

**SENTIMENT ANALYSIS PADA TWEET TENTANG PPKM DI
INDONESIA
(SKRIPSI)**

Oleh

**MAULANA TAUFIQURROHMAN
1817051030**



**JURUSAN ILMU KOMPUTER
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
2023**

**SENTIMENT ANALYSIS PADA TWEET TENTANG PPKM DI
INDONESIA**

Oleh :

Maulana Taufiqurrohman

Skripsi

Sebagai Salah Satu Syarat untuk Memperoleh Gelar

SARJANA ILMU KOMPUTER

Pada

Jurusan Ilmu Komputer

Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung



**JURUSAN ILMU KOMPUTER
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG**

2023

ABSTRACT

SENTIMENT ANALYSIS ON TWEET ABOUT PPKM IN INDONESIA

By

MAULANA TAUFIQURROHMAN

The PPKM policy in Indonesia is very troubling for Indonesian people from the lower middle class to the less fortunate people who are affected by this policy, giving rise to many pro and con opinions on this policy; the public gives this opinion through social media, especially on social media tweets because on the media socially, there is much talk about complaints from the public about the government, these opinions become criticisms and suggestions that contain negative and positive opinions for the Indonesian government which then from these opinions are made into a dataset which is taken through the method of scraping tweet data using the search tweet API tweet, data that has been taken will then be analyzed through several selected algorithms, namely RNN-LSTM, Naive Bayes, and Logistic Regression. This algorithm was chosen based on progress in today's modern era, namely deep learning and machine learning; the data obtained will go through several stages, starting from data labeling, preprocessing, data splitting, training, testing, sentiment classification, evaluation and results. After going through several processes, the three algorithms produce 100% accuracy for RNN-LSTM, 82.83% for Naive Bayes, and 90.5% for logistic regression.

Keywords: ppkm, opinions, sentiment, rnn-lstm, naive bayes, logistic regression.

ABSTRAK

SENTIMENT ANALYSIS PADA TWEET TENTANG PPKM DI INDONESIA

Oleh

MAULANA TAUFIQURROHMAN

Kebijakan PPKM di Indonesia sangat meresahkan masyarakat Indonesia dari masyarakat menengah kebawah sampai ke masyarakat yang kurang mampu terkena dampak dari kebijakan ini, sehingga menimbulkan banyak opini pro dan kontra pada kebijakan tersebut, masyarakat memberikan opini ini melalui sosial media terutama pada sosial media tweet karena pada media sosial ini banyak sekali perbincangan keluh kesah bagi masyarakat terhadap pemerintah, opini ini menjadi kritik dan saran yang berisikan opini negatif dan positif untuk pemerintah Indonesia yang kemudian dari opini ini dijadikan suatu dataset yang diambil melalui metode scraping data tweet menggunakan api tweet searchtweet, data yang sudah diambil kemudian akan di analisis dengan melalui beberapa algoritma yang dipilih yaitu RNN-LSTM, naive bayes, dan Logistic Regression algoritma ini dipilih berdasarkan kemajuan pada era modern sekarang ini yaitu deep learning dan machine learning, data yang sudah didapatkan akan melalui beberapa tahapan dimulai dari data labeling, preprocessing, data splitting, training, testing, sentiment classification, evaluasi dan hasil. setelah melalui beberapa proses tersebut dari ketiga algoritma tersebut menghasilkan akurasi 100% untuk RNN-LSTM, 82,83% untuk Naive Bayes, dan 90,5% untuk logistic regression.

Kata kunci: *ppkm, sentiment, opini, rnn-lstm, naive bayes, logistic regression.*

Judul Skripsi : SENTIMENT ANALYSIS PADA TWEET
TENTANG PPKM DI INDONESIA

Nama Mahasiswa : Maulana Taufiqurrohman

Nomor Pokok Mahasiswa : 1817051030

Program Studi : Ilmu Komputer

Fakultas : Matematika dan Ilmu Pengetahuan Alam



1. Komisi Pembimbing

Kurnia Muludi

Dr. Ir. Kurnia Muludi, M.S.Sc.

NIP. 196406161989021001

2. Ketua Jurusan Ilmu Komputer

Didik Kurniawan

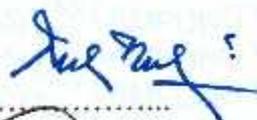
Didik Kurniawan, S.Si., M.T.

NIP. 198004192005011004

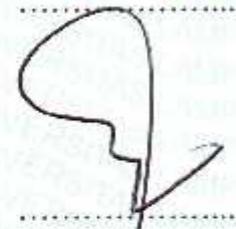
MENGESAHKAN

1. Tim Penguji

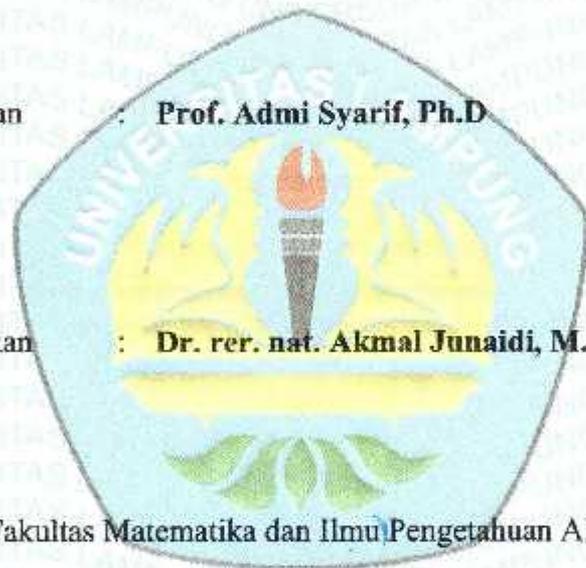
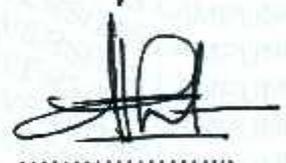
Ketua : **Dr. Ir. Kurnia Muludi, M.S.Sc.**



Penguji I Bukan Pembimbing : **Prof. Admi Syarif, Ph.D**



Penguji II Bukan Pembimbing : **Dr. rer. nat. Akmal Junaidi, M.Sc.**



2. Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam



Dr. Eng. Heri Satria, S.Si., M.Si.

NIP. 197110012005011002

Tanggal lulus ujian skripsi: 08 Agustus 2023

PERNYATAAN

Saya yang bertanda tangan di bawah ini:

Nama : Maulana Taufiqurrohman

NPM : 1817051030

menyatakan bahwa skripsi saya yang berjudul **"Sentiment Analysis Pada Tweet Tentang Ppkm Di Indonesia"** merupakan karya saya sendiri dan bukan karya orang lain. Semua tulisan yang tertuang dalam skripsi ini telah mengikuti kaidah penulisan karya ilmiah Universitas Lampung. Apabila dikemudian hari terbukti skripsi saya merupakan hasil penjiplakan atau dibuat orang lain, maka bersedia menerima sanksi berupa pencabutan gelar yang telah saya terima.

Bandar Lampung, 22 Agustus 2023



Maulana Taufiqurrohman

NPM. 1817051030

RIWAYAT HIDUP



Penulis bernama Maulana Taufiqurrohman bertempat lahir di Kota Bandar Lampung pada tanggal 03 Maret 2000, sebagai anak kedua dari tiga bersaudara. Penulis menyelesaikan pendidikan formal di SDN 02 Palapa dan selesai pada tahun 2006-2012. Kemudian melanjutkan pendidikan menengah pertama di MTs Negeri 2 Bandar Lampung yang diselesaikan pada tahun 2015, lalu melanjutkan ke pendidikan menengah atas di MAN 2 Bandar Lampung yang diselesaikan pada tahun 2018.

Pada tahun 2018 penulis terdaftar sebagai mahasiswa Jurusan Ilmu Komputer Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung melalui jalur SBMPTN. Selama menjadi mahasiswa, penulis melakukan beberapa kegiatan antara lain.

1. Menjadi anggota Adapter Himpunan Mahasiswa Jurusan Ilmu Komputer pada periode 2018/2019
2. Menjadi anggota Bso BBQ Rohani Islam Fakultas Matematika dan Ilmu Pengetahuan Alam periode 2020/2021
3. Menjadi anggota Kestari FIF pada tahun 2018

4. Menjadi anggota Bansus Himpunan Mahasiswa Jurusan Ilmu Komputer periode 2018/2019
5. Menjadi anggota Akset Rohani Islam Fakultas Matematika dan Ilmu Pengetahuan Alam periode 2019/2020
6. Menjadi Asisten Dosen Pengantar Organisasi Komputer Pada Tahun 2020
7. Melaksanakan Kuliah Kerja Nyata Mandiri di Kelurahan Durian Payung Kota Bandar Lampung pada Periode 2020/2021
8. Menjadi Anggota KSAN Anemon Periode 2019/2020
9. Melaksanakan Kerja Praktik di Fakultas Ekonomi dan Bisnis Universitas Lampung pada tahun 2021

MOTTO

“Barangsiapa yang tidak bersyukur meski sedikit, maka ia tidak akan mampu mensyukuri sesuatu yang banyak.”

(HR. Ahmad)

“Don’t tell people your plans, show them your results”

(Anonim)

“Allah tidak membebani seseorang melainkan sesuai dengan kesanggupannya”

(QS. Al-Baqarah 2: Ayat 286)

“Karena sesungguhnya sesudah kesulitan itu ada kemudahan”

(QS. Al Insyirah ayat 5)

PERSEMBAHAN

Alhamdulillahirobbilalamin

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 Shallallahu Alaihi Wasallam.

Kupersembahkan karya ini kepada:

Kedua Orang Tuaku Tercinta

Yang senantiasa memberikan yang terbaik, dan melantunkan do'a yang selalu menyertaiku. Kuucapkan pula terima kasih sebesar-besarnya karena telah mendidik dan membesarkanku dengan cara yang dipenuhi kasih sayang, dukungan, dan pengorbanan yang belum bisa terbalaskan.

Seluruh Keluarga Besar Ilmu Komputer 2018

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 kehadiran 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 **“Sentiment Analysis Pada Tweet Tentang Ppkm Di Indonesia”** dengan baik dan lancar.

