

**ANALISIS SENTIMEN PADA MEDIA SOSIAL DENGAN
MENGUNAKAN ALGORITMA *RADIAL BASIS FUNCTION*
NEURAL NETWORKS (RBFNN)**

SKRIPSI

OLEH :

ADE RIAN

188160004



PROGRAM STUDI TEKNIK INFORMATIKA

FAKULTAS TEKNIK

UNIVERSITAS MEDAN AREA

MEDAN

2024

UNIVERSITAS MEDAN AREA

© Hak Cipta Di Lindungi Undang-Undang

Document Accepted 9/7/24

1. Dilarang Mengutip sebagian atau seluruh dokumen ini tanpa mencantumkan sumber
2. Pengutipan hanya untuk keperluan pendidikan, penelitian dan penulisan karya ilmiah
3. Dilarang memperbanyak sebagian atau seluruh karya ini dalam bentuk apapun tanpa izin Universitas Medan Area

Access From (repository.uma.ac.id)9/7/24

**ANALISIS SENTIMEN PADA MEDIA SOSIAL DENGAN
MENGUNAKAN ALGORITMA RADIAL BASIS FUNCTION
NEURAL NETWORKS (RBFNN)**

SKRIPSI

Diajukan sebagai Salah Satu Syarat untuk Memperoleh Gelar Sarjana di Fakultas
Teknik Universitas Medan Area



Oleh :

ADE RIAN

188160004

PROGRAM STUDI TEKNIK INFORMATIKA

FAKULTAS TEKNIK

UNIVERSITAS MEDAN AREA

MEDAN

2024

UNIVERSITAS MEDAN AREA

© Hak Cipta Di Lindungi Undang-Undang

1. Dilarang Mengutip sebagian atau seluruh dokumen ini tanpa mencantumkan sumber
2. Pengutipan hanya untuk keperluan pendidikan, penelitian dan penulisan karya ilmiah
3. Dilarang memperbanyak sebagian atau seluruh karya ini dalam bentuk apapun tanpa izin Universitas Medan Area

Document Accepted 9/7/24

Access From (repository.uma.ac.id)9/7/24

HALAMAN PENGESAHAN

Judul : Analisis Sentimen Pada Media Sosial Dengan Menggunakan Algoritma *Radial Basis Function Neural Networks* (RBFNN)
Nama : Ade Rian
NPM : 188160004
Fakultas : Teknik

Disetujui Oleh
Komisi Pembimbing


Dr. Rahmad Syah, S.Kom, M.Kom
Pembimbing



Supriatno, S.T, M.T.
Dekan Fakultas Teknik



Rizki Muliono, S.Kom, M.Kom
Ka. Prodi

Tanggal Lulus :

HALAMAN PERNYATAAN

Saya menyatakan bahwa skripsi yang saya tulis, sebagai syarat memperoleh gelar sarjana merupakan hasil karya tulis saya sendiri. Adapun bagian-bagian dalam penulisan skripsi ini yang saya kutip dari hasil karya orang lain dan telah dituliskan sumbernya secara jelas sesuai norma, kaidah dan etika penulisan ilmiah.

Saya bersedia menerima sanksi pencabutan gelar akademik yang saya peroleh dan sanksi lainnya sesuai dengan peraturan yang berlaku, apabila dikemudian hari ditemukan adanya plagiat dalam skripsi ini.

Medan, 13 Juni 2024



Ade Rian
188160004

HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI UNTUK KEPENTINGAN AKADEMIS

Sebagai sivitas akademik Universitas Medan Area, saya yang bertanda tangan dibawah ini :

Nama : Ade Rian
NPM : 181860004
Program Studi : Teknik Informatika
Fakultas : Teknik
Jenis karya : Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Medan Area Hak Bebas Royalti Noneksklusif (Non Exclusive Royalty Free-Right) atas karya ilmiah saya yang berjudul :

Analisis Sentimen Pada Media Sosial Dengan Menggunakan Algoritma Radial Basis Function Neural Networks (RBFNN).

Beserta perangkat yang ada (jika diperlukan). Dengan hak Bebas Royalti Noneksklusif ini Universitas Medan Area berhak menyimpan, mengalihmedia /formatkan, mengelola dalam bentuk pangkalan data (database), merawat, dan mempublikasikan skripsi saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Medan
Pada tanggal : 13 Juni 2024
Yang menyatakan



(Ade Rian)

ABSTRAK

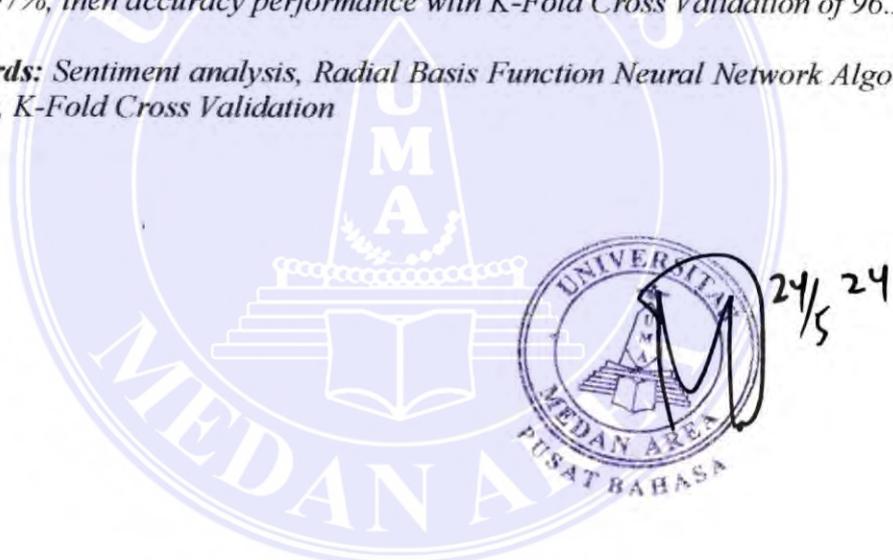
Analisis sentimen atau *opinion mining* merupakan prosedur mengetahui, mengekstraksi, dan mengolah informasi teks secara otomatis untuk memperoleh data sentimen yang terdapat pada kalimat opini. Prosedur analisis data bisa dilakukan melewati teknik *machine learning*. Pada penelitian ini, analisis sentimen dilakukan guna mengetahui pemahaman ataupun trend persepsi kepada perbankan syariah. Analisis sentimen membantu mengidentifikasi komentar atau pendapat yang mempunyai kecenderungan sentimen negatif atau positif, yang dapat dijadikan acuan untuk pembaruan layanan. Dalam penelitian ini, data yang diperoleh sebanyak 1837 dengan membagi 1768 positif dan 69 negatif dari *Twitter* dengan teks berbahasa Indonesia yang terkait dengan bank syariah Indonesia. Hasilnya dengan menggunakan proporsi data *training* 60% dan *testing* 40% menghasilkan performa akurasi dengan algoritma RBFNN sebesar 96,8%, presisi 96,8%, dan *recall* 96,8%, selanjutnya performa akurasi dengan *K-Fold Cross Validation* sebesar 96,2%. Proporsi data *training* 70% dan *testing* 30% menghasilkan performa akurasi dengan algoritma RBFNN sebesar 96,3%, presisi 96,3%, dan *recall* 96,3%, selanjutnya performa akurasi dengan *K-Fold Cross Validation* sebesar 96,2%. Kemudian menggunakan proporsi data *training* 80% dan *testing* 20% menghasilkan performa akurasi dengan algoritma RBFNN sebesar 97%, presisi 97%, dan *recall* 97%, selanjutnya performa akurasi dengan *K-Fold Cross Validation* sebesar 96,2%.

Kata Kunci : Analisis sentimen, Algoritma *Radial Basis Function Neural Network*, *Twitter*, *K-Fold Cross Validation*

ABSTRACT

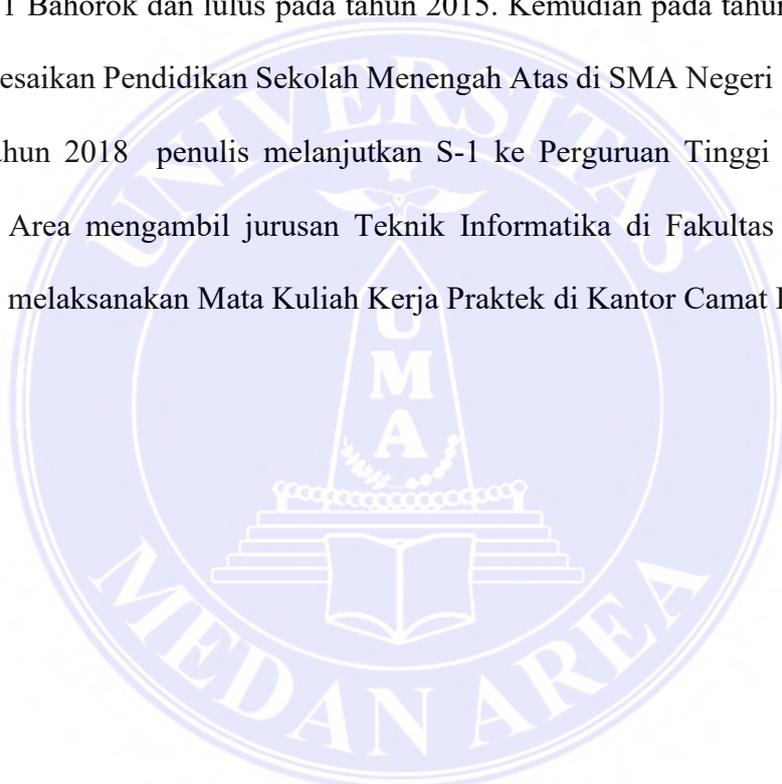
Sentiment analysis or opinion mining is the procedure of automatically knowing, extracting, and processing text information to obtain sentiment data contained in opinion sentences. The data analysis procedure could be done through machine learning techniques. In this research, sentiment analysis was conducted to determine the understanding or trend of perception towards Islamic banking. Sentiment analysis helps identify comments or opinions that have a tendency of negative or positive sentiment, which can be used as a reference for service updates. In this study, 1837 data were obtained by dividing 1768 positive and 69 negative from Twitter with Indonesian text related to Indonesian Islamic banks. The results using the proportion of 60% training and 40% testing data resulted in accuracy performance with the RBFNN algorithm of 96.8%, precision of 96.8%, and recall of 96.8%, then accuracy performance with K-Fold Cross Validation of 96.2%. The proportion of 70% training data and 30% testing resulted in accuracy performance with the RBFNN algorithm of 96.3%, precision of 96.3%, and recall of 96.3%, then accuracy performance with K-Fold Cross Validation of 96.2%. Then using the proportion of 80% training data and 20% testing resulted in accuracy performance with the RBFNN algorithm of 97%, precision 97%, and recall 97%, then accuracy performance with K-Fold Cross Validation of 96.2%.

Keywords: *Sentiment analysis, Radial Basis Function Neural Network Algorithm, Twitter, K-Fold Cross Validation*



RIWAYAT HIDUP

Penulis dilahirkan di Binjai, pada tanggal 16 Mei 2000. Merupakan anak Keempat dari empat bersaudara, dari pasangan Pasti dan Rehulina br Sitepu. Pada tahun 2012, penulis menyelesaikan Pendidikan Sekolah Dasar di SD Negeri 055969, penulis melanjutkan Pendidikan Sekolah Menengah Pertama di SMP Negeri 1 Bahorok dan lulus pada tahun 2015. Kemudian pada tahun 2018 penulis menyelesaikan Pendidikan Sekolah Menengah Atas di SMA Negeri 1 Bahorok dan pada tahun 2018 penulis melanjutkan S-1 ke Perguruan Tinggi di Universitas Medan Area mengambil jurusan Teknik Informatika di Fakultas Teknik. Lalu, Penulis melaksanakan Mata Kuliah Kerja Praktek di Kantor Camat Bahorok.



KATA PENGANTAR

Puji syukur penulis panjatkan kepada Allah SWT, karena atas nikmat dan rahmat-Nya sehingga penulis dapat menyelesaikan skripsi ini. Penulisan skripsi ini dilakukan dalam rangka memenuhi salah satu syarat untuk mencapai gelar Sarjana Komputer Program Studi Teknik Informatika Fakultas Teknik Universitas Medan Area. Proses penyelesaian skripsi ini tidak lepas dari berbagai bantuan, dukungan, saran, dan kritik yang telah penulis dapatkan, oleh karena itu dalam kesempatan ini penulis ingin mengucapkan terima kasih kepada:

1. Tuhan Yang Maha Esa, karena berkat-Nya yang melimpah sehingga penulis dapat menyelesaikan skripsi ini dengan baik.
2. Kedua Orang Tua yang mendoakan serta mendukung penulis dalam mengerjakan skripsi.
3. Bapak Prof. Dr. Dadan Ramdan, M.Eng, M.Sc selaku Rektor Universitas Medan Area.
4. Bapak Dr. Eng., Supriatno, S.T, M.T selaku Dekan Fakultas Teknik Universitas Medan Area.
5. Bapak Rizki Muliono, S.Kom, M.Kom selaku Ketua Program Studi Teknik Informatika.
6. Bapak Dr. Rahmad Syah, S.Kom, M.Kom selaku Dosen Pembimbing yang telah memberikan bimbingan, motivasi, dan arahan kepada penulis sehingga skripsi ini bisa selesai dengan baik.
7. Seluruh Dosen, Staf Karyawan Fakultas Teknik, khususnya Program Studi Teknik Informatika yang telah memberikan bantuan dan kerjasama dari awal perkuliahan.
8. Kepada teman seperjuangan Teknik Informatika angkatan 2018 yang sudah membantu penulis dalam menyelesaikan skripsi ini, terima kasih atas dukungannya. Semoga kita bisa lebih baik lagi dan sukses di masa yang akan datang.
9. Seluruh pihak yang secara langsung maupun tidak langsung membantu penulis dalam menyelesaikan skripsi ini.

