

***SENTIMENT ANALYSIS SATU TAHUN KEPEMIMPINAN WALIKOTA  
BANDARLAMPUNG PERIODE 2021-2026 MENGGUNAKAN  
ALGORITME SUPPORT VECTOR MACHINE (SVM) DAN NAÏVE BAYES***

**(Skripsi)**

**Oleh**

***Elsinta Milenia***  
**NPM 1857051012**



**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS LAMPUNG  
BANDAR LAMPUNG  
2022**

## ABSTRAK

### ***SENTIMENT ANALYSIS SATU TAHUN KEPEMIMPINAN WALIKOTA BANDARLAMPUNG PERIODE 2021-2026 MENGGUNAKAN ALGORITME SUPPORT VECTOR MACHINE (SVM) DAN NAÏVE BAYES***

Oleh

**ELSHINTA MILENIA**

Analisis sentimen adalah bagian dari penelitian komputasi yang mengklasifikasikan polaritas teks yang diberikan pada tingkat dokumen, kalimat, atau fitur apakah opini yang diungkapkan dalam dokumen, kalimat atau fitur tersebut bernilai positif atau negatif. Kemudahan akses internet dan perkembangan *gadget* yang semakin canggih menjadikan perangkat elektronik ini sebagai salah satu kebutuhan utama masyarakat untuk mengakses informasi. Belakangan ini media sosial menjadi sangat menarik, baik sebagai sumber data maupun sebagai sarana penyampaian wawasan dan kehidupan sehari-hari. Berbagai penilaian atau opini masyarakat Bandarlampung yang berbeda dapat ditemukan di media sosial Twitter dan salah satu topik populer masyarakat Bandarlampung adalah gaya kepemimpinan. Gaya kepemimpinan adalah perilaku dan energi kooperatif, yang merupakan kombinasi dari penalaran, kemampuan, atribut, dan praktik yang diterapkan oleh seorang pemimpin untuk mempengaruhi kinerja para pengikutnya. Gaya kepemimpinan yang akhir-akhir ini menyita perhatian dan perbincangan adalah mengenai gaya kepemimpinan Hj. Eva Dwiana, S.E adalah Walikota Bandarlampung periode 2021-2026. Penelitian ini membandingkan dua algoritme: *Support Vector Machine* (SVM) dan *Naïve Bayes* dalam analisis sentimen menggunakan data Walikota Bandarlampung dari Twitter. Berdasarkan hasil penelitian, *Support Vector Machine* (SVM) menunjukkan kinerja terbaik dengan nilai akurasi sebesar 87,38% dibandingkan dengan *Naïve Bayes* dengan nilai akurasi 85,28%.

Kata kunci: Analisis Sentimen, *Support Vector Machine* (SVM), *Naïve Bayes*, Kepemimpinan Walikota

## **ABSTRACT**

### **SENTIMENT ANALYSIS ONE YEAR LEADERSHIP OF BANDARLAMPUNG MAYOR 2021-2026 PERIOD USING SUPPORT VECTOR MACHINE (SVM) AND NAIVE BAYES ALGORITHM**

**By**

**ELSHINTA MILENIA**

Sentiment analysis is a part of computational research that classifies the polarity of a given text at the document, sentence, or feature level, whether the opinion expressed in the document, sentence, or feature is positive or negative. The ease of internet access and the development of increasingly sophisticated gadgets have made this electronic device one of the main needs of the community to access information. Recently, social media has become very interesting, both as a data source and a means of conveying insights and everyday life. Various judgments or opinions of the different Bandarlampung people can be found on social media and Twitter, and one of the popular topics of the people of Bandarlampung is leadership style. Leadership style is cooperative behavior and energy, a combination of reasoning, abilities, attributes, and practices applied by a leader to influence the performance of his followers. The leadership style that has recently garnered attention and discussion is the leadership style of Hj. Eva Dwiana, SE is the Mayor of Bandarlampung for the period 2021-2026. This study compares two algorithms: Support Vector Machine (SVM) and Naïve Bayes, in sentiment analysis using data from the Mayor of Bandarlampung on Twitter. Based on the research results, the Support Vector Machine (SVM) shows the best performance with an accuracy value of 87.38%, compared to Naïve Bayes with an accuracy value of 85.28%.

**Keywords:** Sentiment Analysis, Support Vector Machine (SVM), Naive Bayes, Mayor's Leadership

***SENTIMENT ANALYSIS SATU TAHUN KEPEMIMPINAN WALIKOTA  
BANDARLAMPUNG PERIODE 2021-2026 MENGGUNAKAN  
ALGORITME SUPPORT VECTOR MACHINE (SVM) DAN NAÏVE BAYES***

Oleh

*Elshinta Milenia*

**Skripsi**

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

Pada

Jurusan Ilmu Komputer  
Fakultas Matematika dan Ilmu Pengetahuan Alam  
Universitas Lampung



**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS LAMPUNG  
BANDAR LAMPUNG  
2022**

Judul Skripsi : **SENTIMENT ANALYSIS SATU TAHUN  
KEPEMIMPINAN WALIKOTA BANDARLAMPUNG  
PERIODE 2021-2026 MENGGUNAKAN  
ALGORITME SUPPORT VECTOR  
MACHINE (SVM) DAN NAÏVE BAYES**

Nama Mahasiswa : **Elshinta Milenia**

Nomor Induk Mahasiswa : **1857051012**

Jurusan : **Ilmu Komputer**

Fakultas : **Matematika dan Ilmu Pengetahuan Alam**



**Prof. Admi Syarif, Ph.D.**  
NIP 19670103 199203 1 003

**Dewi Asiah Shofiana, S.Komp., M.Kom.**  
NIP 19950929 202012 2 030

2. Ketua Jurusan Ilmu Komputer

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



**MENGESAHKAN**

**1. Tim Penguji**

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

**Penguji I  
Sekretaris : Dewi Asiah Shofiana, S.Komp., M.Kom.**

**Penguji II  
Bukan Pembimbing : Dr. Ir. Kurnia Muludi, M.S.Sc.**



---

---

---

**2. Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam**



**Dr Eng. Surtpto Dwi Yuwono, S.Si., M.T.**  
NIP 19740705 200003 1 001

**Tanggal Lulus Ujian Skripsi : 11 Agustus 2022**

## PERNYATAAN

Saya yang bertanda tangan di bawah ini:

Nama: Elshinta Milenia

NPM: 1857051012

Dengan ini menyatakan bahwa skripsi saya yang berjudul “*SENTIMENT ANALYSIS SATU TAHUN KEPEMIMPINAN WALIKOTA BANDARLAMPUNG PERIODE 2021-2026 MENGGUNAKAN ALGORITME SUPPORT VECTOR MACHINE (SVM) DAN NAÏVE BAYES*” adalah benar hasil karya sendiri dan bukan orang lain. Seluruh tulisan yang tertuang dalam skripsi ini telah mengikuti kaidah penulisan karya ilmiah Universitas Lampung. Jika dikemudian hari terbukti skripsi saya adalah hasil penjiplakan atau dibuat oleh orang lain, maka saya bersedia menerima sanksi sesuai dengan peraturan yang berlaku.

Bandar Lampung, 17 Agustus 2022

Penulis



Elshinta Milenia

NPM. 1857051012

## RIWAYAT HIDUP



Penulis dilahirkan di Bandarlampung pada tanggal 05 September 2000, sebagai anak kedua dari tiga bersaudara, dari Ayah Akhmad Sapri, S.Kep., M.H. dan Ibu Eliya Afrita. Penulis menyelesaikan pendidikan formal pertama kali di Taman Kanan-Kanak (TK) Al-Bustan Bandarlampung pada tahun 2005. Pendidikan Sekolah Dasar (SD) di SD Negeri 2 PWK diselesaikan pada tahun 2012, Sekolah Menengah Pertama (SMP) di MTsN 2 Bandarlampung pada tahun 2015, dan Sekolah Menengah Atas (SMA) di SMA Negeri 15 Bandarlampung pada tahun 2018.

Pada tahun 2018, penulis terdaftar sebagai mahasiswa Jurusan Ilmu Komputer Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung melalui jalur SMM PTN-BARAT. Selama menjadi mahasiswa beberapa kegiatan yang dilakukan penulis antara lain:

1. Menjadi anggota Adapter Himpunan Mahasiswa Jurusan Ilmu Komputer pada periode 2018/2019.
2. Menjadi anggota Eksternal Himpunan Mahasiswa Jurusan Ilmu Komputer pada periode 2018/2019.
3. Menjadi Asisten Dosen Jurusan Ilmu Komputer untuk mata kuliah Matematika Diskrit pada periode semester genap tahun ajaran 2020/2021.
4. Pada bulan Februari 2021 sampai dengan bulan Maret 2021 penulis melaksanakan Kerja Praktik di Dinas Bina Marga dan Bina Konstruksi Provinsi Lampung.
5. Pada bulan September 2021 penulis melaksanakan kegiatan Kuliah Kerja Nyata (KKN) di Desa Sukarame II, Kecamatan Teluk Betung Barat, Kabupaten Bandarlampung.



## MOTTO

"Cukuplah Allah bagiku, tidak ada Tuhan selain Dia. Hanya kepadaNya aku bertawakal."

**(At Taubah: 129)**

"Seseorang bertindak tanpa ilmu ibarat bepergian tanpa petunjuk. Dan sudah banyak yang tahu kalau orang seperti itu kiranya akan hancur, bukan selamat."

**(Hasan Al Bashri)**

"Tidaklah suatu kegalauan, kesedihan, kebimbangan, kekalutan yang menimpa seorang mukmin atau bahkan tertusuk duri sekalipun, melainkan karenanya Allah akan menggugurkan dosa-dosanya".

**(HR Bukhari dan Muslim)**

## PERSEMBAHAN

### *Alhamdulillahirobbilalamin*

Puji syukur kepada Allah SWT yang telah memberikan rahmat dan hidayah-Nya sehingga dapat menyelesaikan penulisan skripsi ini. Sholawat dan salam saya sanjungkan kepada Nabi Muhammad SAW.

Aku persembahkan karya ini kepada:

### **Ayah dan Ibu**

Sebagai tanda terimakasihku kepada Ayah dan Ibuku yang tercinta dan yang tersayang. Terima kasih telah mendidik dan membesarkanku dengan kasih sayang kalian. Terima kasih selalu mendukungku dan mendoakanku dalam segala pilihanku. Terima kasih atas semua pengorbanan, perjuangan kalian yang tiada hentinya. Terima kasih Ayah dan Ibu.

### **Kakakku dan Adikku**

Terima kasih telah memberikan semangat, dukungan, dan doa.

**Seluruh Keluarga Besar, Sahabat, dan Teman-teman** yang selalu memberikan semangat dan dukungan.

**Almamater Tercinta, Universitas Lampung**

## SANWACANA

Puji syukur kehadirat Allah SWT, karena telah memberikan rahmat dan hidayahNya kepada saya sehingga saya dapat menyelesaikan skripsi dengan judul “*Sentiment Analysis* Satu Tahun Kepemimpinan Walikota Bandarlampung Periode 2021-2026 Menggunakan Algoritme *Support Vector Machine* (SVM) Dan *Naïve Bayes*” dengan baik.

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

1. Ayah dan Ibu tercinta, Akhmad Sapri, S.Kep., M.H. dan Ibu Eliya Afrita yang selalu memberi dukungan, memotivasi, dan menyemangati saya selama proses perkuliahaan sampai dengan penyusunan skripsi. Semoga Allah SWT selalu menyertai, memberkati, dan memberi kesehatan dan kebahagiaan yang berlimpah.
2. Bapak Prof. Admi Syarif, Ph.D. selaku pembimbing utama yang telah meluangkan banyak waktu dan dengan sabar membimbing saya, serta memberikan banyak dukungan, motivasi, dan dorongan untuk menyelesaikan skripsi ini. Terima kasih juga saya ucapkan atas kritik dan saran yang membangun sehingga penulisan skripsi ini dapat diselesaikan.
3. Ibu Dewi Asiah Shofiana, S. Komp., M. Kom. selaku pembimbing kedua yang telah membimbing saya dalam memberikan ide, kritik, dan saran sehingga penulisan skripsi ini dapat diselesaikan dengan baik.
4. Bapak Dr. Ir. Kurnia Muludi, M.S.Sc. selaku pembahas yang telah memberikan masukan yang sangat bermanfaat dalam penulisan dan perbaikan skripsi ini.
5. Bapak Didik Kurniawan, S.Si., M.T. sebagai pembimbing akademik dan selaku

Ketua Jurusan Ilmu Komputer Universitas Lampung yang telah membimbing penulis selama proses perkuliahan, serta memberikan masukan dan dukungan hingga skripsi ini dapat diselesaikan.