Selesainya skripsi ini tidak terlepas dari bantuan, bimbingan, dan dukungan dari berbagai pihak. Oleh karena itu, ucapan terima kasih ditujukan kepada:

1. Allah *Subhanahu Wa Ta'ala* yang telah memberikan rahmat dan karunia-Nya selama menjalani penelitian hingga laporan dapat diselesaikan dengan baik;
2. Ummi, Abi, Tete, Adik dan Keluarga Besar Abah Surya selaku keluarga yang selalu memberikan doa, kasih sayang, dan dukungannya sehingga penelitian ini terselesaikan.
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 Favorisen R. Lumbanraja S.Si., M.Si., Ph.D, selaku dosen pembimbing akademik dan yang telah memberikan arahan dan bimbingan setiap semester hingga penelitian ini selesai.
6. Bapak Dr. Ir. Kurnia Muludi, M.S.Sc., selaku dosen pembimbing utama dalam penelitian ini, yang selalu memberikan bimbingan, ilmu, dan saran;
7. Bapak Prof. Admi Syarif, Ph.D, selaku dosen pembimbing pembahas I, yang telah memberikan ilmu dan saran dalam penelitian ini;
8. Bapak Dr. rer. nat. Akmal Junaidi, M.Sc.. selaku Wakil Dekan III FMIPA Unila dan dosen pembahas II, yang telah memberikan ilmu dan saran dalam penelitian ini.
9. Ibu Anie Rose Irawati, ST, M.Cs. Selaku sekretaris Jurusan Ilmu Komputer Universitas Lampung dan Koordinator Skripsi.

10. Seluruh Bapak dan Ibu dosen Jurusan Ilmu Komputer yang telah memberikan ilmunya selama masa perkuliahan.
11. Teman seperjuangan di kala suka dan duka selama perkuliahan Jonathan Michael, Eka Intan Sari, Nia Nur Atika, Arbi Hidayatulloh, M. Arsyi Shobirin yang selalu memberi semangat dan membantu mencari jalan keluar setiap ada masalah.
12. M. Azriel, Suci Hasanah Berta, Rika Ningtyas, Vindo, Ka Adhi yang tidak pernah lelah menjadi tempat untuk bertanya mengenai informasi yang diperlukan.
13. Arman Efendi, Galang Trimedia, Shela, Dwi yang selalu memberi dukungan dari MAN sampai menyelesaikan skripsi ini.
14. Aini, Alfa, Lutfi, Ana, Isa, Nanda, dan Viani yang selalu mendukung dari jarak jauh sampai skripsi ini selesai.
15. Teman-teman seangkatan Jurusan Ilmu Komputer 2018 di Universitas Lampung.

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

Bandar Lampung, 22 Agustus 2023



Maulana Taufiqurrohman

NPM. 1817051030

DAFTAR ISI

	Halaman
DAFTAR ISI.....	i
DAFTAR TABEL.....	iii
DAFTAR GAMBAR	v
I. Pendahuluan	1
1.1. Latar Belakang.....	1
1.2. Rumusan Masalah	2
1.3. Batasan Masalah.....	3
1.4. Manfaat Penelitian.....	3
1.5. Tujuan Penelitian.....	3
II. Tinjauan Pustaka	4
2.1. Penelitian Terdahulu.....	4
2.2. Dasar Teori	11
2.2.1. Sentiment analysis.....	11
2.2.2. Recurrent Neural Network (RNN).....	11
2.2.3. Logistic Regression.....	12
2.2.4. Naive Bayes	13
2.2.5. Python	15
III. METODOLOGI PENELITIAN.....	19
3.1. Tempat dan Waktu Penelitian	19
3.1.1. Tempat Penelitian.....	19
3.1.2. Waktu Penelitian	19
3.2. Alat dan Objek Penelitian.....	20
3.2.1. Alat Penelitian.....	20
3.2.2. Objek Penelitian	21
3.3. Sumber Data dan Metode Pengumpulan data	21

3.3.1. Sumber Data.....	21
3.3.2. Metode Pengumpulan data.....	21
3.4. Metode Penelitian.....	21
3.5. Tahapan Penelitian	22
IV. HASIL DAN PEMBAHASAN.....	31
4.1. Hasil.....	31
4.1.1. Pengumpulan Data	31
4.1.2. Data Labeling	32
4.1.3. Penghapusan Stopword	33
4.1.4. Tokenisasi Teks.....	33
4.1.5. Data Spliting (Pemisahan Data).....	34
4.1.6. Data Training (data latih).....	34
4.1.7. Sentiment Clasification	35
4.1.8. Metode Oversampling dan Undersampling	42
4.1.9. Evaluasi dan Hasil Analisis.....	44
4.2. Pembahasan	46
V. SIMPULAN DAN SARAN.....	47
5.1. Simpulan.....	47
5.2. Saran.....	47
DAFTAR PUSTAKA	48

DAFTAR TABEL

	Halaman
Table 1 Penelitian Terdahulu	8
Table 2 Perbedaan Api Tweet	17
Table 3 Alur Penelitian	19
Table 4 Confusion Matrix	29
Table 5 Hasil Sentiment Manual	33
Table 6 Stopword Removal	33
Table 7 Tokenisasi Teks	33
Table 8 Confusion Matrix RNN-LSTM Data Test	37
Table 9 Confusion Matrix RNN-LSTM Data Train	37
Table 10 Confusion Matrix RNN-LSTM Data 3000	38
Table 11 Confusion Matrix RNN-LSTM Data Train OverSampling	38
Table 12 Confusion Matrix RNN-LSTM Data Train UnderSampling	38
Table 13 Confusion Matrix RNN-LSTM Data 3000 UnderSampling	38
Table 14 Confusion Matrix Logistic Regression Data Test	39
Table 15 Confusion Matrix Logistic Regression Data Train	39
Table 16 Confusion Matrix Logistic Regression Data 3000	39
Table 17 Confusion Matrix Logistic Regression Data Train OverSampling	40
Table 18 Confusion Matrix Logistic Regression Data Test OverSampling	40
Table 19 Confusion Matrix Logistic Regression Data 3000 OverSampling	40

Table 20 Confusion Matrix Logistic Regression Data Train UnderSampling.....	40
Table 21 Confusion Matrix Logistic Regression Data Test UnderSampling	40
Table 22 Confusion Matrix Logistic Regression Data 3000 UnderSampling	40
Table 23 Confusion Matrix Naive Bayes Data Test	41
Table 24 Confusion Matrix Naive Bayes Data Train.....	41
Table 25 Confusion Matrix Naive Bayes Data 3000	41
Table 26 Confusion Matrix Naive Bayes Data Train OverSampling	41
Table 27 Confusion Matrix Naive Bayes Data Test OverSampling	42
Table 28 Confusion Matrix Naive Bayes Data 3000 OverSampling	42
Table 29 Confusion Matrix Naive Bayes Data Train UnderSampling	42
Table 30 Confusion Matrix Naive Bayes Data Test UnderSampling	42
Table 31 Confusion Matrix Naive Bayes Data 3000 UnderSampling	42
Table 32 Hasil Analisis data Label Manual Tanpa Sampling (Data Train).....	44
Table 33 Hasil Analisis data menggunakan Oversampling (Data Train)	44
Table 34 Hasil Analisis data menggunakan Undersampling (Data Train)	44
Table 35 Hasil Analisis data Label Manual tanpa Sampling (Data Test)	45
Table 36 Hasil Analisis data menggunakan Oversampling (Data Test)	45
Table 37 Hasil Analisis data menggunakan Undersampling (Data Test)	45
Table 38 Table 11 Hasil Analisis data Label Manual Tanpa Sampling (Data 3000)	45
Table 39 Hasil Analisis data menggunakan Oversampling (Data 3000)	46
Table 40 Hasil Analisis data menggunakan Undersampling (Data 3000)	46

DAFTAR GAMBAR

	Halaman
Gambar 1 LSTM model (dari Medium).....	12
Gambar 2 CBOW Model Sumber : (Mikolov et al., 2013).....	25
Gambar 3 RNN Framework Sumber: Denny Britz, 2015.....	25
Gambar 4 Long short term memory sumber : (Alwehaibi & Roy, 2019).....	26
Gambar 5 Arsitektur RNN-LSTM	27
Gambar 6 Logistic regression (Arista, 2022; Majumder et al., 2021)	28
Gambar 7 kode program untuk search tweet.....	31
Gambar 8 data tweet hasil scraping	32
Gambar 9 tweet hasil labeling manual	32
Gambar 10 Hasil Data Spliting	34
Gambar 11 Model word2vec.....	35
Gambar 12 Hasil Model RNN-LSTM.....	36
Gambar 13 grafik acc dan val_acc	37
Gambar 14 Model Logistic regression.....	39
Gambar 15 Hasil Model Naive Bayes.....	41
Gambar 16 Hasil Kode Program Oversampling	43
Gambar 17 Hasil Kode Program Undersampling	43

I. PENDAHULUAN

1.1. Latar Belakang

Kebijakan PPKM di Indonesia sangat meresahkan masyarakat di Indonesia terutama bagi para pedagang kecil dan pedagang keliling lain nya yang hanya bergantung pada penghasilan yang pas-pasan seperti penghasilan harian untuk bisa memenuhi kebutuhan hidupnya dan keluarga sehingga hal ini memicu banyak opini public mengenai kebijakan PPKM di Indonesia apakah baik atau tidak terkait opini public terhadap pemerintah nya, sehingga dalam penelitian ini mengambil opini – opini dari publik kemudian difilter mana yang baik, buruk.

Opini mining atau analisis sentimen adalah metode penelitian komputasional untuk menganalisis opini, sentimen, dan emosi yang diungkapkan dalam bentuk teks oleh publik. Saat ini, jika suatu organisasi, perusahaan, atau individu ingin mendapatkan pendapat publik tentang produk, citra, atau layanan mereka, mereka tidak perlu lagi melakukan survei konvensional atau fokus grup yang mahal dan dapat merugikan para responden.

Deep Learning merupakan bidang baru dalam penelitian Machine Learning, yang bertujuan untuk memperkuat hubungan antara Machine Learning dan Artificial Intelligence (AI), salah satu tujuan utama dari Machine Learning. Deep Learning melibatkan pembelajaran beberapa tingkat representasi dan abstraksi untuk memfasilitasi pemahaman data seperti gambar, suara, dan teks secara efisien dan mudah.