Akhir kata, penulis menyadari bahwa dalam penyajian skripsi ini masih jauh dari sempurna. Apabila ada kebenaran dari penulisan ini maka kebenaran tersebut datangnya dari Allah, tetapi apabila ada kesalahan dalam penulisan ini maka kesalahan ini berasal dari penulis. Semoga skripsi ini membawa manfaat bagi pengembangan ilmu. Penulis berharap Allah SWT berkenan membalas segera kebaikan semua pihak yang telah membantu segala usaha kita.

Medan, 13 Juni 2024



Ade Rian
188160004



DAFTAR ISI

HALAMAN PENGESAHAN	Error! Bookmark not defined.
HALAMAN PERNYATAAN	Error! Bookmark not defined.
ABSTRAK	iv
ABSTRACT	Error! Bookmark not defined.
RIWAYAT HIDUP	vi
KATA PENGANTAR	vii
DAFTAR ISI	ix
DAFTAR GAMBAR	xi
DAFTAR TABEL	xii
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Tujuan Penelitian.....	3
1.4 Manfaat Penelitian.....	3
1.5 Batasan Masalah.....	3
BAB II TINJAUAN PUSTAKA	5
2.1 Analisis Sentimen.....	5
2.2 Sosial Media.....	5
2.3 <i>Cross Validation</i>	6
2.4 <i>Text Mining</i>	6
2.5 <i>Text Preprocessing</i>	7
2.5.1 <i>Case Folding</i>	7
2.5.2 <i>Tokenization</i>	7
2.5.3 <i>Stopwords</i>	7
2.5.4 <i>Stemming</i>	8
2.5.5 <i>TF-IDF</i>	8
2.6 <i>Algoritma Radial Basis Function Neural Networks</i>	9
2.7 <i>Confusion Matrix</i>	11
2.8 Penelitian Terdahulu	13

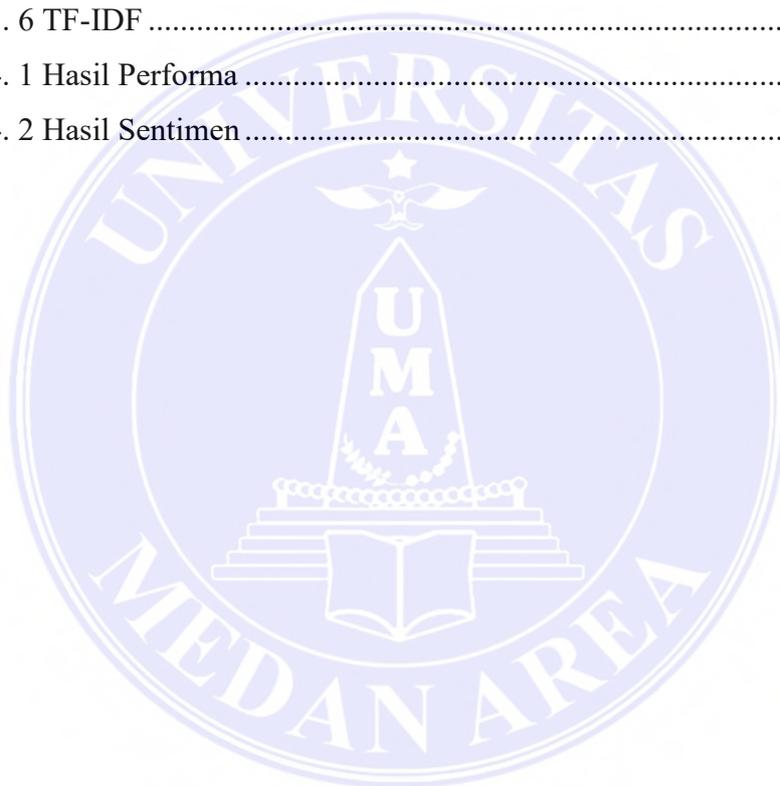
BAB III METODOLOGI PENELITIAN	19
3.1 Data dan Sumber Data.....	19
3.1.1 Pengumpulan Data.....	20
3.2 Teknik Analisis Data.....	21
3.3 <i>Text Preprocessing</i>	22
3.3.1 <i>Case Folding</i>	22
3.3.2 <i>Tokenization</i>	23
3.3.3 <i>Stopwords</i>	23
3.3.4 <i>Stemming</i>	24
3.3.5 TF-IDF	25
3.4 Melatih Algoritma RBFNN Menggunakan <i>Backpropagation</i>	27
BAB IV HASIL DAN PEMBAHASAN	33
4.1 Hasil	33
4.1.1 Evaluasi.....	34
4.2 Pembahasan.....	36
BAB V SIMPULAN DAN SARAN	41
5.1 Simpulan.....	41
5.2 Saran.....	41
DAFTAR PUSTAKA.....	42
LAMPIRAN.....	47

DAFTAR GAMBAR

Gambar 2. 1 Cross Validation.....	6
Gambar 2. 2 Arsitektur RBFNN	11
Gambar 3. 1 Alur Penelitian.....	19
Gambar 3. 2 Developer Portal API	20
Gambar 3. 3 Tampilan Data	21
Gambar 3. 4 Hasil Case Folding	23
Gambar 3. 5 Hasil Tokenization	23
Gambar 3. 6 Hasil Stopwords	24
Gambar 3. 7 Hasil Stemming.....	25
Gambar 3. 8 Hasil TF-IDF	26
Gambar 3. 9 Arsitektur RBFNN	27
Gambar 3. 10 Pseudocode RBFNN	31
Gambar 3. 11 Algoritma RBFNN	31
Gambar 3. 12 K-Fold Cross Validation	32
Gambar 4. 1 Kurva Performa	33
Gambar 4. 2 Confusion Matrix 60% : 40%	34
Gambar 4. 3 Confusion Matrix 70% : 30%	35
Gambar 4. 4 Confusion Matrix 80% : 20%	36
Gambar 4. 5 Teknik Scraping Data Melalui API.....	37
Gambar 4. 6 Codingan Case Folding	38
Gambar 4. 7 Codingan Tokenization	38
Gambar 4. 8 Codingan Stopwords	39
Gambar 4. 9 Codingan Stemming.....	39
Gambar 4. 10 Codingan TF-IDF	40
Gambar 4. 11 Codingan Confusion Matrix.....	40
Gambar 4. 12 Visualisasi Sentimen Positif dan Negatif.....	40

DAFTAR TABEL

Tabel 2. 1 Confusion Matrix	12
Tabel 2. 2 Penelitian Terdahulu	13
Tabel 3. 1 Data Sampel	20
Tabel 3. 2 Case Folding	22
Tabel 3. 3 Tokenization.....	23
Tabel 3. 4 Stopwords	24
Tabel 3. 5 Stemming	24
Tabel 3. 6 TF-IDF	25
Tabel 4. 1 Hasil Performa	33
Tabel 4. 2 Hasil Sentimen.....	41



BAB I

PENDAHULUAN

1.1 Latar Belakang

Perbankan ialah instansi keuangan yang memberikan pelayanan dan berperan selaku penghubung antara pihak yang kelebihan anggaran serta pihak yang kesulitan anggaran. Dalam fungsinya menjadi perantara, bank berfungsi mengelola anggaran tabungan masyarakat untuk disalurkan terhadap pihak yang membutuhkan (Nurlina, 2019). Dalam struktur perbankan Indonesia, terdapat dua jenis sistem perbankan yang dikenal, yaitu bank konvensional dan bank syariah. Berdasarkan Undang-Undang Nomor 21 Tahun 2008 tentang Perbankan Syariah, sistem perbankan syariah berlandaskan prinsip-prinsip hukum Syariah dan prinsip-prinsip hukum Islam yang telah diatur dalam Fatwa Majelis Ulama Indonesia. Prinsip-prinsip ini mencakup prinsip keadilan dan keseimbangan (*adl wa tawazun*), manfaat bersama (*maslahah*), universalisme (*alamiyah*), serta larangan praktik-praktik berisiko tinggi seperti *gharar*, *masyir*, bunga, penindasan, dan barang haram. (Ramadani & Nurmala Hadiani, 2022). Bank syariah merupakan bank yang melakukan kegiatan usaha sesuai menggunakan kaidah syariah, yaitu kesepakatan berlandaskan syariah antara bank dengan pihak lain untuk menyimpan dan/atau membiayai kegiatan usaha ataupun kegiatan berbeda yang dinyatakan selaras dengan syariah (Saputra, 2021). Penerapan bunga bank konvensional tidak selaras dengan nilai mendasar mekanisme keuangan Islam. (Muba dkk., 2023) Sebagai pengganti bunga, bank syariah mempraktekkan bagi hasil (rugi dan bagi hasil). Secara umum bank dapat dipahami sebagai instansi keuangan yang jenis kegiatan utamanya adalah mengumpulkan dana dari masyarakat selanjutnya memberikan produk pelayanan lainnya terhadap pihak yang memerlukan berupa kredit ataupun pembiayaan. (Fatihudin & Firmansyah, 2019).

Analisis sentimen atau *opinion mining* merupakan prosedur mengetahui, mengekstraksi, dan mengolah informasi teks secara otomatis untuk memperoleh data sentimen yang terdapat pada kalimat opini. Prosedur analisis data bisa dilakukan melewati teknik *machine learning*. (Widowati & Sadikin, 2021).

Analisis sentimen membantu mengidentifikasi komentar atau pendapat yang mempunyai kecenderungan sentimen negatif atau positif, yang dapat dijadikan acuan untuk pembaruan layanan. (Suryono & Taufiq Luthfi, 2021). Penelitian terdahulu yang berkaitan dengan analisis sentimen terhadap Bank Syariah Indonesia yaitu (Arifiyanti dkk., 2022) menggunakan data ulasan berbahasa Indonesia dari *Google Play Store* pada aplikasi *mobile* PT Bank syariah Indonesia dengan metode *supervised learning* mendapatkan akurasi tertinggi sebesar 82%. Penelitian lain seperti (Mustofa dkk., 2022) Menggunakan metode *Convenience Sampling*, data primer dari 100 responden digunakan untuk analisis. Analisis data dilakukan dengan menerapkan metode regresi linier berganda. Hasil uji t menunjukkan bahwa variabel potensi, ketika dianalisis secara parsial, tidak memiliki pengaruh signifikan pada persepsi mengenai merger BSI. Sebaliknya, preferensi dan tingkat religiusitas, ketika dianalisis secara parsial, terbukti memiliki pengaruh signifikan pada persepsi masyarakat.

Berlandaskan permasalahan di atas, peneliti mengkaji sentimen positif dan negatif opini *Twitter* kepada Bank Syariah di Indonesia dengan menerapkan algoritma jaringan syaraf tiruan *Radial Basis Function*, dengan mempelajari akurasi dan persentase masing-masing sentimen. Di Indonesia, *Twitter* tidak hanya digunakan sebagai platform untuk mengungkapkan perasaan, tetapi juga sebagai sarana untuk memberikan saran dan kritik terhadap objek, layanan, atau kebijakan tertentu. (Rosdiana dkk., 2019). Pengguna *Twitter* juga mengemukakan pendapat mereka tentang kebijakan tersebut, yang kemudian menjadi topik yang sedang populer. Dalam penelitian ini, data yang diperoleh dari *Twitter* digunakan untuk mengklasifikasikan teks berbahasa Indonesia yang terkait dengan bank syariah Indonesia. Jaringan Syaraf Tiruan (*Artificial Neural Network*) merupakan salah satu teknik dalam ilmu kecerdasan buatan yang bertujuan untuk meniru kemampuan berpikir manusia dengan menerapkannya pada mesin atau komputer. Jaringan syaraf tiruan mampu melakukan tugas-tugas seperti yang dilakukan oleh manusia dalam konteks tertentu. Ada berbagai model Jaringan Syaraf Tiruan yang telah diterapkan secara luas dalam berbagai aplikasi, termasuk klasifikasi dan prediksi. Salah satu jenis Jaringan Syaraf Tiruan yang sering digunakan adalah *Radial Basis Function Neural Network* (RBFNN). (Hayqal dkk., 2022). *Radial Basis Function*

(RBF) merupakan pengembangan dari jaringan saraf *feedforward*. Jaringan Saraf Tiruan dengan Fungsi Basis Radial (RBFNN) adalah jenis jaringan saraf yang memanfaatkan fungsi basis radial dalam lapisan tersembunyi untuk melakukan tugas klasifikasi dan peramalan. (AISYAH dkk., 2021). Jaringan RBF terdiri dari tiga lapisan, yaitu lapisan *input*, lapisan tersembunyi, dan lapisan *output* (Hashemi dkk., 2019). RBFNN memiliki struktur jaringan yang sederhana namun memiliki kemampuan konvergensi yang sangat cepat serta kemampuan pendekatan non-linier yang tinggi (Chen dkk., 2019). Adapun alasan peneliti menggunakan algoritma RBFNN adalah arsitekturnya yang populer digunakan dalam klasifikasi. Berdasarkan latar belakang yang telah diuraikan, penelitian tertarik membahas mengenai “Analisis Sentimen Pada Media Sosial Dengan Menggunakan Algoritma *Radial Basis Function Neural Networks* (RBFNN)”.

1.2 Rumusan Masalah

Berdasarkan latar belakang diatas, maka dapat dibuat rumusan masalah yaitu bagaimana hasil performa pada tingkat akurasi, presis, dan *recall* serta hasil analisis sentimen terhadap Bank Syariah Indonesia pada *Twitter* dengan menggunakan algoritma *Radial Basis Function Neural Networks* ?

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah diatas dapat dibuat tujuan penelitian yaitu dapat mengetahui hasil analisis sentimen terhadap Bank Syariah Indonesia pada *Twitter* dengan menggunakan algoritma *Radial Basis Function Neural Networks*.

1.4 Manfaat Penelitian

Adapun manfaat dari penelitian ini adalah sebagai berikut :

1. Sebagai acuan kepada Bank Syariah Indonesia agar dapat mengetahui bagaimana pendapat nasabah terhadap layanannya.
2. Dapat memberikan alternatif algoritma dalam melakukan analisis sentimen menggunakan *Radial Basis Function Neural Network*.

