

**IMPLEMENTASI METODE *SUPPORT VECTOR MACHINE* (SVM)
DALAM KLASIFIKASI JUDUL BERITA HOAKS POLITIK**

(Skripsi)

Oleh

**CINDY ANISYA
NPM 2017031074**



**JURUSAN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2024**

ABSTRACT

IMPLEMENTATION OF SUPPORT VECTOR MACHINE (SVM) METHOD IN THE CLASSIFICATION POLITICAL HOAX NEWS HEADLINES

By

CINDY ANISYA

Support Vector Machine (SVM) is a method in machine learning with the concept of finding the optimal hyperplane in the input space so that it can classify binary or multiclass classification problems. This study aims to apply the SVM method to the classification of political hoax news titles, and determine the level of accuracy in each kernel function used in the SVM method. The kernel function used in this research is linear kernel, Radial Basis Function (RBF) kernel, polynomial kernel, and sigmoid kernel with a comparison between training data and testing data of 80%: 20%, 90%: 10%, and 95%: 5%. The results of this study show that the best accuracy value in the classification process is to use the RBF kernel function at a ratio of 80%:20%, and 95%:5% with an accuracy value of 92%.

Keywords: Classification, Machine Learning, Machine Vector Support, News
Headline.

ABSTRAK

IMPLEMENTASI METODE *SUPPORT VECTOR MACHINE* (SVM) DALAM KLASIFIKASI JUDUL BERITA HOAKS POLITIK

Oleh

CINDY ANISYA

Support Vector Machine (SVM) merupakan sebuah metode dalam *machine learning* dengan konsep menemukan *hyperplane* optimal dalam ruang input sehingga dapat mengklasifikasikan masalah klasifikasi biner atau multikelas. Penelitian ini bertujuan menerapkan metode SVM pada klasifikasi judul berita hoaks politik, serta mengetahui tingkat akurasi pada setiap fungsi kernel yang digunakan pada metode SVM. Adapun fungsi kernel yang digunakan adalah kernel linear, kernel *Radial Basis Function* (RBF), kernel polinomial, dan kernel sigmoid dengan perbandingan antara *data training* dan *data testing* sebesar 80%:20%, 90%:10%, dan 95%:5%. Hasil dari penelitian ini menunjukkan bahwa nilai akurasi terbaik pada proses klasifikasi yaitu dengan menggunakan fungsi kernel RBF pada perbandingan 80%:20%, dan 95%:5% dengan nilai akurasi sebesar 92%.

Kata kunci: Klasifikasi, *Machine Learning*, *Support Vector Machine*, Judul Berita.

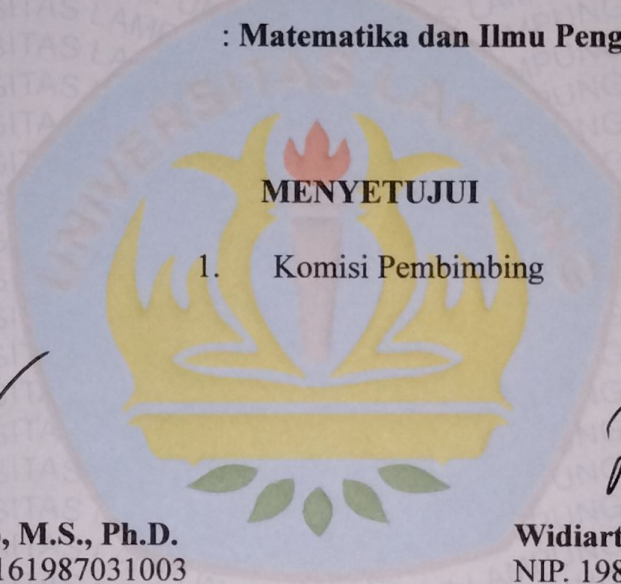
Judul Skripsi : **IMPLEMENTASI METODE *SUPPORT VECTOR MACHINE* (SVM) DALAM KLASIFIKASI JUDUL BERITA HOAKS POLITIK**

Nama Mahasiswa : **Cindy Anisya**

Nomor Pokok Mahasiswa : **2017031074**

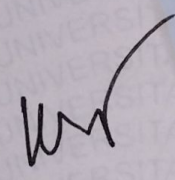
Jurusan : **Matematika**

Fakultas : **Matematika dan Ilmu Pengetahuan Alam**

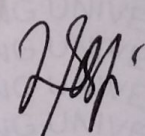


MENYETUJUI

1. Komisi Pembimbing

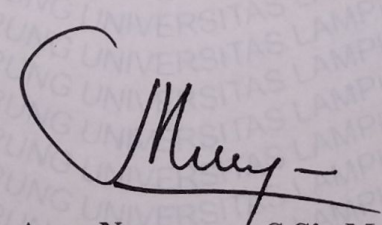


Ir. Warsono, M.S., Ph.D.
NIP. 196302161987031003



Widiarti, S.Si., M.Si.
NIP. 198005022005012003

2. Ketua Jurusan Matematika

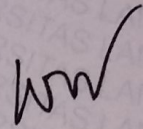


Dr. Aang Nuryaman, S.Si., M.Si.
NIP. 197403162005011001

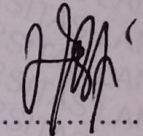
MENGESAHKAN

1. Tim Penguji

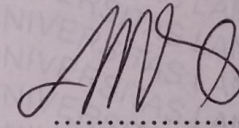
Ketua : Ir. Warsono, M.S., Ph.D.



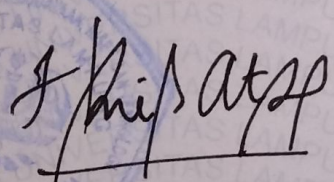
Sekretaris : Widiarti, S.Si., M.Si.



