

***SENTIMENT ANALYSIS PADA TWEET MASYARAKAT MENGENAI  
ROKOK ELEKTRIK MENGGUNAKAN RECURRENT NEURAL  
NETWORK DAN SUPPORT VECTOR MACHINE***

**Skripsi**

**Oleh**

**BAYU ADHI PRABOWO**



**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS LAMPUNG  
BANDAR LAMPUNG  
2023**

***SENTIMENT ANALYSIS PADA TWEET MASYARAKAT MENGENAI  
ROKOK ELEKTRIK MENGGUNAKAN RECURRENT NEURAL  
NETWORK DAN SUPPORT VECTOR MACHINE***

**Abstrak**

**Oleh**

**Bayu Adhi Prabowo**

Penerapan teknik analisis sentimen pada tweet orang tentang rokok elektrik menggunakan *Recurrent Neural Network* (RNN) dan *Support Vector Machine* (SVM). Kedua teknik tersebut digunakan untuk mengidentifikasi sentimen positif dan negatif dalam permintaan tweet terkait rokok elektrik. Analisis sentimen adalah teknik analisis data yang dapat digunakan untuk mengidentifikasi dan mengukur apakah suatu kalimat atau paragraf mengandung emosi positif dan negatif. *Twitter* adalah platform media sosial yang memungkinkan pengguna untuk menulis dan berbagi pesan singkat yang disebut *tweet* yang dibatasi hingga 280 karakter. Pengguna juga dapat mengikuti orang lain, melihat dan menyukai tweet, dan mengirim pesan langsung. Rokok elektronik merupakan alternatif dari rokok tradisional yang menggunakan baterai untuk memanaskan cairan menjadi uap. Hasil penelitian diperoleh dari hasil klasifikasi *Recurrent Neural Network* (RNN) dengan pelabelan manual menunjukkan klasifikasi distribusi 90% pelatihan data dan 10% pengujian data menghasilkan akurasi sebesar 72% dan *Support Vector Machine* (SVM) menghasilkan akurasi sebesar 97%. Hasil penelitian ini diharapkan dapat memberikan informasi yang bermanfaat bagi pihak-pihak yang berkepentingan untuk mengetahui pendapat masyarakat tentang rokok elektrik.

**Kata Kunci :** Rokok Elektrik, *Reccurent Neural Network*, Analisis Sentiment, *Support Vector Machine*, Twitter.

# SENTIMENT ANALYSIS OF PUBLIC TWEETS REGARDING ELECTRIC CIGARETTES USING RECURRENT NEURAL NETWORK AND SUPPORT VECTOR MACHINE

Abstract

By

**Bayu Adhi Prabowo**

*Application of sentiment analysis techniques to people's tweets about e-cigarettes using Recurrent Neural Network (RNN) and Support Vector Machine (SVM). Both techniques are used to identify positive and negative sentiments in tweet requests related to e-cigarettes. Sentiment analysis is a data analysis technique that can be used to identify and measure whether a sentence or paragraph contains positive or negative emotions. Twitter is a social media platform that enables users to write and share short messages called tweets which are limited to 280 characters. Users can also follow others, view and like tweets, and send direct messages. Electronic cigarettes are an alternative to traditional cigarettes that use batteries to heat liquid into vapor. The results obtained from the results of classification of Recurrent Neural Network (RNN) with manual labeling show that the distribution classification of 90% training data and 10% testing of data produces an accuracy of 72% and the Support Vector Machine (SVM) produces an accuracy of 97%. The results of this study are expected to provide useful information for interested parties to find out what the public thinks about e-cigarettes.*

**Keywords:** *E-Cigarette, Recurrent Neural Network, Sentiment Analysis, Support Vector Machine, Twitter.*

***SENTIMENT ANALYSIS PADA TWEET MASYARAKAT MENGENAI  
ROKOK ELEKTRIK MENGGUNAKAN RECURRENT NEURAL  
NETWORK DAN SUPPORT VECTOR MACHINE***

Oleh

**Bayu Adhi Prabowo**

**Skripsi**

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

**Pada**

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



**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS LAMPUNG  
BANDAR LAMPUNG  
2023**

Judul Skripsi : *Sentiment Analysis* pada *Tweet* Masyarakat Mengenai  
Rokok Eleketrik Menggunakan *Recurrent Neural  
Network* dan *Support Vector Machine*

Nama : Bayu Adhi Prabowo

Nomor Pokok Mahasiswa : 1657051011

Program Studi : Ilmu Komputer

Fakultas : Matematika dan Ilmu Pengetahuan Alam



1. Pembimbing Utama,

**Dr. Ir. Kurnia Muludi, M.S.Sc.**  
NIP. 19791031 200604 2 002

2. Mengetahui,  
Ketua Jurusan Ilmu Komputer,

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

MENGESAHKAN

1. Tim Penguji

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

Penguji I : **Prof. Admi Syarif**

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

2. Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam



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

Tanggal Lulus Ujian Skripsi : **10 Maret 2023**

## PERNYATAAN

Saya yang bertanda tangan di bawah ini, menyatakan bahwa skripsi saya yang berjudul **“Sentiment Analysis Pada Tweet Masyarakat Mengenai Rokok Elektrik Menggunakan Recurrent Neural Network Dan Support Vector Machine”** ini merupakan hasil karya sendiri dan bukan hasil karya orang lain. Semua tulisan yang tertuang dalam skripsi ini telah mengikuti kaidah penulisan karya ilmiah Universitas Lampung. Apabila di kemudian hari terbukti bahwa skripsi ini merupakan hasil salinan atau dibuat oleh orang lain, maka saya bersedia menerima sanksi berupa pencabutan gelar akademik yang telah saya terima.

Bandar Lampung, 10 Maret 2023



Bayu Adhi Prabowo  
1657051011

## **RIWAYAT HIDUP**

Penulis dilahirkan pada 23 November 1997 di Kota Bandar Lampung, merupakan putra kedua dari pasangan Bapak Bambang Sigit Pramono dan Ibu Sri Nur Asih. Penulis menyelesaikan pendidikan formal pertamanya di Taman Kanak-Kanak (TK) Al – Kautsar Bandar Lampung pada tahun 2004, melanjutkan sekolah dasar di SD Swasta Al – Kautsar Bandar Lampung dan selesai pada tahun 2010. Kemudian, penulis melanjutkan pendidikan sekolah menengah pertama di SMP Swasta Al – Kautsar Bandar Lampung yang diselesaikan pada tahun 2013, serta menyelesaikan pendidikan sekolah menengah atas pada tahun 2016 di SMA Al – Kautsar Bandar Lampung dengan Ilmu Pengetahuan Alam.

Pada tahun 2016, penulis terdaftar sebagai mahasiswa Jurusan Ilmu Komputer Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung. Selama menjadi mahasiswa, penulis mengikuti beberapa kegiatan antara lain:

1. Anggota Abacus Himpunan Mahasiswa Jurusan Ilmu Komputer periode 2016/2017.
2. Anggota bidang Internal Himpunan Mahasiswa Jurusan Ilmu Komputer (HIMAKOM) periode 2017.
3. Pada bulan Januari 2017 penulis melaksanakan karya wisata ilmiah di Desa Margosari, Kecamatan Pagelaran Utara, Kabupaten Pringsewu
4. Pada bulan Desember 2018 penulis melaksanakan kerja praktik di Bank BTN Bandar Lampung.
5. Pada bulan Juli 2019 penulis melaksanakan KKN di Desa Kedaton, Kecamatan Abung Tinggi, Kabupaten Lampung Utara.



## **MOTTO DAN PERSEMBAHAN**

“Alasan nomor satu kenapa banyak orang gagal dalam hidup adalah karena mereka terlalu mendengarkan teman, keluarga, dan tetangga”

- Napoleon Hill-

Skripsi ini saya persembahkan kepada:

Allah SWT yang telah memberikan rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan skripsi ini

Terima kasih kepada kedua orang tua, Bapak dan Ibu, yang selalu memberikan kasih sayang, dukungan, semangat, dan doa.

Terima kasih kepada Bapak dan Ibu dosen di Jurusan Ilmu Komputer yang senantiasa membagikan ilmu dan memberikan nasihat yang memotivasi.

Terima kasih kepada teman-teman Ilmu Komputer 2016 yang juga selalu mendukung dan berjuang bersama dalam meraih cita-cita.

## SANWANCANA

Puji syukur kehadirat Allah Subhanahu wa ta'ala yang telah melimpahkan rahmat, hidayah serta inayah-Nya sehingga penulis dapat menyelesaikan skripsi ini berjudul "Sentiment Analysis Pada Tweet Masyarakat Mengenai Rokok Elektrik Menggunakan Recurrent Neural Network Dan Support Vector Machine". Tidak lupa shalawat serta salam kepada Nabi Muhammad Shallahu 'alaihi wasallam, yang kita nanti-nantikan syafaatnya di yaumul akhir kelak.