1.5 Batasan Masalah

Adapun batasan masalah pada penelitian in, yaitu:

1. Teknik pengumpulan data yaitu *scraping data* dengan bantuan API *Twitter*.

2. Jumlah data sebanyak 1837 dengan menggunakan kata kunci “bank syariah indonesia” dari tanggal 11 bulan juni tahun 2023.
3. Sentimen yang diambil merupakan layanan dari bank syariah indonesia.
4. Menggunakan algoritma *Radial Basis Function Neural Networks*.
5. Tahapan *text preprocessing* yaitu *case folding*, *tokenization*, *stopwords*, *stemming*, dan TF-IDF.
6. Membagi sentimen menjadi 2 kelas yaitu positif dan negatif.
7. Proporsi data *training* 60% dan *testing* 40%, data *training* 70% dan *testing* 30%, serta data *training* 80% dan *testing* 20%.
8. Menggunakan bahasa pemrograman *Python*.
9. *Tools* yang digunakan adalah *google colaboratory*.



BAB II TINJAUAN PUSTAKA

2.1 Analisis Sentimen

Analisis sentimen adalah salah satu cabang penelitian yang memanfaatkan metode analisis untuk menganalisis perasaan ataupun tanggapan seseorang terhadap kejadian maupun peristiwa yang dibicarakan secara luas dalam data teks yang diperoleh melalui metode *text mining* (Rusardi dkk., 2022). Analisis sentimen adalah teknik untuk mengekstraksi data pendapat, mengerti serta mengolah data teks secara otomatis untuk melihat sentimen yang terdapat dalam opini (Sari & Wibowo, 2019). Tujuan analisis sentimen adalah mengevaluasi opini dan arah pandangan terhadap suatu topik, termasuk yang bersifat negatif dan positif. Teknik analisis sentimen dapat diterapkan pada berbagai konteks, seperti dalam ranah ekonomi, politik, sosial, dan hukum. Dalam konteks media sosial seperti *Twitter*, platform ini membuka peluang bagi para peneliti untuk menggali pemahaman tentang emosi, mood, serta pandangan masyarakat dengan menggunakan analisis sentimen. (Arsi & Waluyo, 2021).

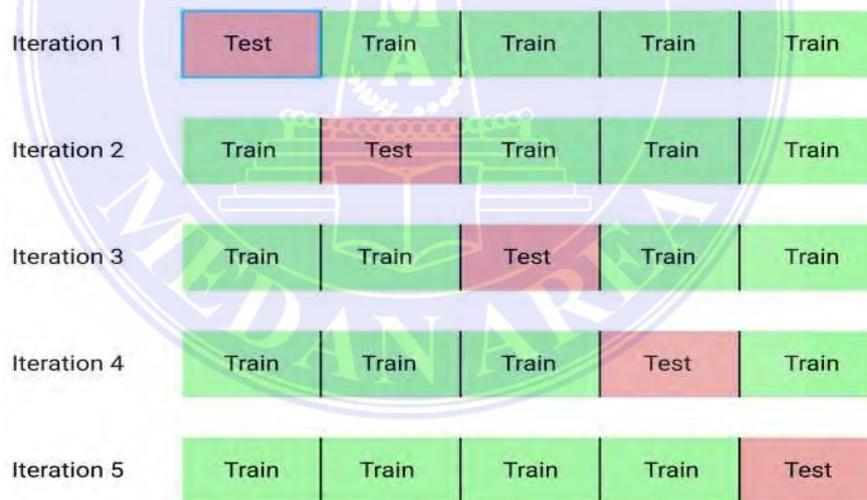
2.2 Sosial Media

Media sosial atau biasa kita sebut dengan singkatan SOSMED adalah sarana elektronik yang biasa digunakan untuk berhubung dengan berbagai pengguna, berbagi salinan seperti foto dan video (Haryanto, 2020). Media sosial adalah media digital yang digunakan untuk mempresentasikan diri, berbagi informasi, berhubung dan bekerja sama dengan orang lain (Rahayu dkk., 2022). Pengguna internet di Indonesia memperoleh 76,8% pada Maret 2021 yang berarti 212,35 juta dari total penduduk 276,2 juta (Yulita, 2021). *Twitter* merupakan salah satu aplikasi sarana sosial paling terkenal di Indonesia. *Twitter* merupakan alat *microblogging*, tempat untuk berbagi pengetahuan tanpa kendala serta secara *real time* (Mufidah dkk., 2022). Mengakses *Twitter* itu mudah karena *tweet* pada akun pengguna yang direkam bersifat umum. Hal ini dapat menyebabkan *Twitter* sebagai sumber data opini, bermanfaat dalam berbagai keperluan seperti sarana belajar, sarana bertukar pandangan, aksi politik, media komunikasi terkait kepuasan, kualitas produk, kualitas layanan, dan lain-lain, yang sangat bernilai

Mengerti persepsi pelanggan beserta pelanggan karakteristik bagi perusahaan sebab dapat dinilai untuk mengenali langkah-langkah strategis untuk meningkatkan layanan dan memahami kebutuhan pelanggan (Fiktri dkk., 2020).

2.3 Cross Validation

Cross Validation yang juga dikenal sebagai teknik estimasi bergilir, merupakan metode validasi model yang digunakan untuk mengukur sejauh mana hasil analisis statistik dapat diaplikasikan secara umum pada kumpulan data yang independen. Teknik ini terutama berguna untuk mengukur kinerja model prediksi dan memperkirakan akurasi prediksi ketika diimplementasikan dalam situasi praktis. Salah satu pendekatan yang umum digunakan adalah validasi silang dengan metode k-fold, yang membagi data menjadi K subset dengan ukuran yang setara. Dalam penelitian ini, kami mengadopsi pendekatan validasi silang k-fold dengan K=5 untuk mengurangi potensi bias dalam data. Pelatihan dan pengujian model dilakukan dalam K iterasi (Azis dkk., 2020). Ilustrasi pendekatan ini dapat dilihat dalam Gambar 2.1.



Gambar 2. 1 *Cross Validation*

Sumber: (Azis dkk., 2020)

2.4 Text Mining

Text mining merupakan teknik yang bisa digunakan untuk mengatasi masalah klasifikasi, pengelompokan, ekstraksi data, serta pencarian informasi. Penambangan teks ialah tahap pencarian informasi di mana pengguna berhubung dengan sekumpulan dokumen dari waktu ke waktu memanfaatkan seperangkat

alat analisis. Prosedur mendasar *text mining* adalah mendapatkan kata-kata yang bisa menggantikan isi dokumen, kemudian melewati prosedur *review* untuk melihat *link* dari setiap dokumen (Afdal & Rahma Elita, 2022).

2.5 Text Preprocessing

Text processing dilakukan untuk seleksi data yang akan diproses pada dokumen. Guna memaksimalkan output *preprocessing* dari penelitian sebelumnya. (Arsi & Waluyo, 2021). Tahap ini merupakan persiapan data untuk dianalisis lebih mendalam di dalam prosedur *text mining*. *Preprocessing* ini mempunyai beberapa tahap, antara lain *case folding*, *tokenization*, *stopwords*, dan *stemming* (Agustina dkk., 2021).

2.5.1 Case Folding

Langkah pengubahan huruf kecil (*case folding*) adalah proses penting dalam proses pembersihan data yang bertujuan untuk menghilangkan karakter yang tidak relevan dari teks dokumen. Karakter-karakter yang dihapus mencakup tanda baca, *username*, *URL*, *mention*, *hashtag*, *retweet*, dan juga simbol atau karakter numerik, seperti: (,,"~&?!><#%{ }([0-9]+;:"') [1122]. Pemakaian kata-kata di dalam dokumen, tidak semua huruf tertulis dikapitalisasi secara konsisten. Oleh karena itu, diperlukan prosedur pelipatan huruf besar guna mengganti semua teks huruf besar di semua tanggapan yang ada dalam dokumen menjadi huruf kecil. *Case folding* cuma mendeteksi huruf 'a' hingga huruf 'z' (Jabal Tursina, 2019).

2.5.2 Tokenization

Tahap tokenisasi adalah fase dimana string input dipotong sesuai dengan kata-kata penyusun (Nugraha dkk., 2020). Dalam fase tokenisasi ini, kalimat diiris menjadi kata-kata individual. Prosedur ini dilakukan untuk memperoleh kata-kata yang bernilai. Tokenisasi juga memudahkan untuk menemukan frekuensi informasi di dalam suatu korpus. (Jabal Tursina, 2019).

2.5.3 Stopwords

Tahap penghapusan kata-kata yang kurang berarti atau tidak relevan dalam data dokumen dikenal sebagai tahap eliminasi stop words. Stop words mencakup kata-kata yang sering digunakan tetapi tidak memberikan deskripsi atau nilai tambah yang signifikan, sehingga kata-kata tersebut dapat dihilangkan dari analisis,

seperti kata “yang”, “itu”, “atau”, “dan”, “dengan”, “di”, “dari” dan sebagainya (Jabal Tursina, 2019).

2.5.4 Stemming

Stemming adalah tahap dimana pencarian kata dasar dilakukan dari setiap kata yang dihasilkan dengan metode *filtering*. Pada tahap *stemming*, seluruh imbuhan yang terdapat pada data dokumen dihilangkan. Penghilangan afiks mencakup awalan, akhiran, dan kombinasi awalan (konfiks) (Jabal Tursina, 2019).

- Prefiks adalah imbuhan pada kata dasar bahasa Indonesia bagian pertama, antara lain: per-, ke-, se-, me-, ter-, di-.
- Akhiran (sufiks) adalah awalan yang dibubuhkan di akhir kata bahasa Indonesia dan meliputi: -an, -i, -kan, dan -nya.
- Kombinasi awalan dan akhiran bahasa Indonesia dari konfiks (kata majemuk) meliputi: di-kan, di-i, ter-kan, ter-i, ke-an, se-nya, ber-kan, ber-an, pe -an, per-an, per-kan, per-i, me-kan, me-i, dll.

2.5.5 TF-IDF

Pembobotan TF-IDF (*Term Frequency-Inverse Document Frequency*) adalah prosedur penghitungan bobot setiap kata yang terdapat pada data dokumen. Data dokumen akan dikonversi menjadi vektor sebanyak kata (istilah) untuk klasifikasi. TF-IDF digunakan untuk dapat melakukan analisis menggunakan algoritma RBFNN. Adapun rumus TF-IDF ialah sebagai berikut (Mayang Sari, 2021):

$$idf = \log \frac{N}{df} \quad (2.1)$$

$$w(k, d) = tf(k, d) * idf$$

$$w(k, d) = tf(k, d) * \log \frac{N}{df} \quad (2.2)$$

Keterangan :

$W(k,d)$ = Bobot kata (*term*) yang tersedia dalam dokumen

$tf(k,d)$ = jumlah frekuensi munculnya kata didalam dokumen

N = jumlah seluruh dokumen yang tersedia didalam *database*

df = jumlah dokumen yang mengandung *term*

2.6 Algoritma *Radial Basis Function Neural Networks*

Jaringan syaraf tiruan *Radial Basis Function* merupakan salah satu teknik pengembangan jaringan syaraf tiruan yang pelatihannya memakai teknik pembelajaran terawasi. Jenis jaringan pada RBFNN hampir mirip dengan jaringan syaraf tiruan *Multilayer Perceptron* (Wahyuningrum, 2020). Keunggulan dari RBFNN ini merupakan mempunyai perulangan komputasi yang lebih cepat dibandingkan dengan teknik lain untuk jaringan syaraf tiruan. RBFNN memakai fungsi linier antara input dan lapisan tersembunyi. Jumlah lapisan tersembunyi di RBFNN diperbarui secara berurutan hingga kesalahan target tercapai. Untuk setiap neuron pada lapisan tersembunyi yang ditambahkan, jarak antara vektor *input* dan vektor *output* dihitung menerapkan norma *Euclidean*. Kemudian, neuron diaktifkan menggunakan fungsi aktivasi sebagai jenis umum dari radial basis fungsi (RBF) (Madhiarasan, 2020).

RBFNN merupakan jaringan saraf tiga lapis dengan banyak neuron di lapisan tersembunyi. Pada RBFNN, fungsi aktivasi basis radial nonlinier ditugaskan ke lapisan tersembunyi, sedangkan fungsi linier ditugaskan ke lapisan keluaran. Menyerupai dengan MLPNN, RBFNN dapat dilatih menggunakan algoritma *backpropagation* yang diawasi (Kermani dkk., 2020). Adapun tahapan propagasi balik adalah sebagai berikut :

1. Inisialisasi bobot dengan nilai acak range kecil.
2. Fase Propagasi maju :
 - a.) Hitung keluaran unit tersembunyi (z_j):

$$z_{net_j} = v_{j0} + \sum_{i=1}^n x_i v_{ji} \quad (2.3)$$

$$z_i = f(z_{net_j}) = \frac{1}{1 + e^{-z_{net_j}}} \quad (2.4)$$

- b.) Hitung keluaran unit y_k :

$$y_{net_k} = w_{k0} + \sum_{j=1}^p z_j w_{kj} \quad (2.5)$$

$$y_k = f(y_{net_k}) = \frac{1}{1 + e^{-y_{net_k}}} \quad (2.6)$$

3. Fase Propagasi Mundur :

Hitung faktor δ di unit keluaran y_k

$$\delta_k = (t_k - y_k) f'(y_{net_k}) = (t_k - y_k) y_k (1 - y_k), \quad t_k = \text{Target} \quad (2.7)$$

$$\Delta w_{kj} = \alpha \delta_k z_j, \quad k = 1, 2, \dots, m; j = 1, 2, \dots, p \quad (2.8)$$

4. Hitung penjumlahan kesaalahn dari unit tersembunyi (δ):

$$\delta_{net_j} = \sum_{k=1}^m \delta_k w_{kj} \quad (2.9)$$

Faktor kesalahan δ di unit tersembunyi :

$$\delta_j = \delta_{net_j} f'(z_{net_j}) = \delta_{net_j} z_j (1 - z_j) \quad (2.10)$$

Suku perubahan bobot ke unit tersembunyi :

$$\Delta v_{ji} = \alpha \delta_j x_i, \quad j = 1, 2, \dots, p; i = 1, 2, \dots, n \quad (2.11)$$