Recurrent Neural Network (RNN) adalah suatu pengembangan dari Deep Learning yang menggunakan metode atau algoritma tumpukan (stack) yang

bertujuan untuk mengekstraksi fitur secara optimal dari sumber daya yang tersedia. Deep Learning sendiri mencakup algoritma supervised dan unsupervised learning sehingga memungkinkan penggunaannya dengan data yang berlabel atau tanpa label. Hal ini sangat penting untuk menjalankan algoritma yang digunakan. Salah satu pendekatan yang sering digunakan untuk mengimplementasikan Deep Learning adalah menggunakan metode grafis atau model grafis multi-layer seperti Recurrent Neural Network.

Sumber data Twitter berbahasa Indonesia dipilih karena penelitian dalam Bahasa Indonesia belum banyak dilakukan pernyataan ini saya ambil dari penelitian (Surya Akbar et al., 2021; Zulfa & Winarko, 2017), dan penelitian dengan tema ppkm secara umum di tweet belum ada yang melakukan penelitian tersebut, sedangkan RNN dipilih karena metode ini sangat efektif dalam kasus *sentiment* analisis dan diharapkan dapat mencapai kinerja yang lebih baik dari metode – metode sebelumnya. Metode RNN diharapkan bisa digunakan dalam pengimplementasian Tweet untuk keperluan data *training* dikumpulkan dari beragam data yang dianotasikan kelas *sentiment*nya secara otomatis dengan akun media sosial di Twitter. Akun media sosial merupakan teks yang hanya berisi pernyataan atas suatu fakta atau kejadian dan tidak mengekspresikan emosi tertentu. Apakah nantinya karakteristik RNN ini sesuai jika diaplikasikan pada data Twitter. Dan untuk memastikan perbedaan hasil kinerja dari metode RNN dengan metode lainnya, saya membuat perbandingan antara metode RNN dan metode lainnya dalam penelitian ini. Berdasarkan hal tersebut akan dibuat *Sentiment* Analysis pada tweet tentang ppkm di Indonesia menggunakan metode RNN.

1.2.Rumusan Masalah

Dari Latar Belakang Masalah di atas dapat diambil rumusan masalah sebagai berikut

1. Bagaimana hasil prediksi algoritma RNN, logistik Regresion dan Naive Bayes dalam menentukan sentiment positif negatif dari tweet

2. Bagaimana perbandingan akurasi, f1-score, presisi dari RNN, logistik regresi dan *naive bayes* dalam melakukan *sentiment analysis*.

1.3. Batasan Masalah

Batasan masalah Dari penelitian ini adalah sebagai berikut :

1. Algoritma yang digunakan RNN, *logistic regression*, dan *naive bayes*.
2. Sumber Data yang digunakan pada penelitian ini berasal dari scraping data tweet kemudian dari banyak nya data tweet yang diambil akan di label secara manual untuk bisa digunakan pada algoritma yang akan dipakai.
3. Objek pada penelitian ini yaitu tweet yang mengandung opini positif dan negatif yang berkaitan tentang ppkm yang terjadi di Indonesia berdasarkan persepsi masyarakat terhadap kebijakan ppkm.

1.4. Manfaat Penelitian

Manfaat dari penelitian ini adalah sebagai berikut :

1. Mengetahui perbandingan opini dari masyarakat Indonesia di tweet baik itu opini positif atau negatif mengenai ppkm yang terjadi di Indonesia.
2. Mengetahui perbandingan algoritma antara RNN, logistik regresi dan *naive bayes* dari sisi akurasi .

1.5. Tujuan Penelitian

Penelitian ini dilakukan untuk mengevaluasi kinerja dari algoritma RNN, *naive bayes* dan logistik regresi dalam melakukan sentimen analysis.

II. TINJAUAN PUSTAKA

2.1. Penelitian Terdahulu

Penelitian ini tidak terlepas dari penelitian sebelumnya yang telah dilakukan, Oleh peneliti, pada penelitian sebelumnya dijadikan sebagai perbandingan dan referensi. Selain itu, untuk menghindari anggapan kesamaan dengan penelitian ini maka peneliti mencantumkan hasil-hasil penelitian terdahulu sebagai berikut :

1. *Sentiment* Analysis Terhadap Tweet Bernada Sarkasme Berbahasa Indonesia

Penelitian ini dilakukan oleh (Septiani & Sibaroni, 2019) yang bertujuan untuk *sentiment* analysis tweet yang bernada sarkasme yang menggunakan metode *naive bayes* dan SVM serta masing-masing ada yang menggunakan fitur injeksi dan ada yang tidak, hasilnya akurasi yang di peroleh menggunakan *naive bayes* lebih tinggi dibanding menggunakan SVM dengan akurasi *naive bayes* tanpa fitur injeksi mendapatkan akurasi 82,49% sedangkan yang menggunakan fitur injeksi mendapatkan akurasi 91,39% dan untuk SVM yang tidak menggunakan fitur injeksi mendapatkan akurasi sebesar 43,62% sedangkan yang menggunakan fitur injeksi mendapatkan akurasi sebesar 57,91% dari sini dapat dibuktikan bahwa penggunaan fitur injeksi dapat mempengaruhi tingkat akurasi pada *naive bayes* mendapatkan kenaikan akurasi sebesar 8% sedangkan SVM mendapatkan kenaikan akurasi sebesar 13%. Kelebihan dari penelitian ini adalah dengan menggunakan

fitur injeksi kita dapat mengukur kenaikan akurasi yang diperoleh dari setiap algoritma dan kelemahan dari penelitian ini untuk akurasi algoritma yang tidak menggunakan fitur injeksi akan mengalami penurunan akurasi dibandingkan menggunakan fitur injeksi.

2. Analisis *Sentiment* Pada Twitter Mahasiswa Menggunakan Metode Backpropagation

Penelitian ini dilakukan oleh (Habibi et al., 2016) yang bertujuan untuk *sentiment analysis* twitter mahasiswa yang mengandung *sentiment* positif dan negatif dengan menggunakan metode Backpropagation dengan mendapatkan nilai akurasinya sebesar 85%, nilai precision 0,930, nilai recall 0,899, dan nilai F-measure 0,901, serta hasil *sentiment* yang di dapat *sentiment* positif 33,330%, negatif 13,330%, dan netral 53,330%. Kelebihan dari penelitian ini hasil akurasinya yang sangat bagus dan pengukuran lainnya yang bagus untuk nilai dari precision, recall dan F-measure, kelemahan dari penelitian ini adalah banyaknya hasil *sentiment* yang netral pada tweet yang di analisis jadi cenderung di tengah tidak memihak ke positif maupun negatif.

3. *Sentimentt* Analysis Of Energy Independence Tweets Using Simple Recurrent Neural Network

Penelitian ini dilakukan oleh (Surya Akbar et al., 2021) bertujuan untuk *sentiment analysis* tweet yang berhubungan dengan kemandirian energi pada negara republik indonesia banyak yang bertentangan pendapat di tweet, metode yang digunakan adalah Simple RNN, LSTM dan Bernauli Naive Bayes dengan masing-masing menghasilkan akurasi, precision, recall, dan F1-score sebagai berikut Simple RNN medapatkan akurasi sebesar 78%, precision sebesar 78%, recall sebesar 78% dan F1-score

sebesar 77%, LSTM mendapatkan akurasi sebesar 75%, precision sebesar 74%, recall sebesar 75%, dan F1-score sebesar 73%, dan Bernauli Naive Bayes mendapatkan akurasi sebesar 67%, precision sebesar 58%, recall sebesar 67%, dan F1-score sebesar 56%.

4. *Sentimentt* Analysis on Social Distancing and Physical Distancing on Twitter Social Media using Recurrent Neural Network (RNN) Algorithm

Penelitian ini dilakukan oleh (Nugraha et al., 2020) yang bertujuan untuk *sentiment* analysis tweet tentang Social distancing dan physical distancing, metode yang digunakan adalah RNN dengan hasil *Training* menggunakan dataset *Training* nya dengan akurasi 89%, recall 89%, precision 89%, dan F1-Score 89% dan hasil Test menggunakan dataset testing nya dengan akurasi 80%, recall 79%, precision 81%, dan F1-score 80% dan untuk hasil klasifikasi dengan keyword Social Distancing 71,8% Positif dan 28,2% negatif *sentiment* nya sedangkan untuk Physical Distancing 58,9% positif dan 41,1% negatif *sentiment* nya.

5. *Sentiment* Analisis Tweet Berbahasa Indonesia Dengan Deep Belief Network

Penelitian ini dilakukan oleh (Zulfa & Winarko, 2017) yang bertujuan untuk melakukan *sentiment* analysis dengan beberapa topik yang sedang ramai di sosmed twitter dan metode yang digunakan Deep Belief Network (DBN), Support Vector Machine (SVM), dan Naive Bayes dengan hasil pengujian klasifikasi tiap metode nya sebagai berikut untuk DBN presisinya 93%, recall 93%, F1-score 93% dan akurasi 93,31%, untuk SVM presisinya 92%, recall 92%, F1-score 92% dan akurasi 92,18%, untuk Naive Bayes presisinya 79%, recall 79%, F1-score 79% dan akurasi 79,10% Kelebihan dari penelitian ini dari semua algoritma yang dipakai hasil akurasinya sangat

bagus selain *naive bayes* hasil akurasi dari pengujian nya sangat bagus kelemahan dari penelitian ini banyak klasifikasi tweet yang cenderung netral untuk yang positif dan negatif kurang mendominasi.