Penulis mengucapkan terimakasih kepada semua pihak yang telah membantu dan memiliki peran besar dalam penyusunan skripsi ini, yaitu:

1. Kedua orang tuaku tercinta, bapakku Bambang Sigit Pramono dan ibuku Sri Nur Asih yang selalu memberikan semua dukungan, kasih sayang dan segala doa yang tiada hentinya.
2. Kakakku Galuh Dwiki Pradana yang selalu mendukung dan membantu adikmu ini dalam menyelesaikan skripsi.
3. Bapak Dr. Ir. Kurnia Muludi, M.S.Sc., selaku dosen pembimbing utama atas kesediaannya dan kesabarannya untuk memberikan dukungan, bimbingan, kritik, dan saran dalam proses penyelesaian skripsi.
4. Bapak Prof. Admi Syarif, Ph. D., selaku dosen penguji satu atas kesediaannya telah memberikan saran dan masukan guna penyempurnaan penulisan skripsi.
5. Bapak Dr. rer. nat. Akmal Junaidi, M.Sc., selaku dosen penguji kedua skripsi sekaligus Sekertaris Jurusan Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lampung, yang telah memberikan saran dan masukan guna penyempurnaan penulisan skripsi.
6. Bapak Didik Kurniawan, S.Si., M.T., selaku Ketua Jurusan Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lampung.

7. Bapak Dekan Dr. Eng. Suropto Dwi Yuwono, S.Si., M.T., selaku Dekan FMIPA Universitas Lampung.
8. Bapak dan Ibu Dosen Jurusan Ilmu Komputer Universitas Lampung yang telah memberikan ilmu dan pengetahuan hidup selama penulis menjadi mahasiswa.
9. Elva Juasta, Silfia Fitriyana, Friska Daesy, Moh Surya Akbar, Arief Achmadi Yusra, Dwi Amalia yang telah memberkan motivasi serta semangat dalam menyusun skripsi ini.
10. Semua pihak yang secara langsung maupun tidak langsung yang telah membantu menyelesaikan skripsi ini.

Penulis menyadari bahwa skripsi ini masih jauh dari kata sempurna, akan tetapi semoga skripsi ini dapat membawa manfaat dan keberkahan bagi perkembangan ilmu pengetahuan terutama bagi semua civitas Ilmu Komputer Universitas Lampung.

Bandar Lampung, 10 Maret 2023

Bayu Adhi Prabowo  
1657051011

## DAFTAR ISI

Halaman

|                                             |            |
|---------------------------------------------|------------|
| <b>DAFTAR ISI</b> .....                     | <b>iii</b> |
| <b>DAFTAR TABEL</b> .....                   | <b>v</b>   |
| <b>DAFTAR GAMBAR</b> .....                  | <b>vi</b>  |
| <b>I. PENDAHULUAN</b> .....                 | <b>1</b>   |
| <b>1.1. Latar Belakang</b> .....            | <b>1</b>   |
| <b>1.2. Rumusan Masalah</b> .....           | <b>3</b>   |
| <b>1.3. Batasan Masalah</b> .....           | <b>3</b>   |
| <b>1.4. Tujuan</b> .....                    | <b>3</b>   |
| <b>1.5. Manfaat</b> .....                   | <b>3</b>   |
| <b>II. TINJAUAN PUSTAKA</b> .....           | <b>5</b>   |
| <b>2.1. Text Mining</b> .....               | <b>5</b>   |
| <b>2.2. Sentiment Analysis</b> .....        | <b>5</b>   |
| 2.2.1. Metode dan Fitur .....               | <b>6</b>   |
| <b>2.3. Deep Learning</b> .....             | <b>7</b>   |
| <b>2.4. Recurrent Neural Network</b> .....  | <b>8</b>   |
| <b>2.5. Twitter</b> .....                   | <b>9</b>   |
| <b>2.6. Word2vec</b> .....                  | <b>10</b>  |
| <b>2.7. POS Tagging</b> .....               | <b>11</b>  |
| <b>2.8. Bahasa Pemrograman Python</b> ..... | <b>11</b>  |
| <b>2.9. Gensim</b> .....                    | <b>12</b>  |
| <b>2.10. Machine Learning</b> .....         | <b>12</b>  |

|                                                                                                                  |           |
|------------------------------------------------------------------------------------------------------------------|-----------|
| 2.11. <i>Support Vector Machine</i> .....                                                                        | 13        |
| 2.12. <i>Vaporizer</i> .....                                                                                     | 13        |
| 2.13. <b>Penelitian Terdahulu</b> .....                                                                          | 14        |
| 1. <i>Twitter Sentiment Analysis Terhadap Brand Reputation: Studi Kasus PT XL Axiata Tbk.</i> (Vidya, 2015)..... | 14        |
| 2. Sentiment Analysis menggunakan <i>Support Vector Machine (SVM)</i> (Nomleni, 2015).....                       | 14        |
| 3. Analisis Sentimen Calon Gubernur DKI Jakarta 2017 Di <i>Twitter</i> (Buntoro, 2017) .....                     | 14        |
| <b>III. METODE PENELITIAN</b> .....                                                                              | <b>18</b> |
| <b>3.1. Tempat dan Waktu Penelitian</b> .....                                                                    | <b>18</b> |
| <b>3.2. Alat dan Bahan</b> .....                                                                                 | <b>18</b> |
| 3.2.1 Perangkat Keras ( <i>Hardware</i> ).....                                                                   | 18        |
| 3.2.2 Perangkat Lunak ( <i>Software</i> ).....                                                                   | 18        |
| <b>3.3. Tahapan Penelitian</b> .....                                                                             | <b>19</b> |
| <b>IV. HASIL DAN PEMBAHASAN</b> .....                                                                            | <b>23</b> |
| <b>4.1. Pengumpulan Data</b> .....                                                                               | <b>23</b> |
| <b>4.2. Pembagian data</b> .....                                                                                 | <b>24</b> |
| <b>4.3. Normalisasi Teks</b> .....                                                                               | <b>24</b> |
| <b>4.4. Tokenisasi dan <i>Lower Case</i></b> .....                                                               | <b>25</b> |
| <b>4.5. Pembuangan <i>Stopword</i></b> .....                                                                     | <b>25</b> |
| <b>4.6. <i>Training</i> Klasifikasi</b> .....                                                                    | <b>26</b> |
| <b>4.7. Evaluasi Hasil Klasifikasi</b> .....                                                                     | <b>39</b> |
| <b>V. SIMPULAN DAN SARAN</b> .....                                                                               | <b>51</b> |
| <b>5.1 Simpulan</b> .....                                                                                        | <b>51</b> |
| <b>5.2 Saran</b> .....                                                                                           | <b>51</b> |
| <b>DAFTAR PUSTAKA</b> .....                                                                                      | <b>52</b> |

## DAFTAR TABEL

| Tabel                                                                                  | Halaman |
|----------------------------------------------------------------------------------------|---------|
| 1. Penelitian Terdahulu. ....                                                          | 16      |
| 2. <i>Confusin matrix</i> .....                                                        | 21      |
| 3. Pembagian Data Twitter. ....                                                        | 24      |
| 4. <i>Confusion Matrix Simple RNN epoch 8</i> . ....                                   | 39      |
| 5. <i>Confusion Matrix Simple RNN epoch 10</i> . ....                                  | 39      |
| 6. <i>Confusion Matrix Simple RNN epoch 12</i> . ....                                  | 39      |
| 7. <i>Confusion Matrix LSTM epoch 8</i> . ....                                         | 39      |
| 8. <i>Confusion Matrix LSTM epoch 10</i> . ....                                        | 40      |
| 9. <i>Confusion Matrix LSTM epoch 12</i> . ....                                        | 40      |
| 10. <i>Confusion Matrix SVM kernel Linear</i> . ....                                   | 40      |
| 11. <i>Confusion Matrix SVM kernel Rbf</i> . ....                                      | 40      |
| 12. <i>Data Training Simple RNN Epoch 8</i> . ....                                     | 41      |
| 13. <i>Data Training Simple RNN epoch 10</i> . ....                                    | 42      |
| 14. <i>Data Training RNN Epoch 12</i> . ....                                           | 43      |
| 15. <i>Data Training LSTM Epoch 8</i> . ....                                           | 44      |
| 16. <i>Data Training LSTM Epoch 10</i> . ....                                          | 45      |
| 17. <i>Data Training LSTM Epoch 12</i> . ....                                          | 46      |
| 18. <i>Data Training SVM Kernel Rbf</i> . ....                                         | 47      |
| 19. <i>Data Training SVM Kernel Linear</i> . ....                                      | 48      |
| 20. Hasil Perbandingan dan Perhitungan Klasifikasi <i>Simple RNN</i> dan LSTM<br>..... | 49      |
| 21. Hasil Perbandingan dan Perhitungan Klasifikasi SVM. ....                           | 49      |