6. Bapak Dr. Eng. Suropto Dwi Yuwono, S.Si., M.T. selaku Dekan FMIPA Universitas Lampung.
7. Bapak Dr. rer. nat. Akmal Junaidi, M. Sc. selaku Sekretaris Jurusan Ilmu Komputer FMIPA Universitas Lampung.
8. Bapak dan Ibu Dosen Jurusan Ilmu Komputer FMIPA Universitas Lampung yang tak bisa disebutkan satu per satu, atas bimbingan dan pengajarannya selama penulis menjadi mahasiswa FMIPA Universitas Lampung.
9. Seluruh staf dan karyawan Fakultas MIPA Universitas Lampung: Ibu Ade Nora Maela, Bang Zainuddin, Mas Sam, Mas Ardi Novalia, dan lainnya yang tidak bisa penulis sebutkan satu per satu, yang telah membantu segala urusan administrasi penulis.
10. Saudara-saudari tersayang, M. Alfero Kesuma, Elsaria Safira, dan Hilda Indhira yang senantiasa memberikan semangat, doa, dan dukungan selalu selama perkuliahan berlangsung.
11. M. Syahputra Prima Negara, S.Kom. seseorang yang selalu memberikan semangat dukungan dan doa sehingga penulis selalu semangat selama penyusunan skripsi.
12. Teman-teman terdekat, Amara Indah Pancarani, Yasmin Hasna, Atika Indah, Arafia Isnayu Akaf, Arfina Shella Meilany, Hendun Naura Syifa, Nadya Aristiawati Sitorus, Amira Inastia, Mustika Rahmi, Elfina Aprilia, Sasya Salsabilla, Rahmayanti Kurniasih, Lita Amelia, Ahmad Julio Rizki, Arsyi Sobirin, dan Muhamad Surya Akbar yang telah memberikan banyak dukungan moril, segala bentuk bantuan, dan selalu menemani dari awal perkuliahan.
13. Keluarga Ilmu Komputer 2018 serta kakak tingkat dan adik tingkat yang tidak bisa penulis sebutkan satu per satu.
14. Almamater tercinta, Universitas Lampung yang sudah memberi banyak wawasan dan pengalaman berharga.



Semoga skripsi ini dapat berguna dan bermanfaat bagi agama, masyarakat, bangsa, negara, para mahasiswa, akademisi, serta pihak-pihak lain yang membutuhkan terutama penulis. Saran dan kritik yang bersifat membangun sangat diharapkan. Akhir kata penulis ucapkan terima kasih. Semoga Allah SWT senantiasa memberikan perlindungan dan kebaikan bagi kita semua.

Bandar Lampung, 17 Agustus 2022

Penulis,

Elshinta Milenia

NPM. 1857051012

## DAFTAR ISI

<b>DAFTAR ISI</b> .....	<b>ii</b>
<b>DAFTAR GAMBAR</b> .....	<b>iv</b>
<b>DAFTAR TABEL</b> .....	<b>v</b>
<b>DAFTAR KODE PROGRAM</b> .....	<b>vi</b>
<b>I. PENDAHULUAN</b> .....	<b>1</b>
A. Latar Belakang.....	1
B. Rumusan Masalah.....	5
C. Batasan Masalah .....	5
D. Tujuan Penelitian.....	5
E. Manfaat Penelitian .....	6
<b>II. TINJAUAN PUSTAKA</b> .....	<b>7</b>
A. <i>Text mining</i> .....	7
B. <i>Supervised &amp; Unsupervised Learning</i> .....	8
C. Analisis sentimen.....	9
D. Twitter .....	10
E. Python .....	11
F. <i>Word2vec</i> .....	13
G. <i>K-Fold Cross Validation</i> .....	15

H. <i>Jupyter Notebook</i> .....	18
I. <i>Support Vector Machine (SVM)</i> .....	18
J. <i>Naïve Bayes</i> .....	21
K. Kepemipinan Walikota .....	22
L. Penelitian Terdahulu .....	24
<b>III. METODOLOGI PENELITIAN .....</b>	<b>28</b>
A. Waktu dan Tempat Penelitian .....	28
B. Perangkat Penelitian. ....	28
1. Perangkat Keras .....	28
2. Perangkat Lunak .....	28
C. Alur Penelitian .....	29
<b>IV. HASIL DAN PEMBAHASAN.....</b>	<b>35</b>
A. Pengumpulan Data .....	35
B. Tahapan Percobaan .....	39
1. <i>Import Data</i> .....	40
2. <i>Preprocessing</i> .....	40
3. Pembagian Data .....	41
4. <i>Word2vec</i> .....	44
5. Klasifikasi <i>Naïve Bayes</i> .....	45
6. Klasifikasi <i>Support Vector Machine (SVM)</i> .....	45
C. Normalisasi Teks.....	46
D. Tokenisasi .....	47
E. Pembuangan <i>Stopword</i> .....	48
F. Hasil <i>Word2vec</i> .....	49
G. <i>Training</i> Klasifikasi .....	51
H. Evaluasi Hasil Klafisifikasi.....	51

<b>V. SIMPULAN DAN SARAN .....</b>	<b>67</b>
A. Simpulan .....	67
B. Saran.....	68
<b>DAFTAR PUSTAKA .....</b>	<b>68</b>
<b>LAMPIRAN.....</b>	<b>70</b>



## DAFTAR GAMBAR

Gambar	Halaman
1. Ilustrasi <i>Supervised Learning</i> (Roihan et al., 2020).....	8
2. Ilustrasi <i>Unsupervised Learning</i> (Roihan et al., 2020) .....	9
3. Contoh <i>K-Fold Cross Validation</i> (Refaeilazadeh et al., 2008) .....	15
4. Contoh <i>Hold-Out Cross Validation</i> (Yadav & Shukla, 2016). .....	17
5. Contoh <i>Leave One-Out Cross Validation</i> .....	17
6. SVM berusaha menemukan <i>hyperplane</i> terbaik yang memisahkan kedua <i>class</i> – 1 dan +1 (Nugroho et al., 2003) .....	18
7 . Tahapan penelitian .....	29
8. Hasil <i>labelling</i> manual .....	36
9. <i>Dataset</i> yang digunakan untuk penelitian .....	36
10. Hasil normalisasi teks .....	47
11. Hasil dari tahap tokenisasi .....	48
12. Hasil dari pembuangan <i>Stopword</i> .....	49
13. Nilai vektor dari hasil latih <i>word2vec</i> .....	50
14. Hasil perbandingan akurasi <i>Naïve Bayes</i> dan SVM.....	58
15. Hasil perbandingan presisi <i>Naïve Bayes</i> dan SVM.....	59
16. Hasil perbandingan <i>recall Naïve Bayes</i> dan SVM.....	60
17. Hasil perbandingan <i>f1-score Naïve Bayes</i> dan SVM .....	61

## DAFTAR TABEL

<b>Tabel</b>	<b>Halaman</b>
1. Jenis <i>kernel</i> pada <i>Support Vector Machine</i> .....	20
2. Penelitian terdahulu.....	25
3. Tabel <i>confusion matrix</i> .....	33
4. Pengumpulan data Twitter .....	35
5. Sentimen positif .....	37
6. Sentimen negatif.....	39
7. Pembagian data latih dan data uji .....	41
8. Pembagian data latih dan data uji .....	43
9. Hasil pengujian berdasarkan <i>kernel</i> .....	51
10. Kinerja metode <i>Support Vector Machine (SVM)</i> .....	53
11. <i>Confusion matrix Support Vector Machine (SVM)</i> .....	53
12. Kinerja metode <i>Naive Bayes</i> .....	54
13. <i>Confusion matrix Naive Bayes</i> .....	54
14. Perbandingan kinerja klasifikasi <i>Naive Bayes</i> dan <i>Support Vector Machine (SVM)</i> .....	56
15. <i>Confusion matrix</i> Skenario 1.....	62
16. <i>Confusion matrix</i> Skenario 2.....	62
17. <i>Confusion matrix</i> Skenario 3.....	63
18. <i>Confusion matrix</i> Skenario 4.....	63
19. <i>Confusion matrix</i> Skenario 5.....	63
20. Hasil uji coba pada <i>Naive Bayes</i> dan <i>Support Vector Machine (SVM)</i> .....	64

## DAFTAR KODE PROGRAM

<b>Kode Program</b>	<b>Halaman</b>
1. Memasukkan data ke Python .....	40
2. Tahapan <i>Preprocessing</i> .....	41
3. Kode pembagian data .....	42
4. <i>Vocab size Word2vec</i> .....	44
5. <i>Total words Word2vec</i> .....	44
6. Model <i>Naïve Bayes</i> .....	45
7. <i>Confusion matrix Naïve Bayes</i> .....	45
8. Model <i>Support Vector Machine (SVM)</i> .....	46
9. <i>Confusion matrix SVM</i> .....	46

## I. PENDAHULUAN

### A. Latar Belakang

Teknologi informasi saat ini berkembang pesat. Kemudahan akses internet dan perkembangan gadget yang semakin canggih menjadikan perangkat elektronik ini sebagai salah satu kebutuhan utama masyarakat untuk mengakses informasi. Berbagai kemudahan dalam mengakses informasi melalui internet, mulai dari sekedar membaca berita, mencari referensi tugas, mencari tutorial, hingga berbagi aktivitas sehari-hari melalui media sosial. Media sosial akhir-akhir ini menjadi sesuatu yang sangat menarik, baik sebagai sumber data, maupun sebagai sarana penyampaian wawasan dan kehidupan sehari-hari. Penilaian yang berbeda atau berbagai opini masyarakat Bandarlampung dapat ditemukan di media sosial, dan salah satu bahasan masyarakat Bandarlampung yang populer yaitu gaya kepemimpinan.

Kepemimpinan adalah tindakan seorang pemimpin dalam menggerakkan atau memengaruhi orang lain/bawahan dengan penuh kesadaran dan kewajiban untuk mencapai tujuan yang telah ditetapkan bersama. Salah satu tugas kepala daerah adalah sebagai pemimpin karena kepemimpinan merupakan sifat dan karakter yang membedakan pemimpin dan pengikutnya. Kepemimpinan sebagai perbuatan atau kegiatan pemimpin dan kepemimpinan yang diidentikkan dengan konsekuensi pelaksanaan dan kepuasan kerja. Kepemimpinan yang efektif adalah hasil dan perpaduan yang tepat antara keadaan yang menyenangkan dan gaya kepemimpinan (Tuti & Adawiyah, 2020).

Gaya kepemimpinan adalah perilaku dan energi kooperatif, yang merupakan perpaduan antara penalaran, kemampuan, atribut, dan praktik yang diterapkan oleh seorang pemimpin untuk memengaruhi kinerja para pengikutnya (Tuti &



Adawiyah, 2020).

Gaya kepemimpinan yang akhir-akhir ini menyita perhatian dan perbincangan adalah mengenai gaya kepemimpinan Hj. Eva Dwiana, S.E atau yang lebih dikenal sebagai Bunda Eva. Beliau adalah istri dari Walikota Bandarlampung periode 2010-2015 dan 2016-2021 yaitu Herman HN. Sebelum menjabat sebagai Walikota Bandarlampung, Hj. Eva Dwiana, S.E lebih dikenal masyarakat sebagai pimpinan Majelis Taklim Rachmat Hidayat. Nama Eva cukup dikenal di kalangan ibu-ibu pengajian karena sering menggelar kegiatan sosial.