Berikut Tabel untuk penelitian terdahulu

Table 1 Penelitian Terdahulu

Nama Peneliti	Judul Penelitian	Metode Penelitian	Tujuan Penelitian	Hasil Penelitian	Kelebihan dan Kelemahan
(Septiani & Sibaroni, 2019)	<i>Sentimentt</i> Analysis Terhadap Tweet Bernada Sarkasme Berbahasa Indonesia	metode naive bayes dan SVM dengan di setiap masing-masing algoritma menggunakan fitur injeksi	<i>sentiment</i> analysis tweet yang bernada sarkasme yang menggunakan metode naive bayes dan SVM serta masing-masing ada yang menggunakan fitur injeksi dan ada yang tidak	naive bayes tanpa fitur injeksi akurasi 82,49% menggunakan fitur injeksi mendapatkan akurasi 91,39% SVM yang tidak menggunakan fitur injeksi mendapatkan akurasi 43,62% menggunakan fitur injeksi mendapatkan akurasi sebesar 57,91% naive bayes mendapatkan kenaikan akurasi sebesar 8% sedangkan SVM mendapatkan kenaikan akurasi sebesar 13%.	Kelebihan dari penelitian ini adalah dengan menggunakan fitur injeksi kita dapat mengukur kenaikan akurasi yang di peroleh dari setiap algoritma dan kelemahan dari penelitian ini untuk akurasi algoritma yang tidak menggunakan fitur injeksi akan mengalami penurunan akurasi dibandingkan menggunakan fitur injeksi.
(Habibi et al., 2016)	Analisis <i>Sentiment</i> Pada Twitter Mahasiswa Menggunakan Metode Backpropagation	metode Backpropagation	<i>sentiment</i> analysis twitter mahasiswa yang mengandung <i>sentiment</i> positif dan negatif dengan menggunakan metode Backpropagation	Backpropagation mendapatkan nilai akurasi 85%, nilai precision 0,930, nilai recall 0,899, dan nilai F-measure 0,901, hasil <i>sentiment</i> yang di dapat <i>sentiment</i> positif 33,330%, negatif 13,330%, dan netral 53,330%.	Kelebihan dari penelitian ini hasil akurasinya yang sangat bagus dan pengukuran lainnya yang bagus untuk nilai dari precision, recall dan F-measure, kelemahan dari penelitian ini adalah banyak nya hasil <i>sentiment</i> yang netral pada tweet yang di analisis jadi cenderung di tengah tidak memihak ke positif maupun negatif.

(Surya Akbar et al., 2021)	<i>Sentimentt</i> Analysis Of Energy Independence Tweets Using Simple Recurrent Neural Network	metode yang digunakan adalah Simple RNN, LSTM dan Bernauli Naive Bayes	<i>sentiment</i> analysis tweet yang berhubungan dengan kemandirian energi pada negara republik indonesia banyak yang bertentangan pendapat di tweet, metode yang digunakan adalah Simple RNN, LSTM dan Bernauli Naive Bayes dan memprediksi hasil <i>Training</i> data dari masing-masing algoritma dengan data Test nya	RNN medaptkan akurasi sebesar 78%, precision sebesar 78%, recall sebesar 78% dan F1-score sebesar 77%, LSTM mendapatkan akurasi sebesar 75%, precision sebesar 74%, recall sebesar 75%, dan F1-score sebesar 73%, dan Bernauli Naive Bayes mendapatkan akurasi sebesar 67%, precision sebesar 58%, recall sebesar 67%, dan F1-score sebesar 56%.	Kelebihan dari penelitian ini pada setiap algoritma yang dipakai prediksi yang berhasil sesuai dengan data Test yang dipakai yaitu untuk label positif dan kelemahan dari penelitian ini untuk label negatif terlalu banyak prediksi yang kurang tepat dan akurasi yang paling kurang terdapat pada algoritma naive bayes
(Nugraha et al., 2020)	<i>Sentimentt</i> Analysis on Social Distancing and Physical Distancing on Twitter Social Media using Recurrent Neural Network (RNN) Algorithm	metode yang digunakan adalah RNN	<i>sentiment</i> analysis tweet tentang Social distancing dan physical distancing, metode yang digunakan adalah RNN	RNN dengan hasil <i>Training</i> Nya dengan akurasi 89%, recall 89%, precision 89%, dan F1-Score 89% dan hasil Test Nya dengan akurasi 80%, recall 79%, precision 81%, dan F1-score 80% dan untuk hasil klasifikasi dengan keyword Social Distancing 71,8% Positif dan 28,2% negatif sedangkan untuk Physical Distancing 58,9% positif dan 41,1% negatif	Kelebihan dari penelitian ini untuk tingkat akurasinya sangat baik dilihat dari hasil prediksi yang hampir 100 akurat dan kelemahan dari penelitian ini masih ada hasil prediksi yang tidak tepat dalam penelitian ini data yang diambil masih terlalu sedikit.

(Zulfa & Winarko, 2017)	<i>Sentiment</i> Analisis Tweet Berbahasa Indonesia Dengan Deep Belief Network	metode yang digunakan Deep Belief Network (DBN), Support Vector Machine (SVM), dan Naive Bayes	<i>sentiment</i> analysis dengan beberapa topik yang sedang ramai di sosmed twitter dan metode yang digunakan Deep Belief Network (DBN), Support Vector Machine (SVM), dan Naive Bayes	DBN presisinya 93%, recall 93%, F1-score 93% dan akurasi 93,31%, untuk SVM presisinya 92%, recall 92%, F1-score 92% dan akurasi 92,18%, untuk Naive Bayes presisinya 79%, recall 79%, F1-score 79% dan akurasi 79,10%.	Kelebihan dari penelitian ini dari semua algoritma yang dipakai hasil akurasinya sangat bagus selain naive bayes hasil akurasi dari pengujian nya sangat bagus kelemahan dari penelitian ini banyak klasifikasi tweet yang cenderung netral untuk yang positif dan negatif kurang mendominasi akurasi dari algoritma naive bayes kurang bagus di bandingkan dengan DBN .
-------------------------	--	--	--	--	--

Dari Tabel di atas dapat di simpulkan dari beberapa penelitian terdahulu yang saya ambil sebagai acuan penelitian terdapat beberapa kekurangan dari masing-masing penelitan pada penelitian (Septiani & Sibaroni, 2019) kelemahan yang dimiliki dari penelitian ini terdapat perbedaan dari penggunaan fitur injeksi yang digunakan pada algoritma, apabila tidak menggunakan fitur injeksi maka akurasi yang didapatkan berkurang, data yang diambil masih sangat sedikit pada penelitian tersebut, kemudian pada penelitian (Habibi et al., 2016) kelemahan yang dimiliki dari penelitian ini banyak nya hasil sentiment yang netral pada tweet yang di analisis jadi cenderung di tengah tidak memihak ke positif maupun negatif, kemudian pada penelitian (Surya Akbar et al., 2021) kelemahan yang dimiliki dari penelitian ini untuk label negatif terlalu banyak prediksi yang kurang tepat dan akurasi yang paling kurang terdapat pada algoritma naive bayes, pada penelitian (Nugraha et al., 2020) kelemahan yang dimiliki dari penelitian ini masih ada hasil prediksi yang tidak tepat dalam penelitian ini data yang diambil masih terlalu sedikit, pada penelitian (Zulfa & Winarko, 2017) kelemahan yang dimiliki dari penelitian ini banyak klasifikasi tweet yang cenderung netral untuk yang positif dan negatif kurang mendominasi akurasi dari algoritma naive bayes kurang bagus di bandingkan dengan DBN.

2.2. Dasar Teori

2.2.1. *Sentiment analysis*

Sentiment analysis adalah cabang dari penambangan teks yang mengidentifikasi teks dan kemudian mengekstraknya informasi dari teks yang diidentifikasi menjadi informasi subjektif (Soong et al., 2019; Surya Akbar et al., 2021). Selain mengekstraksi informasi, analisis *sentiment* juga berfokus pada pengelola opini yang mengandung polaritas, yaitu nilai *sentiment* positif atau negatif dari sebuah teks (Novantirani et al., 2015; Surya Akbar et al., 2021). Masalah dalam analisis *sentiment* biasanya sulit untuk didefinisikan, termasuk konsep masalah, sub masalah, dan tujuan yang dijadikan sebagai kerangka kinerja dalam berbagai penelitian (Joshi et al., 2017; Surya Akbar et al., 2021).

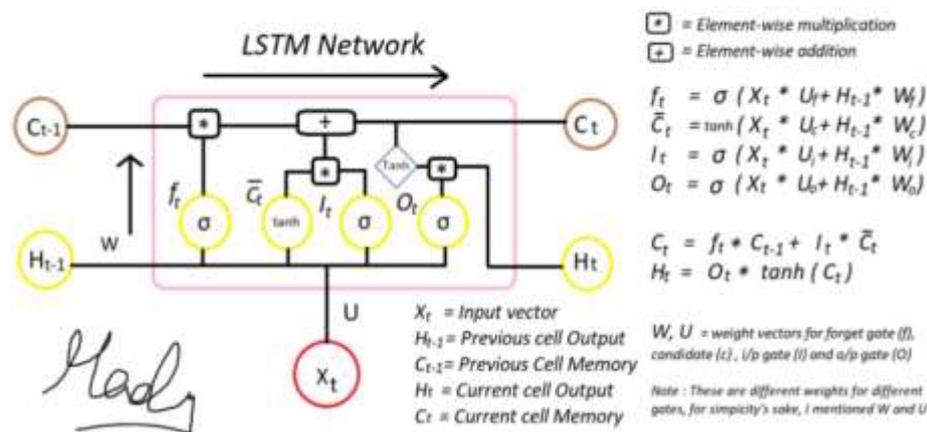
2.2.2. *Recurrent Neural Network (RNN)*

Algoritma RNN adalah jaringan saraf multi-layer yang dapat digunakan untuk menganalisis input sekuensial, seperti teks, ucapan atau video, untuk tujuan klasifikasi dan prediksi. Selain itu, algoritma RNN tidak dibatasi oleh panjang input dan dapat menggunakan konteks temporal untuk memprediksi makna dengan lebih baik (Kowsari et al., 2017; Nugraha et al., 2020). Metode RNN juga memiliki akurasi yang lebih tinggi dalam mengolah hasil dibandingkan dengan metode lainnya (Nugraha et al., 2020; Nurrohmat & SN, 2019).