5. Hitung semua perubahan bobot :

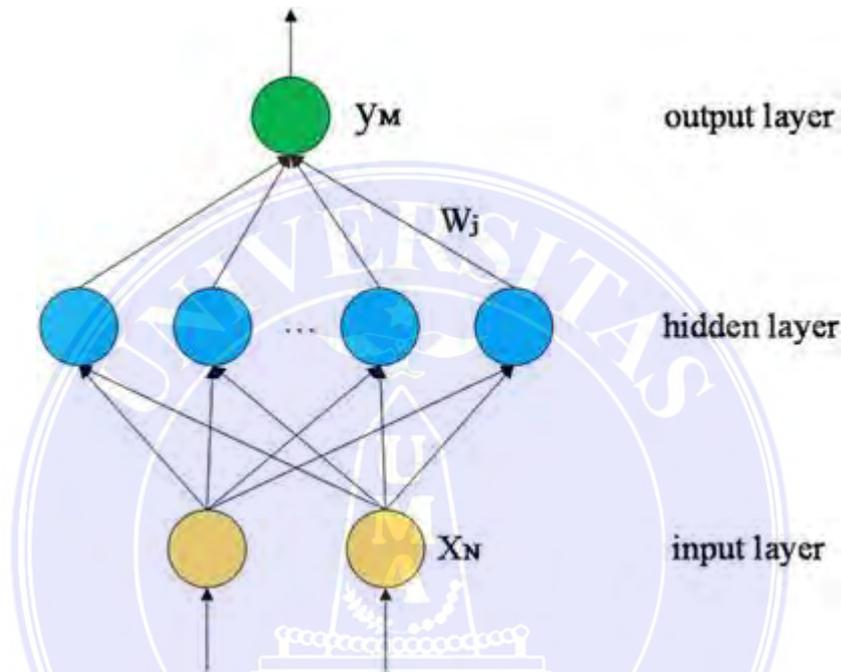
Perubahan bobot unit keluaran :

$$w_{kj}(\text{baru}) = w_{kj}(\text{lama}) + \Delta w_{kj}, \quad (k = 1, 2, \dots, m; j = 1, 2, \dots, p) \quad (2.12)$$

Perubahan bobot unit tersembunyi :

$$v_{ji}(\text{baru}) = v_{ji}(\text{lama}) + \Delta v_{ji}, \quad (j = 1, 2, \dots, p; i = 1, 2, \dots, n) \quad (2.13)$$

RBFNN mempunyai 3 lapisan, yaitu lapisan masukan, satu lapisan tersembunyi dan lapisan keluaran. RBFNN mempunyai bobot antara lapisan tersembunyi dan lapisan keluaran, fungsi aktivasi dapat menghasilkan nilai berupa persamaan nonlinier pada lapisan tersembunyi, sebaliknya nilai yang dihasilkan oleh metode RBFNN pada lapisan keluaran berupa persamaan linier (Prastowo dkk., 2021).



Gambar 2. 2 Arsitektur RBFNN

Sumber: (Hashemi dkk., 2019)

Ukuran dimensi lapisan tersembunyi menjadi besar karena setiap neuron dalam lapisan input terhubung secara langsung ke lapisan tersebut, dan tiap neuron tersembunyi dalam lapisan tersebut memiliki parameter, seperti lebar dan posisi tengah. Selain itu, semua neuron tersembunyi dalam lapisan tersembunyi juga menggunakan fungsi aktivasi. (Sembiring & Sondang, 2019).

2.7 Confusion Matrix

Confusion matrix merupakan matriks yang digunakan untuk menguji prosedur model klasifikasi dalam hal total data uji yang benar dan salah. Menggunakan matriks ini, kita dapat memahami kinerja dari model klasifikasi. (Normawati & Prayogi, 2021). Matriks ini berisi perhitungan data target

dibandingkan dengan data target sebenarnya. Data prediksi adalah nilai yang terdapat dari hasil pemodelan *machine learning*, sebaliknya data aktual ialah nilai nyata yang Anda miliki. Adanya *confusion matrix* untuk memahami seberapa baik pembelajaran mesin bekerja seperti yang diharapkan. *Confusion matrix* berisi berbagai properti ternilai seperti akurasi, presisi, dan *recall* guna melihat seberapa baik pemodelan tersebut. (Irwansyah, 2022). Contoh *confusion matrix* (Normawati & Prayogi, 2021).

Tabel 2. 1 *Confusion Matrix*
Sumber: (Normawati & Prayogi, 2021)

Nilai Prediksi	Nilai Aktual	
	<i>Positive</i>	<i>Negative</i>
<i>Positive</i>	<i>True Positive</i>	<i>False Negative</i>
<i>Negative</i>	<i>False Positive</i>	<i>True Negative</i>

Keterangan :

TP (*True Positive*) = jumlah data nilai aktual kelas positif dan nilai prediksi kelas positif

TN (*True Negative*) = jumlah data nilai aktual negatif dan nilai prediksi negatif

FP (*False Positive*) = jumlah data nilai aktual positif dan nilai prediksi negatif

FN (*False Negative*) = jumlah data nilai aktual negatif dan nilai prediksi positif

Akurasi merupakan nilai proporsi *tweet* terhadap data yang terdeteksi pada pengujian. Nilai akurasi dapat memperlihatkan seberapa dekat nilai prediksi sistem dengan nilai prediksi manusia (Azhari dkk., 2021). Berikut rumusnya:

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \quad (2.14)$$

$$Precision = \frac{TP}{TP + FP} \times 100\% \quad (2.15)$$

$$Recall = \frac{TP}{TP + FN} \times 100\% \quad (2.16)$$

2.8 Penelitian Terdahulu

Tabel 2. 2 Penelitian Terdahulu

No	Judul Penelitian	Penulis	Metode	Hasil
1.	Analisis Sentimen Ulasan Pengguna BSI Mobile Pada Google Play Dengan Pendekatan <i>Supervised Learning</i>	(Arifiyanti dkk., 2022)	SVM, KNN, <i>Decision Tree</i> , dan <i>Naïve Bayes</i>	Jumlah data yang digunakan untuk pembuatan model klasifikasi berjumlah 2554 baris data. Dengan persentase jumlah sentimen positif sebesar 59,2% dan sentimen negatif sebesar 40,8%. Hasilnya, model klasifikasi yang dibuat dengan algoritma <i>Multinomial Naïve Bayes</i> menghasilkan hasil evaluasi terbaik dengan nilai ROC area sebesar 0,84. Hasil ini mengungguli hasil evaluasi nilai ROC area dari SVM, <i>Decision Tree</i> , dan KNN yang secara berurutan hasilnya sebesar 0,82; 0,78; dan 0,77.
2.	Analisis Sentimen Pada Media Sosial Twitter Terhadap Undang-Undang Cipta Kerja	(Sandryan dkk., 2021)	<i>Backpropagation</i>	Data yang digunakan berasal dari Twitter, jumlah tweet yang

	<p>Menggunakan Algoritma <i>Backpropagation</i> dan <i>Term Frequency-Inverse Document Frequency</i></p>			<p>digunakan adalah sebanyak 1000 data. <i>Preprocessing</i> yang dilakukan meliputi <i>case folding</i>, <i>tokenization</i>, normalisasi, <i>stopword removal</i>, dan <i>stemming</i>. Performa terbaik yang didapatkan oleh JST adalah akurasi sebesar 95%, <i>precision</i> sebesar 98%, <i>recall</i> sebesar 92.4%, dan <i>f-measure</i> sebesar 95.1%. Nilai tersebut diperoleh ketika Jaringan Syaraf Tiruan melakukan pelatihan sebanyak 1500 <i>epoch</i>, 1 <i>hidden layer</i> dengan 10 <i>hidden node</i>, serta <i>learning rate</i> sebesar 0.2.</p>
3.	<p>Analysis of The Multilayer Perceptron Algorithm on Twitter User's Sentiment Towards The COVID-19 Vaccine</p>	<p>(Pasaribu dkk., 2023)</p>	<p><i>Multilayer Perceptron</i></p>	<p>Data yang digunakan berjumlah 228.208 dengan membagi 3 kelas yaitu positif, negatif, dan netral. Metode yang digunakan dalam penelitian ini meliputi pengumpulan data, <i>text preprocessing</i>, TF-IDF,</p>

				<p>algoritma <i>Multilayer Perceptron</i>, dan pengujian dengan <i>confusion matrix</i>. Hasil performa dari penelitian tersebut diperoleh akurasi tertinggi sebesar 81,2%, presisi sebesar 83,8%, dan <i>recall</i> sebesar 71,2%. Hasil sentiment analisis dapat dilihat dari opini masyarakat terhadap vaksin COVID-19 yang terbagi menjadi tiga, yaitu opini positif sebesar 35%, opini negatif sebesar 16.3% dan opini netral sebesar 48,7%.</p>
4.	<p>Analisis Sentimen Dompot Elektronik Pada <i>Twitter</i> Menggunakan Metode <i>Naïve Bayes Classifier</i></p>	(Putra dkk., 2020)	<i>Naïve Bayes Classifier</i>	<p>Data diperoleh secara langsung dari situs <i>microblogging twitter</i> dengan hashtag pencarian #GoPay, #OVO dan #Dana, dari data tersebut, diambil 100 data dengan rincian 70 data latih dan 30 data uji. Hasil penelitian ini menunjukkan bahwa dompet elektronik <i>GoPay</i> lebih banyak dinilai positif oleh pengguna</p>

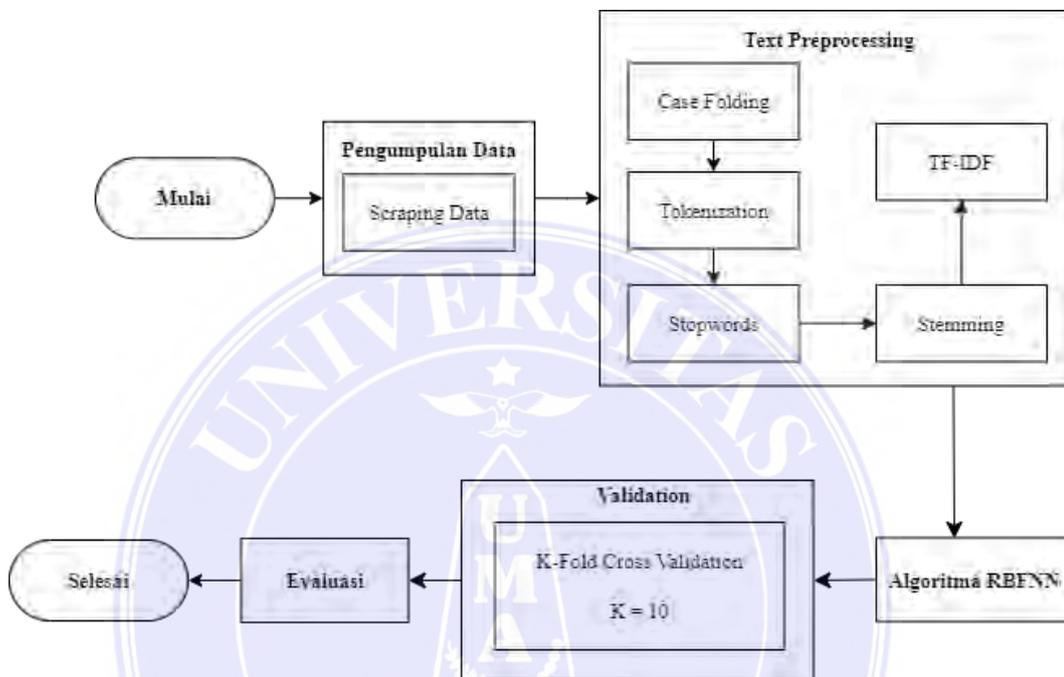
				<p><i>twitter</i> yaitu sebesar 46,67% diikuti oleh Dana sebesar 37,50% dan <i>OVO</i> sebesar 16,67%. <i>Dompet elektronik OVO</i> memiliki nilai negatif yang lebih tinggi yaitu sebesar 63,33% diikuti oleh <i>GoPay</i> sebesar 53,33% dan Dana sebesar 30,00%, sehingga hasil performa akurasi yang didapatkan dari hasil analisa Naive Bayes adalah 96,67%.</p>
5.	<p>Analisis Sentimen Pada Data Saran Mahasiswa Terhadap Kinerja Departemen Di Perguruan Tinggi Menggunakan <i>Convolutional Neural Network</i></p>	<p>(Yuliska dkk., 2021)</p>	<p><i>Convolutional Neural Network</i></p>	<p>Data yang digunakan pada penelitian ini adalah data hasil <i>Quality Assurance (QA)</i> penilaian performa layanan unit kerja atau departemen di Politeknik Caltex Riau di semester genap Tahun Ajaran 2018/2019. Yang digunakan pada penelitian ini yaitu data saran atau <i>review</i> dari mahasiswa. Jumlah data saran di masing-masing unit adalah mulai dari 1470 hingga 1500, namun hanya 1200 data saran saja yang</p>

				<p>akan digunakan, sedangkan 300 data saran lainnya dieliminasi berdasarkan banyak kata yang terkandung di dalamnya dan juga demi keseimbangan data saran di masing-masing kelas (kelas data saran dengan label positif, negatif dan netral). Berdasarkan hasil pengujian, <i>DoubleMax</i> CNN dapat mengklasifikasi sentimen pada saran mahasiswa dengan sangat baik, yaitu mencapai Akurasi tertinggi sebesar 98%, <i>Recall</i> 97%, <i>Precision</i> 98% dan <i>F1-Score</i> 98%.</p>
6.	<p>Analisis Sentimen Twitter untuk Menilai Opini Terhadap Perusahaan Publik Menggunakan Algoritma <i>Deep Neural Network</i></p>	<p>(Hidayat dkk., 2021)</p>	<p><i>Deep Neural Network</i></p>	<p>Data <i>Tweet record</i> sebanyak 5504 didapat dengan melakukan <i>crawling</i> melalui <i>Application Programming Interface</i> (API) <i>Twitter</i> yang selanjutnya dilakukan <i>preprocessing</i> (<i>cleansing, case folding,</i></p>

				<p>formalisasi, <i>stemming</i>, dan tokenisasi). Proses labeling dilakukan untuk 3902 <i>record</i> dengan memanfaatkan aplikasi <i>Sentiment Strength Detection</i>. Tahap pelatihan model dilakukan menggunakan algoritma DNN dengan variasi jumlah <i>hidden layer</i>, susunan <i>node</i>, dan nilai <i>learning rate</i>. Eksperimen dengan proporsi data <i>training</i> dan <i>testing</i> sebesar 90:10 memberikan hasil performa terbaik. Model tersusun dengan 3 <i>hidden layer</i> dengan susunan <i>node</i> tiap layer pada model tersebut yaitu 128, 256, 128 <i>node</i> dan menggunakan <i>learning rate</i> sebesar 0.005, model mampu menghasilkan nilai akurasi mencapai 88.72%.</p>
--	--	--	--	--

BAB III METODOLOGI PENELITIAN

Pada bab ini merupakan proses gambaran umum terkait alur penelitian yang akan dilakukan dalam pengerjaan tugas akhir ini dari awal hingga akhir.