**Penguji
Bukan Pembimbing : Dr. Ahmad Faisol, S.Si., M.Sc.**



2. Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam



Dr. Eng. Heri Satria, S.S., M.Si.
NIP. 197110012005011002

Tanggal Lulus Ujian Skripsi: 12 Juli 2024

PERNYATAAN SKRIPSI MAHASISWA

Yang bertanda tangan di bawah ini:

Nama Mahasiswa : **Cindy Anisya**

Nomor Pokok Mahasiswa : **2017031074**

Jurusan : **Matematika**

Judul Skripsi : **IMPLEMENTASI METODE *SUPPORT VECTOR MACHINE* (SVM) DALAM KLASIFIKASI JUDUL BERITA HOAKS POLITIK**

Dengan ini menyatakan bahwa penelitian ini adalah hasil pekerjaan saya sendiri, bukan hasil orang lain, dan semua hasil yang tertuang dalam skripsi ini telah mengikuti kaidah penulisan karya ilmiah Universitas Lampung. Apabila kemudian hari terbukti bahwa skripsi ini merupakan hasil salinan atau dibuat oleh orang lain, maka saya bersedia menerima sanksi sesuai dengan ketentuan akademik yang berlaku.

Bandar Lampung, 15 Juli 2024
Penulis



Cindy Anisya
NPM. 2017031074

**IMPLEMENTASI METODE *SUPPORT VECTOR MACHINE* (SVM)
DALAM KLASIFIKASI JUDUL BERITA HOAKS POLITIK**

Oleh

CINDY ANISYA

Skripsi

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

Pada

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



**JURUSAN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2024**

RIWAYAT HIDUP

Penulis bernama lengkap Cindy Anisya, anak pertama dari dua bersaudara yang lahir pada tanggal 4 Januari 2002 di Mulyosari, Lampung Selatan dari pasangan Bapak Joniarto dan Ibu Tri Astuti.

Penulis menempuh pendidikan Taman Kanak-Kanak (TK) di Taman Kanak-Kanak Islam Terpadu Al-Amanah, yang diselesaikan pada tahun 2008, kemudian melanjutkan pendidikan di Sekolah Dasar (SD) di Sekolah Dasar Negeri Taman Cibodas yang diselesaikan pada tahun 2014. Kemudian menempuh pendidikan Sekolah Menengah Pertama (SMP) di Sekolah Menengah Pertama Negeri 1 Tanjungsari yang diselesaikan pada tahun 2017, dan kemudian melanjutkan pendidikan Sekolah Menengah Atas (SMA) di Sekolah Menengah Atas Negeri 5 Bandar Lampung yang diselesaikan pada tahun 2020, dan pada tahun 2020, penulis diterima sebagai mahasiswa S-1 Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung melalui jalur Seleksi Bersama Masuk Perguruan Tinggi Negeri (SBMPTN).

Selama menjadi mahasiswi penulis pernah bergabung pada Himpunan Mahasiswa Jurusan Matematika (HIMATIKA) sebagai anggota Bidang Dana dan Usaha periode 2021, dan Badan Eksekutif Mahasiswa Fakultas Matematika dan Ilmu Pengetahuan Alam (BEM-FMIPA) Universitas Lampung Periode 2022, sebagai anggota Departemen Bidang Komunikasi dan Informasi (Kominfo).

Pada tahun 2023, di bulan Januari-Februari penulis melaksanakan Kerja Praktik (KP) di PT PLN UID Lampung, dan di bulan Juni-Agustus penulis melaksanakan Kuliah Kerja Nyata (KKN) di Desa Bina Karya Jaya, Kecamatan Putra Rumbia, Kabupaten Lampung Tengah.

KATA INSPIRASI

“Tidak ada manusia yang baik-baik saja di dunia ini, semua sedang berjuang dengan ujiannya masing-masing. Maka, ingatlah selalu...

Hasbunaallahu wa ni'malwakiil

Cukuplah Allah, SWT. sebagai penolong kami, dan Allah SWT. adalah sebaik-baik pelindung”

(Q.S Ali-Imran : 173)

PERSEMBAHAN

*Bismillahirrahmanirrahim
Alhamdulillahirobbil'alamiin*

Dengan mengharakan rahmat dan ridho Allah SWT, kupersembahkan karya sederhana yang penuh dengan perjuangan ini kepada:

Ayah dan Mama Tercinta

Yang tak pernah lelah menyayangi, merawat, dan mendidikku hingga saat ini dengan sepenuh hati. Yang selalu mendukung semua kegiatanku dan senantiasa berdoa untuk keberhasilanku. Terima kasih karena selalu ada untukku di saat suka maupun duka. Semoga karya ini menjadi langkah awal kesuksesanku agar dapat membuat Ayah dan Mama bangga kepadaku.

Adik, dan Keluarga Tersayang

Yang senantiasa menyayangiku, memberikan dukungan, dan motivasi kepadaku agar aku lebih bersemangat dalam menyelesaikan pendidikan..

Dosen Pembimbing dan Penguji

Yang telah meluangkan waktu untuk memberikan bimbingan dan ilmu yang bermanfaat kepadaku.

Almamater Tercinta, Universitas Lampung

SANWACANA

Alhamdulillahirobbil'alaamiin puji dan syukur penulis panjatkan kehadirat Allah subhanahu wata'ala. Tuhan Semesta Alam atas segala rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan skripsi dengan judul "**IMPLEMENTASI METODE *SUPPORT VECTOR MACHINE* (SVM) DALAM KLASIFIKASI JUDUL BERITA HOAKS POLITIK**" dengan baik dan tepat pada waktunya.