## DAFTAR GAMBAR

| Gambar                                                                   | Halaman |
|--------------------------------------------------------------------------|---------|
| 1. Arsitektur RNN.....                                                   | 9       |
| 2. Tahapan Penelitian.....                                               | 19      |
| 3. Dataset yang Digunakan Dalam Penelitian.....                          | 24      |
| 4. Hasil Normalisasi Teks.....                                           | 25      |
| 5. Hasil Tokenisasi dan <i>Lower Case</i> .....                          | 25      |
| 6. Hasil <i>Stopword</i> .....                                           | 26      |
| 7. Grafik <i>Training Simple Recurrent Neural Network Epoch 8</i> .....  | 31      |
| 8. Grafik <i>Training Simple Recurrent Neural Network Epoch 10</i> ..... | 32      |
| 9. Grafik <i>Training Simple Recurrent Neural Network Epoch 12</i> ..... | 33      |
| 10. Grafik <i>Training LSTM Epoch 8</i> .....                            | 34      |
| 11. Grafik <i>Training LSTM Epoch 10</i> .....                           | 35      |
| 12. Grafik <i>Training LSTM Epoch 12</i> .....                           | 36      |

## I. PENDAHULUAN

### 1.1. Latar Belakang

Rokok adalah barang yang mudah ditemui dimanapun di Indonesia. Semakin berkembangnya teknologi rokok yang saat ini sudah ada dalam bentuk elektrik, atau disebut juga sebagai rokok elektrik. Rokok elektrik adalah sebuah inovasi dari bentuk rokok konvensional menjadi rokok modern. Rokok elektrik pertama kali dikembangkan pada tahun 2003. Rokok elektrik sehari-hari dikenal juga sebagai *vape*, *vape* sendiri adalah alat yang digunakan untuk menguapkan zat untuk inhalasi. *Vape* sudah mulai masuk Indonesia mulai pada tahun 2012 namun pada saat itu *vape* tidak langsung dikenal oleh orang Indonesia. Pada tahun 2014 *vape* terkena banyaknya isu negatif di Indonesia namun isu negatif tentang *vape* tidak terbukti hingga selisih 1 tahun pada tahun 2015 menjadi tahun yang baik karena mengalami kenaikan penggunaannya. Mulai tahun 2018 *vape* resmi dilegalkan oleh pemerintah Indonesia (Yusuf, 2019).

Sekarang adalah zaman dimana hampir semua orang memiliki akun media sosial sendiri. Media sosial adalah sebuah media dalam jaringan, dengan para penggunaannya dapat dengan mudah berpartisipasi, berbagi, dan menciptakan isi blog, jejaring sosial, *Wikipedia*, forum, dan dunia virtual. Blog, Facebook, Twitter dan *Wikipedia* merupakan bentuk media sosial yang paling umum digunakan oleh masyarakat di seluruh dunia. Berbagai macam *sentiment* atau pendapat dari masyarakat Indonesia, tentang rokok elektrik (*vape*) yang beredar di Indonesia dapat ditemui diberbagai media sosial.

*Sentiment analysis* merupakan salah satu cabang dari *text mining* yang melakukan identifikasi teks dan kemudian mengekstrak informasi dari teks yang telah diidentifikasi menjadi informasi subjektif dalam sumber (Soong et al., 2019). *Sentiment analysis* ini pun dapat membantu pihak individu atau



kelompok (organisasi) untuk memahami *sentiment* sosial dari merek, produk, atau layanan mereka saat memantau percakapan *online*. Dengan menggunakan *sentiment analysis*, individu atau organisasi dapat mengetahui respon pengguna umum terhadap merek, produk, atau layanan yang telah mereka buat. Untuk mendapatkan informasi tentang *sentiment analysis* dapat diperoleh dari mana saja, salah satunya dengan menggunakan sosial media seperti *Twitter*. *Sentiment analysis* dapat membantu dalam menyimpulkan suatu persepsi *netizen* dan pengguna media sosial terhadap informasi, status, *tweet*, komentar, dan suatu reaksi atau respon dalam media sosial dengan melakukan analisis dari setiap kalimat atau data yang didapat.

*Sentiment analysis* dalam pengumpulan data statistik menggunakan berbagai algoritme yang berasal dari cabang ilmu *artificial intelligence*, seperti *deep learning* dan algoritme *machine learning* lainnya seperti *Naïve bayes*, *Support Vector Machine (SVM)*, *Artificial Neural Network (ANN)*, regresi dan sebagainya. *Support Vector Machines (SVM)* telah menjadi metode klasifikasi dan regresi yang populer untuk masalah linear dan nonlinear. Keistimewaan dari *SVM* karena kemampuan untuk menerapkan pemisahan linear pada input data nonlinear berdimensi tinggi, dan ini diperoleh dengan menggunakan fungsi kernel yang diperlukan. Efektifitas *Support Vector Machines* sangat dipengaruhi oleh jenis fungsi kernel yang dipilih dan diterapkan berdasarkan karakteristik data (Haddi et al., 2013).

*Deep learning* dapat digunakan dalam *sentiment analysis*. Hal itu karena *deep learning* dapat dengan mudah diaplikasikan ke dalam *natural language processing (NLP)*. *Natural language processing* sangat tepat untuk *sentiment analysis* karena berhubungan dengan pemahaman bahasa tulisan, pemahaman bahasa alami, dan pengenalan suatu teks atau tulisan. Tetapi kelemahan dari *deep learning* adalah membutuhkan data dalam jumlah yang besar atau banyak untuk dijadikan data *training* dan data *testing*.

Pada penelitian ini, *deep learning* akan diterapkan pada *sentiment analysis* untuk menganalisa dan mengklasifikasikan respon dari masyarakat Indonesia

apakah berpendapat positif, negatif atau netral mengenai rokok elektrik di Indonesia dari media sosial Twitter. Kemudian dalam penelitian ini pun juga akan dilakukan perbandingan akurasi antara algoritme *deep learning* menggunakan metode *Recurrent Neural Network* (RNN) dengan algoritme *Support Vector Machine* pada *sentiment analysis*.

## 1.2. Rumusan Masalah

1. Bagaimana pengklasifikasian *sentiment* negatif, *sentiment* positif terhadap data yang didapatkan dari Twitter dengan menggunakan algoritme *Recurrent Neural Network* (RNN).
2. Bagaimana perbandingan nilai *confusion matrix* dari tiga metode *Recurrent Neural Network* (RNN), *Long Short Term Memory*(LSTM), dan algoritme *Support Vector Machine* (SVM) dalam *sentiment analysis* pada data Twitter.

## 1.3. Batasan Masalah

Penelitian ini dibatasi oleh beberapa hal, yang ditentukan sebagai berikut:

1. Menggunakan studi kasus mengenai rokok elektrik di Indonesia.
2. Data *tweet* yang digunakan hanya *tweet* yang menyebutkan kata kunci yang berhubungan dengan rokok elektrik di Indonesia.
3. *Tweet* yang digunakan berbahasa Indonesia.
4. Algoritme klasifikasi yang digunakan adalah *Recurrent Neural Network* dan *Support Vector Machine*.

## 1.4. Tujuan

Tujuan dari penelitian ini adalah sebagai berikut.

1. Melakukan pengklasifikasian menggunakan *Recurrent Neural Network* (RNN) dan *Support Vector Machine* (SVM).
2. Melakukan perbandingan nilai *confusion matrix* dari 3 metode yaitu *Simple RNN*, LSTM, dan SVM.

## 1.5. Manfaat

Manfaat dari penelitian ini yaitu:

1. Mengetahui cara mengklasifikasikan sentimen pada data *tweet* menggunakan algoritme *Recurrent Neural Network* dalam melakukan analisis sentimen.
2. Mengetahui keunggulan antara dua algoritme (*Recurrent Neural Network* dan *Support Vector Machine* (SVM) dalam melakukan analisis sentiment.

## II. TINJAUAN PUSTAKA

### 2.1. *Text Mining*

*Text mining* memiliki definisi menambang data yang berupa teks dimana sumber data biasanya didapatkan dari dokumen, dan tujuannya adalah mencari kata-kata yang dapat mewakili isi dari dokumen sehingga dapat dilakukan analisa keterhubungan antar dokumen (Raymond J Mooney, 2006). *Text mining* merupakan penambangan pengetahuan dari data yang berupa teks dan data jenis ini sifatnya tidak terstruktur. Banyak data teks yang bisa kita temui dalam kehidupan sehari-hari, dan data tersebut bisa kita olah sesuai dengan tujuan penelitian kita. Data teks itu seperti artikel di media *online*, *chat grup whatsapp*, *status* atau *tweet* di media sosial dan lain sebagainya. Menurut penelitian, terdapat lebih dari 80% data yang ada di internet bersifat tidak terstruktur, seperti data teks, video, audio, image dan lainnya (Multinity, 2020).