Recurrent Neural Network (RNN) adalah salah satu metode deep learning dan bagian dari *neural network* yang dapat digunakan untuk menganalisis

sentiment dengan hasil keluaran berdasarkan hasil perhitungan sebelumnya menggunakan informasi sekuensial. RNN merupakan model yang efisien untuk analisis *sentiment* karena menggunakan sel memori yang mampu menangkap informasi sekuensial tanpa dibatasi oleh panjang input (Kurniasari & Setyanto, 2020; Nugraha et al., 2020).

pada penelitian ini akan digunakan RNN-LSTM



Gambar 1 LSTM model (dari Medium)

2.2.3. Logistic Regression

Logistic regression (Regresi Logistik) merupakan analisis regresi yang tepat untuk dilakukan ketika variabel dependen bersifat dikotomis (biner). Seperti semua analisis regresi, regresi logistik adalah analisis prediktif. Regresi logistik digunakan untuk menggambarkan data dan menjelaskan hubungan antara satu variabel biner dependen dan satu atau lebih variabel independen nominal, ordinal, interval atau rasio (O'Connell, 2016).

Regresi Logistik digunakan dalam ilmu biologi pada awal abad kedua puluh. Itu kemudian digunakan dalam banyak aplikasi ilmu sosial. Regresi Logistik digunakan ketika variabel dependen (target) adalah kategoris (Saishruthi, 2018).

Jenis Regresi Logistik

1. Regresi Logistik Biner

Tanggapan kategoris hanya memiliki dua 2 kemungkinan hasil. Contoh: Spam atau Tidak

2. Regresi Logistik Multinomial

Tiga atau lebih kategori tanpa memesan. Contoh: Memprediksi makanan mana yang lebih disukai (Vegetarian, Non-Vegetarian, Vegan)

3. Regresi Logistik Ordinal

Tiga atau lebih kategori dengan pemesanan. Contoh: Peringkat film dari 1 hingga 5.

Pada penelitian ini yang digunakan adalah logistik regression biner dalam algoritma yang dipakai

Berikut rumus untuk logistik regression

$$\log(p) = \log\left(\frac{p(y-1)}{1-(p-1)}\right) = \beta_0 + \beta_{1-x_\alpha} + \beta_{2-x_\alpha} + \dots + \beta_{p-x_\alpha}$$

2.2.4. *Naive Bayes*

Naive Bayes adalah salah satu algoritma yang paling sederhana dan kuat untuk klasifikasi berdasarkan Teorema *Bayes* dengan asumsi independensi di antara prediktor. Model *Naive Bayes* mudah dibuat dan sangat berguna untuk kumpulan data yang sangat besar (Keshari, 2020).

Naive Bayes Classifier merupakan sebuah metoda klasifikasi yang berakar pada teorema *Bayes*. Metode pengklasifikasian dengan menggunakan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas *Bayes*, yaitu memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya sehingga dikenal sebagai Teorema *Bayes*. Ciri utama dari *Naive Bayes Classifier* ini adalah asumsi yang sangat kuat

(naïf) akan independensi dari masing-masing kondisi / kejadian (Mochammad Haldi Widiyanto, 2019).

Keuntungan penggunaan metode ini yaitu hanya membutuhkan jumlah data pelatihan (*training data*) yang kecil untuk menentukan estimasi parameter yang diperlukan dalam proses pengklasifikasian. Karena yang diasumsikan sebagai variabel independent, maka hanya varians dari suatu variabel dalam sebuah kelas yang dibutuhkan untuk menentukan klasifikasi, bukan keseluruhan dari matriks kovarians (Keshari, 2020).

Pada penelitian ini metode yang digunakan pada naive bayes menggunakan model multinomial, Multinomial Naive Bayes merupakan metode pembelajaran probabilistik yang banyak digunakan dalam Natural Language Processing (NLP). Algoritme didasarkan pada teorema Bayes dan memprediksi tag teks seperti sepotong email atau artikel surat kabar. Ini menghitung probabilitas setiap tag untuk sampel yang diberikan dan kemudian memberikan tag dengan probabilitas tertinggi sebagai output (upGrad, 2021).

Multinomial Naïve Bayes mempertimbangkan vektor fitur di mana istilah yang diberikan mewakili berapa kali muncul atau sangat sering yaitu frekuensi (Great Learning, 2020).

Teorema Bayes, dirumuskan oleh Thomas Bayes, menghitung probabilitas suatu peristiwa terjadi berdasarkan pengetahuan sebelumnya tentang kondisi yang terkait dengan suatu peristiwa. Ini didasarkan pada rumus berikut:

$$P(A|B) = P(A) * P(B|A)/P(B)$$

Dimana kita menghitung probabilitas kelas A ketika prediktor B sudah tersedia.

$P(B)$ = probabilitas sebelumnya dari B

$P(A)$ = peluang awal kelas A

$P(B|A)$ = kemunculan prediktor B dengan probabilitas kelas A

Rumus ini membantu dalam menghitung probabilitas tag dalam teks.

2.2.5. *Python*

Bahasa pemrograman *Python* memantapkan dirinya sebagai salah satu bahasa paling populer untuk komputasi ilmiah. Berkat sifatnya yang interaktif tingkat tinggi dan ekosistem perpustakaan ilmiahnya yang matang, ini adalah pilihan yang menarik untuk pengembangan algoritmik dan analisis data eksplorasi (Dubois et al., 2007; Millman & Aivazis, 2011; Pedregosa et al., 2011).

Python adalah bahasa pemrograman general yang dapat digunakan secara efektif untuk membangun hampir semua jenis program yang tidak memerlukan akses langsung ke perangkat keras komputer. *Python* tidak optimal untuk program yang memiliki batasan keandalan yang tinggi (karena pemeriksaan semantik statisnya yang lemah) atau yang dibangun dan dipelihara oleh banyak orang atau dalam jangka waktu yang lama (sekali lagi karena pemeriksaan semantik statis yang lemah) (Guttag, 2013).

Python merupakan bahasa pemrograman tingkat tinggi (high level language) dengan filosofi perancangan yang berfokus pada tingkat keterbacaan kode. *Python* dirancang untuk memberikan kemudahan kepada programmer melalui berbagai segi yaitu kemudahan dalam pengembangan, efisiensi waktu, multifungsi, dan kompatibilitas dengan sistem (Saragih, 2016).

Beberapa fitur dan Kelebihan yang dimiliki *Python* menurut Jubilee Enterprise (Chilmi, 2021; Enterprise, 2017) , yaitu:

1. Memiliki koleksi library yang banyak. Telah tersedia modul siap pakai untuk berbagai keperluan
2. Memiliki struktur bahasa yang jelas, sederhana dan mudah dipelajari,

3. *Python* adalah pemrograman berorientasi objek (OOP). Data dalam *Python* adalah sebuah objek yang dibuat dari kelas (class).

4. Memiliki sistem pengelolaan memori otomatis (Garbage Collection).

5. Bersifat modular, sehingga mudah dikembangkan dengan menciptakan modul baru.

Selain memiliki beberapa kelebihan, *Python* juga mempunyai beberapa library. Sebagian besar library nya dapat di-install dengan menggunakan “PIP”, dimana PIP merupakan Standard Package Manager untuk Python yang dapat digunakan untuk membantu melakukan instalasi dan mengatur package tambahan yang bukan merupakan dari Python Standard Library. Beberapa library Python, yaitu:

1. Search Tweets

Search tweets adalah metode scraping data tweet yang menggunakan api tweet v1 digunakan dalam python dengan menggunakan fitur premium search yang ada pada developer tweet. dimana pada fitur premium search ini berbeda dengan fitur search biasa karena bisa menggunakan full Archive dan search 30 hari.

Pada penelitian ini api yang digunakan adalah full Archive karena data yang di ambil adalah data pada tahun 2021 dari bulan Januari sampai bulan Desember.

Berikut perbedaan antara api 30 hari, full Archive dan recent

Table 2 Perbedaan Api Tweet

Jenis API	Perbedaannya
30 hari	<ol style="list-style-type: none"> 1. Data tweet yang bisa di ambil hanya 30 hari terakhir yang terbaru dari hari pada saat data itu di tarik 2. Batas penarikan data bisa 25k tweet perbulan untuk sandbox akun 3. Wajib elevated akun untuk bisa akses sandbox akun nya
Full archive	<ol style="list-style-type: none"> 1. Data tweet yang di tarik bisa dipilih sesuai tanggal yang di inginkan dengan maksimal 30 hari sesuai tanggal pertama dan terakhir nya 2. Maksimal data tweet yang bisa di ambil 5k perbulan untuk sanbox akun 3. Wajib elevated akun untuk bisa akses sandbox akun nya
Recent	<ol style="list-style-type: none"> 1. Bukan premium akun atau sanbox akun hanya elevated akun biasa 2. Hanya bisa menarik tweet 7 hari terbaru saja 3. Batas penarikan tweet Nya tergantung dari seberapa banyak tweet yang sesuai dari keyword yang di input.

2. Numpy

NumPy array adalah struktur data yang umum digunakan dalam Python untuk menyimpan data sebagai grid atau matriks. Dalam Python, struktur data adalah objek yang mengatur dan memanipulasi data dengan mendefinisikan hubungan antara nilai data yang disimpan dalam struktur data dengan fungsi-fungsi yang dijalankan pada struktur data. Baris NumPy terdiri dari nilai yang berurutan yang disebut elemen dan menggunakan indeks untuk mengatur dan memanipulasi elemen dalam baris NumPy. Ciri-ciri utama dari NumPy array adalah semua elemen dalam array harus memiliki tipe yang sama, misalnya semua bilangan bulat, string, dan lain sebagainya. NumPy array berbeda dengan list, jika list memerlukan packages Python tertentu, NumPy array didefinisikan menggunakan fungsi `array()` pada NumPy Packages.

3. Scikit-learn

Scikit-learn adalah perpustakaan pembelajaran mesin open source yang mendukung pembelajaran yang diawasi dan tidak diawasi. Ini juga menyediakan berbagai alat untuk pemasangan model, pra-pemrosesan data, pemilihan model, evaluasi model, dan banyak utilitas lainnya.

4. Keras

Keras adalah API deep learning yang ditulis dalam bahasa pemrograman Python, dan berjalan diatas platform Machine learning TensorFlow. API ini dikembangkan dengan fokus utama yaitu memungkinkan komputasi yang cepat. Mampu pergi dari ide ke hasil secepat mungkin merupakan tujuan dari API Keras ini. Keras merupakan High – level API sehingga command yang ada pada keras dapat kita baca dengan mudah. Saat ini, API Keras sangatlah populer digunakan baik dalam perusahaan ataupun organisasi bahkan perusahaan seperti NASA, Youtube menggunakan Keras sebagai API deep learning nya.