Skripsi ini dapat penulis susun dengan baik dan lancar tidak lain karena dukungan dan bantuan dari banyak pihak. Oleh karena itu, penulis mengucapkan rasa terima kasih kepada:

1. Bapak Ir. Warsono, M.S., Ph.D. selaku dosen pembimbing I yang telah bersedia memberikan bimbingan, bantuan, arahan, dan dukungan dalam proses menyelesaikan skripsi ini.
2. Ibu Widiarti, S.Si., M.Si. selaku dosen pembimbing kedua atas kesediaan waktu dalam memberikan bimbingan, bantuan, arahan, dan dukungan dalam proses penyusunan skripsi ini.
3. Bapak Dr. Ahmad Faisol, S.Si., M.Sc. selaku dosen penguji yang telah memberikan kritik dan saran yang membangun dalam proses penyusunan skripsi ini.
4. Bapak Dr. Aang Nuryaman, S.Si, M.Si., selaku Ketua Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam.
5. Bapak Dr. Agus Sutrisno, S.Si., M.Si., selaku dosen pembimbing akademik yang telah memberikan bimbingan, dan arahan selama masa perkuliahan.
6. Bapak Dr. Eng. Heri Satria, S.Si., M.Si. selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.

7. Seluruh dosen, staf, dan karyawan Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung yang telah memberikan ilmu dan bantuan kepada penulis.
8. Ayah, mama, adik, dan seluruh keluarga tercinta yang selalu mendoakan, memberi dukungan dan motivasi untuk selalu semangat dan fokus dalam mengejar cita-cita.
9. Maya, Gustina, Happy, Salsa, Meryam, Nada, teman baik dan seperjuangan penulis yang selalu memberikan semangat penulis selama masa perkuliahan.
10. A. Fabiyo Aufadilla Fasya yang senantiasa memberi semangat, dan dukungannya untuk penulis.
11. Arini, Ayu, Mayang, teman-teman seperbimbingan, teman KKN Bina Karya Jaya, teman baik penulis yang senantiasa selalu menemani setiap proses penulis dalam menyelesaikan skripsi ini.
12. Teman-teman jurusan matematika angkatan 2020 yang memberikan rasa kekeluargaan dan dukungan baiknya.
13. Semua pihak yang tidak bisa penulis sebutkan satu persatu yang telah membantu dalam menyelesaikan skripsi ini.

Penulis menyadari bahwa skripsi ini masih jauh dari kata sempurna. Kritik dan saran sangat penulis harapkan agar dapat digunakan sebagai bahan perbaikan ke depannya. Semoga skripsi ini dapat memberikan manfaat kepada kita.

Bandar Lampung, 15 Juli 2024

Penulis

Cindy Anisya

DAFTAR ISI

	Halaman
DAFTAR TABEL	xiii
DAFTAR GAMBAR	xv
I. PENDAHULUAN	1
1.1 Latar Belakang dan Masalah	1
1.2 Tujuan Penelitian	3
1.3 Manfaat Penelitian	3
II. TINJAUAN PUSTAKA	5
2.1 <i>Machine Learning</i>	5
2.2 <i>Text Mining</i>	6
2.3 Klasifikasi Teks	6
2.4 <i>Preprocessing Data</i>	7
2.5 Ekstraksi Fitur	8
2.6 <i>Support Vector Machine</i>	9
2.7 Evaluasi Model	14
2.7.1 <i>Confusion Matrix</i>	15
2.7.2 <i>Classification Report</i>	15
2.8 Berita Hoaks	16
III. METODOLOGI PENELITIAN	18
3.1 Waktu dan Tempat Penelitian	18
3.2 Data Penelitian	18
3.3 Metode Penelitian	20
IV. HASIL DAN PEMBAHASAN	22
4.1 Karakteristik Data	22

4.2 Hasil <i>Preprocessing</i> Data.....	22
4.3 Hasil Perhitungan Ekstraksi Fitur	27
4.4 Proses Klasifikasi.....	33
4.4.1 <i>Splitting Data Training</i> dan <i>Data Testing</i>	33
4.4.2 Proses <i>Training</i>	33
4.5 Proses Klasifikasi menggunakan Program.....	40
4.5.1 Hasil Ekstraksi Fitur menggunakan Program.....	41
4.5.2 Hasil Klasifikasi SVM menggunakan Program.....	42
4.5.3 Hasil Evaluasi Model.....	43
IV. KESIMPULAN DAN SARAN	46
5.1 Kesimpulan	46
5.2 Saran	47
DAFTAR PUSTAKA	48
LAMPIRAN	52

DAFTAR TABEL

1.	<i>Confusion Matrix</i>	15
2.	Data Penelitian	19
3.	Hasil <i>Text Cleaning</i>	23
4.	Hasil <i>Case Folding</i>	24
5.	Hasil <i>Tokenizing</i>	25
6.	Hasil <i>Stopword Removal</i>	26
7.	Hasil <i>Stemming</i>	27
8.	Contoh Dokumen pada Data <i>Training</i>	28
9.	Contoh Dokumen pada Data <i>Testing</i>	28
10.	Hasil Perhitungan Nilai TF dan Nilai IDF Contoh Dokumen pada Data Training dan Data Testing	29
11.	Hasil Perhitungan Nilai TF-IDF Contoh Dokumen pada Data Training dan Data Testing	31
12.	Hasil <i>Splitting</i> pada Seluruh Data	33
13.	Kelas Contoh Dokumen pada Data <i>Training</i>	34
14.	Hasil Perhitungan Nilai <i>Error</i> Iterasi Pertama	35
15.	Hasil Perhitungan Nilai Delta Alpha Iterasi Pertama	36
16.	Hasil Perhitungan Nilai Alpha Iterasi Pertama.....	36
17.	Hasil Perhitungan Nilai Error Iterasi Kedua.....	36
18.	Hasil Perhitungan Nilai Delta Alpha Iterasi Kedua.....	37
19.	Hasil Perhitungan Nilai Alpha Iterasi Kedua	37
20.	Hasil Perhitungan Nilai Bobot Contoh Dokumen pada Data <i>Training</i>	38
21.	Hasil Perhitungan Nilai Bobot Contoh Dokumen pada Data <i>Testing</i>	39
22.	Hasil TF-IDF Seluruh Data menggunakan Program	40
23.	Hasil Klasifikasi SVM menggunakan Program.....	42
24.	Hasil Evaluasi Model menggunakan <i>Confusion Matrix</i>	43