Karier politik Eva Dwiana cukup bagus, mulai dari Ketua Dewan Pimpinan Cabang (DPC) Demokrat Kota Bandarlampung, hingga menjabat sebagai Dewan Perwakilan Daerah (DPD) Partai Demokrasi Indonesia Perjuangan (PDIP) Lampung pada konferda 8 Maret 2015. Beliau menjadi Ketua Bidang Perempuan dan Anak Dewan Perwakilan Daerah (DPD) Partai Demokrasi Indonesia Perjuangan (PDIP) Lampung. Eva Dwiana juga merupakan anggota Dewan Perwakilan Rakyat Daerah (DPRD) Provinsi Lampung. Beliau tercatat sebagai Komisi III Dewan Perwakilan Rakyat Daerah (DPRD) Provinsi Lampung setelah meraih suara terbanyak yaitu sebesar 19.818 suara, Eva Dwiana meraih urutan kedua suara terbanyak dari para caleg 12 parpol pada Pemilu 2014 untuk Dewan Perwakilan Rakyat Daerah (DPRD) Lampung dari Kota Bandarlampung.

Pada tahun 2019, Eva kembali memperoleh suara terbanyak. Bahkan separuh suara Partai Demokrasi Indonesia Perjuangan (PDIP) Bandarlampung menduduki kursi Dewan Perwakilan Rakyat Daerah (DPRD) Provinsi Lampung. Hasil rekapitulasi suara di Komisi Pemilihan Umum (KPU) Bandarlampung, suara Eva Dwiana mencapai 86.258 dari suara partai sebanyak 146.294. Tingginya perolehan suara Eva Dwiana, saat pleno Komisi Pemilihan Umum (KPU) Lampung berakhir, empat caleg Demokrasi Indonesia Perjuangan (PDIP) dari Daerah Pemilihan (Dapil) 1 Bandarlampung duduk di Dewan Perwakilan Rakyat Daerah (DPRD) Lampung yaitu, Kostiana (13.437), AR Suparno (4.714), dan Apriliaty (4.304). Selanjutnya Eva Dwiana pun

mencalonkan diri di Pemilihan Walikota Bandarlampung 2020 yang merupakan bagian dari Pemilihan Kepala Daerah (PILKADA) serentak 2020. Eva Dwiana mencalonkan diri sebagai Walikota didampingi Deddy Amarullah sebagai Calon Wakil Walikota 2020. Mereka mendapat nomor urut tiga pada tanggal 9 Desember 2020, pasangan Eva-Deddy unggul berdasarkan hitung cepat (Maulida & Kurniawan, 2021).

Setiap orang memiliki jenis atau model gaya kepemimpinan yang berbeda-beda, yang secara langsung akan memengaruhi hubungan pemimpin dan masyarakatnya dalam suatu wilayah tertentu. Ada banyak media sosial yang populer dan banyak diminati masyarakat saat ini, diantaranya adalah Instagram, Facebook, dan Twitter. Twitter merupakan media komunikasi yang dapat digunakan masyarakat untuk berekspresi dan lebih diminati oleh masyarakat dunia. Di Twitter ada istilah *tweet* yaitu pengguna Twitter dapat memberikan kabar terbaru, berekspresi, beraspirasi, dan beropini yang ditulis oleh pengguna Twitter lainnya terutama topik atau hal yang sedang menjadi perbincangan utama (Darwis et al., 2020).

Keanekaragaman *tweet* tersebut serta banyaknya penggunaan bahasa yang tidak baku pada *tweet* menjadi alasan diperlukan analisis sentimen. Analisis sentimen merupakan daerah penelitian perhitungan untuk mengekstraksi polaritas pendapat antar kelas (positif dan negatif) dari dokumen teks (Antinasari et al., 2017). Analisis sentimen juga berfokus pada pengamatan opini yang mengandung polaritas, yang memiliki nilai opini positif atau negatif. Isu-isu yang ada dalam pemeriksaan opini biasanya sulit untuk dikarakterisasi, dicirikan oleh ide-ide isu dan tujuan yang berfungsi sebagai kerangka kinerja dalam berbagai penelitian.

Analisis sentimen dalam pengumpulan data statistik menggunakan berbagai algoritme yang berasal dari cabang ilmu *Artificial Intelligence*, misalnya perhitungan *Deep Learning* dan *Machine Learning* yang terdiri dari *Naive Bayes*, *Support Vector Machine (SVM)*, *Artificial Neural Network (ANN)*, regresi, dan lain-lain (Roihan et al., 2020). Penelitian ini akan menerapkan algoritme *Support Vector Machine (SVM)* dan *Naive Bayes* untuk

menganalisis kecenderungan informasi mengenai suatu topik apakah cenderung positif atau negatif dan mengklasifikasikan data dengan topik *tweet* mengenai gaya kepemimpinan Walikota Bandarlampung dari media sosial Twitter. Algoritme *Support Vector Machine* (SVM) ini menghasilkan model klasifikasi yang cukup baik meskipun dilatih dengan menggunakan himpunan data yang sedikit dan merupakan model yang berasal dari hipotesis pembelajaran faktual yang akan memberikan hasil yang lebih baik dibandingkan menggunakan metode lainnya (Darwis et al., 2020). Kelebihan SVM dibandingkan metode yang lain terletak pada kemampuannya untuk menemukan *hyperplane* terbaik yang memisahkan dua buah *class* pada *feature space* yang ditunjang oleh strategi *Structural Risk Minimization* (SRM) (Nugroho et al., 2003).

Dalam penelitian ini, *classifier* lain yang dipilih adalah *Naive Bayes*, ciri utama dari algoritme *Naive Bayes* adalah asumsi yang sangat kuat (naif) akan independensi dari masing-masing kondisi atau kejadian. Kelebihan dari *Naive Bayes* adalah proses klasifikasi data dapat disesuaikan dengan sifat dan kebutuhan masing-masing (Edukasi et al., 2018). *Naive Bayes* adalah sebuah metode sederhana yang dikembangkan berdasarkan aturan Bayes dengan melihat kondisi-kondisi yang ada dan peluang-peluang setiap kondisinya. Penelitian sebelumnya mengenai analisis sentimen dengan menggunakan *Naive Bayes* sebagai *clasifier*-nya telah banyak dilakukan, diantaranya adalah penelitian menggunakan data Twitter berjudul “Analisis Sentimen Tentang Opini Film pada Dokumen Twitter Berbahasa Indonesia Menggunakan *Naive Bayes* dengan Perbaikan Kata Tidak Baku” (Antinasari et al., 2017).

Dari permasalahan yang telah diuraikan, pada penelitian ini akan dianalisis data *tweet* hasil pencarian dengan kata kunci “Walikota Bandarlampung Eva Dwiana”, kemudian data tersebut akan diklasifikasi menjadi dua kelas sentimen, yakni positif atau negatif dengan menggunakan dua metode yaitu algoritme *Support Vector Machine* (SVM) dan *Naive Bayes*, lalu akan dilakukan perbandingan akurasi, presisi, *recall*, dan *f1-score* antara dua algoritme tersebut. Diharapkan penelitian ini dapat membantu pihak-pihak

yang membutuhkan informasi mengenai gaya kepemimpinan Walikota Bandarlampung dan untuk menjadi pembelajaran kinerja yang lebih baik untuk Walikota Bandarlampung di periode selanjutnya atau pada masa yang akan datang.

## **B. Rumusan Masalah**

Rumusan masalah pada penelitian ini adalah bagaimana mengklasifikasi sentimen negatif dan sentimen positif terhadap data yang didapatkan dari Twitter dengan menggunakan algoritme *Support Vector Machine* (SVM) dan *Naive Bayes*.

## **C. Batasan Masalah**

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

1. Penelitian ini menggunakan studi kasus mengenai gaya kepemimpinan Walikota Bandarlampung.
2. Proses pembuatan model menggunakan data *tweet* menggunakan Twitter API yang merupakan *library* untuk menarik *tweet*.
3. Data *tweet* yang digunakan hanya *tweet* yang menyebutkan kata kunci yang berhubungan dengan gaya kepemimpinan Walikota Bandarlampung.
4. *Tweet* yang digunakan berbahasa Indonesia.
5. Algoritme klasifikasi yang digunakan adalah *Support Vector Machine* (SVM) dan *Naive Bayes*.

## **D. Tujuan Penelitian**

Tujuan penelitian yang akan dicapai sebagai berikut:

1. Mengimplementasikan algoritme *Support Vector Machine* (SVM) dan

*Naive Bayes* untuk melakukan sentimen analisis terhadap data dari Twitter tentang kepemimpinan Walikota Bandarlampung.

2. Menguji dan membandingkan hasil akurasi, presisi, *recall*, dan *f1-score* algoritme *Support Vector Machine* (SVM) dan *Naive Bayes* pada analisis sentimen.
3. Membandingkan performansi kecepatan algoritme *Support Vector Machine* (SVM) dan *Naive Bayes*.

### **E. Manfaat Penelitian**

Manfaat yang ingin dicapai dalam penelitian ini yaitu:

1. Untuk membantu pihak-pihak yang membutuhkan informasi mengenai gaya kepemimpinan Walikota Bandarlampung.
2. Untuk menjadi pembelajaran kinerja yang lebih baik untuk Walikota Bandarlampung di periode selanjutnya atau pada masa yang akan datang.
3. Untuk mengetahui keunggulan antara dua algoritme (*Support Vector Machine* dan *Naive Bayes*) dalam melakukan analisis sentimen.

## II. TINJAUAN PUSTAKA

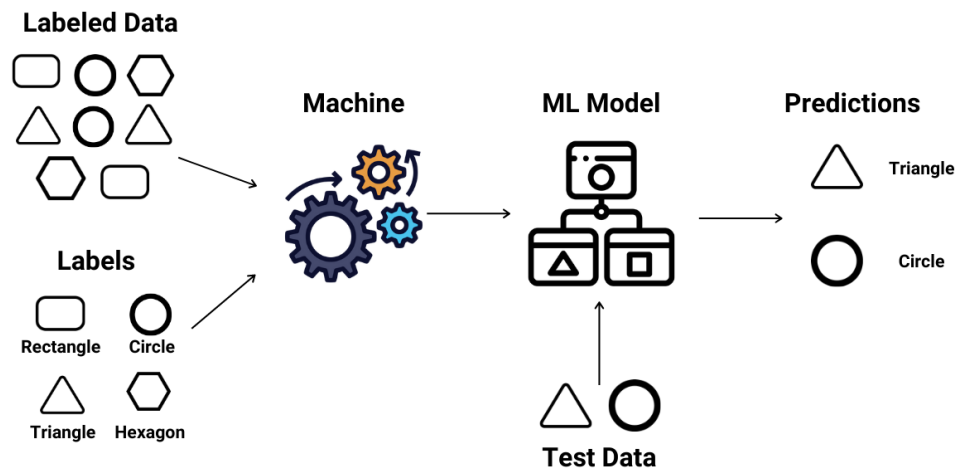
### A. *Text mining*

*Text mining* adalah subjek penelitian yang sangat baru dan mulai menjadi penting dan digunakan banyak orang. Dalam menangani masalah, *text mining* umumnya digabungkan dengan beberapa subjek yang berbeda seperti *Data Mining*, *Natural Language Processing*, dan lain-lain. Dalam *text mining*, terdapat tahapan-tahapan, misalnya memisahkan teks dengan menggunakan metode tertentu, pemrosesan teks yang biasa disebut *text*, pembobotan atau pengurutan teks, seperti halnya menganalisis sebuah teks (Pravina & Adikara, 2019).

*Text mining* adalah rangkaian data, sosial, dan realitas yang tersembunyi di dalam teks ketika dilakukan pemrosesan dan analisis data dalam jumlah besar, Struktur teks yang membingungkan dan tidak memadai, aspek tinggi, dan informasi yang *noise*. Ada begitu banyak kegunaan *text mining* yang diperlukan dalam kehidupan sehari-hari (Pravina & Adikara, 2019). Analisis teks melibatkan pencarian informasi, analisis makna kata untuk mempelajari distribusi frekuensi kata, pengenalan pola, penandaan / anotasi, ekstraksi informasi, teknik penambangan data, termasuk analisis tautan dan sosiasi, visualisasi, dan analitik prediktif. Tujuannya adalah untuk mengubah teks menjadi data untuk dianalisis. Untuk data yang dianalisis dapat melalui aplikasi *Natural Language Processing* (NLP), berbagai jenis algoritme, dan metode analitik. Fase penting dari proses ini adalah penafsiran informasi (Pravina & Adikara, 2019).