Gambar 3. 1 Alur Penelitian

3.1 Data dan Sumber Data

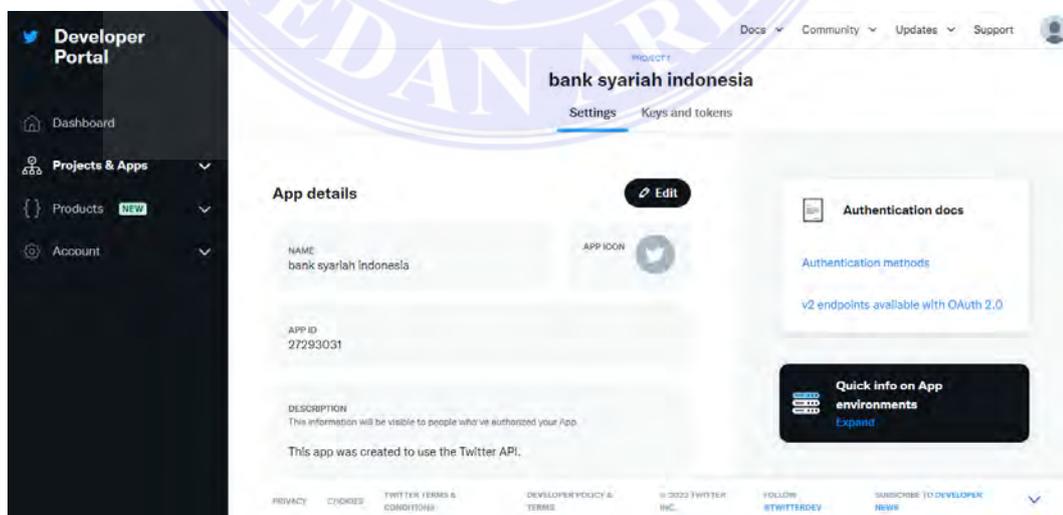
Penelitian ini memanfaatkan data primer yakni media sosial *twitter* melalui pemanfaatan API (*Application Programming Interface*) yang merupakan program yang disediakan oleh beberapa pengembang aplikasi agar pengembang aplikasi lainnya dapat mengakses aplikasi tersebut dengan lebih mudah. API berperan sebagai sistem kontak yang dibuat dengan *twitter*. Untuk dapat mengakses API, diperlukan *key*, *secret key*, *acces token* dan *secret acces token*, yang bertindak sebagai kunci agar *twitter* dapat memahami apa yang dilakukan aplikasi serta bagaimana melakukannya dengan mendaftarkan aplikasi *twitter* API di <https://developer.twitter.com/>. Kemudian gunakan bahasa pemrograman *Python* untuk mencari data dari *twitter* dan mencari data berlandaskan kata kunci.

Tabel 3. 1 Data Sampel

No	Date	Id	Teks
1.	2023-02-14	0102021	Ransomware LockBit mengklaim menjadi dalang dan bertanggung jawab atas pencurian data Bank Syariah Indonesia (BSI).
2.	2023-02-15	0102022	Muhammadiyah berencana tarik dana dari bank syariah indonesia (BSI)
...
1837	2023-06-02	1664638385212780000	silakan Kakak dapat membuat laporan pengaduan melalui BSI Mobile di menu customare care

3.1.1 Pengumpulan Data

Untuk mengaksesnya dibutuhkan sebuah kode yang didapatkan dari *Twitter* API dengan cara mendaftar ke portal *developer Twitter* yang berfungsi untuk mengkonfirmasi kepada pihak *Twitter* dalam memberikan ijin menjelajahi terkait data yang berkaitan dengan *Twitter*. Untuk lebih jelas dapat dilihat pada Gambar 3.2.

Gambar 3. 2 *Developer Portal* API

Kemudian setelah berhasil registrasi dengan *Twitter* API, akan didapatkan beberapa kode diantaranya *consumer key*= 'QYtVJY5kibBG6NZRlqqTUgSdN', *consumer secret*= 'tcoS7qfAxmUiWxETkjiMDndyxri4di2iBxkGLqfMgC086dzvab', *access token*= '1539625531591004163-z7S5pVvUlwf9YnKORbr1s6SmWzaUY', *access_token_secret*= 'rLKgD9onjSfHAcRGVnVILWRfnqFIyLIQrDexAX7RrtNhI'. Kode tersebut yang akan digunakan untuk proses integrasi antara *Twitter* API dengan bahasa pemrograman *Python*.

	date	id	teks
0	2023-06-11 14:30:02+00:00	1667902010308587520	Bank Indonesia Sumut Laksanakan Grand Launchin...
1	2023-06-11 13:10:27+00:00	1667881983907872768	@ayang2_ku Harap selalu waspada dan berhati-ha...
2	2023-06-11 12:53:03+00:00	1667877604567945217	@ayang2_ku atas segala bentuk modus penipuan y...
3	2023-06-11 12:27:47+00:00	1667871248494505985	@Vtaenana mengatasmakan Bank Syariah Indones...
4	2023-06-11 12:09:30+00:00	1667866646072143872	@cibichie biaya pergantiannya sebesar Rp 5.000...
...
1832	2023-06-02 15:13:26+00:00	1664651444350521345	@mingsosera melalui layanan Hasanah Card di no...
1833	2023-06-02 14:38:45+00:00	1664642716590907409	@yayaaasks dengan membawa KTP Asli, Buku Tabun...
1834	2023-06-02 14:38:38+00:00	1664642685120987136	@yayaaasks silakan Kakak dapat membuat laporan...
1835	2023-06-02 14:21:52+00:00	1664638464699015169	@nenkfitri_k dengan membawa KTP Asli, Buku Tab...
1836	2023-06-02 14:21:33+00:00	1664638385212780550	@nenkfitri_k silakan Kakak dapat membuat lapor...

1837 rows x 3 columns

Gambar 3. 3 Tampilan Data

Setelah semuanya selesai seperti pada Gambar 3.3 maka menampilkan data yang sudah didapatkan berdasarkan kata kunci dan jumlah data yang ditentukan. Gambar 3.3 mendapatkan data sebanyak 1837 dan memiliki 1768 positif dan 69 negatif.

3.2 Teknik Analisis Data

Teknik analisis data adalah sebuah kegiatan berupa menganalisis suatu penelitian dengan mengontrol seluruh data yang berasal dari instrumen penelitian, seperti dokumen, catatan, rekaman dan lain – lain. Tujuannya adalah untuk memahami data dan memperoleh kesimpulan dengan lebih mudah. Teknik analisis data yang digunakan sebagai berikut :

1. Menganalisis data secara menyeluruh untuk mengetahui gambaran umum mengenai data sentimen masyarakat *twitter* terhadap Bank Syariah

Indonesia, serta menentukan variabel apa saja yang digunakan pada penelitian ini.

2. Membagi data yaitu data *training* sebesar 60%, 70%, dan 80%, serta data *testing* sebesar 40%, 60, dan 20%.
3. Membuat model klasifikasi dari data *training* menggunakan algoritma RBFNN beserta *K-Fold Cross Validation*.
4. Menghitung ketepatan klasifikasi dengan cara menghitung akurasi dalam proses pengklasifikasian.

3.3 Text Preprocessing

Tahapan ini adalah proses seleksi pemeriksaan pada teks untuk melakukan pembersihan, memperbaiki kesalahan pada teks serta menyederhanakan teks, sehingga teks tersebut dapat diproses lebih lanjut. Adapun tahapan pada *text preprocessing* sebagai berikut.

3.3.1 Case Folding

Case folding adalah proses dalam membersihkan dan menghilangkan karakter yang tidak diperlukan pada setiap data dokumen. Dapat dilihat pada Tabel 3.2.

Tabel 3. 2 *Case Folding*

Teks	<i>Case Folding</i>
Ransomware LockBit mengklaim menjadi dalang dan bertanggung jawab atas pencurian data Bank Syariah Indonesia (BSI).	Ransomware lockbit mengklaim menjadi dalang dan bertanggung jawab atas pencurian data bank syariah indonesia bsi
Muhammadiyah berencana tarik dana dari bank syariah indonesia (BSI)	muhammadiyah berencana tarik dana dari bank syariah indonesia bsi

Gambar 3.4 adalah tahapan *case folding* yang berfungsi menghapus seluruh tanda baca karena tidak digunakan dan tidak memiliki nilai penting. *Python* mempunyai *library* tersendiri dalam mengatasi tahapan ini misalnya *library string* dan *library re*. Kemudian seluruh teks diubah menjadi huruf kecil untuk menyeragamkan dengan menggunakan fungsi *lower()*.

	date	id	teks	case_folding
0	2023-06-11 14:30:02+00:00	1667902010308587520	Bank Indonesia Sumut Laksanakan Grand Launchin...	bank indonesia sumut laksanakan grand launchin...
1	2023-06-11 13:10:27+00:00	1667881983907872768	@ayang2_ku Harap selalu waspada dan berhati-ha...	yang ku harap selalu waspada dan berhati hati...
2	2023-06-11 12:53:03+00:00	1667877604567945217	@ayang2_ku atas segala bentuk modus penipuan y...	yang ku atas segala bentuk modus penipuan yan...
3	2023-06-11 12:27:47+00:00	1667871248494505985	@Vtaenana mengatasmamakan Bank Syariah Indones...	taenana mengatasmamakan bank syariah indonesia...
4	2023-06-11 12:09:30+00:00	1667866646072143872	@cibichie biaya pergantiannya sebesar Rp 5.000...	cibichie biaya pergantiannya sebesar rp ...

Gambar 3. 4 Hasil *Case Folding*

3.3.2 Tokenization

Tahapan *tokenization* ini, dilakukan pemotongan kalimat sehingga menjadi kata-kata tunggal. Adapun proses *tokenization* dapat dilihat pada Tabel 3.3.

Tabel 3. 3 *Tokenization*

<i>Case Folding</i>	<i>Tokenization</i>
Ransomware LockBit mengklaim menjadi dalang dan bertanggung jawab atas pencurian data Bank Syariah Indonesia (BSI).	['ransomware'], ['lockbit'], ['mengklaim'], ['menjadi'], ['dalang'], ['dan'], ['bertanggung'], ['jawab'], ['atas'], ['pencurian'], ['data'], ['bank'], ['syariah'], ['indonesia'], ['bsi']
Muhammadiyah berencana tarik dana dari bank syariah indonesia (BSI)	['muhammadiyah'], ['berencana'], ['tarik'], ['dana'], ['dari'], ['bank'], ['syariah'], ['indonesia'], ['bsi']

Tahapan selanjutnya merupakan tahap tokenisasi berfungsi menjadikan sebuah kalimat tersebut menjadi kata per kata atau token. *Python* sendiri memiliki library untuk mengatasi hal tersebut contohnya library *NLTK (Natural Language Toolkit)* dengan modul *word tokenize*. *NLTK* sendiri dikembangkan untuk memproses data yang berbentuk teks. Hasilnya dapat dilihat pada Gambar 3.5.

	date	id	teks	case_folding	tokenization
0	2023-06-11 14:30:02+00:00	1667902010308587520	Bank Indonesia Sumut Laksanakan Grand Launchin...	bank indonesia sumut laksanakan grand launchin...	[bank, indonesia, sumut, laksanakan, grand, la...
1	2023-06-11 13:10:27+00:00	1667881983907872768	@ayang2_ku Harap selalu waspada dan berhati-ha...	yang ku harap selalu waspada dan berhati hati...	[yang, ku, harap, selalu, waspada, dan, berhati...
2	2023-06-11 12:53:03+00:00	1667877604567945217	@ayang2_ku atas segala bentuk modus penipuan y...	yang ku atas segala bentuk modus penipuan yan...	[yang, ku, atas, segala, bentuk, modus, penipu...
3	2023-06-11 12:27:47+00:00	1667871248494505985	@Vtaenana mengatasmamakan Bank Syariah Indones...	taenana mengatasmamakan bank syariah indonesia...	[taenana, mengatasmamakan, bank, syariah, indo...
4	2023-06-11 12:09:30+00:00	1667866646072143872	@cibichie biaya pergantiannya sebesar Rp 5.000...	cibichie biaya pergantiannya sebesar rp ...	[cibichie, biaya, pergantiannya, sebesar, rp, ...

Gambar 3. 5 Hasil *Tokenization*

3.3.3 Stopwords

Stopwords adalah tahap dimana menghapuskan kata yang tidak memiliki makna atau kurang penting pada data dokumen. Dapat dilihat pada Tabel 3.4.