Fungsi dari *text mining* sangat beragam karena cakupannya adalah teks, maka apapun yang kita ingin digali dari teks tersebut dapat kita lakukan dengan teknik Text Mining. Salah satunya adalah menganalisis sentimen berdasarkan chat/status/*tweet* dari banyak orang terhadap suatu kasus tertentu. Misalnya menganalisis sentimen para pengguna aplikasi *game online Mobile Legend*, lebih banyak mana orang yang suka atau orang yang tidak suka.

### 2.2. *Sentiment Analysis*

*Sentiment Analysis* (juga dikenal sebagai *opinion mining*) mengacu pada penggunaan pemrosesan bahasa alami, analisis teks, linguistik komputasi, dan biometrik untuk mengidentifikasi, mengekstraksi, mengukur, dan mempelajari status afektif dan informasi subjektif secara sistematis (Saranya & Jayanthi, 2018). *Sentiment Analysis* secara luas diterapkan untuk

menyuarakan materi pelanggan seperti ulasan dan tanggapan survei, media online dan sosial, dan bahan perawatan kesehatan untuk aplikasi yang berkisar dari pemasaran hingga layanan pelanggan hingga kedokteran klinis.

Tugas dasar dalam *Sentiment Analysis* adalah mengklasifikasikan polaritas teks yang diberikan pada tingkat dokumen, kalimat, atau fitur/aspek apakah opini yang diungkapkan dalam dokumen, kalimat atau fitur/aspek entitas adalah positif, negatif, atau netral.

### 2.2.1. Metode dan Fitur

Pendekatan yang ada untuk *sentiment analysis* dapat dikelompokkan ke dalam tiga kategori utama, yaitu *knowledge-based techniques*, *statistical methods*, dan *hybrid approaches* (Cambria et al., 2013). *Knowledge-based techniques* mengklasifikasikan teks dengan memengaruhi yang tidak ambigu seperti senang, sedih, takut dan bosan (Ortony et al., 2015). Beberapa *knowledge-based* tidak hanya mencantumkan kata-kata yang mempengaruhi secara jelas, tetapi juga menetapkan kata-kata yang mungkin berhubungan dengan emosi tertentu.

*Statistical method* memanfaatkan elemen dari *machine learning* seperti *support vector machines* (SVM), “*bag-of-words*”, *latent semantic analysis*, “*pointwise mutual information*” untuk *semantic operations*, dan *deep learning*. Metode yang lebih canggih mencoba untuk mendeteksi pelaku *sentiment* dan target diteliti (Kim et al., 2006). Untuk menambang pendapat dalam konteks dan mendapatkan fitur tentang pendapat pembicara, hubungan tata bahasa kata digunakan. Hubungan ketergantungan *grammatical* diperoleh dengan penguraian teks yang dalam.

*Hybrid approaches* memanfaatkan *machine learning* dan elemen dari representasi pengetahuan seperti ontologi dan jaringan semantik untuk mendeteksi *semantic* yang diekspresikan secara halus, misalnya melalui analisis konsep yang tidak secara eksplisit

menyampaikan informasi yang relevan, tetapi yang secara implisit terkait ke konsep lain yang melakukannya.

### 2.3. *Deep Learning*

*Deep Learning* merupakan salah satu algoritme dalam pembelajaran mesin (*Machine Learning*). Yang terdiri dari beberapa lapisan tersembunyi dari jaringan saraf tiruan. Metodologi *deep learning* menerapkan transformasi nonlinier dan abstraksi model tingkat tinggi dalam basis data besar. Kemajuan terbaru dalam arsitektur pembelajaran di berbagai bidang telah memberikan kontribusi signifikan dalam kecerdasan buatan (*Artificial Intelligence*). Kecerdasan buatan (*Artificial Intelligence/AI*) seperti kecerdasan yang ditunjukkan oleh mesin telah menjadi pendekatan yang efektif untuk pembelajaran dan penalaran manusia. Pada tahun 1950, "*The Turing Test*" diusulkan sebagai penjelasan yang memuaskan tentang bagaimana komputer dapat melakukan penalaran kognitif manusia (Vargas & Ruiz, 2018).

Konsep *Deep Learning* muncul untuk pertama kalinya pada tahun 2006 sebagai bidang penelitian baru dalam pembelajaran mesin. Ini pertama kali dikenal sebagai pembelajaran hierarkis di awal, dan banyak bidang diterapkan dalam penelitian yang berkaitan dengan pengenalan pola (Lecun et al., 2015). *Deep learning* terutama mempertimbangkan dua faktor utama: pemrosesan non linear dalam berbagai lapisan atau tahapan dan pembelajaran yang diawasi atau tidak diawasi. Pemrosesan nonlinear dalam banyak lapisan mengacu pada algoritme di mana lapisan saat ini mengambil output dari lapisan sebelumnya sebagai input. Hierarki dibangun di antara lapisan-lapisan untuk mengatur pentingnya data agar dianggap berguna atau tidak. (Bengio, 2014) Di sisi lain, pembelajaran yang diawasi dan tidak diawasi terkait dengan label target kelas, ketersediaannya berarti sistem yang diawasi, sedangkan ketidakhadirannya berarti sistem yang tidak diawasi.

Untuk melakukan sentiment analysis, aplikasi *deep learning* menggunakan adalah *Natural Language Processing* (NLP) sebagai dasarnya. *Neural*

*networks* telah digunakan untuk mengimplementasikan model bahasa sejak awal 2000-an. Penyisipan kata, seperti *word2vec*, dapat dianggap sebagai lapisan representasional dalam arsitektur *deep learning* yang mengubah kata atom menjadi representasi posisi kata relatif terhadap kata lain dalam *dataset* posisi direpresentasikan sebagai titik dalam ruang vektor. Menggunakan kata *embedding* sebagai lapisan input RNN memungkinkan jaringan untuk mengurai kalimat dan frasa menggunakan tata bahasa vektor komposisi yang efektif. Tata bahasa vektor komposisi dapat dianggap sebagai tata bahasa bebas konteks probabilistik atau *probabilistic context free grammar* (PCFG) yang diterapkan oleh RNN.

#### **2.4. Recurrent Neural Network**

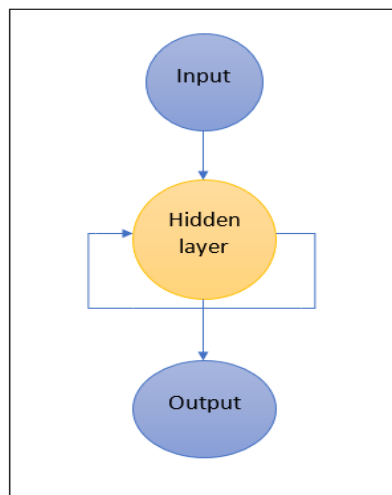
*Recurrent Neural Network* (RNN) termasuk dalam jenis jaringan pemodelan Neural Network. Selain *recurrent neural network* (RNN), *Multilayer Perceptron* (MPL) dan *feedforward neural network* (FNN) termasuk dalam jenis pemodelan *Neural Network* (Budiarti, 2006). FNN telah banyak diaplikasikan untuk melakukan prediksi atau klasifikasi data selama bertahun-tahun yang lalu. Salah satu tipe dari *neural network* yang dikembangkan dari dasar pemikiran FNN adalah RNN.

RNN yang juga disebut jaringan umpan balik adalah jenis jaringan pada neural network terdapat *loop* sebagai koneksi umpan balik dalam jaringan (Rizal & Soraya, 2018). Jaringan RNN adalah jaringan yang mengakomodasi *output* jaringan untuk menjadi input pada jaringan tersebut yang kemudian digunakan untuk menghasilkan *output* yang baru.

Pembuatan RNN sebenarnya dibuat untuk data-data yang bersifat sequential atau bertahap. Penggunaan *neural network* biasanya semua *input* dan *output* tidak bergantung satu sama lain, maka akan terjadi penumpukan tugas pada neural network sangat banyak dan bertumpuk, sedangkan jika ingin membangun sebuah aplikasi yang mengelola data ilmiah dengan data sampel berbentuk *time series* diperlukan algoritme yang dapat menangani data tersebut dengan cepat dan tepat, maka RNN dapat melakukan hal itu karena

RNN melakukan tugas yang sama pada setiap elemen data di sebuah urutan, lalu memproses outputnya yang mengacu pada hasil komputasi sebelumnya. Maka secara teori RNN mampu menggunakan informasi yang telah direkam sebelumnya yang panjang urutannya beragam-ragam (Jain et al., 2016). Jaringan node pada RNN dimasukkan ke dalam *layer* yang berurutan. Setiap *node* dalam lapisan tertentu terhubung dengan koneksi terarah ke setiap *node* lain dilapisan layer berikutnya secara beturut-turut, setiap node juga memiliki aktivasi yang bernilai nyata yang bervariasi tergantung waktu.

