimah, M. (2016). Analisis Pengaruh Kinerja Pelayanan Frontliner Dan Kepuasan Nasabah Terhadap Loyalitas Nasabah Prioritas Dengan Pendekatan Metode Regresi Linear Multiple. *Sinergi*, 20(1), 21. <https://doi.org/10.22441/sinergi.2016.1.004>
- Lestari, S., Mupaat, M., & Erfina, A. (2022). *Analisis Sentimen Masyarakat Indonesia terhadap Pindahan Ibu Kota Negara Indonesia pada Twitter*. 8(1).
- Maulana, S., & Muhajirin, M. (2021). Analisis Kualitas Pelayanan Terhadap Kepuasan Pelanggan Pada Aisy Cabang Bima. *Jurnal Disrupsi Bisnis*, 4(4), 372. <https://doi.org/10.32493/drj.v4i4.11678>
- Muhammad Romzi, & Kurniawan, B. (2020). Pembelajaran Pemrograman Python Dengan Pendekatan Logika Algoritma. *JTIM: Jurnal Teknik Informatika Mahakarya*, 03(2), 37–44.
- Muis, I. A., Affandes, M., Muis, I. A., Studi, P., Informatika, T., & Sains, F. (2015). *Penerapan Metode Support Vector Machine (SVM) Menggunakan Kernel Radial Basis Function (RBF) Pada Klasifikasi Tweet*. 12(2), 189–197.
- Muludi, K., Akbar, M. S., Shofiana, D. A., & Syarif, A. (2021). Sentiment Analysis Of Energy Independence Tweets Using Simple Recurrent Neural Network. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, 15(4), 339. <https://doi.org/10.22146/ijccs.66016>
- Nasution, M. R. A., & Hayaty, M. (2019). Perbandingan Akurasi dan Waktu Proses Algoritma KNN dan SVM dalam Analisis Sentimen Twitter. *Jurnal*

Informatika, 6(2), 226–235. <https://doi.org/10.31311/ji.v6i2.5129>

- Novantirani, A., Sabariah, M. K., & Effendy, V. (2015). Analisis Sentimen pada Twitter untuk Mengenai Penggunaan Transportasi Umum Darat Dalam Kota dengan Metode Support Vector Machine. *E-Proceeding of Engineering*, 2(1), 1177–1183.
- Nugraha, P. A. (2013). *Perbandingan Metode Probabilistik Naive Bayesian Classifier dan Jaringan Syaraf Tiruan Learning Vector Quantization dalam Kasus Klasifikasi Penyakit Kandungan*. 2(2), 20–33.
- Nurjannah, M., & Astuti, I. F. (2013). *PENERAPAN ALGORITMA TERM FREQUENCY-INVERSE DOCUMENT FREQUENCY (TF-IDF) UNTUK TEXT MINING*. 8(3), 110–113.
- Ogi, E., Pratiwi1, I., & Yustanti2, W. (2021). Analisis Sentimen Kualitas Layanan Teknologi Pembayaran Elektronik pada Twitter (Studi Kasus Ovo dan Dana). *Jeisbi*, 02(03), 47–54.
- Pakpahan, D., & Widyastuti, H. (2014). Aplikasi Opinion Mining dengan Algoritma *Naïve Bayes* untuk Menilai Berita Online. *Jurnal Integrasi*, 6(1), 1–10. <https://sci-hub.do/https://jurnal.polibatam.ac.id/index.php/JI/article/view/105>
- Paulus Hariyono. (2007). Pengertian dan Ruang Lingkup. *Pengertian Dan Ruang Lingkup*, 1, 1–51.
- Pisner, D. A., & Schnyer, D. M. (2019). Support vector machine. *Machine Learning: Methods and Applications to Brain Disorders*, 101–121. <https://doi.org/10.1016/B978-0-12-815739-8.00006-7>
- Regaer, R., Areros, W., & Rogahang, J. (2016). Analisis Pemberian Kredit Mikro Terhadap Peningkatan Nasabah. *Jurnal Administrasi Bisnis 2016*, 4(4), 1–7. <https://ejournal.unsrat.ac.id/index.php/jab/article/view/13827>
- Retnoningsih, E., & Pramudita, R. (2020). Mengenal Machine Learning Dengan Teknik Supervised Dan Unsupervised Learning Menggunakan Python. *Bina Insani Ict Journal*, 7(2), 156. <https://doi.org/10.51211/biict.v7i2.1422>

- Risnantoyo, R., Nugroho, A., & Mandara, K. (2020). *JITE (Journal of Informatics and Telecommunication Engineering)*. 4(1), 86–96.
- Romadoni, F., Umaidah, Y., & Sari, B. N. (2020). Text Mining Untuk Analisis Sentimen Pelanggan Terhadap Layanan Uang Elektronik Menggunakan Algoritma Support Vector Machine. *Jurnal Sisfokom (Sistem Informasi Dan Komputer)*, 9(2), 247–253. <https://doi.org/10.32736/sisfokom.v9i2.903>
- Sadewo, M. G., Windarto, A. P., & Damanik, I. S. (2019). Algoritma *Naïve Bayes* Dalam Memprediksi Kepuasan Nasabah. *Prosiding Seminar Nasional Riset Information Science, September*, 318–327.
- Sari, E. D. N., & Irhamah, I. (2020). Analisis Sentimen Nasabah pada Layanan Perbankan Menggunakan Metode Regresi Logistik Biner, *Naïve Bayes Classifier* (NBC), dan Support Vector Machine (SVM). *Jurnal Sains Dan Seni ITS*, 8(2). <https://doi.org/10.12962/j23373520.v8i2.44565>
- Siringoringo, R. (2018). Klasifikasi data tidak Seimbang menggunakan algoritma SMOTE dan k-nearest neighbor. *Jurnal ISD*, 3(1), 44–49.
- Sodik, F., & Kharisudin, I. (2021). Analisis Sentimen dengan SVM, *NAIVE BAYES* dan KNN untuk Studi Tanggapan Masyarakat Indonesia Terhadap Pandemi Covid-19 pada Media Sosial Twitter. *Prisma*, 4, 628–634.
- Verawardina, U., Edi, F., & Watrianthos, R. (2021). Analisis Sentimen Pembelajaran Daring Pada Twitter di Masa Pandemi COVID-19 Menggunakan Metode *Naïve Bayes*. *Media Informatika Budidarma*, 5, 157–163. <https://doi.org/10.30865/mib.v5i1.2604>
- Vidyandari, A. F., Hutagaol, M. P., & Affandi, J. (2021). Analisis Kepuasan Nasabah Terhadap Pelayanan Frontliner Bank X Kantor Cabang Khusus Sudirman. *Jurnal Aplikasi Bisnis Dan Manajemen*, 7(1), 1–10. <https://doi.org/10.17358/jabm.7.1.1>
- Watanabe, H. (2009). Historical background. *Contemporary Interventional Ultrasonography in Urology*, 1–6. https://doi.org/10.1007/978-1-84800-217-3_1

Wibowati, J. (2021). Pengaruh Kualitas Pelayanan Terhadap Kepuasan Pelanggan Pada Pt Muarakati Baru Satu Palembang. *Jurnal Manajemen*, 8(2), 15–31. <https://doi.org/10.36546/jm.v8i2.348>