## B. Supervised & Unsupervised Learning

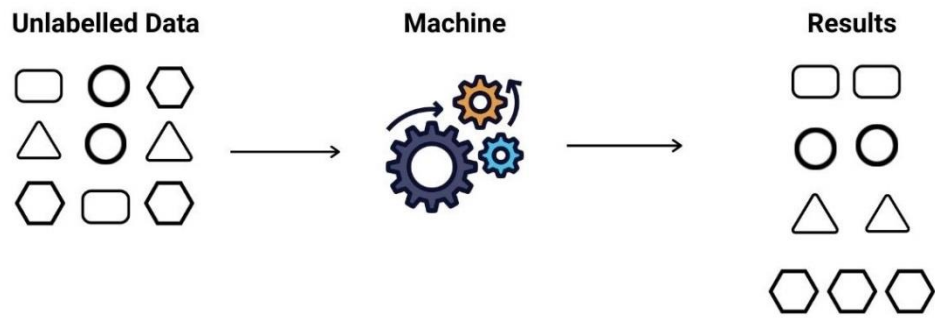
*Supervised learning* adalah suatu pembelajaran yang terarah ataupun terawasi, dengan tujuan untuk melakukan klasifikasi (*classification*).



Gambar 1. Ilustrasi *Supervised Learning* (Roihan et al., 2020)

Pada Gambar 1 *supervised learning* adalah metode klasifikasi di mana kumpulan data sepenuhnya diberikan label untuk mengklasifikasikan kelas yang tidak dikenal (Roihan et al., 2020). Dengan kata lain, cara kerja *supervised learning* adalah terdapat data yang dilatih dan terdapat variabel yang ditargetkan dengan tujuan untuk mengelompokkan suatu data ke data yang sudah ada. Beberapa contoh model yang digunakan pada *supervised learning* yaitu *naive bayes*, KNN, SVM, dan *random forest*.

*Unsupervised learning* adalah suatu pembelajaran yang tidak terarah ataupun tidak diawasi, dengan tujuan untuk melakukan *clustering*.



Gambar 2. Ilustrasi *Unsupervised Learning* (Roihan et al., 2020)

Pada Gambar 2 tersebut dalam jenis pembelajaran *unsupervised learning*, sistem disediakan dengan beberapa input sampel tetapi tidak ada output yang hadir. Karena tidak ada output yang diinginkan di sini kategorisasi dilakukan sehingga algoritma membedakan dengan benar antara kumpulan data. Ini adalah tugas mendefinisikan fungsi untuk menggambarkan struktur yang tersembunyi dari data yang tidak berlabel, maka dari itu *unsupervised learning* harus mencari distribusi asli data  $q(x)$  berdasarkan beberapa sampel data. Definisi lain dari *unsupervised learning* adalah mengingat persamaan yang terdapat *input* dan parameter (Roihan et al., 2020). Salah satu contoh model yang digunakan pada *supervised learning* yaitu *K-Means*.

### C. Analisis sentimen

Analisis sentimen dilakukan dengan dua cara, pertama dengan pelabelan otomatis mencocokkan kamus data yang disediakan berisi kata positif dan negatif dengan kalimat untuk dianalisis secara sentimen, kedua dengan pelabelan kalimat secara manual. Alur proses analisis sentimen untuk mendapatkan sentimen positif dan negatif. Urutan proses pada sistem adalah memasukkan *tweet documents*, *sentiment analysis*, kemudian mengeluarkan hasil sentimen positif dan negatif. Berikut ini adalah contoh pelabelan otomatis dan pelabelan manual (Ardhiansyah et al., 2019).



a. *Automatic Labelling*

Imam sukses mencetak goal ke gawang arsenal

Kata **sukses** ada pada kamus data positif, maka kalimat tersebut menjadi kalimat positif.

b. *Manual Labelling*

Kiper arsenal dapat menepis tendangan benzema (Positif)

Analisis sentimen (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 (Ardhiansyah et al., 2019). Analisis sentimen secara luas diterapkan untuk memberikan informasi seperti ulasan dan tanggapan survei, media *online* dan sosial, dan bahan perawatan kesehatan untuk aplikasi yang berkisar dari pemasaran, layanan pelanggan hingga kedokteran klinis.

Tugas dasar dalam analisis sentimen adalah mengklasifikasikan polaritas teks yang diberikan pada tingkat dokumen, kalimat, atau fitur apakah opini yang diungkapkan dalam dokumen, kalimat atau fitur entitas adalah positif atau negatif (Ardhiansyah et al., 2019). Ada berbagai jenis analisis sentimen seperti analisis sentimen berbasis aspek, analisis sentimen penilaian (positif dan negatif), serta analisis sentimen multibahasa dan deteksi emosi. Pada penelitian ini, *knowledge-based techniques* dan *statistical methods* akan digunakan untuk membuat model sentimen karena dapat melakukan kategorisasi sentimen. Teknik klasifikasi *knowledge-based* dan *statistical methods* dapat melakukan kategori berdasarkan kemunculan kata dan menganalisis struktur bahasa.

## D. Twitter

Twitter adalah media sosial yang memungkinkan klien untuk menawarkan sudut pandang dan sentimen mereka tentang banyak isu atau permasalahan

(Pravina & Adikara, 2019). Berbeda dengan media *online* lainnya, yang harus menjadi teman sebelum mereka dapat berkolaborasi, Twitter mengizinkan klien untuk tetap terkait walaupun mereka tidak saling berteman (Pravina & Adikara, 2019). Pada penelitian ini, Twitter dimanfaatkan sebagai media tempat pengunduhan informasi yang akan digunakan. Untuk pengambilan informasi akan memanfaatkan Twitter API yang merupakan perpustakaan Twitter untuk pemrograman yang dapat mengambil informasi dari Twitter.

## E. Python

Python merupakan salah satu bahasa pemrograman dengan banyak paradigma yang mendukung objek, prosedural, dan pemrograman berorientasi fungsional. Hal ini memungkinkan program untuk ditulis dalam beberapa pendekatan sekaligus. Misalnya, antarmuka grafis dibuat dalam bentuk orientasi objek, sedangkan pengolahannya dibuat dalam bentuk fungsional atau prosedural. Python diatur oleh Guido van Rossum dimulai pada akhir dekade 1980 di Centrum voor Wiskunde en Informatica, Belanda. Sejak itu, Python terkenal di kalangan komunitas pemrograman (Analysis et al., n.d.).

Python diketik dan dikumpulkan secara dinamis. Hal ini mendukung beberapa paradigma pemrograman, termasuk pemrograman prosedural, berorientasi objek, dan fungsional. Python sering digambarkan sebagai bahasa "termasuk baterai" karena pustaka standarnya yang komprehensif (Analysis et al., n.d.). Python merupakan bahasa pemrograman *multi-paradigm* dengan maksud untuk meraih berbagai jenis tujuan dengan metode yang beragam. Pemrograman berorientasi objek dan pemrograman terstruktur sepenuhnya didukung, dan banyak fitur-fiturnya mendukung pemrograman fungsional dan pemrograman berorientasi aspek termasuk oleh *metaprogramming* dan *metaobjects*. Banyak paradigma lain didukung melalui ekstensi, termasuk desain berdasarkan kontrak dan pemrograman logika (Analysis et al., n.d.).

Python yang digunakan yaitu versi 3.9. Ada beberapa *library* yang digunakan pada penulisan ini. Untuk *install library*, cukup menggunakan perintah `pip`.

**a. Tweepy**

*Tweepy* merupakan *library* Python untuk mengakses Twitter API dan mengambil data dari Twitter.

**b. TextBlob**

TextBlob merupakan *library* Python yang digunakan untuk memproses data tekstual. Textblob menyediakan API sederhana untuk mengerjakan tugas *Natural Language Processing* (NLP), antara lain:

1. Pemberian *tag* kata;
2. Ekstraksi kata benda;
3. Sentimen Analisis;
4. Klasifikasi;
5. Terjemahan; dan sebagainya.

**c. Pandas**

Pandas merupakan *library* Python untuk mengolah dan menganalisis data-data yang terstruktur. Untuk mengambil fungsi pandas dengan menuliskan “`import pandas as pd`”. Format yang sering digunakan pada pandas yaitu DataFrame.

**d. Numerical Python**

*Numerical Python* atau Numpy merupakan salah satu *library* Python yang berfungsi untuk melakukan proses komputasi dalam operasi vektor dan matriks.

**e. *Regular Expression***

*Regular Expression* merupakan *library* Python untuk mendefinisikan pola pencarian dengan pengenalan karakter untuk mendapatkan data yang sudah dimanipulasi sedemikian rupa, sehingga data menjadi bersih.

**f. *Matplotlib***

Matplotlib merupakan *library* Python untuk menampilkan hasil analisis berupa grafik dan *chart*.

**g. *Wordcloud***

*Wordcloud* merupakan *library* Python untuk menampilkan hasil analisis berupa gambar *wordcloud*. Dalam *library wordcloud* juga menyediakan *stopwords* yang berfungsi untuk menghapus kata-kata yang sering muncul (*common words*) namun kata-kata tersebut tidak memiliki arti untuk hasil analisis data.

**h. *Pillow***

*Pillow* merupakan *Python Imaging Library* (PIL) yang berfungsi untuk membuka, menggunakan dan menyimpan gambar.

**i. *CSV***

CSV (*Comma Separated Value*) merupakan *library* Python sebagai tempat untuk menyimpan data dengan format “.csv” (Nofiyanti et al., 2021).

**F. *Word2vec***

*Word2vec* adalah sekelompok model yang saling terkait yang digunakan untuk menghasilkan vektor kata. *Word2vec* mengambil kata-kata dari dokumen atau kalimat sebagai masukan kumpulan teks yang banyak dan menghasilkan ruang

vektor, biasanya beberapa ratus dimensi, dengan setiap kata unik dalam kumpulan tulisan ditugaskan vektor yang sesuai dalam ruang. Vektor kata diposisikan dalam ruang vektor sehingga kata-kata yang berbagi konteks umum dalam kumpulan tulisan terletak berdekatan satu sama lain dalam ruang tersebut (Mikolov et al., 2013).

*Word2vec* dibuat dan diterbitkan pada tahun 2013 oleh tim peneliti yang dipimpin oleh Tomas Mikolov di Google dan dipatenkan. Algoritme ini kemudian dianalisis dan dijelaskan oleh peneliti lain. Vektor kata yang dibuat menggunakan algoritme *Word2vec* memiliki banyak keunggulan dibandingkan dengan algoritme sebelumnya seperti *latent semantic analysis*. Keunggulan dari *word2vec* adalah memiliki ekstensi untuk menyimpan kata-kata dari seluruh kalimat yang disebut dengan *paragraf2vec* atau *doc2vec*. *Word2vec* juga telah diimplementasikan dalam beberapa bahasa pemrograman yaitu C, Python, dan Java yang dapat mendukung inferensi penyimpanan dokumen pada dokumen tak terlihat (Mikolov & Com, 2014).

*Word2vec* dapat menggunakan salah satu dari dua model arsitektur untuk menghasilkan representasi kata-kata yang terdistribusi: bagan kata-kata kontinu (CBOW) atau skip-gram kontinu. 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, 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 et al., 2013). Pada penelitian ini, *Word2vec* digunakan untuk penghitungan kemunculan kata dan menyimpan kata yang muncul dari keseluruhan data. *Word2vec* hanya melakukan penghitungan kata yang muncul dari setiap data dan yang paling sering muncul. *Word2vec* dapat membantu dalam pengkategorian nilai sentimen dari data.

## G. *K-Fold Cross Validation*

*Cross validation* adalah sebuah metode statistika yang digunakan untuk mengevaluasi dan membandingkan algoritme pembelajaran dengan membaginya menjadi dua bagian yaitu untuk melakukan pembelajaran dan pelatihan untuk model yang lain, dan juga digunakan dalam memvalidasi model (Refaeilazadeh et al., 2016). Berikut adalah jenis-jenis dari *cross validation*.