Tabel 3. 4 *Stopwords*

<i>Tokenization</i>	<i>Stopwords</i>
['ransomware'], ['lockbit'], ['mengklaim'], ['menjadi'], ['dalang'], ['dan'], ['bertanggung'], ['jawab'], ['atas'], ['pencurian'], ['data'], ['bank'], ['syariah'], ['indonesia'], ['bsi']	['ransomware'], ['lockbit'], ['mengklaim'], ['dalang'], ['bertanggung'], ['jawab'], ['pencurian'], ['data'], ['bank'], ['syariah'], ['indonesia']
['muhammadiyah'], ['berencana'], ['tarik'], ['dana'], ['dari'], ['bank'], ['syariah'], ['indonesia'], ['bsi']	['muhammadiyah'], ['berencana'], ['tarik'], ['dana'], ['bank'], ['syariah'], ['indonesia']

Selanjutnya tahap *stopwords* berfungsi menghapus kata yang tidak memiliki arti, contohnya seperti kata hubung atau kata ganti. Adapaun *library* yang digunakan juga berasal dari *nlk* sendiri dengan modul *stopwords*. Gambar 3.6 adalah hasil setelah dilakukan proses *stopwords*.

	date	id	teks	case_folding	tokenization	Stopwords
0	2023-06-11 14:30:02+00:00	1667902010308587520	Bank Indonesia Sumut Laksanakan Grand Launchin...	bank indonesia sumut laksanakan grand launchin...	[bank, indonesia, sumut, laksanakan, grand, la...	[bank, indonesia, sumut, laksanakan, grand, la...
1	2023-06-11 13:10:27+00:00	1667881983907872768	@ayang2_ku Harap selalu waspada dan berhati-ha...	yang ku harap selalu waspada dan berhati hall...	[yang, ku, harap, selalu, waspada, dan, berhati...	[ku, harap, waspada, berhati, hall, ya, bentuk...
2	2023-06-11 12:53:03+00:00	1667877604567945217	@ayang2_ku alas segala bentuk modus penipuan y...	yang ku alas segala bentuk modus penipuan yan...	[yang, ku, alas, segala, bentuk, modus, penipu...	[ku, bentuk, modus, penipuan, mengatasnamakan,...
3	2023-06-11 12:27:47+00:00	1667871248494505985	@Vtaenana mengatasnamakan Bank Syariah Indones...	taenana mengatasnamakan bank syariah indonesia.	[taenana, mengatasnamakan, bank, syariah, indo...	[taenana, mengatasnamakan, bank, syariah, indo...
4	2023-06-11 12:09:30+00:00	1667866646072143872	@cibichie biaya pergantian nya sebesar Rp 5.000...	cibichie biaya pergantian nya sebesar rp ...	[cibichie, biaya, pergantian nya, sebesar, rp, ...	[cibichie, biaya, pergantian nya, rp, ya, harap...

Gambar 3. 6 Hasil *Stopwords*

3.3.4 *Stemming*

Stemming adalah proses untuk menemukan kata dasar dari setiap kata yang telah melewati tahap *filtering*. Dalam proses *stemming*, kata-kata tersebut diubah untuk menghilangkan semua imbuhan yang terdapat pada data dokumen. Dapat dilihat pada Tabel 3.5.

Tabel 3. 5 *Stemming*

<i>Stopwords</i>	<i>Stemming</i>
['ransomware'], ['lockbit'], ['mengklaim'], ['menjadi'], ['dalang'], ['dan'], ['bertanggung'], ['jawab'], ['atas'], ['pencurian'], ['data'], ['bank'], ['syariah'], ['indonesia'], ['bsi']	['ransomware'], ['lockbit'], ['klaim'], ['dalang'], ['tanggung'], ['jawab'], ['curi'], ['data'], ['bank'], ['syariah'], ['indonesia']

['muhammadiyah'], ['berencana'], ['tarik'], ['dana'], ['dari'], ['bank'], ['syariah'], ['indonesia'], ['bsi']	['muhammadiyah'], ['rencana'], ['tarik'], ['dana'], ['bank'], ['syariah'], ['indonesia']
---	--

Gambar 3.7 merupakan tahapan *stemming* yang bertujuan mengambil kata dasar dari sebuah kata tersebut (sufiks dan prefiks). *Library* yang digunakan adalah *sastrawi* dengan modul *stemmer*. *Library* tersebut khusus digunakan dengan data yang bersifat bahasa Indonesia.

	date	id	teks	case_folding	tokenization	Stopwords	stemming
0	2023-06-11 14:30:02+00:00	1667902010308587520	Bank Indonesia Sumut Laksanakan Grand Launchin	bank indonesia sumut laksanakan grand launchin	[bank, indonesia, sumut, laksanakan, grand, la	[bank, indonesia, sumut, laksanakan, grand, la	bank indonesia sumut laksana grand launching p
1	2023-06-11 13:10:27+00:00	166788196390782768	@ayang2_ku Harap selalu waspada dan berhati-ha...	yang ku harap selalu waspada dan berhati ha...	[yang, ku, harap, selalu, waspada, dan, berhati...	[ku, harap, waspada, berhati, hati, ya, bentuk...	ku harap waspada hati hati ya bentuk modus tip...
2	2023-06-11 12:53:03+00:00	1667877604567945217	@ayang2_ku atas segala bentuk modus penipuan y...	yang ku atas segala bentuk modus penipuan yan	[yang, ku, atas, segala, bentuk, modus, penipu...	[ku, bentuk, modus, penipuan, mengatasnamakan...	ku bentuk modus lipu mengatasnamakan bank syar
3	2023-06-11 12:27:47+00:00	1667871248494505885	@Vtaenana mengatasnamakan Bank Syariah Indones...	taenana mengatasnamakan bank syariah indonesia...	[taenana, mengatasnamakan, bank, syariah, indo...	[taenana, mengatasnamakan, bank, syariah, indo...	taenana mengatasnamakan bank syariah indonesia...
4	2023-06-11 12:09:30+00:00	1667866646072143872	@cibichie biaya pergantiannya sebesar Rp 5 000	cibichie biaya pergantiannya sebesar rp	[cibichie, biaya, pergantiannya, sebesar, rp...	[cibichie, biaya, pergantiannya, rp, ya, harap	cibichie biaya ganti rp ya harap waspada hati

Gambar 3. 7 Hasil *Stemming*

3.3.5 TF-IDF

Menghitung *term frequency* (TF) dilakukan untuk mendapatkan jumlah setiap kata atau fitur yang ada di masing-masing dokumen. Perhitungan *term frequency* dilakukan dengan menghitung kemunculan setiap kata dalam dokumen. Proses menghitung TF-IDF selanjutnya melibatkan *document frequency*(DF) yang merupakan banyaknya dokumen dimana suatu kata atau fitur muncul, kemudian *inverses document frequency* (IDF) merupakan hasil *inverse* dari DF dimana D adalah jumlah keseluruhan dokumen dan W merupakan hasil dari TF-IDF yaitu bobot kata yang dihasilkan dari TF dikali dengan IDF. Berikut adalah hasil perhitungan TF-IDF.

Tabel 3. 6 TF-IDF

Term	TF		df	IDF	TF-IDF	
	D1	D2			D1	D2
ransomware	1	0	1	0,301	0,301	0
lockbit	1	0	1	0,301	0,301	0
klaim	1	0	1	0,301	0,301	0
dalang	1	0	1	0,301	0,301	0

tanggung	1	0	1	0,301	0,301	0
jawab	1	0	1	0,301	0,301	0
curi	1	0	1	0,301	0,301	0
data	1	0	1	0,301	0,301	0
bank	1	1	2	0	0	0
syariah	1	1	2	0	0	0
indonesia	1	1	2	0	0	0
muhammadiyah	0	1	1	0,301	0	0,301
rencana	0	1	1	0,301	0	0,301
tarik	0	1	1	0,301	0	0,301
dana	0	1	1	0,301	0	0,301

Tahapan terakhir adalah TF-IDF berfungsi mengubah data teks menjadi numerik agar dapat dilakukan perhitungan terhadap algoritma yang digunakan. *Library* yang digunakan berasal dari *sklearn* dengan modul *TfidfVectorizer*. Data dengan teks “bank indonesia sumut laksana grand launching pekan syariah rangkai fesyar sumatera”, maka hasil TF-IDF adalah 0.350, 0.350, 0.350, 0.082, 0.350, 0.350, 0.350, 0.350, 0.350, 0.086, 0.073. Hasilnya dapat dilihat pada Gambar 3.8.

```

print(dataset.stemming[0])
print(vector[0])

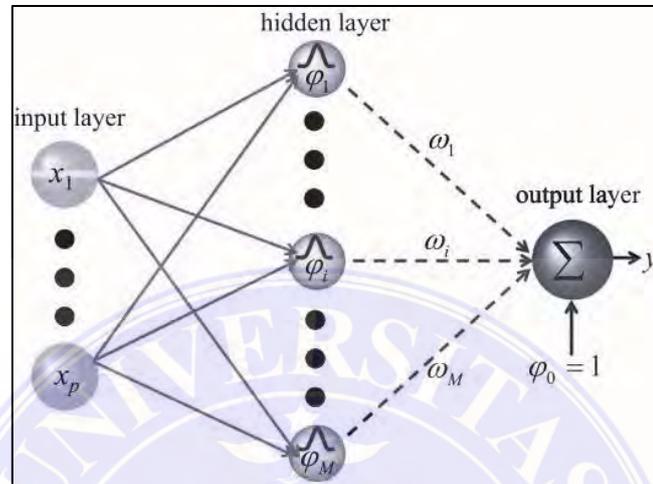
bank indonesia sumut laksana grand launching pekan syariah rangkai fesyar sumatera
(0, 2039) 0.3500325732372399
(0, 564) 0.3500325732372399
(0, 1742) 0.3500325732372399
(0, 2061) 0.0827500122049047
(0, 1630) 0.3500325732372399
(0, 1138) 0.3500325732372399
(0, 634) 0.3500325732372399
(0, 1119) 0.3500325732372399
(0, 2040) 0.3500325732372399
(0, 829) 0.08673635551215897
(0, 123) 0.07380258461522114

```

Gambar 3. 8 Hasil TF-IDF

3.4 Melatih Algoritma RBFNN Menggunakan *Backpropagation*

Pada tahapan ini dilakukan perhitungan RBFNN menggunakan *backpropagation* dengan menggunakan hasil TF-IDF sebagai inputan.



Gambar 3. 9 Arsitektur RBFNN

Pada Gambar 3.9 merupakan arsitektur RBFNN yang mana memiliki Input layer yang akan dihitung berdasarkan hasil TF-IDF dari Tabel 3.6, X_1 sebagai D1 dan X_2 sebagai D2.

1. Fase *Fordward Propagation* :
 - a.) Hitung keluaran unit tersembunyi (z_j) :

Z_j merupakan nilai dari hasil keluaran *hidden layer*.

$$\begin{aligned}
 Z_{\text{net}_1} &= (0,301 * 0,021) + (0,301 * 0,072) + (0,301 * 0,034) + (0,301 * \\
 &0,103) + (0,301 * (-0,012)) + (0,301 * (-0,091)) + (0,301 * (-0,021)) + \\
 &(0,301 * 0,102) + (0 * (-0,021)) + (0 * (-0,202)) + (0 * (-0,022)) + (0,301 * \\
 &(-0,040)) + (0,301 * (-0,060)) + (0,301 * 0,103) + (0,301 * (-0,092)) = 0,035
 \end{aligned}$$

$$z_{net_2} = (0,301 * (-0,011)) + (0,301 * 0,022) + (0,301 * 0,19) + (0,301 * 0,111) + (0,301 * 0,122) + (0,301 * (-0,091)) + (0,301 * (-0,073)) + (0,301 * 0,032) + (0 * (-0,052)) + (0 * (-0,202)) + (0 * (-0,120)) + (0,301 * 0,130) + (0,301 * (-0,101)) + (0,301 * 0,023) + (0,301 * (-0,024)) = 0,099$$

Menghitung Sigmoid :

Fungsi Aktivasi *Sigmoid Biner* Menghasilkan nilai *output* pada interval 0 hingga 1.

$$f(z_{net_1}) = \frac{1}{1 + e^{-0,035}} = 0,508$$

$$f(z_{net_2}) = \frac{1}{1 + e^{-0,099}} = 0,524$$

b.) Hitung keluaran unit y_k :

y_k merupakan nilai fungsi aktivitas pada lapisan keluaran.

$$y_{net_k} = w_{k0} + \sum_{j=1}^p z_j w_{kj} = 0,020 + (0,508)(0,017) + (0,524)(0,030) = 0,043$$

$$y_k = f(y_{net_k}) = \frac{1}{1 + e^{-0,043}} = 0,510$$

w_{kj} merupakan nilai bobot yang menghubungkan masukan node 2 dengan lapisan keluaran 2.

2. Fase Propagasi Mundur :

Hitung faktor δ di unit keluaran y_k

$$\delta_k = (t_k - y_k) f'(y_{net_k}) = (t_k - y_k) y_k (1 - y_k) = (1 - 0,510) 0,510 (1 - 0,510) = 0,122$$

$$\Delta w_{10} = \alpha \delta_k z_0 = 0,2 (0,122) (1) = 0,024$$

$$\Delta w_{11} = \alpha \delta_k z_1 = 0,2 (0,122) (0,508) = 0,012$$

$$\Delta w_{12} = \alpha \delta_k z_2 = 0,2 (0,122) (0,524) = 0,012$$

δ_k adalah nilai *error* lapisan keluaran.

T_k adalah nilai keluaran yang diharapkan.