III. METODOLOGI PENELITIAN

3.1. Tempat dan Waktu Penelitian

3.1.1. Tempat Penelitian

Penelitian ini dilaksanakan di Jurusan Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lampung yang beralamatkan di Jl. Soemantri Brojonegoro No. 1 Gedong Meneng, Bandar Lampung.

3.1.2. Waktu Penelitian

Penelitian ini dilaksanakan mulai bulan Maret 2022 sampai bulan Februari tahun 2023. Adapun alur waktu penelitian dapat dilihat pada Tabel berikut :

Table 3 Alur Penelitian

Tahapan	Kegiatan	2022				2023
		Maret	April	Mei	Juni	Februari
Penelitian Awal	Studi Literatur	√				
	Penentuan Tema	√				
	Pengambilan Data		√			
	Penyusunan Draft (Bab I-III)			√		
Penelitian Lanjutan	Seminar Usul				√	
	Data Labeling dan				√	

	Preprocessing data		
	Analisis Data	√	
	Penyusunan Draft (Bab IV-V)		√
	Seminar Hasil Penelitian		√
	Sidang Komprehensif		√
Evaluasi	Revisi Skripsi		√

3.2. Alat dan Objek Penelitian

3.2.1. Alat Penelitian

3.2.1.1. Perangkat Keras (Hardware)

Perangkat Keras yang digunakan dalam penelitian ini adalah sebuah laptop dengan spesifikasi sebagai berikut :

- System Manufacture : ASUSTeK Computer Inc
- Processor : Intel Core I3-6006LI CPU @ 2.00GHz (4 CPUs), ~2.0GHz
- Installed Ram : 8 GB
- VGA : Intel(R) HD Graphics 520
- GPU : NVIDIA GeForce MX 110 VRAM 2 GB

3.2.1.2. Perangkat Lunak (Software)

Perangkat Lunak yang digunakan dalam penelitian ini antara lain :

- Sistem Operasi : Windows 10 Home Single Language
- Web Browser : Google Chrome
- Python 3.7.10
- Library Keras, Scikit Learn, Search tweets,

3.2.2. *Objek Penelitian*

Objek Penelitian yang digunakan yaitu data tweet yang berhubungan tentang kebijakan ppkm yang berlaku di Indonesia.

3.3. Sumber Data dan Metode Pengumpulan data

3.3.1. *Sumber Data*

Sumber data yang digunakan dalam penelitian ini yaitu berupa data tweet yang diambil dari database tweet dengan keyword yang sudah ditetapkan yaitu tentang kebijakan ppkm yang berlaku di Indonesia keyword pencarian pada tweet api menggunakan hastag ppkm.

3.3.2. *Metode Pengumpulan data*

Metode pengumpulan data yang digunakan dalam penelitian ini adalah dalam bentuk *scraping* data tweet yang menggunakan api tweet yang tersedia dari developer tweet yang digunakan khusus eksperimen data tweet, data yang di ambil adalah data tweet tahun 2021.

3.4. Metode Penelitian

Metode penelitian menggunakan eksperimen dengan cara menganalisis *sentiment* menggunakan beberapa algoritma kemudian membandingkan akurasi, presisi, recall dan f1-score dan hasil analisis sentimen positif atau negatif dari masing-masing algoritma yang digunakan.

3.5. Tahapan Penelitian

Tahapan penelitian merupakan teknik atau langkah-langkah yang akan digunakan dalam penelitian. Adapun penjelasan dari masing-masing tahapan yaitu :

1. Identifikasi dan Perumusan Masalah

Tahap ini dilakukan untuk mengidentifikasi algoritma apa yang akan dipakai dalam analisis yang dilakukan dan merumuskan masalah sesuai data dan algoritma yang digunakan.

2. Pengumpulan Data

Tahapan pengumpulan data pada penelitian ini dengan melakukan scraping data tweet yang menggunakan kata kunci ppcm sehingga data yang di ambil sesuai dengan penelitian yang akan dilakukan yaitu tentang kebijakan ppcm yang ada di indonesia.

Data mentah yang di ambil langsung dari tweet sebelum di bersihkan sekitar 20 ribu data kemudian data setelah di filter dari duplikat data sekitar 14000 data untuk keseimbangan data nya kebanyakan data tweet yang diambil tidak mengandung opini dari banyak nya data yang diambil data yang mengandung opini lebih sedikit dibandingkan data yang mengandung opini

3. Data Labeling

Pada tahapan ini data tweet yang akan di analisis terbagi menjadi 2 kelompok data yaitu data tweet yang berisi tweet positif dan negatif karena data yang sudah di ambil dari tweet belum ada label datanya sehingga pada tahapan ini dilakukan labeling data tweet dengan memperhatikan kalimat pada tweet tersebut apakah mengandung opini positif atau negatif atau tidak mengandung opini sama sekali kemudian dipisahkan mana tweet yang berisikan muatan positif dan negatif

sehingga menghasilkan dataset yang bisa digunakan, target labeling data tweet ini adalah 3000 data tweet dari data mentah yang sudah di ambil tetapi apabila tidak bisa mencapai angka tersebut maka data labeling yang akan digunakan sesuai hasil dari labeling data yang ada.

4. Data Processing

Pada tahapan ini Diperlukan preprocessing data untuk mendapatkan data yang bersih dan meningkatkan nilai klasifikasi tweet. Langkah-langkah yang dilakukan saat preprocessing adalah normalisasi teks, tokenisasi teks, dan penghapusan stopword. Fungsi dari tahap preprocessing adalah sebagai berikut:

1. Normalisasi teks

Tujuan dari normalisasi teks adalah untuk memperbaiki struktur kata dan kosa kata dalam kalimat. Pada penelitian ini akan dilakukan normalisasi terhadap semua data yang terkumpul. Proses normalisasi teks meliputi penghapusan URL, tanda baca, karakter khusus, dan simbol berupa huruf alfabet sehingga tidak terjadi kerancuan pada data teks.

2. Tokenisasi teks

Tokenisasi teks adalah proses pemotongan string atau input kalimat berdasarkan setiap kata yang membentuk kalimat dalam data teks. Tokenisasi secara luas memecah sekumpulan karakter dalam teks menjadi unit kata, membedakan karakter tertentu yang dapat diperlakukan sebagai pemisah kata.

3. Penghapusan Stopword

Stopwords adalah kumpulan kata yang sering muncul tetapi tidak mengubah kalimat artinya jika dihapus, atau dengan kata lain menghilangkan kata yang tidak bermakna atau tidak relevan dalam kalimat.

5. Data *Splitting* (Pemisahan data)

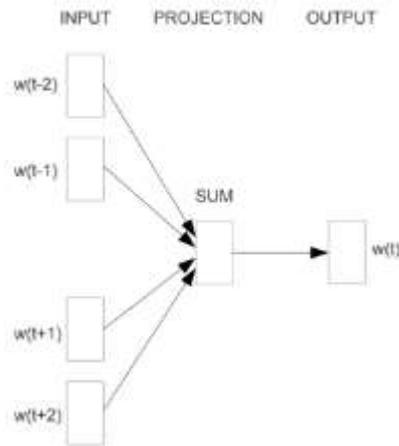
Setelah dilakukan preprocessing data maka akan dilakukan pemisahan data yang dimana pemisahan data akan dibagi menjadi 2 kelompok yaitu 80% data untuk data *Training* dan 20% data untuk data Test, pemisahan data ini dapat dilakukan apabila dataset yang digunakan cukup banyak untuk bisa dipisahkan menjadi dua bagian tetapi apabila dataset yang digunakan tidak cukup maka akan digunakan cara yang lain seperti Cross validasi, pemisahan data ini dilakukan karena dibutuhkan teknik sampling dataset untuk melatih model dari algoritma yang akan dipakai dan untuk mendapatkan hasil yang maksimal dan pada metode ini digunakan random sampling agar dari semua dataset yang ada diambil secara acak dan menyeluruh.

6. Data *Training*

Pada tahapan ini data yang sudah dipisahkan akan di *training* terlebih dahulu dengan word2vec. Kemudian akan dilanjutkan dengan klasifikasi data menggunakan algoritma yang sudah ditentukan yaitu RNN-LSTM, Multinomial Navie bayes, dan *Logistic regression*.

Word2Vec

Word2vec akan melatih dataset dan mendapatkan nilai vektor dari kata yang dilatih. Itu akan juga memperoleh jumlah total kata dan kosa kata yang muncul dari proses pelatihan dataset. *Model CBOW (Continuous Bag of Words)* digunakan untuk pelatihan model word2vec karena model ini menggunakan representasi konteks yang terdistribusi secara kontinu [10]. Untuk arsitektur model CBOW ditunjukkan pada Gambar 2.



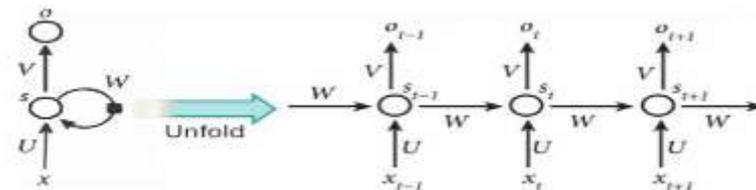
Gambar 2 CBOW Model Sumber : (Mikolov et al., 2013)

7. Sentiment clasification

Pada tahapan ini dilakukan pengujian prediksi pada data set yang sudah di latih menggunakan algoritma RNN, logistik regresi, dan naive bayes kemudian hasil dari pengujian dari algoritma akan ditampilkan dalam bentuk *confusion matrix* mana yang true negatif, false negatif, true positif, dan false positif agar terlihat lebih jelas hasil nya berikut penjelasan dari masing-masing algoritma dan *confusion matrix*.