### 1. *K-Fold Cross Validation*

*K-fold cross validation* adalah teknik umum untuk memperkirakan kinerja pengklasifikasian. Proses *k-fold cross validation* dilakukan dengan tahapan pertama yaitu data akan dipartisi menjadi  $k$  yang sama dengan ukuran iterasi atau lipatan. Selanjutnya  $k$  dalam model pelatihan dan validasi dilakukan pengulangan (iterasi) sedemikian rupa sehingga dalam setiap iterasi, iterasi data yang berbeda tidak berlaku untuk dilakukan validasi (Refaeilazadeh et al., 2016). Berikut contoh *k-fold cross validation* yang dapat dilihat pada Gambar 3.



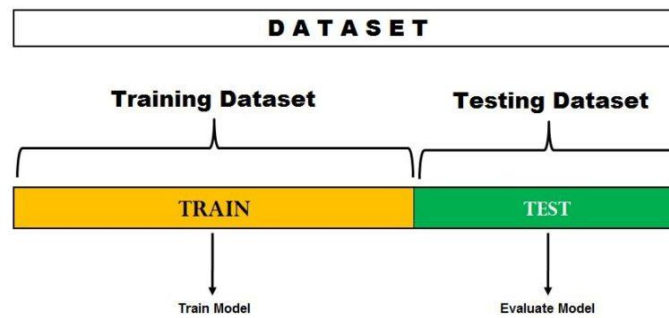
Gambar 3. Contoh *K-Fold Cross Validation* (Refaeilazadeh et al., 2016)

## 2. *Hold-Out Cross Validation.*

*Hold-out cross validation* adalah teknik validasi yang paling sederhana pada metode ini data dibagi secara acak menjadi dua bagian. Bagian ini disebut data pelatihan (*training*) dan data tes (*validation*), dengan rasio tipikal pembagian 60:40 atau 80:20. Teknik ini digunakan untuk melatih model menggunakan data pelatihan dan mengevaluasi model menggunakan data tes. Teknik evaluasi model yang digunakan pada data pelatihan dalam menghitung kesalahan bergantung dari jenis masalah pada penelitian, jika data pelatihan atau pengujian tidak dapat merepresentasikan data secara lengkap maka hasil data dapat menjadi tidak baik (Yadav & Shukla, 2016). Pada kondisi terbatasnya data yang digunakan untuk *training* dan *testing*, diperlukan metode untuk mendapatkan hasil tingkat akurasi dari sebuah metode pada *machine learning*. Salah satu cara untuk validasi adalah dengan menggunakan metode *holdout*. Metode *holdout* adalah metode yang akan menyediakan sejumlah data untuk digunakan sebagai data *testing*, dan sisanya sebagai data *training* (Nugraha, 2013).

Saat proses pengacakan data untuk dibagi sebagai data *training* dan *testing*, sangat mungkin terjadi *overrepresented* pada salah satu atau lebih klasifikasi. Dalam artian bahwa klasifikasi tersebut dominan dibandingkan klasifikasi lainnya, sehingga data *training* dan *testing* yang tercipta menjadi tidak representatif. Maka dari itu diperlukan prosedur *stratification holdout*, dimana dengan prosedur ini dapat dijamin bahwa setiap klasifikasi dapat terwakili pada data *training* dan *testing* yang tercipta secara proporsional (Nugraha, 2013).

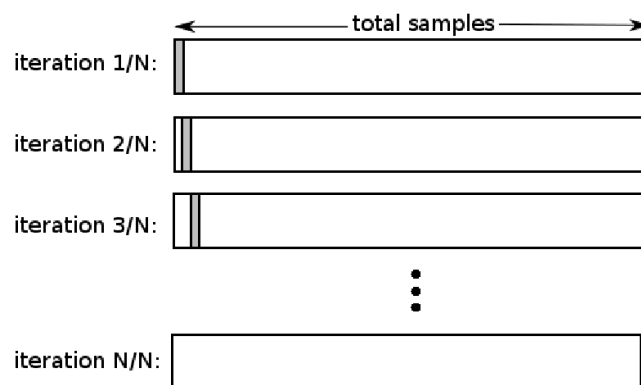
Dilakukan perulangan terhadap seluruh proses *training* dan *testing* beberapa kali dengan data *training* dan *testing* yang teracak. Kemudian diambil nilai rata-ratanya. Prosedur ini dikatakan sebagai *repeated holdout* (Nugraha, 2013). Berikut contoh *holdout cross validation* dapat dilihat pada Gambar 4.



Gambar 4. Contoh *Hold-Out Cross Validation* (Yadav & Shukla, 2016).

### 3. *Leave-One-Out Cross Validation*

*Leave-one-out cross validation* adalah kasus spesial *k-fold cross validation* dimana  $k$  sama dengan jumlah *instance* dalam data. Dengan kata lain disetiap iterasi hampir semua data kecuali untuk observasi data tunggal digunakan untuk pelatihan modelnya dan model yang digunakan diuji dengan pengamatan tunggal tersebut. Metode ini masih banyak digunakan ketika tersedia data yang sangat jarang, terutama pada bidang bioinformatika yang penerapannya masih menggunakan sampel data yang sedikit tersedia (Refaeilazadeh et al., 2016). Berikut contoh *leave-one-out cross validation* yang dapat dilihat pada Gambar 5.



Gambar 5. Contoh *Leave One-Out Cross Validation*

(Refaeilazadeh et al., 2016)

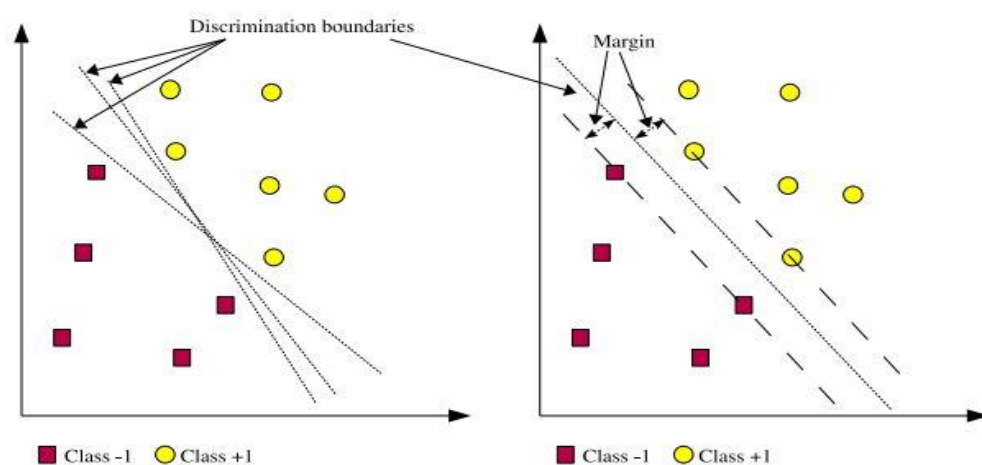


## H. Jupyter Notebook

Jupyter Notebook merupakan aplikasi *open source* yang digunakan sebagai tempat untuk membuat dan berbagi dokumen dari Python. Untuk dapat *install* Jupyter Notebook yang perlu dilakukan hanya cukup menggunakan perintah `pip`, sedangkan untuk mengaksesnya yang dilakukan hanya cukup dengan mengetikkan “Jupyter Notebook” di *command prompt* maka Jupyter Notebook akan dapat diakses dan digunakan (Nofiyanti et al., 2021).

## I. Support Vector Machine (SVM)

*Support Vector Machine* (SVM) secara resmi diperkenalkan oleh Cortes dan Vapnik pada tahun 1995 (Cortes dan Vapnik, 1995) dan telah terbukti menjadi salah satu perhitungan algoritme *supervised machine learning* yang paling banyak digunakan untuk tujuan klasifikasi (Hayati & Alifi, 2021). Metode *Support Vector Machine* (SVM) bertujuan untuk menemukan fungsi pemisah terbaik di antara fungsi yang ada untuk memisahkan dua macam objek (Hendrastuty et al., 2021).



Gambar 6. SVM berusaha menemukan *hyperplane* terbaik yang memisahkan kedua *class* -1 dan +1 (Nugroho et al., 2003)

Konsep *Support Vector Machine* (SVM) dapat dijelaskan secara sederhana sebagai usaha mencari *hyperplane* terbaik yang berfungsi sebagai pemisah dua buah *class* pada *input space*. Gambar 6 memperlihatkan beberapa *pattern* yang merupakan anggota dari dua buah *class* : +1 dan -1. *Pattern* yang tergabung pada *class* -1 disimbolkan dengan warna merah (kotak), sedangkan *pattern* pada *class* +1, disimbolkan dengan warna kuning (lingkaran) (Nugroho et al., 2003).

Masalah klasifikasi dapat diterjemahkan dengan usaha menemukan garis (*hyperplane*) yang memisahkan antara kedua kelompok tersebut. Berbagai alternatif garis pemisah (*discrimination boundaries*) yang dapat dilihat pada Gambar 6. *Hyperplane* pemisah terbaik antara kedua *class* dapat ditemukan dengan mengukur *margin hyperplane* tersebut dan mencari titik maksimalnya. *Margin* adalah jarak antara *hyperplane* tersebut dengan *pattern* terdekat dari masing-masing *class*. *Pattern* yang paling dekat ini disebut sebagai *support vector*. Garis solid pada Gambar 6 menunjukkan *hyperplane* yang terbaik, yaitu yang terletak tepat pada tengah-tengah kedua *class*, sedangkan titik merah dan kuning yang berada dalam lingkaran hitam adalah *support vector*. Usaha untuk mencari lokasi *hyperplane* ini merupakan inti dari proses pembelajaran pada SVM (Nugroho et al., 2003).

Pemodelan menggunakan metode SVM untuk menghasilkan akurasi yang terbaik. Dalam proses pemodelan ini menggunakan fungsi klasifikasi SVM untuk menjalankan algoritme pada python yaitu menggunakan SVM dengan mengimpor *Support Vector Classification* (SVC). SVM merupakan metode *supervised learning* yang digunakan untuk klasifikasi, regresi dan deteksi *outlier*. Metode SVM dipilih karena terbukti memiliki akurasi yang baik untuk menganalisis teks. *Kernel* merupakan seperangkat fungsi matematika yang dimiliki SVM. *Kernel* ini berfungsi untuk mengambil data sebagai input kemudian mengubah data tersebut menjadi ruang dimensi tingkat tinggi (*Kernel Space*). *Kernel* yang populer dalam SVM diantaranya *linear kernel*, *polynomial function*, *sigmoid*, *Gaussian radial basis function* (RBF), *Laplacian radial basis function*, dan *linear spline kernel* yang dapat dilihat

pada Tabel 1. (Wati et al., 2021).

Tabel 1. Jenis *kernel* pada *Support Vector Machine*

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

Metode SVM membangun satu atau sekelompok *hyperplane* dalam ruang dimensi tinggi atau tak berujung, yang dapat digunakan untuk klasifikasi, regresi atau tugas lainnya. Secara naluriah, pemisahan yang baik dicapai oleh *hyperplane* yang memiliki jarak terbesar ke titik data latih terdekat dari kelas mana pun (disebut *margin* fungsional), karena secara keseluruhan semakin besar margin, semakin rendah spekulasi pengklasifikasi (Hayati & Alifi, 2021).

SVM dapat bekerja dengan data nonlinear dengan menggunakan pendekatan *kernel* pada fitur awal himpunan data. Fungsi *kernel* digunakan untuk memetakan dimensi awal (dimensi yang lebih rendah) dari indeks data ke dimensi baru (dimensi yang lebih tinggi). Konsep SVM adalah pencarian *hyperplane* terbaik yang berfungsi sebagai pemisah data dari dua kelas pada input space. *Hyperplane* pemisah terbaik adalah *hyperplane* yang terletak di tengah di antara dua susunan objek dari dua kelas. *Hyperplane* terbaik dapat dilihat dengan memperluas tepi atau jarak dari dua susunan objek dari dua kelas yang berbeda. Dapat diasumsikan bahwa kedua belah kelas dapat terpisah secara sempurna oleh *hyperplane* (*linear separable*). Akan tetapi, pada umumnya dua belah kelas pada *input space* tidak dapat terpisah secara sempurna (*nonlinear separable*) (Auliasari et al., 2021). Penelitian ini menggunakan *kernel* linear dalam proses pengolahan datanya, sehingga pada komposisi *dataset* yang berbeda dapat dilihat perbedaannya dan mencari nilai akurasi yang terbaik.