DAFTAR GAMBAR

1. Konsep *Hyperplane* pada SVM 10
2. Diagram Alir Penelitian 21

I. PENDAHULUAN

1.1 Latar Belakang dan Masalah

Text mining merupakan bagian dalam *data mining* yang dapat dilakukan untuk penambangan informasi penting dari data yang berupa tulisan, dokumen atau teks (Harjanta, 2015). Salah satu teknik yang dapat dilakukan dalam *text mining* adalah teknik klasifikasi. Menurut Zhang, dkk., (2008), klasifikasi teks merupakan proses menetapkan kategori atau kelas yang telah ditentukan sebelumnya, salah satunya pada dokumen teks berita melalui tahap pelatihan (*training*) dan tahap pengujian (*testing*), kemudian berdasarkan model klasifikasi yang dihasilkan dilakukan perbandingan antara hasil perkiraan kelas dengan kelas sebenarnya. Klasifikasi teks saat ini menjadi topik dan penelitian sejak dimulainya dokumen digital. Mengingat jumlahnya yang besar dengan berbagai kategori yang berbeda, maka diperlukan sebuah metode klasifikasi yang mampu menangani permasalahan klasifikasi pada teks (Ikonomakis, dkk., 2005).

Menurut Sebastiani (2001), *machine learning* merupakan perspektif yang secara otomatis dapat melakukan klasifikasi teks dengan mempelajari sekumpulan dokumen teks yang sebelumnya telah dikategorikan. Metode *machine learning* yang populer dan sangat akurat untuk menangani masalah klasifikasi adalah *Support Vector Machine* (SVM). SVM mencoba menemukan *hyperplane* optimal dalam ruang input sehingga dapat mengklasifikasikan masalah klasifikasi biner atau multikelas dengan benar. SVM dianggap sebagai metode klasifikasi yang tidak terlalu rentan terhadap *overfitting* dibandingkan metode pembelajaran lainnya karena kompleksitas model yang dihasilkan tidak bergantung pada

dimensi ruang fitur, dan dapat menangani ruang fitur yang besar dengan akurasi klasifikasi yang sangat baik. Namun dalam implementasinya, kompleksitas tersebut tidak dapat disesuaikan terhadap jumlah dokumen teks (Nidhi dan Gupta, 2011).

Berbagai macam penelitian telah dilakukan untuk mengklasifikasikan dokumen teks. Ropikoh, dkk., (2021) melakukan klasifikasi terhadap berita hoaks covid-19 menggunakan metode SVM dengan melakukan perbandingan dua fungsi kernel yaitu kernel linear dan kernel *Radial Basis Function* (RBF). Hasil penelitian tersebut menunjukkan nilai akurasi tertinggi oleh kernel linear pada 80% data pelatihan dan 20% data pengujian yaitu sebesar 92,90 %. Sedangkan nilai akurasi kernel RBF pada 90% data pelatihan dan 10% data pengujian yaitu sebesar 90,46%. Penelitian selanjutnya adalah Klasifikasi Berita *Online* Menggunakan *Support Vector Machine* (SVM) dengan Perbandingan Pembobotan Fitur *Term Frequency-Inverse Document Frequency* (TF-IDF) dan *Term Frequency Absolute* (TF-ABS) oleh Apriani, dkk., (2023). Berdasarkan penelitian tersebut diperoleh kesimpulan bahwa TF-IDF dapat memberikan hasil yang unggul apabila digabungkan dengan algoritma SVM dengan nilai akurasi yang dihasilkan sebesar 96,63%, dibandingkan dengan TF-ABS yang hanya sebesar 89,66%. Sedangkan Asiyah dan Fithriasari (2016) dalam penelitiannya membandingkan dua metode untuk klasifikasi berita *online* yaitu SVM dan *K-Nearest Neighbor* (KNN). Pada klasifikasi menggunakan SVM digunakan dua fungsi kernel yaitu kernel linear dan polinomial yang menunjukkan hasil akurasi kernel polinomial lebih besar dibandingkan kernel linear yaitu sebesar 93.2%. Sedangkan dalam metode KNN menggunakan 2-NN, 3-NN, dan 5-NN menunjukkan hasil bahwa akurasi terbesar diperoleh oleh 2-NN yaitu sebesar 83.97%. Kemudian dilakukan perbandingan antara SVM dan KNN yang menunjukkan metode SVM lebih baik dibandingkan dengan KNN dilihat dari nilai akurasi, *precision*, *recall*, dan *F-Measure*.

Penyebaran berita hoaks semakin marak seiring berkembangnya teknologi. Kementerian Komunikasi dan Informasi (Kominfo), pada situs *online* resminya memaparkan bahwa isu hoaks di Indonesia pada triwulan pertama tahun 2023,

dari berbagai situs web dan *platform* digital mencapai angka 425, meningkat 8.142% dibandingkan tahun 2022 pada periode yang sama yaitu sebanyak 393 isu (Kominfo, 2023). Sehubungan dengan maraknya berita hoaks di Indonesia, masyarakat harus lebih selektif serta berusaha memastikan berita yang diterima dan disebarluaskan merupakan fakta (Cahyo dan Aesy, 2023). Dengan demikian, diperlukan suatu metode klasifikasi sebagai salah satu cara yang dapat memisahkan berita ke dalam kelas berita hoaks atau fakta.