1. RNN-LSTM

RNN adalah model yang efisien untuk analisis sentimen. RNN menggunakan sel memori yang mampu menangkap informasi tentang urutan panjang, terlihat pada gambar berikut

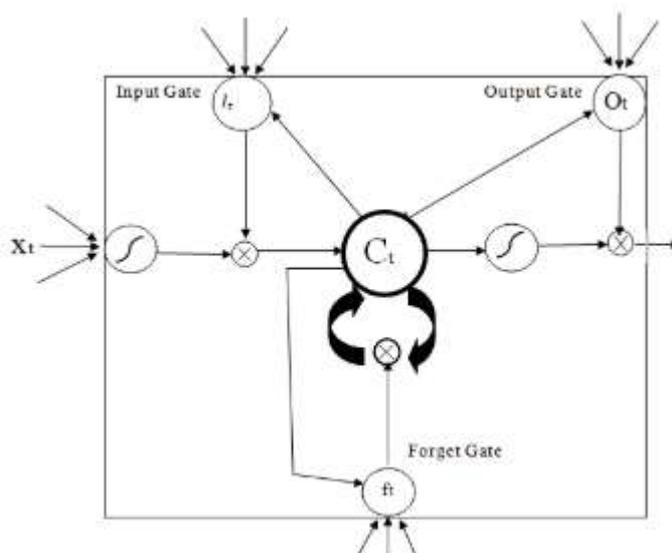


Gambar 3 RNN Framework Sumber: Denny Britz, 2015

Persamaan yang diberikan menunjukkan rumus dasar untuk RNN: $a_t = f(h_{t-1}, x_t)$ Di mana a_t mewakili output dari node sebelumnya, fungsi aktivasi f adalah fungsi tanh dan x_t menunjukkan urutan input

$(x_0, x_1, x_2, \dots, x_t)$. Jaringan saraf rekursif mewakili struktur pohon dalam yang mampu menangkap semantik teks. Proses ini memakan waktu tugas yang merupakan kelemahan dari jaringan saraf rekursif (Kurniasari & Setyanto, 2020; Socher et al., 2011). Jaringan saraf berulang telah meningkatkan kompleksitas waktu. Pada model RNN, analisis dilakukan kata per kata dan membutuhkan waktu yang lama (Kurniasari & Setyanto, 2020; Socher et al., 2011). RNN mampu mempelajari kata-kata terbaru dibandingkan dengan kata-kata sebelumnya yang mengarah pada masalah gradien yang hilang. RNN menggunakan LSTM untuk memecahkan masalah gradien hilang.

LSTM (Long Short Term Memory) adalah bagian dari RNN yang digunakan untuk mempelajari dependensi jarak jauh untuk urutan teks. LSTM berisi blok memori yang juga dikenal sebagai gerbang untuk mengontrol aliran teks. Blok memori berisi tiga gerbang yang disebut sebagai gerbang input; forget gerbang dan gerbang keluaran untuk mengontrol aliran informasi (Kurniasari & Setyanto, 2020; Sak et al., 2014).



Gambar 4 Long short term memory sumber : (Alwehaibi & Roy, 2019)

Pada gambar arsitektur RNN-LSTM di bawah layer pertama dengan jenis layer embedding menggunakan 300 neuron dimana untuk neuron ini bisa di ubah sesuai

angka yang dibutuhkan tidak terpaut pada 300 kemudian 300 itu adalah 300 baris pertama dari data yang digunakan dalam model dan input shape yang berisi 3 dimensi dalam data yang akan dilatih, setelah itu ditambahkan dropout agar menghindari terjadinya overfitting dalam proses testing dan Training, kemudian data masuk ke layer LSTM dengan parameter pengulangan dropout sama seperti sebelumnya hingga ke tahap output nya menggunakan 3 neuron sesuai parameter data yang digunakan dengan menggunakan fungsi aktivasi softmax.

Model: "sequential_3"

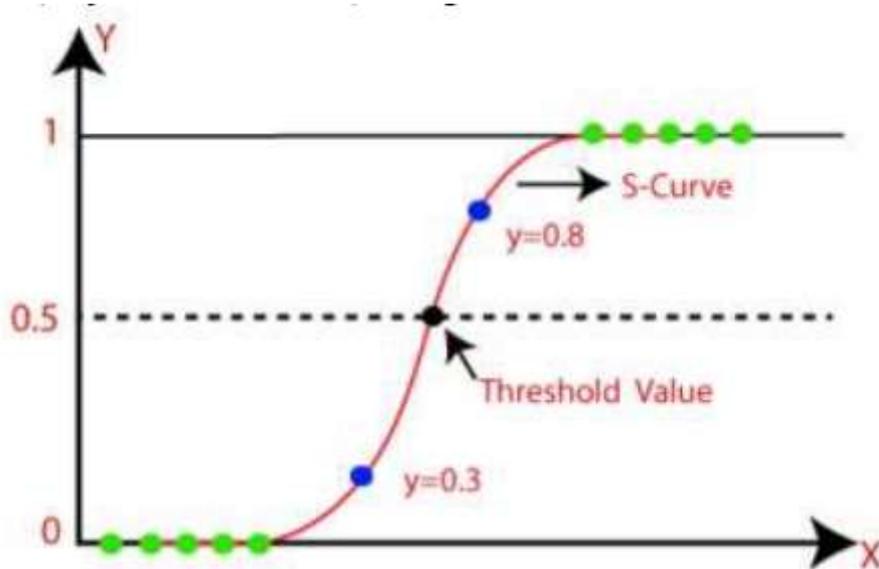
Layer (type)	Output Shape	Param #
embedding_3 (Embedding)	(None, 300, 300)	2099100
dropout_3 (Dropout)	(None, 300, 300)	0
lstm_5 (LSTM)	(None, 300, 128)	219648
lstm_6 (LSTM)	(None, 128)	131584
dense_5 (Dense)	(None, 16)	2064
dense_6 (Dense)	(None, 2)	34
Total params: 2,452,430		
Trainable params: 353,330		
Non-trainable params: 2,099,100		

Gambar 5 Arsitektur RNN-LSTM

2. Logistic Regression

Regresi Logistik (LR) adalah pendekatan klasifikasi lain yang menonjol. LR adalah teknik pembelajaran mesin yang dapat diterapkan untuk masalah klasifikasi. Pierre Francais Verhulst mendefinisikan fungsi logistik dan atributnya dalam makalah yang diterbitkan dalam Prosiding Akademi Kerajaan Belgia dengan menetapkan tiga parameter dan kurva yang mengalir melaluinya. Ini adalah metode pembelajaran mesin yang relatif sederhana yang banyak digunakan. Pendekatan statistik untuk memprediksi kelas biner adalah regresi logistik. Dalam hal ini, variabel terikat memiliki distribusi Bernoulli. Fungsi sigmoid, juga dikenal sebagai fungsi

logistik, adalah kurva berbentuk 'S' yang mengambil nilai antara 0 dan 1. Jika kurva bergerak ke tak terhingga positif, 1 akan diantisipasi, dan jika menuju tak terhingga negatif, 0 akan menjadi diprediksi (Arista, 2022; Majumder et al., 2021). Berikut gambar kurva yang di bentuk dari algoritma logistik regression



Gambar 6 *Logistic regression* (Arista, 2022; Majumder et al., 2021)

3. Naive Bayes

Model Naive Bayes yang akan digunakan pada penelitian ini adalah multinomial naive bayes karena dengan model ini dapat mengurangi terjadi nya error pada saat pemrosesan uji data latihan dan prediksi Multinomial Naïve Bayes juga dapat menangani ukuran kosakata yang besar dan mengurangi tingkat kesalahan yang tinggi, keuntungan lain dari menggunakan algoritma naive bayes adalah implementasi algoritma yang mudah dan performa algoritma nya sangat baik dalam melakukan pemrosesan data untuk di klasifikasi nya (Farisi et al., 2019; Qomariah, 2021).

8. Menggunakan metode under sampling dan over sampling

Tahap ini dilakukan karena data tweet yang didapatkan tidak seimbang atau (imbalance data) maka digunakan teknik sampling under dan over

untuk membandingkan hasil analisis yang dilakukan antara data manual dengan data yang menggunakan teknik sampling tersebut.

Teknik undersampling merupakan teknik dimana memotong data yang lebih dominan untuk disesuaikan dengan data yang lebih sedikit dalam hal ini data yang lebih dominan adalah tweet yang bernilai positif maka data tweet yang bernilai positif dikurangi hingga berjumlah sama dengan data tweet yang negatif baru kemudian dilatih menggunakan algoritma yang dipakai agar mendapatkan hasilnya.

Teknik oversampling merupakan teknik dimana data yang tidak dominan dilipat gandakan agar setara dengan data yang lebih dominan dalam hal ini data tweet yang bernilai negatif sangat sedikit dari data tweet yang bernilai positif maka data tweet yang bernilai negatif digandakan agar sama dengan data tweet yang positif kemudian dilatih menggunakan algoritma yang dipakai agar mendapatkan hasilnya.

9. Evaluasi dan hasil

Pada tahap ini dilakukan perhitungan akurasi, presisi, recall, dan f1-score dari masing-masing algoritma yang sudah di ujikan kemudian dilakukan perbandingan antara masing-masing algoritma untuk evaluasi kedepannya berikut rumus untuk akurasi, presisi, recall, dan f1-score dari Table *confusion matrix*.

1. Confusion Matrix

Hasil dari setiap algoritma klasifikasi dievaluasi menggunakan matriks konfusi. Nilai akurasi, presisi, recall, dan f1-score diturunkan dari *confusion matrix*. Berikut adalah Table dari *confusion matrix*

Table 4 Confusion Matrix

True Label	Prediction Label	
	Negative	Positive
Negative	TN (True Negative)	FP (False Positive)
Positive	FN (False Negative)	TP (True Positive)

Confusion matrix dalam masalah klasifikasi dengan dua kelas ditunjukkan pada Tabel 1. Dapat dapat disimpulkan dari gambar bahwa ada kemungkinan empat hasil prediksi yang berbeda. Hasil yang benar-benar positif dan benar-benar negatif adalah klasifikasi yang benar, sedangkan hasil positif palsu dan negatif palsu adalah dua kemungkinan jenis kesalahan (Dj Novakovi et al., 2017; Surya Akbar et al., 2021). Nilai True Negative (TN) adalah banyaknya data negatif yang terdeteksi dengan benar, sedangkan False Negative (FN) adalah data positif yang terdeteksi negatif. Nilai True Positive (TP) adalah data positif yang terdeteksi dengan benar, sedangkan False Positive (FP) adalah data negatif tetapi terdeteksi positif. Tabel matriks konfusi digunakan untuk mengukur kinerja suatu metode klasifikasi dengan menghitung nilai akurasi, presisi, recall, dan f1-score.