## J. Naïve Bayes

Metode *Naïve Bayes* adalah metode klasifikasi dalam penambangan teks yang digunakan dalam analisis sentimen. Metode ini berpotensi baik dalam klasifikasi dalam hal presisi dan komputasi data. *Naïve Bayes* banyak digunakan dalam teknik klasifikasi terutama Twitter menggunakan beberapa metode seperti *Unigram Naïve Bayes*, *Multinomial Naïve Bayes*, dan *Maximum Entropy Classification*. Fitur utama klasifikasi *Naïve Bayes* untuk mendapatkan hipotesis yang kuat dari setiap kondisi atau peristiwa (Verawardina et al., 2021).

Perhitungan kategori probabilitas dalam *Naïve Bayes* menggunakan pendekatan algoritma *Bayes* menggunakan persamaan (1):

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

Persamaan (1) menunjukkan Y suatu kelas spesifik, X merupakan data pada dengan kelas yang belum diketahui,  $P(Y|X)$  adalah probabilitas hipotesis Y berdasarkan kondisi X, sedangkan  $P(Y)$  adalah hipotesis Y dan  $P(X|Y)$  merupakan probabilitas X berdasarkan kondisi hipotesis Y, serta  $P(X)$  merupakan probabilitas X. Dalam klasifikasi *Naïve Bayes*, persamaan 1 kemudian dikembangkan lagi menjadi persamaan (2) berikut:

$$P(Y + X_1, X_2, \dots, X_n) = \frac{P(X_1, X_2, \dots, X_n | Y) P(Y)}{P(X_1, X_2, \dots, X_n)} = \frac{P(X_1 | Y) P(X_2 | Y) \dots P(X_n | Y) P(Y)}{P(X_1, X_2, \dots, X_n)} \dots \dots \dots (2)$$

Dimana  $(Y|X_1, X_2, \dots, X_n)$  merupakan hasil hitung dari semua probabilitas posterior pada nilai X untuk semua nilai di Y, sehingga klasifikasi *Naïve Bayes* akan membuat prediksi berdasarkan probabilitas maksimum dari probabilitas *posterior laplace*, dimana c merupakan jumlah nilai pada Y, yang ditunjukkan pada persamaan (3) berikut:

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

Teknik *opinion mining* melalui Twitter menggunakan metode *Naïve Bayes*. *Crawling* data dilakukan dengan memberikan kata kunci dalam periode tertentu. Proses *labelling* untuk menentukan sentimen diselesaikan setelah data terkumpul. Tahap berikutnya *preprocessing* untuk seleksi data dan mengubahnya terstruktur. Pada tahap ini dilakukan proses *cleaning* untuk mengurangi *noise* dan *remove stopword* untuk menghapus kata-kata yang tidak bermakna seperti ‘saya’, ‘dan’, ‘atau’. Proses *tokenization* digunakan dalam identifikasi kata-kata dan memecah kalimat menjadi istilah berdasarkan spasi dan tanda baca. Tahap terakhir dalam *preprocessing* yaitu *stemming* merubah kata imbuhan menjadi kata dasar. Tahap ketiga dalam *opinion mining* yaitu ekstrak fitur untuk mempermudah klasifikasi *Naïve Bayes*. Tahap ini menghasilkan model dan digunakan untuk menunjukkan ketepatan hasil klasifikasi.

Data dalam bentuk *tweet* di-*crawling* dari Twitter dan disimpan dalam bentuk CSV file. Data dibagi menjadi dua *dataset* yaitu data latih dan data uji. Pelabelan akan diberikan untuk membedakan *tweet* positif dan negatif. Metode *Naïve Bayes* digunakan pada tahap klasifikasi sentimen dan interpretasi hasil analisis sentimen (Verawardina et al., 2021).

## K. Kepemimpinan Walikota

Kepemimpinan merupakan suatu hal yang dimiliki oleh setiap pemimpin dalam organisasi. Kepemimpinan perempuan berbeda dengan kepemimpinan laki-laki menurut (Tuti & Adawiyah, 2020). Hal ini terjadi karena bagi perempuan kekuasaan itu dimaknai sebagai keinginan untuk mensejahterahkan orang lain, negara-negara dengan jumlah perempuan yang banyak tentang perumusan kebijakan perempuan yang lebih maju, seperti terjadi di Swedia, Finlandia,

Denmark, Norwegia, Jerman, Belanda, dan sebagainya.

Di Indonesia, lahirnya dorongan terhadap perempuan dalam menduduki posisi kepala daerah atau kepemimpinan kepala daerah sangat kuat dan relevan. Dukungan melalui kebijakan pemerintah dalam mewujudkan kesetaraan dan keadilan *gender* telah ditetapkan dalam Instruksi Presiden Nomor 9 Tahun 2000 tentang Pengarusutamaan *Gender* (PUG). Di samping itu, pemerintah juga sedang membuat Rancangan Undang-undang Kesetaraan dan Keadilan *Gender* (RUU KKG) yaitu salah satu RUU yang akan dibahas dalam Program Legislasi Nasional (Prolegnas) 2009-2014 (Tuti & Adawiyah, 2020).

Dulu, kata “pemimpin” identik dengan kata sifat kompetitif, agresif dan dominan, yang biasanya diasosiasikan dengan maskulinitas. Seorang pemimpin dianggap cocok dengan maskulinitas, yaitu kompeten, tangguh dan dingin. Hal yang sama disampaikan Freud dan Nocielson bahwa kita tidak dapat memberi konsep baru pada kata maskulin dan *feminine*. Konsep kepemimpinan akan selalu berkenaan dengan aktivitas kekuasaan, “kekuasaan identik dengan sifat-sifat maskulinitas yakni ketegaran, kekuatan dan kemampuan mempengaruhi orang lain” (Tuti & Adawiyah, 2020).

Walikota adalah kepala daerah untuk daerah kota atau kotamadya. Seorang walikota sejajar dengan bupati, yakni kepala daerah untuk daerah kabupaten. Pada dasarnya, Walikota memiliki tugas dan wewenang memimpin penyelenggaraan daerah berdasarkan kebijakan yang ditetapkan bersama DPRD Kota. Walikota dipilih dalam satu paket pasangan dengan wakil walikota melalui Pilkada. Walikota merupakan jabatan politis, dan bukan Pegawai Negeri Sipil. Ketentuan mengenai kepala daerah dan wakil kepala daerah secara umum diatur dalam Undang-Undang Nomor 23 Tahun 2014 tentang Pemerintahan Daerah sebagaimana yang telah diubah beberapa kali, terakhir dengan Undang-Undang Nomor 9 Tahun 2015 (Maulida & Kurniawan, 2021). Pada penelitian ini, kepemimpinan walikota digunakan sebagai topik dari penelitian analisis sentimen ini. Data dari Twitter yang berkaitan dengan sistem kepemimpinan Walikota akan digunakan untuk melakukan klasifikasi analisis sentimen.

## **L. Penelitian Terdahulu**

Penelitian terkait dengan *text mining* untuk masalah analisis sentimen telah banyak dilakukan sebelumnya dengan metode yang berbeda-beda. Berikut ini adalah penelitian terdahulu yang berkaitan dengan analisis sentimen, yang dapat dilihat pada Tabel 2.

Tabel 2. Penelitian terdahulu

<b>Perbandingan teori</b>	<b>Peneliti (tahun penelitian)</b>	<b>Tujuan</b>	<b>Langkah-langkah</b>	<b>Metode</b>	<b>Output</b>
Analisis Sentimen Tentang Opini Maskapai Penerbangan pada Dokumen Twitter Menggunakan Algoritme <i>Support Vector Machine</i> (SVM)	Pravina & Adikara (2019)	Mengklasifikasikan sentimen pada data <i>tweet</i> ke dalam kelas positif dan negative, sehingga dengan diperolehnya hasil klasifikasi tersebut, dapat membantu kebutuhan perusahaan maupun masyarakat dalam memilih maskapai penerbangan yang tepat untuk keperluan pribadi, organisasi, maupun perusahaan.	<ol style="list-style-type: none"> <li>a. Pengumpulan data <i>tweet</i></li> <li>b. Penentuan sentimen secara manual</li> <li>c. Pembagian data</li> <li>d. <i>Indexing</i> data latih dan data uji</li> <li>e. Seleksi fitur</li> <li>f. Fungsi klasifikasi <i>Support Vector Machine</i> (SVM)</li> <li>g. Evaluasi</li> </ol>	Metode klasifikasi <i>Support Vector Machine</i> (SVM) dengan fitur <i>Lexicon Based</i>	Didapatkan nilai parameter <i>learning rate</i> (gamma) sebesar 0,03 dan nilai C sebesar 10 sebagai nilai parameter paling optimal. Didapatkan tingkat akurasi paling baik sebesar 40%, <i>precision</i> sebesar 40%, <i>recall</i> sebesar 100%, dan <i>f-measure</i> sebesar 57,14%. Tingkat akurasi tersebut didapatkan dengan jumlah iterasi maksimum sebanyak 50 kali dengan diimplementasikannya fitur <i>lexicon based</i> .



Tabel 2 (lanjutan)

<b>Perbandingan teori</b>	<b>Peneliti (tahun penelitian)</b>	<b>Tujuan</b>	<b>Langkah-langkah</b>	<b>Metode</b>	<b>Output</b>
Analisis Sentimen Program Acara di SCTV pada Twitter Menggunakan Metode <i>Naive Bayes</i> dan <i>Support Vector Machine</i>	Ramadhan & Setiawan (2019)	Mengetahui tingkat sentimen positif kepuasan penonton, menggunakan metode <i>Naive Bayes</i> dan <i>Support Vector Machine</i> (SVM) untuk mencari <i>trending event</i> agar dapat dipertahankan dan diperbaiki lagi agar lebih menarik hati dari para penonton televisi.	<ol style="list-style-type: none"> <li>a. Pengumpulan awal data</li> <li>b. Pengolahan <i>dataset</i> dan data <i>training</i></li> <li>c. Kategorisasi data <i>testing</i></li> <li>d. Perhitungan reputasi <i>Brand</i></li> </ol>	<i>Naive Bayes</i> dan <i>Support Vector Machine</i>	Akurasi terbaik dari Metode <i>Naive Bayes</i> dan Metode <i>Support Vector Machine</i> (SVM) adalah <i>Support Vector Machine</i> (SVM) dengan seluruh program acara didapatkan hasil akurasi 88,57%.

Tabel 2 (lanjutan)

<b>Perbandingan teori</b>	<b>Peneliti (tahun penelitian)</b>	<b>Tujuan</b>	<b>Langkah-langkah</b>	<b>Metode</b>	<b>Output</b>
Sistem Analisis Sentimen pada Ulasan Produk Menggunakan Metode <i>Naive Bayes</i>	Edukasi et al., (2018)	Dengan adanya sistem analisis sentimen ini diharapkan dapat membantu perusahaan mengetahui umpan balik terhadap merk dagangnya dan masyarakat dalam menilai sebuah produk berdasarkan opini dan <i>review</i> yang ada.	<ol style="list-style-type: none"> <li>a. Pengumpulan data <i>tweet</i></li> <li>b. <i>Data pre-processing</i></li> <li>c. <i>Labelling</i></li> <li>d. klasifikasi <i>Naive Bayes</i></li> <li>e. Perancangan system</li> <li>f. Pengujian system</li> <li>g. Analisis Hasil Pengujian</li> <li>h. Evaluasi</li> </ol>	Metode <i>Naive Bayes</i>	Sistem analisis sentimen pada ulasan produk <i>online</i> menggunakan metode <i>Naive Bayes</i> menghasilkan nilai akurasi terendah pada pengujian lima kelas menggunakan <i>dataset</i> 80% latih dan 20% data uji sebesar 52.66%, sedangkan pada pengujian tiga kelas menggunakan <i>dataset</i> 90% data latih dan 10% data uji memiliki akurasi tertinggi sebesar 77.78%.