Pada Gambar 1, merupakan diagram arsitektur pada pemrosesan RNN, terlihat bahwa lingkaran hijau merupakan *input layer* atau masuknya data yang akan dilakukan komputasi pada RNN, setelah itu masuk ke lingkaran biru yang merupakan *hidden layer* terlihat ada garis looping yang kembali pada lingkaran biru tersebut, pola *looping* ini memungkinkan RNN untuk menyimpan ingatan sementara yang nantinya akan dipakai dalam pemrosesan data berikutnya, dan juga yang terakhir ada *output layer* yang nantinya adalah hasil dari proses komputasi RNN.



Gambar 1. Arsitektur RNN.

## 2.5. Twitter

Twitter adalah layanan microblogging dan jejaring sosial di mana pengguna memposting dan berinteraksi dengan pesan yang dikenal sebagai "*tweets*". *Tweet* awalnya dibatasi hingga 140 karakter, tetapi pada 7 November 2017,



batas ini digandakan menjadi 280 untuk semua bahasa kecuali Cina, Jepang, dan Korea. Pengguna terdaftar dapat memposting, menyukai, dan me-*retweet tweet*, tetapi pengguna yang tidak terdaftar hanya dapat membacanya. Pengguna mengakses Twitter melalui antarmuka situs webnya, melalui Layanan Pesan Singkat (SMS) atau perangkat lunak aplikasi perangkat selulernya ("aplikasi"). Twitter, Inc. berbasis di San Francisco, California, dan memiliki lebih dari 25 kantor di seluruh dunia (Kelly, 2009).

Twitter dibuat pada Maret 2006 oleh Jack Dorsey, Noah Glass, Biz Stone, dan Evan Williams, diluncurkan pada Juli tahun itu. Layanan ini dengan cepat mendapatkan popularitas di seluruh dunia. Pada tahun 2012, lebih dari 100 juta pengguna memposting 340 juta *tweet* per hari, dan layanan ini menangani rata-rata 1,6 miliar permintaan pencarian per hari. Pada 2013, itu adalah salah satu dari sepuluh situs web yang paling banyak dikunjungi dan telah digambarkan sebagai "SMS Internet". Pada 2018, Twitter memiliki lebih dari 321 juta pengguna aktif bulanan (Holton et al., 2014). Sejak 2015, Twitter telah menjadi sarang perdebatan dan berita yang meliput politik Amerika Serikat. Selama pemilihan presiden A.S. 2016, Twitter adalah sumber berita terbesar pada hari itu, dengan 40 juta *tweet* terkait pemilu dikirim pada pukul 10:00 malam. (Waktu Bagian Timur) hari itu.

## 2.6. *Word2vec*

*Word2vec* adalah sekelompok model terkait yang digunakan untuk menghasilkan *word embeddings*. Model-model ini dangkal, jaringan saraf dua lapis yang dilatih untuk merekonstruksi konteks kata-kata linguistik. *Word2vec* menggunakan input corpus besar dan menghasilkan ruang vektor, biasanya beberapa ratus dimensi, dengan setiap kata unik dalam corpus ditugaskan vektor yang sesuai dalam ruang. Vektor kata diposisikan dalam ruang vektor sehingga kata-kata yang berbagi konteks umum dalam korpus terletak berdekatan satu sama lain dalam ruang tersebut (Mikolov, Corrado, et al., 2013).

*Word2vec* dibuat dan diterbitkan pada 2013 oleh tim peneliti yang dipimpin oleh Tomas Mikolov di Google dan dipatenkan. Algoritme ini kemudian dianalisis dan dijelaskan oleh peneliti lain. Penggunaan vektor yang dibuat menggunakan algoritme *Word2vec* memiliki banyak keunggulan dibandingkan dengan algoritme sebelumnya seperti *latent semantic analysis*.

*Word2vec* dapat menggunakan salah satu dari dua model arsitektur untuk menghasilkan representasi kata-kata yang terdistribusi: *Continuous Bag-of-Word* (CBOW) atau *skip-gram*. Dalam arsitektur *bag-of-words* yang berkelanjutan, model memprediksi kata saat ini dari jendela kata konteks sekitarnya. Urutan kata konteks tidak memengaruhi prediksi (asumsi *bag-of-words*). Dalam arsitektur *skip-gram* terus menerus, model menggunakan kata saat ini untuk memprediksi jendela kata konteks di sekitarnya. Arsitektur *skip-gram* menimbang kata-kata konteks terdekat lebih berat daripada kata-kata konteks lebih jauh (Mikolov, Sutskever, et al., 2013). *Continuous Bag-of-Word* (CBOW) lebih cepat sementara *skip-gram* lebih lambat tetapi melakukan pekerjaan dengan lebih baik pada kata-kata yang jarang.

## 2.7. *POS Tagging*

Dalam linguistik corpus, *part-of-speech tagging* (*POS tagging* atau *PoS tagging* atau *POST*), juga disebut penandaan gramatikal atau disambiguasi kata-kategori, adalah proses menandai kata dalam teks (*corpus*) sesuai dengan bagian tertentu pidato, berdasarkan definisi dan konteksnya yaitu hubungannya dengan kata-kata yang berdekatan dan terkait dalam frasa, kalimat, atau paragraf. Bentuk yang disederhanakan dari ini umumnya diajarkan kepada anak-anak usia sekolah, dalam mengidentifikasi kata-kata sebagai kata benda, kata kerja, kata sifat, kata keterangan, dan lain-lain (Güngör, 2006).

## 2.8. Bahasa Pemrograman Python

*Python* adalah bahasa pemrograman tujuan umum tingkat tinggi yang ditafsirkan. Diciptakan oleh Guido van Rossum dan dirilis pertama kali pada tahun 1991, filosofi desain *Python* menekankan keterbacaan kode dengan

penggunaan spasi spasi yang signifikan. Konstruksi bahasanya dan pendekatan berorientasi objek bertujuan untuk membantu programmer menulis kode yang jelas dan logis untuk proyek skala kecil dan besar (Guttag, 2016).

*Python* diketik dan dikumpulkan secara dinamis. Ini mendukung beberapa paradigma pemrograman, termasuk pemrograman prosedural, berorientasi objek, dan fungsional. *Python* sering digambarkan sebagai bahasa "termasuk baterai" karena pustaka standarnya yang komprehensif.

## 2.9. *Gensim*

*Gensim* atau *Generate Similar* adalah *open-source library* untuk *unsupervised topic modeling* dan *natural language processing* menggunakan *modern statistical machine learning*. *Gensim* diimplementasikan dalam Python dan Cython. *Gensim* dirancang untuk menangani koleksi teks besar menggunakan streaming data dan algoritme online tambahan, yang membedakannya dari sebagian besar paket perangkat lunak *machine learning* lainnya yang hanya menargetkan pemrosesan dalam memori (Rehurek & Sojka, 2010). Python *gensim* sendiri dapat berjalan di semua platform (seperti Windows, Mac OS, Linux) yang mendukung *Python* dan *Numpy*.

## 2.10. *Machine Learning*

*Machine Learning* atau pembelajaran mesin adalah sebuah studi tentang algoritme komputer yang meningkatkan pengalaman secara otomatis menggunakan data yang sudah ada (Mitchell, 1997). *Machine Learning* mula-mula diperkenalkan oleh Arthur Samuel pada tahun 1959. Menurut *machine learning* adalah tempat belajar yang membuat komputer berkemampuan untuk belajar tanpa diprogram secara eksplisit (Simon, 2013). Sistem pembelajaran dapat memanfaatkan contoh (data) untuk menangkap ciri yang diperlukan dari probabilitas yang mendasarinya (yang tidak diketahui). *Machine Learning* menggunakan beberapa jenis algoritme yang secara berulang-ulang menuntun komputer untuk belajar dari data, teknik *machine learning* sendiri terdiri dari *supervised learning*, *unsupervised learning*, *semi-supervised learning* dan *reinforcement learning*.