Berdasarkan penelitian yang telah dilakukan sebelumnya dan permasalahan yang telah diuraikan, metode SVM cukup efektif digunakan untuk klasifikasi teks. Hal ini dilihat dari baiknya akurasi yang merupakan nilai hasil perbandingan jumlah prediksi benar dengan total prediksi, dan juga dilihat dari nilai *presicion*, *recall*, *f1-score*, dan *confusion matrix*. Pada penelitian ini penulis melakukan klasifikasi terhadap judul berita hoaks menggunakan metode SVM yang berfokus pada pengaplikasian beberapa fungsi kernel.

1.2 Tujuan Penelitian

Adapun tujuan dari penelitian ini antara lain sebagai berikut:

1. Menerapkan metode SVM pada klasifikasi judul berita hoaks.
2. Mengetahui tingkat akurasi pada setiap fungsi kernel yang digunakan pada metode SVM.

1.3 Manfaat Penelitian

Adapun manfaat yang diharapkan penulis dalam penelitian ini antara lain sebagai berikut:

1. Memberikan informasi kepada pembaca terkait judul berita yang termasuk ke dalam berita hoaks atau berita fakta.

2. Memberikan pengetahuan kepada pembaca tentang penerapan metode SVM pada klasifikasi judul berita hoaks.
3. Menjadi bahan evaluasi pembaca untuk lebih bijak akan konsumsi berita di berbagai media khususnya media *online*.

II. TINJAUAN PUSTAKA

2.2 *Machine Learning*

Machine learning atau pembelajaran mesin merupakan pemrograman komputer untuk mengoptimalkan kinerja kriteria menggunakan contoh data atau pengalaman masa lalu, untuk menghasilkan model prediktif sebagai prediksi di masa depan, dan atau model deskriptif untuk memperoleh pengetahuan dari data (Alpaydin, 2020). Dalam *machine learning* terdapat tiga kategori yaitu:

1. *Supervised Learning*

Supervised learning merupakan masalah pembelajaran dimana terdapat input dan output yang sudah diketahui nilai kebenarannya, dan selanjutnya digunakan algoritma untuk mempelajari fungsi pemetaan dari *input* terhadap output tersebut. *Supervised learning* dikelompokkan lebih lanjut dalam masalah klasifikasi ketika variabel output berbentuk kategori dan regresi ketika variabel *output* kontinu.

2. *Unsupervised Learning*

Dalam *unsupervised learning* hanya terdapat data input, tujuannya adalah untuk memodelkan struktur atau distribusi yang mendasarinya dalam data untuk mempelajari lebih lanjut tentang data. Algoritma diserahkan pada rancangannya sendiri untuk ditemukan dan menyajikan struktur menarik dalam data. *Unsupervised learning* dikelompokkan lebih lanjut dalam masalah *clustering* dan asosiasi.

3. *Reinforcement Learning*

Reinforcement learning merupakan salah satu teknik dalam *machine learning* dimana terdapat sebuah perantara menerima sebuah kondisi, selanjutnya

perantara tersebut melakukan sebuah tindakan. Kemudian kondisi tersebut akan menghasilkan perantara baru yang akan menghasilkan *reward*. Proses ini dilakukan secara berulang, hingga perantara belajar memaksimalkan kumulatif *reward* yang didapat dengan kata lain belajar berdasarkan pengalaman yang terjadi pada perantara tersebut.

2.2 *Text Mining*

Text mining atau dapat disebut sebagai *knowledge discovery from text*, pertama kali disebutkan oleh Feldmann dan Dagan pada tahun 1995 sebagai analisis teks yang didukung oleh mesin menggunakan teknik berupa pengambilan informasi, ekstraksi informasi, serta pemrosesan bahasa alami (*natural language processing*), dan selanjutnya akan dihubungkan dengan algoritma *data mining*, *machine learning*, dan statistik. Tujuan dari *text mining* adalah mengungkapkan informasi tersembunyi melalui metode yang mampu mengatasi sejumlah besar kata dan struktur dalam bahasa alami dan memungkinkan untuk menangani ketidakpastian dan ketidakjelasan (Hotho, dkk., 2005).

2.3 **Klasifikasi Teks**

Menurut Darujati dan Gumelar (2012), klasifikasi teks dianggap sebagai proses membentuk golongan atau kelas pada dokumen berdasarkan kelas kelompok yang sudah ditentukan sebelumnya (*supervised*). Tujuan dari klasifikasi teks adalah untuk memperoleh suatu model atau fungsi yang diperoleh dari data pelatihan (*training*) yang dideskripsikan oleh himpunan atribut atau variabel. Salah satu atribut mendeskripsikan kelas yang diikuti oleh suatu contoh hingga disebut atribut kelas, dan atribut lain dinamakan atribut independen atau prediktor. Proses klasifikasi teks terbagi dalam dua fase, yaitu:

1. Fase *Information Retrieval* (IR)

Fase ini bertujuan untuk mendapatkan data numerik dari dokumen teks melalui *feature extraction* dengan berbagai pendekatan salah satunya distribusi frekuensi kata. Nilai numerik yang diperoleh dapat berupa berapa kali suatu kata muncul di dalam dokumen. Nilai 1 menunjukkan kata ada dalam dokumen, dan nilai 0 jika tidak ada kata dalam dokumen. Beberapa pendekatan pada *feature extraction* yaitu seperti menghapus *stop-words*, stemming, statistical filtering, dan sebagainya.