### III. METODOLOGI PENELITIAN

#### A. Waktu dan Tempat Penelitian

Penelitian ini dilakukan di Jurusan Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung yang beralamat di Jalan Soemantri Brojonegoro No.1 Gedung Meneng, Bandar Lampung. Penelitian ini dilaksanakan mulai semester ganjil tahun ajaran 2021/2022 hingga bulan Mei tahun 2022.

#### B. Perangkat Penelitian.

Perangkat pendukung yang digunakan dalam penelitian ini adalah sebagai berikut.

##### 1. Perangkat Keras

Perangkat keras yang digunakan yaitu laptop dengan spesifikasi:

- Prosesor : Intel(R) Core(TM) i3-7020U CPU @ 2.30GHz
- RAM : 8,00 GB
- *Operating System* : Windows 10 Ultimate 64-bit Operating System

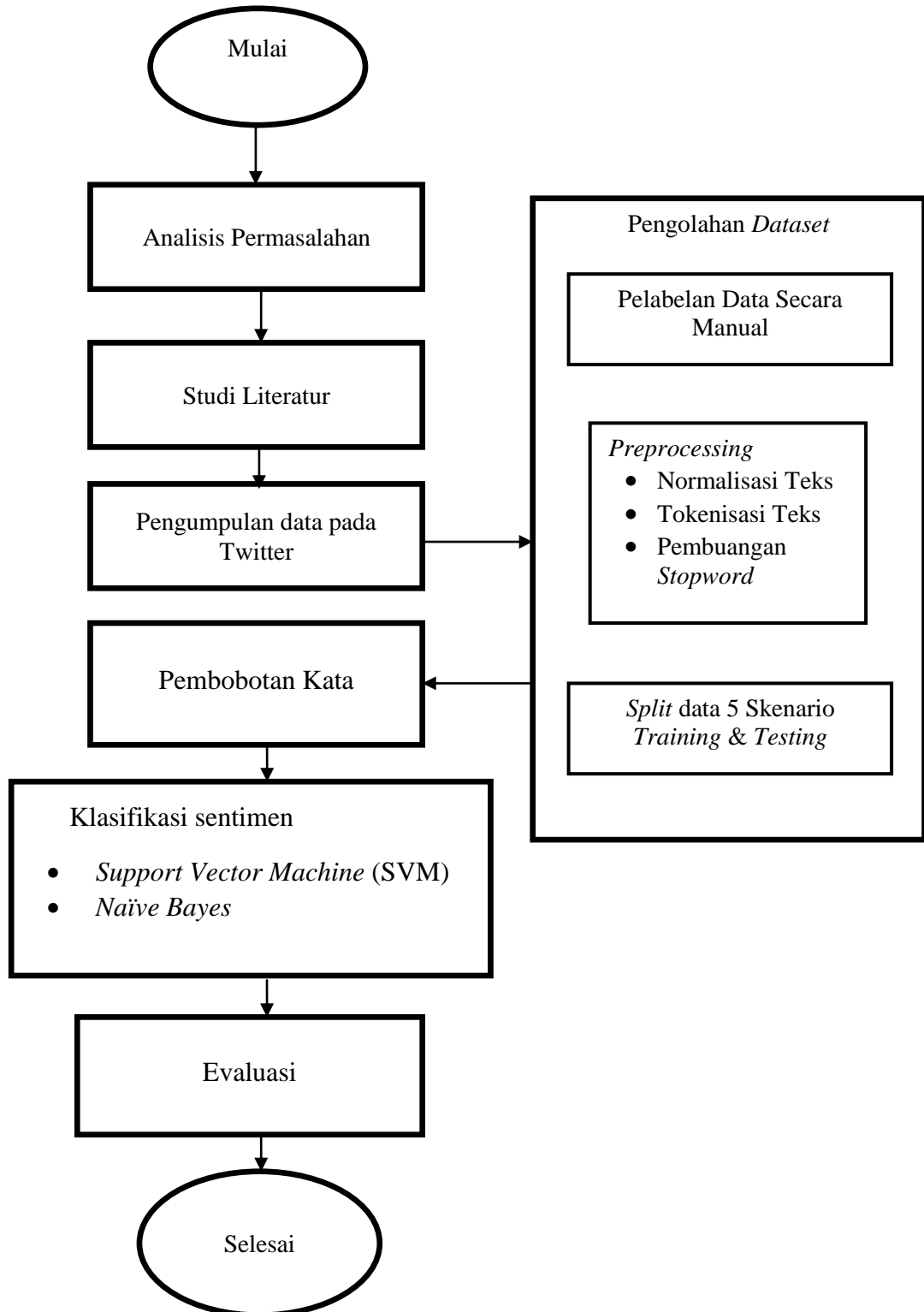
##### 2. Perangkat Lunak

Perangkat lunak yang digunakan adalah sebagai berikut:

- *Web Browser*, digunakan untuk pengunduhan data.
- Python 3, digunakan sebagai bahasa pemrograman.
- *Jupyter Notebook*, digunakan untuk membuat *script* program.

### C. Alur Penelitian

Penelitian ini melalui beberapa tahapan yang diilustrasikan pada Gambar 7.



Gambar 7. Tahapan penelitian

Gambar 7 menunjukkan tahapan-tahapan yang dilalui dalam penelitian. Berikut ini adalah penjabaran dari masing-masing tahapan penelitian yang dilakukan.

### 1. Analisis Permasalahan

Pada tahap ini dilakukan analisis permasalahan yang ada yaitu mengenai metode algoritme yang mendapatkan nilai akurasi yang terbaik mengenai analisis sentimen. Topik penelitian ini adalah analisis sentimen untuk pengklasifikasian sentimen negatif dan sentimen positif terhadap data yang didapatkan dari Twitter dengan menggunakan algoritme *Support Vector Machine* (SVM) dan *Naïve Bayes* tentang kepemimpinan Walikota Bandarlampung dan untuk mengetahui persepsi masyarakat terhadap Walikota Bandarlampung guna membantu pihak-pihak yang membutuhkan informasi mengenai gaya kepemimpinan Walikota Bandarlampung dan untuk menjadi pembelajaran kinerja yang lebih baik untuk Walikota Bandarlampung di periode selanjutnya atau pada masa yang akan datang. Kedua algoritme ini merupakan salah satu algoritme yang paling banyak digunakan untuk melakukan analisis sentimen dan pada penelitian ini akan dilakukan perbandingan antara algoritme *Support Vector Machine* (SVM) dan *Naive Bayes*. Algoritme *Support Vector Machine* (SVM) ini menghasilkan model klasifikasi yang cukup baik meskipun dilatih dengan menggunakan himpunan data yang sedikit.

### 2. Studi Literatur

Pada tahap ini dilakukan pengumpulan informasi dan penelitian terdahulu yang berkaitan dengan topik permasalahan atau penelitian yang sama dan mempelajari berbagai metode penelitian. Untuk penelitian terdahulu mengenai analisis sentimen, terdapat tiga penelitian terdahulu dengan topik yang berbeda yang dijadikan sebagai studi literatur untuk penelitian ini yang ditampilkan pada Tabel 2. Algoritme yang digunakan dari ketiga penelitian terdahulu berbeda-beda.

### 3. Pengumpulan Data

Pada 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. Data yang dikumpulkan dari Twitter merupakan data teks yang tidak memiliki nilai sentimen positif ataupun negatif. Kata kunci untuk pengumpulan data dari Twitter menggunakan kata “Walikota Bandarlampung Bunda Eva Dwiana”. Untuk pengumpulan datanya sendiri termasuk lama karena dilakukan hampir per minggu karena data di Twitter yang selalu *update*. Untuk pengumpulan datanya sendiri menggunakan program yang berbeda dengan program klasifikasi.

### 4. Pengolahan *Dataset*

Pada tahap ini dilakukan pengolahan *dataset* yang telah diperoleh dari Twitter. Proses tahapan yang dilakukan yaitu tokenisasi teks, normalisasi teks, dan pembuangan *stopword*. Fungsi dari tahapan adalah sebagai berikut:

#### a. Normalisasi Teks

Normalisasi teks berfungsi untuk melakukan penormalan kata atau kalimat dalam suatu data. Untuk normalisasi teks ini tujuannya adalah untuk memperbaiki struktur kata dan kosakata dalam kalimat.

Pada penelitian ini normalisasi akan dilakukan pada seluruh data teks penelitian yang digunakan. Normalisasi teks yang dilakukan adalah penghapusan URL, penghapusan tanda baca, karakter khusus, dan simbol-simbol dengan bentuk huruf abjad sehingga tidak terjadi ambuigitas dalam data teks.

#### b. Tokenisasi Teks

Tahap tokenisasi teks adalah tahap pemotongan *string* atau kalimat input berdasarkan tiap kata yang menyusun kalimat dalam data teks.

Tokenisasi secara garis besar memecah sekumpulan karakter dalam suatu teks ke dalam satuan kata, bagaimana membedakan karakter-karakter tertentu yang dapat diperlakukan sebagai pemisah kata atau bukan.

c. Pembuangan *Stopword*

*Stopwords* merupakan kumpulan kata-kata yang sering muncul tetapi jika dihapus tidak mengubah makna dari *tweet* tersebut. Pembuangan *stopword* dimaksudkan untuk mengetahui suatu kata yang tidak memiliki arti atau tidak relevan. Kata yang diperoleh dari tahap tokenisasi diperiksa dalam suatu *stopword*, apabila ada sebuah kata masuk di dalam *stopword* maka kata tersebut tidak akan diproses lebih lanjut.

Tahapan ini dilakukan proses klasifikasi sentimen dengan masing-masing algoritme. Selanjutnya, data-data yang telah dikumpulkan dilakukan pelabelan nilai sentimen pada data secara manual apakah data tersebut memiliki nilai positif atau negatif. Tahapan pelabelan data ini dilakukan karena pada saat pengumpulan data, data-data yang telah didapatkan tidak memiliki nilai sentimen positif ataupun negatif.

5. Klasifikasi Sentimen

Pada tahap ini dilakukan klasifikasi sentimen menggunakan dua algoritme yaitu algoritme *Support Vector Machine* (SVM) dan *Naïve Bayes*. Data yang diklasifikasi adalah data *tweet* yang telah melewati tahap pengolahan data. Data-data tersebut diklasifikasi dan dianalisis menggunakan dua metode yaitu algoritme *Support Vector Machine* (SVM) dan *Naive Bayes*. Sebelum itu dilakukan pembagian data menjadi dua, yaitu sebagai data latih dan data uji. Data latih digunakan untuk melatih algoritme atau metode, sedangkan data uji digunakan untuk mengetahui performa algoritme yang sudah dilatih sebelumnya ketika menemukan data baru yang belum pernah ditemui

sebelumnya. Setelah data dibagi menjadi data latih dan data uji, dilakukan pelatihan data menggunakan data latih yang telah dibagi menggunakan *word2vec*. Selanjutnya adalah melakukan klasifikasi dengan *Support Vector Machine* (SVM) dan *Naive Bayes* yang program klasifikasinya telah dibuat. Hasil akhir yang didapatkan dari klasifikasi berupa prediksi sentimen positif dan negatif dalam bentuk *confusion matrix* dan *classification report* yang isinya berupa akurasi, presisi, *recall*, dan *f1-score* hasil dari melakukan klasifikasi.

## 6. Evaluasi

Evaluasi yang dilakukan adalah melakukan perbandingan nilai sentimen dari pelabelan manual dan analisis sentimen menggunakan kedua algoritme tersebut. Setelah itu dilakukan perbandingan antara hasil dari algoritme *Support Vector Machine* (SVM) dan algoritme *Naive Bayes*. Hasil klasifikasi yang didapatkan adalah *confusion matrix* dan *classification report* berupa akurasi, presisi, *recall*, dan *f1-score*.

Untuk *confusion matrix* sendiri berisi informasi hasil prediksi klasifikasi dari data aktual yang dilakukan oleh sistem klasifikasi. Kinerja sistem klasifikasi umumnya dihitung menggunakan data dalam tabel *confusion matrix* seperti pada Tabel 3.