### 2.11. *Support Vector Machine*

*Support Vector Machine* (SVM) merupakan salah satu metode klasifikasi dengan menggunakan machine learning (*supervised learning*) yang memprediksi kelas berdasarkan model atau pola dari hasil proses training. Klasifikasi dilakukan dengan mencari *hyperplane* atau garis pembatas (decision boundary) yang memisahkan antara suatu kelas dengan kelas lain, yang dalam kasus ini garis tersebut berperan memisahkan *tweet* bersentimen positif (berlabel +1) dengan *tweet* bersentimen negatif (berlabel -1). SVM melakukan pencarian nilai hyperplane dengan menggunakan *support vector* dan nilai *margin* (Han & Kamber, 2013). Dalam penelitian ini, data inputan yang memiliki representasi vektor didapatkan dari proses pembobotan. Dengan dilakukannya training pada klasifikasi SVM, maka akan menghasilkan sebuah nilai atau pola yang akan digunakan pada proses testing untuk proses testing SVM, yang bertujuan memberi label sentimen pada *tweet* (Pang et al., 2019).

### 2.12. *Vaporizer*

Rokok elektronik atau vapor/*vape* ke dalam *Electronic Delivery System* (EDS), yakni alat yang menggunakan listrik dari tenaga baterai untuk memberikan nikotin dalam bentuk uap (Notoatmodjo, 2012). Rokok jenis ini dirancang untuk membantu pecandu rokok tembakau mulai berhenti merokok. Dengan beralih dari rokok tembakau ke rokok elektrik, secara perlahan mereka belajar untuk berhenti merokok. Rokok jenis ini terdapat dalam berbagai bentuk dan ukuran, tetapi terdapat tiga komponen utama dalam rokok elektrik, yaitu baterai, elemen pemanas, dan tabung yang berisi cairan (*cartridge*). Pada awal keberadaan rokok elektronik, produk tersebut dikatakan aman bagi kesehatan karenalarutan nikotin yang terdapat pada rokok elektronik hanya terdiri dari campuran air, propilen glikol, zat penambah rasa, aroma tembakau, dan senyawa-senyawa lain yang tidak mengandung tar, tembakau atau zat-zat toksik lain yang umum terdapat pada rokok tembakau (Trtchounian et al., 2010). Beberapa rokok elektrik memiliki

baterai dan cartridge yang dapat diisi ulang. Bahkan di Indonesia saja orang dari perokok aktif di Indonesia adalah perokok elektrik (*vape*).

### 2.13. Penelitian Terdahulu

Penelitian terkait dengan *text mining* pada umumnya untuk masalah *sentiment analysis* telah banyak dilakukan sebelumnya dengan metode yang berbeda-beda. Tabel 1, menunjukkan penelitian terdahulu.

1. *Twitter Sentiment Analysis Terhadap Brand Reputation: Studi Kasus PT XL Axiata Tbk.*(Vidya, 2015).

Vidya (2015) dalam penelitiannya menunjukkan bahwa Twitter dapat digunakan untuk analisis *brand* untuk kegiatan evaluasi *marketing communication*. Klasifikasi sentiment dan kata kunci dari *account Twitter brand* tersebut, yaitu XL Axiata. Peneliti melakukan pembagian objek yang diteliti menjadi 3 kelompok yaitu dengan metode algoritme klasifikasi *naïve bayes*, *support vector machine* dan *decision tree*.

2. *Sentiment Analysis menggunakan Support Vector Machine (SVM)* (Nomleni, 2015).

Nomeleni (2015) menjelaskan hasil akuisisi dan percobaan data teks yang diolah menjadi *sentiment analysis* menggunakan *Support Vector Machine (SVM)* pengujian yang dilakukan sebanyak 7 kali percobaan memiliki tingkat akurasi rata-rata diatas 80% dan memiliki akurasi tertinggi sebesar 84.4086% dengan data *training* 20% dan data uji sebanyak 80%.

3. *Analisis Sentimen Calon Gubernur DKI Jakarta 2017 Di Twitter* (Buntoro, 2017) .

Untuk proses klasifikasinya menggunakan metode *Naïve Bayes Classifier (NBC)* dan *Support Vector Machine (SVM)*. Data yang digunakan adalah *tweet* dalam bahasa Indonesia dengan kata kunci AHY, Ahok, Anies, dengan jumlah dataset sebanyak 300 *tweet*. Hasil dari penelitian ini adalah analisis sentimen terhadap calon gubernur DKI Jakarta 2017. Akurasi tertinggi didapat saat menggunakan metode

klasifikasi *Naïve Bayes Classifier* (NBC), dengan nilai rata-rata akurasi mencapai 95%, nilai *presisi* 95%, nilai *recall* 95% nilai *TP rate* 96,8% dan nilai *TN rate* 84,6%.

Tabel 1. Penelitian Terdahulu.

| <b>Perbandingan Teori</b>                                                                 | <b>Peneliti dan Tahun Penelitian</b> | <b>Tujuan</b>                                                                                                                                                                                                                                               | <b>Langkah-langkah</b>                                                                                                                                                                                                                                                                           | <b>Metode</b>                                                                                                                                       | <b>Output</b>                                                                                                                                    |
|-------------------------------------------------------------------------------------------|--------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------|
| <i>Twitter Sentiment Analysis Terhadap Brand Reputation: Studi Kasus PT XL Axiata Tbk</i> | Nur Azizah Vidya, 2015               | Mencari algoritme terbaik dari <i>Twitter sentiment analysis</i> dalam mengkategorisasi sentiment positif dan negative yang akan digunakan sebagai variabel perhitungan <i>Net Brand Reputation</i>                                                         | <ol style="list-style-type: none"> <li>Pengumpulan awal data</li> <li>Pengolahan dataset dan data training</li> <li>Kategorisasi data testing</li> <li>Perhitungan reputasi <i>brand</i></li> </ol>                                                                                              | Algoritme klasifikasi <i>naïve bayes, support vector machine dan decision tree</i> dengan perhitungan <i>Net sentiment dan net brand reputation</i> | Penelitian ini menunjukkan bahwa dengan metode <i>Naïve bayes</i> menghasilkan 78.90%, SVM menghasilkan 82.40%, dan <i>Decision Tree: 72.90%</i> |
| Sentiment Analysis menggunakan <i>Support Vector Machine (SVM)</i>                        | Petrix Nomleni, 2015                 | Mengklasifikasikan sentiment pada media <i>center</i> sebagai informasi positif atau negative menggunakan metode <i>support vector machine (SVM)</i> dan memberikan masukan sebagai bahan acuan untuk filterisasi setiap keluhan pada media <i>center</i> . | <ol style="list-style-type: none"> <li>Pengumpulan data <i>tweet</i></li> <li>Penentuan sentiment secara manual</li> <li>Pembagian data</li> <li>Indexing data latih dan data uji</li> <li>Seleksi fitur</li> <li>Fungsi klasifikasi <i>multinomial naïve bayes</i></li> <li>Evaluasi</li> </ol> | <i>Mutual Information (MI) dan Inverse Document Frequency (IDF)</i> dengan metode <i>Multinomial naïve bayes</i>                                    | Penelitian ini menunjukkan bahwa dengan metode MI menghasilkan 71.89%, dan IDF menghasilkan 60.67%                                               |

Tabel 1. Penelitian Terdahulu (Lanjutan)

| Perbandingan Teori                                                  | Peneliti dan Tahun Penelitian | Tujuan                                                                                                                                                                                                                                                   | Langkah-langkah                                                                                                                                                                                                                                                                                                                              | Metode                                                                                                                                                                                                                                                                                                | Output                                                                                              |
|---------------------------------------------------------------------|-------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------|
| Analisis Sentimen Calon Gubernur DKI Jakarta 2017 Di <i>Twitter</i> | Ghulam Asrofi Buntoro, 2017   | Tujuannya membantu masyarakat menentukan sentimen yang terdapat pada twit opini Bahasa Indonesia yang ada di <i>Twitter</i> . Setelah dilakukan analisis sentimen, terlihat berapa banyak sentimen yang ditujukan kepada calon Gubernur DKI Jakarta 2017 | <ol style="list-style-type: none"> <li>Mengumpulkan data <i>tweet</i></li> <li><i>Preprocessing</i> data</li> <li><i>Tokenisasi</i></li> <li><i>Part of Speech (POS) Tagger</i></li> <li>Penentuan <i>Class Attribute</i></li> <li><i>Load Dictionary</i></li> <li><i>Determine Sentiment</i></li> <li>Klasifikasi Evaluasi Hasil</li> </ol> | Metode yang digunakan untuk proses klasifikasinya menggunakan metode <i>Naïve Bayes Classifier (NBC)</i> dan <i>Support Vector Machine (SVM)</i> . Data yang digunakan adalah <i>tweet</i> dalam bahasa Indonesia dengan kata kunci AHY, Ahok, Anies, dengan jumlah dataset sebanyak 300 <i>tweet</i> | Penelitian ini menunjukkan nilai akurasi metode <i>Naïve bayes</i> sebesar 95%, dan SVM sebesar 90% |



### III. METODE PENELITIAN

#### 3.1. Tempat dan Waktu Penelitian

Penelitian ini dilakukan di Jurusan Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lampung. Penelitian dilaksanakan mulai bulan Juli sampai bulan Desember 2022.