2. Fase Klasifikasi Utama

Fase ini dilakukan ketika algoritma memroses data numerik yang diperoleh pada fase sebelumnya untuk memutuskan ke kategori mana teks ditempatkan. Beberapa algoritma klasifikasi yang dapat diterapkan pada fase ini, di antaranya adalah *naive bayesian*, *rocchio*, *decision tree*, *k-nearest neighbor*, *neural network*, dan *support vector machine*.

2.4 *Preprocessing Data*

Secara umum, *preprocessing data* atau prapemrosesan data merupakan salah satu tahap terpenting dalam *text mining* khususnya dalam klasifikasi teks. Proses ini dapat menghilangkan kebisingan (*noise*) pada teks seperti kesalahan ejaan, pengurangan karakter yang direplikasi, dan akronim yang ambigu (HaCohen-Kerner, dkk., 2020). Berikut adalah beberapa teknik yang digunakan dalam *preprocessing data*:

1. *Text Cleaning*

Text cleaning adalah proses membersihkan teks dari karakter yang bersifat non alfabetis (tidak sesuai dengan abjad) untuk mengurangi *noise*. Proses ini juga membersihkan sekumpulan simbol-simbol seperti tanda baca, dan karakter non alfabetis lainnya (Firmansyah, dkk., 2020).

2. *Case Folding*

Case folding adalah proses mengubah seluruh huruf dalam dokumen menjadi huruf kecil (Firmansyah, dkk., 2020).

3. *Tokenizing*

Tokenizing atau tokenisasi merupakan proses memecah kalimat setelah melalui proses *case folding*. Proses *tokenizing* tersebut dibuat berdasarkan spasi pada kalimat, dan selanjutnya dibentuk *list* yang terdiri dari sekumpulan kata (token) (Firmansyah, dkk., 2020).

4. *Stopword Removal*

Stopword removal merupakan daftar kata umum yang tidak memiliki arti penting dan tidak digunakan. Pada proses ini kata umum akan dihapus untuk mengurangi jumlah kata yang disimpan oleh sistem (Luqyana, dkk., 2018).

5. *Stemming*

Stemming merupakan proses untuk mencari kata dasar (*stem*) yang dihasilkan dari proses penghapusan *stopword*. Terdapat dua aturan dalam melakukan *stemming* yaitu melalui pendekatan kamus dan pendekatan aturan (Luqyana, dkk., 2018).

2.5 Ekstraksi Fitur

Pada klasifikasi, data teks tidak dapat secara langsung digunakan oleh algoritma *machine learning* karena secara umum teks merupakan sekumpulan data yang tidak terstruktur. Oleh karena itu data teks tersebut harus melalui proses ekstraksi fitur, yaitu perubahan dari format tekstual yang tidak terstruktur menjadi terstruktur, sehingga dapat dengan mudah digunakan dan dipahami dalam proses pengklasifikasian ke kelas yang telah ditentukan (Budiman, dkk., 2020). Salah satu pendekatan yang dapat dilakukan untuk ekstraksi fitur adalah *Term Frequency-Inverse Document Frequency* (TF-IDF). Menurut Prihatini (2016), nilai TF dihitung berdasarkan jumlah kemunculan setiap kata dalam tiap dokumen, dan nilai IDF dihitung berdasarkan jumlah kemunculan kata dalam keseluruhan dokumen. Nilai TF dibandingkan terhadap nilai IDF, dan kemudian menghasilkan matriks dimana kolom-kolomnya berisi nilai TF-IDF untuk setiap dokumen. Oleh karena itu, TF-IDF mengurangi dokumen dengan ukuran panjang bervariasi menjadi dokumen dengan ukuran panjang yang tetap.

Berikut merupakan rumus perhitungan TF-IDF :

1. Nilai persamaan untuk TF:

$$TF_{(i,j)} = f_{(i,j)} \quad (2.1)$$

2. Nilai persamaan untuk IDF:

$$IDF_{i,j} = \log \frac{N}{df_i} \quad (2.2)$$

3. Nilai persamaan untuk TF-IDF:

$$TFIDF = TF \times IDF \quad (2.3)$$

dengan:

$f_{(i,j)}$: banyaknya kemunculan kata i pada dokumen j

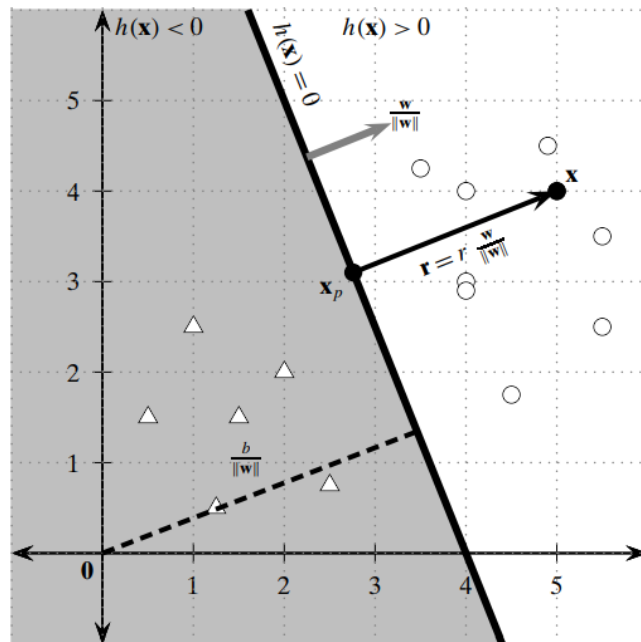
N : jumlah dokumen

df_i : jumlah dokumen yang mengandung kata i

2.6 Support Vector Machine

Support Vector Machine (SVM) diperkenalkan oleh Vapnik pada tahun 1992 sebagai rangkaian harmonis konsep-konsep unggulan pada *pattern recognition*, yang merupakan pemetaan suatu data dalam kelas atau kategori yang telah didefinisikan sebelumnya (Nugroho, dkk., 2003). SVM bertujuan untuk menemukan *hyperplane* optimal yang memiliki margin maksimal, yaitu jarak antara *hyperplane* dengan titik terdekat pada setiap kelas. Menurut Wang, dkk., (2006), SVM merupakan salah satu metode dengan hasil terbaik untuk klasifikasi teks. SVM adalah pengklasifikasi biner universal yang mampu menemukan fungsi ambang batas linear maupun non linear untuk memisahkan contoh kategori tertentu berdasarkan prinsip *Structural Minimization Risk* (SRM) dari teori pembelajaran komputasi, dan memiliki kemampuan mengelola fitur dalam jumlah besar dan kemampuan generalisasi yang tinggi.