Tabel 3. Tabel *confusion matrix*

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



Nilai *True Negative* (TN) merupakan jumlah data negatif yang terdeteksi dengan benar, sedangkan *False Negative* (FN) merupakan data positif yang terdeteksi negatif. Nilai *True Positive* (TP) merupakan data positif yang terdeteksi benar, sedangkan *False Positive* (FP) merupakan data negatif namun terdeteksi positif. Tabel *confusion matrix* digunakan untuk mengukur kinerja suatu metode klasifikasi dengan menghitung nilai akurasi, presisi, *recall*, dan *f1-score* (Muludi et al., 2021).

1. Akurasi (*accuracy*) merupakan rasio prediksi benar (positif, dan negatif) dengan keseluruhan data.

$$\text{Akurasi (accuracy)} = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \dots \dots \dots (4)$$

2. Presisi (*precision*) merupakan rasio prediksi benar positif dibandingkan dengan keseluruhan hasil yang diprediksi positif.

$$\text{Presisi (precision)} = \frac{TP}{TP + FP} \times 100\% \dots \dots \dots (5)$$

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

$$\text{Recall} = \frac{TP}{TP + FN} \times 100\% \dots \dots \dots (6)$$

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

$$\text{F1 Score} = 2 \times \frac{(\text{recall} \times \text{precision})}{(\text{recall} + \text{precision})} \dots \dots \dots (7)$$

## V. SIMPULAN DAN SARAN

### A. Simpulan

Pada penelitian ini dapat disimpulkan bahwa:

1. Penelitian sudah berhasil mengimplementasikan algoritme *Support Vector Machine* (SVM) dan *Naive Bayes*. Hasil dari beberapa pengujian menunjukkan bahwa sentimen masyarakat dari Twitter terhadap kepemimpinan Walikota Bandarlampung cenderung positif. Tingkat akurasi dalam penelitian ini tinggi berada di kisaran 87%. Hasil dari pemberian sentimen secara manual terhadap data dari Twitter tentang kepemimpinan Walikota Bandarlampung menghasilkan 918 *tweet* pada kelas positif dan 150 *tweet* pada kelas negatif yaitu artinya sebanyak 85% opini masyarakat Bandarlampung termasuk ke dalam sentimen positif mengenai gaya kepemimpinan Walikota Bandarlampung dan sebanyak 15% opini masyarakat Bandarlampung termasuk ke dalam sentimen negatif.
2. Perbandingan hasil kinerja dari klasifikasi algoritme *Support Vector Machine* (SVM) menunjukkan nilai klasifikasi yang lebih baik daripada kinerja dari algoritme *Naive Bayes*. Algoritme *Support Vector Machine* (SVM) menghasilkan akurasi tertinggi dengan menggunakan skenario ke-4 dengan *dataset* 80% latih - 20% data uji menghasilkan akurasi sebesar 87,38%, presisi 87%, *recall* 100%, dan *f1-score* 93%, sedangkan algoritme *Naive Bayes* menghasilkan akurasi terbaik dengan menggunakan skenario ke-2 dengan *dataset* 60% latih - 40% data uji menghasilkan akurasi sebesar 85,28%, presisi 86%, *recall* 99%, dan *f1-score* 92%. Untuk hasil *confusion matrix* dari *Support Vector Machine* (SVM) menunjukkan bahwa tidak terlalu banyak nilai prediksi yang *error* pada metode *Support Vector Machine* (SVM). Dari hasil kedua klasifikasi tersebut dapat disimpulkan bahwa kinerja algoritme *Support Vector Machine* (SVM) memiliki hasil klasifikasi analisis sentimen

yang lebih baik dibandingkan dengan kinerja algoritme *Naive Bayes*.

3. Dari pengujian dua kelas yaitu kelas positif dan kelas negatif yang dilakukan dengan lima skenario pembagian *dataset*. Semakin besar data latih yang digunakan tidak selalu meningkatkan nilai akurasi. Hasil pengujian menunjukkan akurasi tertinggi algoritme *Support Vector Machine* (SVM) menggunakan kombinasi data pada 80% data latih dan 20% data uji, sedangkan algoritme *Naive Bayes* menunjukkan akurasi tertingginya menggunakan kombinasi data 60% data latih dan 40% data uji.
4. Perbandingan performansi algoritme *Naive Bayes* lebih baik dibandingkan algoritme *Support Vector Machine* (SVM), waktu komputasi terbaik sistem dalam menyelesaikan klasifikasi dengan menggunakan metode *Naive Bayes* dengan rata-rata waktu yang dibutuhkan yaitu selama 0,0055 detik, sedangkan *Support Vector Machine* (SVM) memiliki rata-rata waktu klasifikasi 0,7897 detik. Dapat disimpulkan *Naive Bayes* memiliki waktu klasifikasi yang lebih singkat sehingga mempercepat proses sistem analisis sentimen.

## **B. Saran**

Terdapat beberapa hal yang dapat ditambahkan atau diperbaiki untuk penelitian selanjutnya, yaitu:

1. Dapat mengembangkan analisis sentimen ini dengan mengombinasikan metode *Deep Learning* seperti *Reccurent Neural Network*, *Deep Neural Networks*, dan *Convolutional Neural Networks* untuk dilakukan pengujian dan mendapatkan nilai akurasi yang lebih baik.
2. Pada penelitian selanjutnya lebih baik menerapkan *dataset* yang sudah dieksplorasi keseimbangannya yang berguna untuk meningkatkan performa klasifikasi.

## DAFTAR PUSTAKA

- Analysis, S., Misyani, N., Rafie, M., Dewi, K., Shafenoer, M., Suryawati, E., Munandar, D., Babafemi, O., Jonathan, D. I., & Ibukun, T. A. (n.d.). 2020. Twitter Sentiment Analysis as an Evaluation and Service Base On Python Textblob. *Journal of Materials Science and Engineering*. <https://doi.org/10.1088/1757-899X/1125/1/012034>
- Antinasari, P., Perdana, R. S., & Fauzi, M. A. 2017. Analisis Sentimen Tentang Opini Film Pada Dokumen Twitter Berbahasa Indonesia Menggunakan Naive Bayes Dengan Perbaikan Kata Tidak Baku. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 1 (12), 1733–1741.
- Ardhiansyah, M. N., Umar, R., & Sunardi. 2019. Analisis Sentimen pada Twitter Menggunakan Metode Support Vector Machine. *Jurnal Nasional Teknologi Fakultas Teknik Universitas Krisnadwipayana*, 1 (1), 739–742. <https://jurnal.teknikunkris.ac.id/index.php/semnastek2019/article/view/343/342>
- Arsya Monica Pravina, I. C., & Putra Pandu Adikara. 2019. Analisis Sentimen Tentang Opini Maskapai Penerbangan pada Dokumen Twitter Menggunakan Algoritme Support Vector Machine (SVM) / *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*. 3 (3), 2789–2797. <https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/4793>
- Auliasari, K., Prasetya, R. P., & Industri, F. T. 2021. Analisis Sentiment Cyberbullying Pada Sosial Media Twitter Menggunakan Metode Support Vector Machine. *Jurnal Mahasiswa Teknik Informatika*, 5 (2), 548–556.

- Darwis, D., Pratiwi, E. S., & Pasaribu, A. F. O. 2020. Penerapan Algoritma Svm Untuk Analisis Sentimen Pada Data Twitter Komisi Pemberantasan Korupsi Republik Indonesia. *Edutic - Scientific Journal of Informatics Education*, 7 (1), 1–11. <https://doi.org/10.21107/edutic.v7i1.8779>
- Edukasi, J., Gunawan, B., Pratiwi, H. S., & Pratama, E. E. 2018. Sistem Analisis Sentimen pada Ulasan Produk Menggunakan Metode Naive Bayes. *Jurnal Edukasi dan Penelitian Informatika*, 4 (2), 113–118.
- Hayati, H., & Alifi, M. R. 2021. Analisis Sentimen Pada Tweet Terkait Vaksin Covid-19 Menggunakan Metode Support Vector Machine. *Jurnal Teknologi Terapan*, 7 (2), 110–119.
- Hendrastuty, N., Isnain, A. R., Rahmadhani, A. Y., Studi, P., Informasi, S., Indonesia, U. T., Informatika, P. S., Indonesia, U. T., & Lampung, K. B. 2021. Analisis Sentimen Masyarakat Terhadap Program Kartu Prakerja Pada Twitter Dengan Metode Support Vector Machine. *Jurnal pengembangan IT (JPIT)*, 6 (3), 150–155.
- Maulida, K., & Kurniawan, R. C. 2021. Strategi Pemenangan Eva Dwiana dan Deddy Amarullah dalam Pemilihan Kepala Daerah Tahun 2020. *Jurnal Ilmiah Ilmu Pemerintahan*, 6 (2). <https://doi.org/10.14710/jiip.v6i2.11708>
- Mikolov, T., & Com, T. G. 2014. Distributed Representations of Sentences and Documents. *NeurIPS Proceedings*, 32. <https://proceedings.neurips.cc/paper/2013/hash/9aa42b31882ec039965f3c4923ce901b-Abstract.html>
- Mikolov, T., Corrado, G., Chen, K., & Dean, J. 2013. *Efficient Estimation of Word Representations in Vector Space*. <https://arxiv.org/abs/1301.3781>

- Muludi, K., Akbar, M. S., Shofiana, D. A., & Syarif, A. 2021. Sentiment Analysis Of Energy Independence Tweets Using Simple Recurrent Neural Network. *Indonesian Journal of Computing and Cybernetics Systems*, *x*, 1–10. <https://doi.org/10.22146/ijccs.xxxx>
- Nofiyanti, E., Maricha, E., Nur, O., & Janabadra, U. 2021. Analisis Sentimen terhadap Penanggulangan Bencana di Indonesia. *Jurnal Ilmiah Sinus*, *19* (2), <https://doi.org/10.30646/sinus.v19i2.563>
- Nugraha, P. A. 2013. Perbandingan Metode Probabilistik Naive Bayesian Classifier dan Jaringan Syaraf Tiruan Learning Vector Quantization dalam Kasus Klasifikasi Penyakit Kandungan. *Jurnal Teknologi dan Informasi*, *2* (2), 20–33.
- Ramadhan, D. A., & Setiawan, E. B. 2019. Analisis Sentimen Program Acara Di SCTV Pada Twitter Menggunakan Metode Naive Bayes Dan Support Vector Machine. *e-Proceeding of Engineering*, *6* (2), 36-42.
- Refaeilazadeh, P., Tang, L., & Liu, H. 2016. Cross-Validation. In M. Liu, L., Özsu (Ed.), *Encyclopedia of Database Systems* (pp. 532–538). Springer. <https://doi.org/10.7829/j.ctv280b7r6.6>
- Roihan, A., Sunarya, P. A., & Rafika, A. S. 2020. Pemanfaatan Machine Learning dalam Berbagai Bidang : Review paper. *Indonesian Journal on Computer and Information Technology*, *5* (1), 75–82.
- Tuti, R. W. D., & Adawiyah, E. 2020. Kepemimpinan Walikota Surabaya Tri Rismaharini Sebagai Model Birokrasi Efektif. *Jurnal Ilmu Komunikasi Politik Dan Komunikasi Bisnis*, *4* (1), 73–82.
- Verawardina, U., Edi, F., & Watrianthos, R. 2021. Analisis Sentimen Pembelajaran Daring Pada Twitter di Masa Pandemi COVID-19

Menggunakan Metode Naïve Bayes. *Jurnal Media Informatika Budidarma*, 5 (1), 157–163. <https://doi.org/10.30865/mib.v5i1.2604>

Wati, R., Ernawati, S., Studi, P., Informasi, S., Bina, U., Informatika, S., Studi, P., Informasi, S., Nusa, U., & Timur, J. 2021. Analisis Sentimen Persepsi Publik Mengenai PPKM Pada Twitter Berbasis SVM Menggunakan Python. *Jurnal Teknik Informatika Unika St. Thomas (JTIUST)*, 6 (2), 240–247.

Yadav, S., & Shukla, S. 2016. Analysis of k-Fold Cross-Validation over Hold-Out Validation on Colossal Datasets for Quality Classification. *Proceedings - 6th International Advanced Computing Conference, IACC 2016*, Cv, 78–83. <https://doi.org/10.1109/IACC.2016.25>