#### 3.2. Alat dan Bahan

Penelitian ini dilakukan dengan menggunakan alat untuk mendukung dan menunjang pelaksanaan penelitian, yaitu sebagai berikut.

##### 3.2.1 Perangkat Keras (*Hardware*)

Perangkat keras yang digunakan dalam penelitian analisis sentimen ini adalah satu unit laptop dengan spesifikasi:

- Prosesor: AMD Ryzen 3 1300X *Quad-Core Processor* (4CPUs), ~3.5GHz
- RAM: 16384 MB

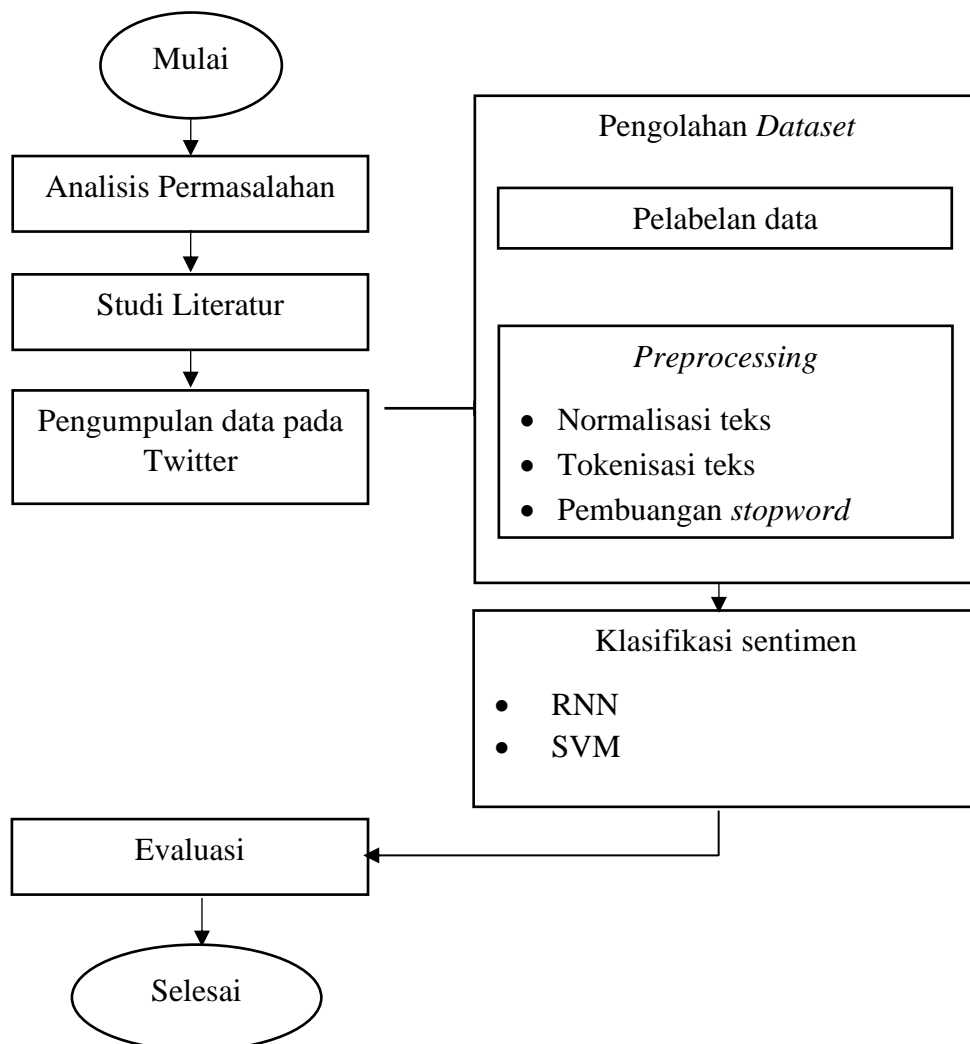
##### 3.2.2 Perangkat Lunak (*Software*)

Perangkat lunak yang digunakan dalam analisis sentimen ini adalah sebagai berikut:

- *Operating System*: Windows 10
- *Web Browser*: Google Chrome, digunakan untuk pengunduhan data.
- Python 3, digunakan sebagai bahasa pemrograman.
- *Jupyter Notebook*, digunakan untuk membuat *script* program.
- *Google Colab* untuk membantu merunning program untuk kacepatan data yang diolah.

### 3.3. Tahapan Penelitian

Penelitian ini melalui beberapa tahapan yang diilustrasikan pada Gambar 2.



Gambar 2. Tahapan Penelitian.

Gambar 2, menunjukkan proses yang dilalui dalam penelitian. Berikut ini adalah penjabaran dari masing-masing tahapan penelitian yang dilakukan sebagai berikut:

#### 1. Analisis Permasalahan

Pada tahap awal melakukan suatu analisis permasalahan yang ada yaitu pendapat masyarakat tentang rokok elektrik menggunakan media sosial Twitter yang digunakan untuk mendapatkan suatu data informasi untuk ke tahap selanjutnya. Topik penelitian ini adalah analisis sentimen untuk

membandingkan pendapat masyarakat mengenai rokok elektrik dengan deep learning menggunakan algoritme *Recurrent Neural Network* (RNN) dan *Support Vector Machines*.

## 2. Studi Literatur

Pada tahap studi literature dilakukan suatu pengumpulan informasi atau data dari penelitian terdahulu yang berkaitan dengan topik permasalahan yang mirip dan mempelajari berbagai metode penelitian yang digunakan dari literature yang dikumpulkan sebagai pedoman untuk penelitian.

## 3. Pengumpulan Data

Tahap ini dilakukan pengumpulan data tweet dari sosial media Twitter menggunakan API key yang disediakan oleh Twitter yaitu *tweepy*. Untuk menggunakan *tweepy* ini, diperlukan akun Twitter yang telah didaftarkan menjadi akun developer Twitter, setelah menjadi akun *developer* dapat mengambil data dengan cara *scrape* pada Api Twitter yang sudah ada.

## 4. Pengolahan Dataset

Tahap selanjutnya yaitu pengolahan dataset yang telah diperoleh dari Twitter dengan melakukan preprocessing data. Tahapan yang dilakukan saat melakukan preprocessing yaitu tokenisasi teks, normalisasi teks, pembuangan stopword, dan seleksi fitur yang menghasilkan data yang bersih untuk dijadikan sebuah data analisis pada tahap selanjutnya.

## 5. Klasifikasi Sentimen

Pada tahap ini dilakukan klasifikasi sentimen menggunakan dua jenis algoritme yaitu Recurrent Neural Network (RNN) dan Support Vector Machine (SVM). Data yang diklasifikasikan adalah data tweet yang telah melwati tahap pengolahan data. Proses klasifikasi menggunakan RNN akan digunakan dengan menggunakan simple RNN dan LSTM, sedangkan proses klasifikasi menggunakan SVM akan dilakukan menggunakan kernel linear dan rbf.

## 6. Evaluasi

Evaluasi yang dilakukan adalah melakukan perbandingan nilai sentimen dari pelabelan manual dan analisis sentimen menggunakan kedua algoritme tersebut. Setelah itu akan dilakukan perbandingan antara hasil

dari algoritme RNN dan algoritme *Support Vector Machines*. Hasil klasifikasi yang didapatkan adalah *confusion matrix* dan *classification report* berupa akurasi, presisi, *recall*, dan *f1-score*. Berikut adalah contoh *confusion matrix* untuk prediksi klasifikasi yang ditunjukkan pada Tabel 2.

Tabel 2. *Confusin matrix*.

| <i>Actual Values</i> | <i>Predicted values</i>    |                            |
|----------------------|----------------------------|----------------------------|
|                      | <i>Negative</i>            | <i>Positive</i>            |
| <i>Negative</i>      | <i>True Negative (TN)</i>  | <i>False Positive (FP)</i> |
| <i>Positive</i>      | <i>False Negative (FN)</i> | <i>True Positive (TP)</i>  |

Adapun beberapa nilai dari *confusion matrix* yang tertera pada Tabel 2, diantaranya:

1. *True Negative (TN)* merupakan nilai data fakta negatif dan diprediksi oleh program data tersebut juga bernilai negative.
2. *False positive (FP)* merupakan data fakta bernilai negatif tetapi diprediksi program data tersebut bernilai positif.
3. *False negative (FN)* merupakan nilai data fakta positif namun diprediksi program data tersebut bernilai negatif.
4. *true positive (TP)* adalah nilai data fakta positif dan diprediksi oleh program bahwa data tersebut juga bernilai positif.