Hyperplane terbaik antara kedua kelas dapat ditemukan dengan mengukur margin atau jarak terdekat antara *hyperplane* dengan anggota terdekat dari masing-masing kelas, yang disebut *support vector*. Usaha untuk mencari *hyperplane* ini merupakan inti dari proses pembelajaran pada SVM (Nugroho, dkk., 2003).



Gambar 1. Konsep Hyperplane pada SVM

Berdasarkan Gambar 1 terdapat dua pola yang menunjukkan kelas atau kategori pada data, yaitu pola lingkaran yang menunjukkan kelas pertama yaitu $+1$ dan pola segitiga yang menunjukkan kelas kedua yaitu -1 . Hyperplane $h(x) = 0$ membagi dua ruang, dimana wilayah yang diarsir terdiri dari semua titik x yang memenuhi $h(x) < 0$, sedangkan wilayah yang tidak diarsir terdiri dari semua titik yang memenuhi $h(x) > 0$. Vektor bobot ditunjukkan oleh $\frac{w}{\|w\|}$ dan ortogonal terhadap hyperplane. Sedangkan $\frac{b}{\|w\|}$ merupakan jarak terarah dari daerah asal ke *hyperplane* (Zaki dan Meira, 2014).

Dimisalkan data pelatihan D dimana setiap input x_i berdimensi d dan berada di salah satu dari dua kelas $y_i \in \{-1, 1\}$, maka bentuk data pelatihan:

$$D = \{x_i, y_i\}, i = 1, \dots, n, y_i \in \{-1, 1\}, x \in R^d$$

Sebuah *hyperplane* pada dimensi d diberikan sebagai himpunan semua titik $\mathbf{x} \in R^D$ yang memenuhi persamaan $h(\mathbf{x}) = 0$, dimana $h(\mathbf{x})$ adalah fungsi *hyperplane* (Zaki dan Meira, 2014). Yang didefinisikan sebagai berikut:

$$h(\mathbf{x}) = (\mathbf{w}^T \cdot \mathbf{x}) + b = 0 \quad (2.4)$$

Hyperplane membagi ruang dimensi d menjadi dua ruang, dan dapat dikatakan terpisahkan secara linear jika setiap ruang hanya mempunyai titik-titik dari satu kelas saja. Berdasarkan *hyperplane* $h(\mathbf{x}) = 0$, untuk semua titik berlabel $y_i = -1$, berlaku $h(\mathbf{x}_i) \leq 0$, dan untuk semua titik berlabel $y_i = +1$, berlaku $h(\mathbf{x}_i) \geq 0$. Maka diperoleh:

$$\mathbf{w}^T \cdot \mathbf{x}_i = -b \geq +1, \text{ untuk } y_i = +1 \quad (2.5)$$

$$\mathbf{w}^T \cdot \mathbf{x}_i = -b \leq -1, \text{ untuk } y_i = -1 \quad (2.6)$$

Persamaan (2.6) dan (2.7) dapat digabungkan menjadi:

$$y_i(\mathbf{w}^T \cdot \mathbf{x}_i - b) - 1 \geq 0, \mathbf{x}_i \in D \text{ untuk } i = 1, 2, \dots, n \quad (2.7)$$

Pendekatan SVM linear dapat digunakan untuk kumpulan data dengan batasan keputusan nonlinier melalui *kernel trick* (Zaki dan Meira, 2014). Konsepnya adalah memetakan titik-titik asli \mathbf{x}_i berdimensi D di ruang masukan ke titik $\phi(\mathbf{x}_i)$ dalam ruang fitur berdimensi tinggi melalui beberapa transformasi nonlinear. Untuk menerapkan *kernel trick* pada klasifikasi SVM nonlinear, kita harus menunjukkan bahwa semua operasi hanya memerlukan fungsi kernel:

$$K(\mathbf{x}_i, \mathbf{x}_j) = \phi(\mathbf{x}_i)^T \phi(\mathbf{x}_j) \quad (2.8)$$

dengan:

K : fungsi kernel

ϕ : fungsi pemetaan ke ruang fitur tinggi

Pada data $D = \{\mathbf{x}_i, y_i\}_{i=1}^n$ akan diterapkan ϕ di setiap titik, dan didapatkan kumpulan data baru di ruang fitur $D_\phi = \{\phi(\mathbf{x}_i), y_i\}_{i=1}^n$.