2. Akurasi adalah rasio prediksi yang benar (positif dan negatif) terhadap keseluruhan data.

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100\%$$

3. Presisi adalah rasio prediksi positif sejati dengan prediksi positif keseluruhan.

$$precision = \frac{TP}{TP + FP} \times 100\%$$

4. Recall adalah rasio prediksi positif benar terhadap keseluruhan data positif benar.

$$recall = \frac{TP}{TP + FN} \times 100\%$$

5. F1-Score adalah perbandingan rata-rata tertimbang dari presisi dan daya ingat.

$$f1\ score = 2 \times \frac{(recall \times precision)}{(recall + precision)}$$

V. SIMPULAN DAN SARAN

5.1. Simpulan

Dari penelitian sentimen analisis yang sudah dilakukan dapat disimpulkan sebagai berikut :

1. Data yang didapatkan dari hasil scraping data yang dilakukan lebih banyak data yang bersifat positif daripada data yang bersifat negatif sehingga dataset yang dibuat termasuk ke dalam imbalance dataset.
2. Dari ketiga algoritma yang sudah diuji coba urutan algoritma yang mendapatkan hasil yang sangat baik sampai dengan yang cukup baik adalah RNN-LSTM, Logistic Regression, dan Naive Bayes.
3. Penggunaan teknik oversampling dan undersampling sangat mempengaruhi hasil analisis dari tiap algoritma yang digunakan terutama pada algoritma naive bayes dan logistic regression dan termasuk juga RNN-LSTM.

5.2. Saran

Dari penelitian sentimen analisis yang sudah dilakukan terdapat beberapa saran yang dapat diambil sebagai berikut :

1. Bisa menggunakan algoritma lain dengan menggunakan dataset yang sudah dibuat sebagai perbandingan hasil analisis.
2. Menggunakan dataset dari penelitian lain untuk menguji hasil dari ketiga algoritma yang sudah digunakan.

DAFTAR PUSTAKA

- Alwehaibi, A., & Roy, K. (2019). Comparison of Pre-Trained Word Vectors for Arabic Text Classification Using Deep Learning Approach. *Proceedings - 17th IEEE International Conference on Machine Learning and Applications, ICMLA 2018*, 1471–1474. <https://doi.org/10.1109/ICMLA.2018.00239>
- Arista, A. (2022). Comparison Decision Tree and Logistic Regression Machine Learning Classification Algorithms to determine Covid-19. *Sinkron*, 7(1), 59–65. <https://doi.org/10.33395/sinkron.v7i1.11243>
- Chilmi, M. L. C. (2021). Latent dirichlet allocation lda untuk mengetahui topik pembicaraan warganet twitter tentang omnibus law. *Repository.Uinjkt.Ac.Id*. [https://repository.uinjkt.ac.id/dspace/handle/123456789/56724%0Ahttps://repository.uinjkt.ac.id/dspace/bitstream/123456789/56724/1/M. LUVIAN CHISNI CHILMI-FST.pdf](https://repository.uinjkt.ac.id/dspace/handle/123456789/56724%0Ahttps://repository.uinjkt.ac.id/dspace/bitstream/123456789/56724/1/M.LUVIAN%20CHISNI%20CHILMI-FST.pdf)
- Dj Novakovi, J., Veljovi, A., Ili, S. S., Zeljko Papi, ˇ, & Tomovi, M. (2017). Evaluation of Classification Models in Machine Learning. *Theory and Applications of Mathematics & Computer Science*, 7(1), 39–46.
- Dubois, P. F., Oliphant, T. E., Pérez, F., Granger, B. E., & Greenfield, P. (2007). *PYTHON : Guest Editor ' s Introduction Python for Scientific Computing IPython : A System for Interactive Scientific Computing Reaching for the Stars with Python* (Issue June).
- Enterprise, J. (2017). *Otodidak Pemrograman Python*. Elex Media Komputindo. <https://books.google.co.id/books?id=K-M8DwAAQBAJ>
- Farisi, A. A., Sibaroni, Y., & Faraby, S. Al. (2019). Sentiment analysis on hotel reviews using Multinomial Naïve Bayes classifier. *Journal of Physics: Conference Series*, 1192(1). <https://doi.org/10.1088/1742->

6596/1192/1/012024

- Great Learning. (2020). *Multinomial Naive Bayes Classifier Algorithm*. Great Learning. <https://www.mygreatlearning.com/blog/multinomial-naive-bayes-explained/>
- Guttag, J. V. (2013). *Introduction to Computation and Programming Using Python*.
- 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>
- Joshi, A., Bhattacharyya, P., & Ahire, S. (2017). *Sentiment Resources: Lexicons and Datasets*. https://doi.org/10.1007/978-3-319-55394-8_5
- Keshari, K. (2020). *Naive Bayes Tutorial | Naive Bayes Classifier in Python | Edureka*. <https://www.edureka.co/blog/naive-bayes-tutorial/>
- Kowsari, K., Brown, D. E., Heidarysafa, M., Jafari Meimandi, K., Gerber, M. S., & Barnes, L. E. (2017). HDLTex: Hierarchical Deep Learning for Text Classification. *Proceedings - 16th IEEE International Conference on Machine Learning and Applications, ICMLA 2017, 2017-Decem*, 364–371. <https://doi.org/10.1109/ICMLA.2017.0-134>
- Kurniasari, L., & Setyanto, A. (2020). Sentiment Analysis using Recurrent Neural Network. *Journal of Physics: Conference Series*, 1471(1), 1–4. <https://doi.org/10.1088/1742-6596/1471/1/012018>
- Majumder, A. B., Gupta, S., Singh, D., & Majumder, S. (2021). An intelligent system for prediction of COVID-19 case using machine learning framework- logistic regression. *Journal of Physics: Conference Series*, 1797(1), 0–9. <https://doi.org/10.1088/1742-6596/1797/1/012011>
- Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *1st International Conference on Learning Representations, ICLR 2013 - Workshop Track Proceedings*, 1–12.

- Millman, K. J., & Aivazis, M. (2011). Python for scientists and engineers. *Computing in Science and Engineering*, 13(2), 9–12.
<https://doi.org/10.1109/MCSE.2011.36>
- Mochammad Haldi Widiyanto. (2019). *Algoritma Naive Bayes / BINUS UNIVERSITY BANDUNG - Kampus Teknologi Kreatif*. Binus.Ac.Id.
<https://binus.ac.id/bandung/2019/12/algoritma-naive-bayes/>
- 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), 1–7.
- Nugraha, F. A., Harani, N. H., Habibi, R., & Fatonah, R. N. S. (2020). Sentiment Analysis on Social Distancing and Physical Distancing on Twitter Social Media using Recurrent Neural Network (RNN) Algorithm. *Jurnal Online Informatika*, 5(2), 195. <https://doi.org/10.15575/join.v5i2.632>
- Nurrohmat, M. A., & SN, A. (2019). Sentiment Analysis of Novel Review Using Long Short-Term Memory Method. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, 13(3), 209.
<https://doi.org/10.22146/ijccs.41236>
- O'Connell, S. S. (2016). *What is Logistic Regression? - Statistics Solutions*. Solutions Statistics. <https://www.statisticssolutions.com/free-resources/directory-of-statistical-analyses/what-is-logistic-regression/>
- Pedregosa, F., Michel, V., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Vanderplas, J., Cournapeau, D., Pedregosa, F., Varoquaux, G., Gramfort, A., Thirion, B., Grisel, O., Dubourg, V., Passos, A., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.
<https://jmlr.csail.mit.edu/papers/v12/pedregosa11a.html>
- Qomariah, N. (2021). Sentiment Analysis on Coffee Consumer Perceptions on Social Media Twitter Using Multinomial Naïve Bayes. *Journal of Intelligent Computing and Health ...*, 2(1).

<https://jurnal.unimus.ac.id/index.php/ICHI/article/view/7241>

- Saishruthi. (2018). Logistic Regression — Detailed Overview | by Saishruthi Swaminathan | Towards Data Science. In *TowardsDataScience*.
<https://towardsdatascience.com/logistic-regression-detailed-overview-46c4da4303bc>
- Sak, H., Senior, A., & Beaufays, F. (2014). *Long Short-Term Memory Based Recurrent Neural Network Architectures for Large Vocabulary Speech Recognition*. Cd. <http://arxiv.org/abs/1402.1128>
- Saragih, R. R. (2016). Pemrograman dan bahasa Pemrograman. *STMIK-STIE Mikroskil, December*, 1–91.
- Septiani, L., & Sibaroni, Y. (2019). Sentiment Analysis Terhadap Tweet Bernada Sarkasme Berbahasa Indonesia. *Jurnal Linguistik Komputasional*, 2(2), 62–67. <https://doi.org/10.26418/jlk.v2i2.23>
- Socher, R., Lin, C. C. Y., Ng, A. Y., & Manning, C. D. (2011). Parsing natural scenes and natural language with recursive neural networks. *Proceedings of the 28th International Conference on Machine Learning, ICML 2011*, 129–136.
- Soong, H., Jalil, N. B. A., Ayyasamy, R. K., & Akbar, R. (2019). The Essential of Sentiment Analysis and Opinion Mining in Social Media. *2019 IEEE 9th Symposium on Computer Applications & Industrial Electronics (ISCAIE)*, 272–277.
- Surya Akbar, M., Shofiana, D. A., Syarif, A., & Muludi, K. (2021). Sentiment Analysis Of Energy Independence Tweets Using Simple Recurrent Neural Network. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, 15(4), 339–348. <https://doi.org/10.22146/ijccs.66016>
- upGrad. (2021). *Multinomial Naive Bayes Explained: Function, Advantages & Disadvantages, Applications in 2021*. Unmetric.
<https://www.upgrad.com/blog/multinomial-naive-bayes-explained/>
- Zulfa, I., & Winarko, E. (2017). Sentimen Analisis Tweet Berbahasa Indonesia

Dengan Deep Belief Network. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, 11(2), 187. <https://doi.org/10.22146/ijccs.24716>