3. Hitung penjumlahan kesalahan dari unit tersembunyi (δ) :

Faktor kesalahan δ di unit tersembunyi :

$$\delta_1 = (0,122)(0,4) = 0,048$$

$$\delta_2 = (0,122)(-0,2) = -0,024$$

$$\delta_1 = \delta_{net_{z1}}(1 - z_1) = (-0,06)(0,508)(1 - 0,508) = -0,014$$

$$\delta_2 = \delta_{net_{z2}}(1 - z_2) = (0,04)(0,524)(1 - 0,524) = 0,009$$

Suku perubahan bobot ke unit tersembunyi :

$$\Delta v_{10} = \alpha \delta_1 = (0,2)(0,004)(1) = 0,001$$

$$\Delta v_{20} = \alpha \delta_1 = (0,2)(-0,007)(1) = -0,001$$

$$\Delta v_{11} = \alpha \delta_1 = (0,2)(0,004)(0,301) = 0$$

$$\Delta v_{21} = \alpha \delta_2 = (0,2)(-0,005)(0,301) = -0,0003$$

$$\Delta v_{12} = \alpha \delta_1 = (0,2)(0,006)(0,301) = 0,0003$$

$$\Delta v_{22} = \alpha \delta_2 = (0,2)(-0,007)(0,301) = -0,0004$$

4. Hitung semua perubahan bobot :

Perubahan bobot unit keluaran :

$$w_{10}(\text{baru}) = w_{10}(\text{lama}) + \Delta w_{10} = 0,5 + 0,024 = 0,524$$

$$w_{11}(\text{baru}) = w_{11}(\text{lama}) + \Delta w_{11} = -0,3 + 0,012 = -0,288$$

$$w_{12}(\text{baru}) = w_{12}(\text{lama}) + \Delta w_{12} = -0,4 + 0,012 = -0,388$$

Perubahan bobot unit tersembunyi :

$$v_{10}(\text{baru}) = v_{10}(\text{lama}) + \Delta v_{10} = -0,01 + 0,001 = -0,009$$

$$v_{20}(\text{baru}) = v_{20}(\text{lama}) + \Delta v_{20} = 0,01 + (-0,001) = 0,009$$

$$v_{11}(\text{baru}) = v_{11}(\text{lama}) + \Delta v_{11} = 0,02 + 0 = 0,02$$

$$v_{21}(\text{baru}) = v_{21}(\text{lama}) + \Delta v_{21} = 0,03 + (-0,0003) = 0,0297$$

$$v_{12}(\text{baru}) = v_{12}(\text{lama}) + \Delta v_{12} = 0,03 + 0,0003 = 0,0303$$

$$v_{22}(\text{baru}) = v_{22}(\text{lama}) + \Delta v_{22} = 0,01 + (-0,0004) = 0,0096$$

Berdasarkan fase nomor 4 dapat disimpulkan bahwa sentimen pada data sampel di Tabel 3.1 sebagai berikut :

1. Ditentukan berdasarkan perubahan bobot unit tersembunyi $V_{10}(\text{baru}) = -0,009$, $V_{11}(\text{baru}) = 0,02$, dan $V_{12}(\text{baru}) = 0,0303$. Hasil tersebut akan ditentukan berdasarkan yang paling dominan apakah mengarah ke positif atau negatif. Maka data sampel nomor 1 bersentimen positif.
2. Ditentukan berdasarkan perubahan bobot unit tersembunyi $V_{20}(\text{baru}) = -0,009$, $V_{21}(\text{baru}) = 0,0297$, dan $V_{22}(\text{baru}) = 0,0096$. Hasil tersebut akan ditentukan berdasarkan yang paling dominan apakah mengarah ke positif atau negatif. Maka data sampel nomor 2 bersentimen positif.

```

Deklarasi :

  Inisialisasi bobot acak range 1, -1
  Hitung keluaran unit tersembunyi ( $z_j$ )
  Hitung keluaran unit  $y_k$ 
  Hitung faktor  $\delta$  di unit keluaran  $y_k$ 
  Hitung penjumlahan kesalahan dari unit tersembunyi ( $\delta$ )
  Hitung semua perubahan bobot

START :

  Mulai Training
  IF Training Error
  ELSE Deklarasi
END

```

Gambar 3. 10 Pseudocode RBFNN

Gambar 3.10 adalah *pseudocode* algoritma RBFNN sebagai dokumentasi sederhana sehingga program tersebut mudah untuk dipahami. Inisialisasi/tentukan bobot acak *range* 1 hingga -1, hitung *output* unit tersembunyi, hitung faktor di unit y_k , hitung *error* dari unit tersembunyi, dan hitung semua perubahan bobot. Jika *training error* maka akan di hitung kembali melalui deklarasi, jika tidak *error* maka *training* selesai.

```

▶ from RBFClassifier import RBFClassifier # panggil Libraries RBFNN
  from sklearn.metrics import accuracy_score # panggil library

  rbfnn = RBFClassifier()
  rbfnn.fit(X_train, y_train)

  y_pred = rbfnn.predict(X_test)

  acc = accuracy_score(y_test, y_pred) # akurasi test

  print('Akurasi Testing :', acc)

```

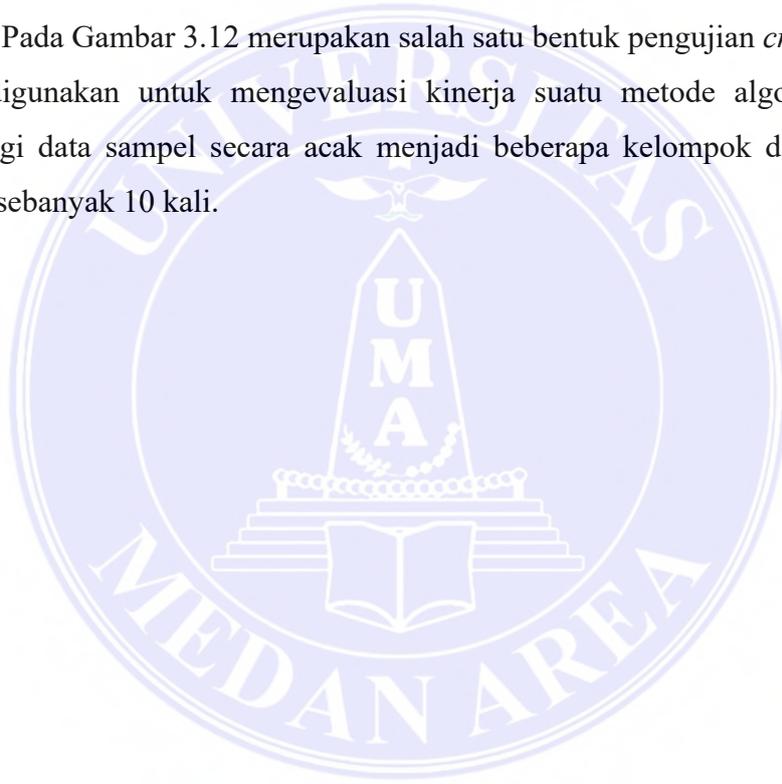
Gambar 3. 11 Algoritma RBFNN

Gambar 3.11 adalah menerapkan algoritma RBFNN dengan memanggil *library* beserta modul *RBFClassifier*, selanjutnya melatih data menggunakan data *training* dan data *testing* sebagai data ujinya. Adapun performa tersebut berasal dari pembelajaran data yang sudah di uji.

```
from sklearn.model_selection import KFold, cross_validate  
  
kfold = KFold(10, shuffle=True)  
scores = cross_validate(rbfnn, X, y, cv=kfold, scoring=['accuracy'])
```

Gambar 3. 12 *K-Fold Cross Validation*

Pada Gambar 3.12 merupakan salah satu bentuk pengujian *cross validation* yang digunakan untuk mengevaluasi kinerja suatu metode algoritma dengan membagi data sampel secara acak menjadi beberapa kelompok dan mengulang proses sebanyak 10 kali.



BAB V

SIMPULAN DAN SARAN

Berdasarkan hasil yang sudah dilakukan maka dibuat kesimpulan dan saran agar dapat dikembangkan oleh peneliti selanjutnya.

5.1 Simpulan

Berdasarkan hasil yang telah dilakukan pada penelitian ini dapat dibuat kesimpulan yaitu :

1. Proporsi data *training* 80% dan *testing* 20% menghasilkan performa akurasi dengan algoritma RBFNN sebesar 97%, presisi 97%, dan *recall* 97%, selanjutnya performa akurasi dengan *K-Fold Cross Validation* sebesar 96,2%.
2. Berdasarkan data *testing* sebanyak 368 data dengan kata kunci “Bank Syariah Indonesia” mendapatkan hasil sentimen positif sebanyak 357 atau 97,01% dan negatif sebanyak 11 atau 2,99%.

5.2 Saran

Saran yang dapat dilakukan untuk penelitian selanjutnya adalah :

1. Menambah jumlah data karena dapat mempengaruhi data prediksinya terhadap data yang diuji.
2. Melakukan pembersihan data secara mendalam agar tidak memiliki kalimat-kalimat yang ambigu dan menggunakan algoritma klasifikasi lainnya seperti *Support Vector Machine*, *Multilayer Perceptron*, *Random Forest*, *Decision Tree*, *C.45*, dan *K-Nearest Neighbor*.

DAFTAR PUSTAKA

- Afdal, M., & Rahma Elita, L. (2022). Penerapan Text Mining Pada Aplikasi Tokopedia Menggunakan Algoritma K-Nearest Neighbor. *Jurnal Ilmiah Rekayasa dan Manajemen Sistem Informasi*, 8(1), 78–87.
- Agustina, D. A. Zukhronah, E. (2021). Implementasi Text Mining Pada Analisis Sentimen Pengguna Twitter Terhadap Marketplace di Indonesia Menggunakan Algoritma Support Vector Machine. *Indonesian Journal of Applied Statistics*, 3(2), 109. <https://doi.org/10.13057/ijas.v3i2.44337>
- AISYAH, S. AMIJAYA, F. (2021). Peramalan Jumlah Titik Panas Provinsi Kalimantan Timur Menggunakan Metode Radial Basis Function Neural Network. *Jambura Journal of Probability and Statistics*, 2(2), 64–74. <https://doi.org/10.34312/jjps.v2i2.10292>
- Arifiyanti, A. A. Syafira, A. O. (2022). *Analisis Sentimen Ulasan Pengguna BSI Mobile Pada Google Play Dengan Pendekatan Supervised Learning*. 283–288.
- Arsi, P., & Waluyo, R. (2021). Analisis Sentimen Wacana Pemindahan Ibu Kota Indonesia Menggunakan Algoritma Support Vector Machine (SVM). *Jurnal Teknologi Informasi dan Ilmu Komputer*, 8(1), 147. <https://doi.org/10.25126/jtiik.0813944>
- Azhari, M. Rosnelly, R. (2021). Perbandingan Akurasi, Recall, dan Presisi Klasifikasi pada Algoritma C4.5, Random Forest, SVM dan Naive Bayes. *JURNAL MEDIA INFORMATIKA BUDIDARMA*, 5(2), 640. <https://doi.org/10.30865/mib.v5i2.2937>
- Azis, H. Putri, I. P. (2020). Performa Klasifikasi K-NN dan Cross Validation Pada Data Pasien Pengidap Penyakit Jantung. *ILKOM Jurnal Ilmiah*, 12(2), 81–86. <https://doi.org/10.33096/ilkom.v12i2.507.81-86>
- Chen, Y. Lin, H. (2019). Bioresource Technology Application of Radial Basis Function Artificial Neural Network to Quantify Interfacial Energies Related

to Membrane Fouling in a Membrane Bioreactor. *Bioresource Technology*, 293.

Fatihudin, A., & Firmansyah, A. (2019). *Manajemen Bank*. Qiara Media.

Fiktri, M. I. Azhar, Y. (2020). Perbandingan Metode Naïve Bayes dan Support Vector Machine pada Analisis Sentimen Twitter. *SMATIKA JURNAL*, 10, 71–76.

Haryanto, A. . (2020). *Riset: Ada 175,2 Juta Pengguna Internet di Indonesia*. detik.com.<https://inet.detik.com/cyberlife/d4907674/riset-ada-1752-juta-penggunainternet-di-indonesia>

Hashemi, A. Abbasi, M. (2019). *Implementation of Multilayer Perceptron (MLP) and Radial Basis Function (RBF) Neural Networks to Predict Solution Gas-Oil Ratio of Crude Oil Systems*.

Hayqal, H. H. Q. Yuana Sukmawaty. (2022). K-Means Clustering dan Principal Component Analysis (PCA) Dalam Radial Basis Function Neural Network (RBFNN) Untuk Klasifikasi Data Multivariat. *Journal of Mathematics: Theory and Applications*, 4(1), 1–7. <https://doi.org/10.31605/jomta.v4i1.1757>

Hidayat, E. Y. Affandy, A. (2021). Analisis Sentimen Twitter untuk Menilai Opini Terhadap Perusahaan Publik Menggunakan Algoritma Deep Neural Network. *Jurnal Nasional Teknologi dan Sistem Informasi*, 7(2), 108–118. <https://doi.org/10.25077/teknosi.v7i2.2021.108-118>

Irwansyah, S. (2022). *Machine Learning Untuk Pemula*. Informatika Bandung.

Jabal Tursina, M. (2019). *Sentimen Analisis Sistem Zonasi Sekolah Pada Media Sosial Youtube Menggunakan Metode K-Nearest Neighbor Dengan Algoritma Levenshtein Distance*. Universitas Islam Negeri Syarif Hidayatullah Jakarta.

Kermani, M. Z. Hinkelmann, R. (2020). *Advanced Engineering Informatics Ensemble data mining modeling in corrosion of concrete sewer : A comparative study of network-based (MLPNN & RBFNN)*.

Madhiarasan, M. (2020). Accurate prediction of different forecast horizons wind

speed using a recursive radial basis function neural network. *Protection and Control of Modern Power Systems*, 5(1). <https://doi.org/10.1186/s41601-020-00166-8>

Mayang Sari, S. (2021). *Analisis Sentimen Terhadap New Normal Di Era COVID-19 Menggunakan Algoritma K-Nearest Neighbor (K-NN)*. Universitas Islam Sumatera Utara Medan.

Muba, M. Sapsuha, I. (2023). *Analisis Preferensi Konsumen dalam Memilih Jasa Bank Syariah Indonesia (BSI) KCP Morotai*. 9(01), 732–738.

Mufidah, F. S. Sani, R. R. (2022). Analisis Sentimen Masyarakat Terhadap Layanan ShopeeFood Melalui Media Sosial Twitter Dengan Algoritma Naïve Bayes Classifier. *JOINS (Journal of Information System)*, 7(1), 14–25. <https://doi.org/10.33633/joins.v7i1.5883>

Mustofa, M. A. Psikologi, F. (2022). Persepsi Masyarakat dalam Kebijakan Merger Bank Syariah Indonesia Studi Kasus Kabupaten Tegal. *Ekonomika dan Bisnis Islam*, 5, 207–216. <https://journal.unesa.ac.id/index.php/jei/article/view/12623>

Normawati, D., & Prayogi, S. A. (2021). Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter. *Jurnal Sains Komputer & Informatika (J-SAKTI)*, 5(2), 697–711. <http://ejournal.tunasbangsa.ac.id/index.php/jsakti/article/view/369>

Nugraha, F. A. Habibi, R. (2020). *Analisis Sentimen Terhadap Pembatasan Sosial Menggunakan Deep Learning* (K. I. Nusantara (ed.)).

Nurlina. (2019). *Persepsi Masyarakat Massenrempulu Terhadap Bank Syariah Di Kota Parepare*. Institut Agama Islam Negeri (IAIN) Parepare.