*Confusion matrix* merupakan hasil informasi prediksi klasifikasi dari aktual data yang dilakukan oleh sistem. *Confusion matrix* digunakan untuk mengukur klasifikasi suatu metode dengan menghitung nilai akurasi, presisi, *recall*, dan *f1-score*.

- a. Akurasi merupakan rasio prediksi benar (positif, dan negatif) dengan keseluruhan data.

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \quad (1)$$

- b. Presisi merupakan rasio prediksi benar positif dibandingkan dengan keseluruhan hasil yang diprediksi positif.

$$precision = \frac{TP}{TP + FP} \times 100\% \quad (2)$$

- c. *Recall* merupakan rasio prediksi benar positif dibandingkan dengan keseluruhan data yang benar positif.

$$recall = \frac{TP}{TP + FN} \times 100\% \quad (3)$$

- d. *F1-Score* merupakan perbandingan rata-rata presisi dan *recall* yang dibobotkan.

$$f1\ score = 2 \times \frac{(recall \times precision)}{(recall + precision)} \quad (4)$$

## V. SIMPULAN DAN SARAN

### 5.1 Simpulan

Penelitian ini dapat disimpulkan sebagai berikut:

1. Hasil klasifikasi sentimen dari metode *support vector machine kernel rbf* menunjukkan nilai klasifikasi terbaik dari pemrosesan data dibandingkan dengan metode *simple RNN* dan metode LSTM yang dapat dilihat nilai terbesar terdapat di metode *support vector machine* dari nilai Akurasi, Presisi, *Recall*, dan *F1-Score* berbanding sedikit dari data *weighted avg* dan nilai terkecil terdapat di metode RNN.
2. Perbandingan nilai *confusion matrix* dari tiga metode yaitu *Simple RNN*, LSTM, dan SVM menunjukkan bahwa tidak terlalu banyak nilai prediksi yang error pada ke tiga metode.

### 5.2 Saran

Terdapat beberapa hal yang dapat ditambahkan atau diperbaiki untuk penelitian selanjutnya sebagai berikut:

1. Data *tweet* yang digunakan dapat ditambahkan sebanyak 5000 ribu data *tweet* agar data latih yang digunakan dapat lebih optimal, dan menyetarakan kelas data positif dan negatif sehingga sistem klasifikasi dapat memiliki nilai akurasi yang lebih baik.
2. Dapat mengembangkan analisis sentiment yang lainnya dengan mengkombinasikan metode *Deep Learning* yang lain selain RNN seperti *Deep Neural Networks*, dan *Convolutional Neural Networks* untuk dilakukan data training atau pengujian, dan mendapatkan nilai akurasi yang lebih baik serta optimal.

## DAFTAR PUSTAKA

- Bengio, Y. (2014). *Learning Deep Architectures for AI Learning Deep Architectures for AI* (Issue January 2009).  
<https://doi.org/10.1561/22000000006>
- Breiman, L. (1996). Bagging predictors. *Machine Learning*, 24(2), 123–140.
- Budiarti, A. (2006). Bab 2 landasan teori. *Aplikasi Dan Analisis Literatur Fasilkom UI*, 4–25.
- Buntoro, G. A. (2017). Analisis Sentimen Calon Gubernur DKI Jakarta 2017 Di Twitter. *Integer Journal Maret*, 1(1), 32–41.
- Cambria, E., Schuller, B., Xia, Y., & Havasi, C. (2013). *New Avenues in Opinion Mining and Sentiment Analysis*. April, 15–21.
- Güngör, T. (2006). HANDBOOK OF NATURAL LANGUAGE PROCESSING SECOND EDITION. In *HLT-NAACL 2006 - Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics, Proceedings of the Main Conference*.  
<https://doi.org/10.3115/1220835.1220877>
- Guttag, J. V. (2016). *Introduction to Computation and Programming using Python with application to understanding data*. *Journal of Chemical Information and Modeling*. (Vol. 53).
- Haddi, E., Liu, X., & Shi, Y. (2013). The role of text pre-processing in sentiment analysis. *Procedia Computer Science*, 17, 26–32.  
<https://doi.org/10.1016/j.procs.2013.05.005>

- Han, J., & Kamber, M. (2013). *Data Mining: Concepts and Techniques Second Edition*.
- Holton, A. E., Baek, K., Coddington, M., & Yaschur, C. (2014). Seeking and Sharing: Motivations for Linking on Twitter. *Communication Research Reports*, 31(1), 33–40. <https://doi.org/10.1080/08824096.2013.843165>
- I. Taufik dan S.A.Pamungkas. (2018). Analisis Sentimen Terhadap Tokoh Publik Menggunakan Algoritma Support Vector Machine (Svm). *Jurnal "LOG!K@,"* 8(1), 71–79.
- Jain, A., Zamir, A. R., Savarese, S., & Saxena, A. (2016). Jain\_Structural-RNN\_Deep\_Learning\_CVPR\_2016\_paper. *Cvpr*, 5308–5317.
- Kelly, R. (2009). Twitter Study – August 2009 Introduction. *New York, 2010(August)*, 1–17.
- Kim, S., Hovy, E., Rey, M., & Edu, I. S. I. (2006). *Automatic Identification of Pro and Con Reasons in Online Reviews*. *July*, 483–490.
- Lecun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444. <https://doi.org/10.1038/nature14539>
- Mikolov, T., Corrado, G., Chen, K., & Dean, J. (2013). *Efficient Estimation of Word Representations in Vector Space*. 1–12.
- Mikolov, T., Sutskever, I., Chen, K., Corrado, G., & Dean, J. (2013). *Distributed Representations of Words and Phrases and their Compositionality*. 1–9.
- Mitchell, T. (1997). *Machine Learning*. New York: McGraw Hill. In *Proceedings of the International Conference on Document Analysis and Recognition, ICDAR*.
- Muludi, K., Akbar, M. S., Shofiana, D. A., & Syarif, A. (2021). Sentiment Analysis Of Energy Independence Tweets Using Simple Recurrent Neural



- Network. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, 15(4), 339. <https://doi.org/10.22146/ijccs.66016>
- Nomleni, P. (2015). Sentiment Analysis Menggunakan Support Vector Machine ( Svm ). *Seminar Nasional Teknologi Dan Komunikasi 2015, 2015*(Sentika), 1–8. <https://doi.org/10.1016/j.eswa.2013.08.047>
- Notoatmodjo, S. (2012). Promosi Kesehatan & Ilmu Perilaku. In *Jakarta: Rineka Cipta*.
- Ortony, A., L. Clore, G., & Collins, A. (2015). *The Cognitive Structure of Emotions*.
- Pang, B., Lee, L., & Shivakumar, V. (2019). Thumbs up? Sentiment Classification using Machine Learning Techniques. *Proceedings of the Institution of Civil Engineers - Transport*, 172(2).
- Raymond J Mooney, Cs. (2006). Machine Learning Text Categorization. *Machine Learning Text Categorization*, 1–6.
- Rehurek, R., & Sojka, P. (2010). Software Framework for Topic Modelling with Large Corpora. *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks, May*, 45–50. <https://doi.org/10.13140/2.1.2393.1847>
- Rizal, A. A., & Soraya, S. (2018). *Multi Time Steps Prediction Dengan Recurrent Neural*. 18(1), 115–124.
- Saranya, K., & Jayanthi, S. (2018). Onto-based sentiment classification using machine learning techniques. *Proceedings of 2017 International Conference on Innovations in Information, Embedded and Communication Systems, ICIIECS 2017, 2018-January*. <https://doi.org/10.1109/ICIIECS.2017.8276047>
- Simon, Phil. W. and. (2013). Too Big to Ignore : The Business Case for Big Data. *Journal of Chemical Information and Modeling*, 53(9), 1689–1699.

Soong, H. C., Jalil, N. B. A., Kumar Ayyasamy, R., & Akbar, R. (2019). The essential of sentiment analysis and opinion mining in social media : Introduction and survey of the recent approaches and techniques. *ISCAIE 2019 - 2019 IEEE Symposium on Computer Applications and Industrial Electronics*, 272–277. <https://doi.org/10.1109/ISCAIE.2019.8743799>

Szandała, T. (2020). *Review and Comparison of Commonly Used Activation Functions for Deep Neural Networks*.

Trtchounian, A., Williams, M., & Talbot, P. (2010). Conventional and electronic cigarettes (e-cigarettes) have different smoking characteristics. *Nicotine & Tobacco Research : Official Journal of the Society for Research on Nicotine and Tobacco*, 12(9), 905–912. <https://doi.org/10.1093/ntr/ntq114>

Vargas, R., & Ruiz, L. (2018). Deep Learning : Previous and Present. *Journal of Awareness*, 2(Special 3), 11–20.

Vidya, N. A. (2015). *TWITTER SENTIMENT ANALYSIS TERHADAP BRAND REPUTATION: STUDI KASUS PT XL AXIATA Tbk*.