Pasaribu, F. H. Syah, R. (2023). *Analysis of The Multilayer Perceptron Algorithm on Twitter User's Sentiment Towards The COVID-19 Vaccine*. 6(January), 403–413.

Prastowo, A. Ritha, N. (2021). Implementasi Algoritma Radial Basis Function Neural Network dalam Memprediksi Jumlah Angka Kelahiran. *Jurnal Sustainable*, 10, 1–6.

- Putra, M. W. A. Herwin. (2020). Analisis Sentimen Dompok Elektronik Pada Twitter Menggunakan Metode Naïve Bayes Classifier. *IT Journal Research and Development*, 5(1), 72–86. [https://doi.org/10.25299/itjrd.2020.vol5\(1\).5159](https://doi.org/10.25299/itjrd.2020.vol5(1).5159)
- Rahayu, I. P. Indra, J. (2022). Analisis Sentimen Terhadap Program Kampus Merdeka Menggunakan Naive Bayes Dan Support Vector Machine. *Jurnal Sistem Komputer dan Informatika (JSON) Hal: 296–, 301(2)*, 25–38.
- Ramadani, M., & Nurmala Hadiani, N. (2022). Analisis Persepsi Konsumen Bank Syariah Indonesia (Studi Pada Mahasiswa Universitas Serang Raya). *Jurnal Manajemen Perusahaan: JUMPA*, 1(2), 9–20. <https://doi.org/10.30656/jumpa.v1i2.5335>
- Rosdiana, R. Muhammad, N. Y. U. (2019). Analisis Sentimen pada Twitter terhadap Pelayanan Pemerintah Kota Makassar. *Proceeding SNTEI*, 87–93.
- Rusardi, M. N. Adikara, P. P. (2022). Analisis Sentimen Masyarakat terhadap Isu New Normal Scenario berdasarkan Opini dari Twitter menggunakan Algoritma Naive Bayes Classifier. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 6(3), 1434–1440.
- Sandryan, M. K. Ratnawati, D. E. (2021). Analisis Sentimen Pada Media Sosial Twitter Terhadap Undang-Undang Cipta Kerja Menggunakan Algoritma Backpropagation dan Term Frequency-Inverse Document Frequency. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 5(12), 5349–5355.
- Saputra, I. (2021). *Persepsi Nasabah Pada Layanan Mobile Banking BRI Syariah (BSI) Kota Bengkulu*. Institut Agama Islam Negeri (IAIN) Bengkulu.
- Sari, F. V., & Wibowo, A. (2019). Analisis Sentimen Pelanggan Toko Online Jd.Id Menggunakan Metode Naïve Bayes Classifier Berbasis Konversi Ikon Emosi. *Jurnal SIMETRIS*, 10(2), 681–686.
- Sembiring, J., & Sondang. (2019). Deteksi Awal Penyakit Maag Melalui Lidah Menggunakan Metode Radial Basis Function Network Pada Smartphone Berbasis Android. *Seminar Nasional Sistem Informasi Indonesia*.

- Suryono, S., & Taufiq Luthfi, E. (2021). Analisis sentimen pada Twitter dengan menggunakan metode Naïve Bayes Classifier. *JNANALOKA*, 81–86. <https://doi.org/10.36802/jnanaloka.2020.v1-no2-81-86>
- Wahyuningrum, V. (2020). Penerapan Radial Basis Function Neural Network dalam Pengklasifikasian Daerah Tertinggal di Indonesia. *Jurnal Aplikasi Statistika & Komputasi Statistik*, 12(1), 37. <https://doi.org/10.34123/jurnalasks.v12i1.250>
- Widowati, T. T., & Sadikin, M. (2021). Analisis Sentimen Twitter terhadap Tokoh Publik dengan Algoritma Naive Bayes dan Support Vector Machine. *Simetris: Jurnal Teknik Mesin, Elektro dan Ilmu Komputer*, 11(2), 626–636. <https://doi.org/10.24176/simet.v11i2.4568>
- Yuliska, Y. ... Najwa, N. F. (2021). Analisis Sentimen pada Data Saran Mahasiswa Terhadap Kinerja Departemen di Perguruan Tinggi Menggunakan Convolutional Neural Network. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 8(5), 1067. <https://doi.org/10.25126/jtiik.2021854842>
- Yulita, W. (2021). Analisis Sentimen Terhadap Opini Masyarakat Tentang Vaksin Covid-19 Menggunakan Algoritma Naïve Bayes Classifier. *Jurnal Data Mining dan Sistem Informasi*, 2(2), 1. <https://doi.org/10.33365/jdmsi.v2i2.1344>

LAMPIRAN

1. Source Code

```

# Scraping data menggunakan API Twitter
import tweepy as tw
import pandas as pd

# Menggunakan token dari twitter
consumer_key= 'QYtVJY5kibBG6NZRlqqTUgSdN'
consumer_secret=
'tcoS7qfAxmUiWxETkjJMDndyxri4di2iBxkGLqfMgC086dzvab'
access_token='1539625531591004163-
z7S5pVvUlwff9YnKORbr1s6SmWzaUY'
access_token_secret='rLKgD9onjSfHAcRGVnVlLWRfnqFIyLIQrDexAX7Rrt
NhI'

auth = tw.OAuthHandler(consumer_key, consumer_secret)
auth.set_access_token(access_token, access_token_secret)

api = tw.API(auth, wait_on_rate_limit=True)

# menentukan query
search_words = "bank syariah indonesia"
new_search = search_words + "-filter:retweets"

tweets = tw.Cursor(api.search_tweets,
                    q = new_search,
                    count = 1900,
                    include_entities = True).items(1900)

users_locs = [[tweet.created_at, tweet.id, tweet.text] for
tweet in tweets]

tweet_text = pd.DataFrame(data = users_locs,
                           columns = ["date", "id",
"teks"])#Variabel

tweet_text.to_csv('bsi.csv', index=False)

#Menghubungkan google drive dengan google collab
from google.colab import drive
drive.mount('/content/drive')

#Membaca Data di csv/excel
import pandas as pd

```

```

df = pd.read_csv('/content/drive/MyDrive/data/bsi.csv')
df

from google.colab import drive
drive.mount('/content/drive')

#Mengecek nilai kosong
df.isna().sum()

df.duplicated().sum()

#Case Folding
import re

def cleaningulasan(teks):
    teks = re.sub(r'@[A-Za-a0-9]+', ' ', teks)
    teks = re.sub(r'#[A-Za-z0-9]+', ' ', teks)
    teks = re.sub(r"http\S+", ' ', teks)
    teks = re.sub(r'[0-9]+', ' ', teks)
    teks = re.sub(r"[-()\\"#/@/;:<>{}'+=~|.!?,_]", " ", teks)
    teks = teks.strip(' ')
    teks = teks.lower()
    return teks
df['case_folding'] = df['teks'].apply(cleaningulasan)
df.head()

#Tokenization
import nltk
nltk.download('punkt')
from nltk.tokenize import word_tokenize

def tokenizingText(teks):
    teks = word_tokenize(teks)
    return teks
df['tokenization'] = df['case_folding'].apply(tokenizingText)
df.head()

#Stopwords
from nltk.corpus import stopwords
nltk.download('stopwords')

daftar_stopword = stopwords.words('indonesian')
daftar_stopword.extend(["yg", "dg", "rt", "dn", "bsi", 'kak', "gt", "d
m", "pin"])
daftar_stopword = set(daftar_stopword)

```

```

def stopwordsText(words):
    return [word for word in words if word not in daftar_stopword]

df['Stopwords'] = df['tokenization'].apply(stopwordsText)
df.head()

#Stemming
#Sastrawi yang mengatasi data yang berbahasa indonesia
from Sastrawi.Stemmer.StemmerFactory import StemmerFactory

factory = StemmerFactory()
stemmer = factory.create_stemmer()

def porterstemmer(text):
    text = ' '.join(stemmer.stem(word) for word in text if word
in text)
    return text

df['stemming'] = df['Stopwords'].apply(porterstemmer)
df.head()

#Labelling
# Mengubah tipe data dari kolom date

df['date'] = pd.to_datetime(df['date'], format='%Y-%m-%d')

import nltk
nltk.download('vader_lexicon')
from nltk.sentiment.vader import SentimentIntensityAnalyzer
#library untuk membuat label otomatis

sentimen = SentimentIntensityAnalyzer()
polaritas = lambda x : sentimen.polarity_scores(x)
nilai_sentimen = df.stemming.apply(polaritas)
nilai_sentimen.head()

# Menentukan sentimen positif=0 dan negatif=1

label = lambda x : 1 if x>0 else 0
sentiment_data['sentimen'] =
sentiment_data.compound.apply(label)
sentiment_data.head()

# Menampilkan jumlah masing-masing kelas

dataset['sentimen'].value_counts()

```

```

#TF-IDF
stem = dataset['stemming'].astype(str)

#Mentransformasi text jadi vector
from sklearn.feature_extraction.text import TfidfVectorizer
#Library tf-idf

vectorizer = TfidfVectorizer()
vector = vectorizer.fit_transform(stem)

# membagi data untuk di seleksi

from sklearn.model_selection import train_test_split

X = vector
y = dataset['sentimen']

X_train, X_test, y_train, y_test = train_test_split(X, y,
test_size=0.4, random_state=21)

from RBFClassifier import RBFClassifier # panggil Libraries
RBFNN
from sklearn.metrics import accuracy_score # panggil library
from sklearn.metrics import precision_score
from sklearn.metrics import recall_score

rbfnn = RBFClassifier()
rbfnn.fit(X_train, y_train)

y_pred = rbfnn.predict(X_test)

acc = accuracy_score(y_test, y_pred) # Akurasi
pre = precision_score(y_test, y_pred, average="micro") # Presisi
re = recall_score(y_test, y_pred, average="micro") # Recall

print('Akurasi :', acc)
print('Presisi :', pre)
print('Recall :', re)

from sklearn.metrics import confusion_matrix

print('Confusion Matrix :\n')
cm = confusion_matrix(y_test, y_pred)
print(cm)

from sklearn.model_selection import KFold, cross_validate

```

```
kfold = KFold(10, shuffle=True) #Percobaan kfold
scores = cross_validate(rbfnn, X, y, cv=kfold,
scoring=['accuracy'])

print('Accuracy : ',scores['test_accuracy'].mean())# Nilai
rata-rata

y_test.value_counts()

import matplotlib.pyplot as plt

data1 = y_test.value_counts()
data1.plot(kind='pie',autopct='%.2f%%')
plt.show()
```

2. Plagiasi



turnitin Similarity Report ID: cid:29477-45437521

PAPER NAME	188160004_Aderian_Terbaru_2.pdf	AUTHOR	Ade Rian
WORD COUNT	6019 Words	CHARACTER COUNT	35733 Characters
PAGE COUNT	38 Pages	FILE SIZE	854.7KB
SUBMISSION DATE	Oct 26, 2023 2:09 PM GMT+7	REPORT DATE	Oct 26, 2023 2:10 PM GMT+7

22% Overall Similarity
The combined total of all matches, including overlapping sources, for each database.

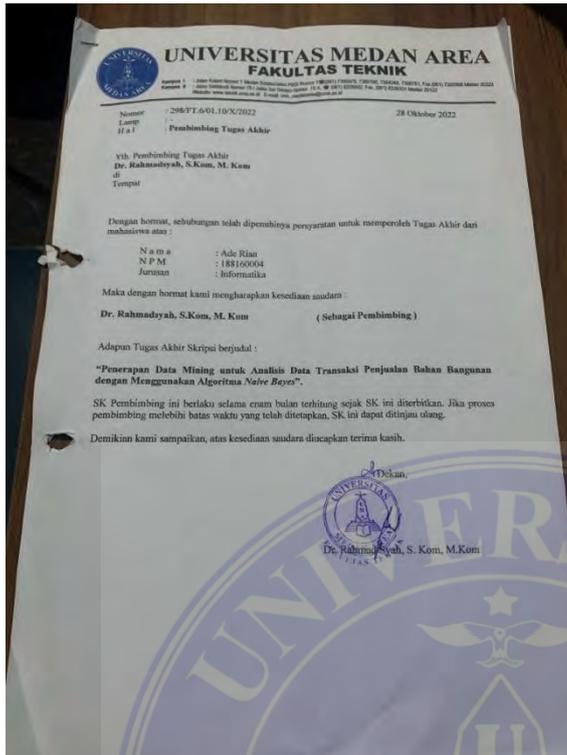
- 15% Internet database
- 3% Publications database
- Crossref database
- Crossref Posted Content database
- 15% Submitted Works database

Excluded from Similarity Report

- Small Matches (Less than 10 words)

Summary

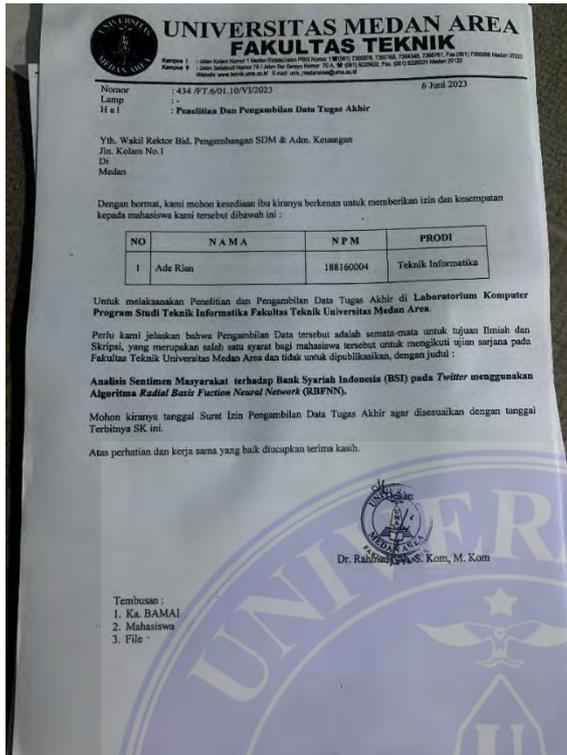
3. Surat Keterangan Pembimbing



4. Surat Izin Melakukan Riset dari Fakultas Teknik



5. Surat izin melakukan penelitian di PUSKOM Universitas Medan Area dari BATRI



6. Surat Selesai Riset