Nilai bobot w dihitung dengan persamaan:

$$w = \sum_{i=1}^n \alpha_i y_i \phi(\mathbf{x}_i), \alpha_i > 0 \quad (2.9)$$

Sementara itu bias b didapatkan dari persamaan:

$$b = y_i - \sum_{i=1}^n \alpha_j y_j K(\mathbf{x}_j, \mathbf{x}_i) \quad (2.10)$$

Selanjutnya kita dapat memprediksi kelas untuk titik z yang baru sebagai berikut:

$$\hat{y} = \text{sgn} \left(\sum_{\alpha_i > 0} \alpha_i y_i K(\mathbf{x}_i, z) + b \right) \quad (2.11)$$

Beberapa fungsi kernel yang secara umum digunakan pada metode SVM (Putra, dkk., 2023) antara lain sebagai berikut:

1. Kernel Linear

$$K(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i \cdot \mathbf{x}_j \quad (2.12)$$

2. Kernel Polinomial

$$K(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i \cdot \mathbf{x}_j + r)^d \quad (2.13)$$

3. Kernel *Radial Basis Function* (RBF) atau Gaussian Kernel

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp \frac{(-\|\mathbf{x}_i - \mathbf{x}_j\|^2)}{2\sigma^2} \quad (2.14)$$

4. Kernel Sigmoid

$$K(\mathbf{x}_i, \mathbf{x}_j) = \tanh(\sigma(\mathbf{x}_i \cdot \mathbf{x}_j) + r) \quad (2.15)$$

dengan:

K : fungsi kernel

$\mathbf{x}_i, \mathbf{x}_j$: vektor fitur \mathbf{x}_i dan \mathbf{x}_j

d : derajat polinomial

\exp : basis logaritma natural

sgn : fungsi tanda

σ : parameter lebar kernel :

r : konstanta

Proses sederhana untuk menemukan *hyperplane* yang terbaik dikembangkan dalam SVM dengan cara merumuskannya ke dalam *Quadratic Programming* (QP) (Vijayakumar & Wu, 1999). Langkah-langkah pada *Sequential Training* adalah sebagai berikut:

1. Menginisialisasi nilai awal parameter alpha (α), gamma (γ), $Cost(C)$, epsilon (ϵ), lambda (λ), dan iterasi maksimum (i_{max}).
2. Menghitung kernel dengan persamaan:

$$K(\mathbf{D}_i, \mathbf{x}_j) = \phi(\mathbf{D}_i)^T \phi(\mathbf{D}_j) \quad (2.16)$$

3. Menghitung nilai $\mathbf{y}_i \mathbf{y}_j$ berdasarkan label kelas yang sebelumnya telah ditentukan.
4. Menghitung nilai matriks hessian dalam persamaan:

$$\mathbf{D}_{ij} = \mathbf{y}_i \mathbf{y}_j (K(\mathbf{D}_i, \mathbf{D}_j) + \lambda^2) \quad (2.17)$$

dengan:

\mathbf{D}_{ij} : matriks *Hessian*

\mathbf{y}_i : kelas data ke- i

\mathbf{y}_j : kelas data ke- j

$K(\mathbf{D}_i, \mathbf{D}_j)$: fungsi kernel

5. Menghitung nilai *error rate* dalam persamaan:

$$E_i = \sum_j^i \alpha_i \mathbf{D}_{ij} \quad (2.18)$$

dengan:

E_i : nilai *error rate* data ke- i

6. Menghitung nilai delta alpha dalam persamaan:

$$\delta \alpha_i = \min (\max [\gamma(10 - E_i), -\alpha_i], C - \alpha_i) \quad (2.19)$$

dengan:

$\delta \alpha_i$: nilai delta alpha data ke- i

7. Menghitung nilai alpha baru untuk digunakan dalam iterasi selanjutnya.

$$\alpha_i + \delta\alpha_i \quad (2.20)$$

8. Melakukan proses iterasi sampai data training mencapai nilai konvergen ($\max(|\delta\alpha_i|) < \varepsilon$).

9. Menghitung nilai $w \cdot x^0$ dan $w \cdot x^1$ untuk memperoleh nilai bias

$$w \cdot x^0 = \alpha_i y_i K(\mathbf{x}_i, \mathbf{x}^0) \quad (2.21)$$

$$w \cdot x^1 = \alpha_i y_i K(\mathbf{x}_i, \mathbf{x}^1) \quad (2.22)$$

maka:

$$b = -\frac{1}{2}(w \cdot x^0 + w \cdot x^1) \quad (2.23)$$

dengan:

$w \cdot x^0$: nilai bobot data x dengan kelas 0

$w \cdot x^1$: nilai bobot data x dengan kelas 1

b : nilai bias

10. Menghitung kernel dan nilai bobot pada data *testing* untuk menghitung fungsi klasifikasi, dan selanjutnya menentukan kelas pada data *testing* yang ditunjukkan pada Persamaan 2.24.

$$f(x) = \operatorname{sgn} \sum_{i=1}^4 (\alpha_i y_i K(\mathbf{U}_i, \mathbf{D}_j) + b) \quad (2.24)$$

2.7 Evaluasi Model

Evaluasi model pada klasifikasi bertujuan untuk menguji apakah suatu model memiliki kinerja yang baik atau tidak pada proses klasifikasi. Model klasifikasi yang dibuat adalah pemetaan suatu data dengan output berupa prediksi kelas atau target dari data tersebut. Beberapa indikator yang dapat dilihat untuk evaluasi model diantaranya *confusion matrix* dan *classification report*.

2.7.1 *Confusion Matrix*

Menurut Sani, dkk., (2022), *confusion matrix* memberikan informasi yang dapat digunakan untuk merangkum kinerja dari klasifikasi yang sehubungan dengan beberapa data uji dengan memvisualisasikan matriks dua dimensi, dimana satu dimensi mewakili kelas sebenarnya dari suatu objek dan dimensi lain mewakili kelas yang diprediksi oleh metode klasifikasi.

Tabel 1. *Confusion Matrix*

		Nilai Prediksi	
		Positif (P)	Negatif (N)
Nilai Aktual	Positif (P)	<i>True Positive (TP)</i>	<i>False Positive (FP)</i>
	Negatif (N)	<i>False Negative (FN)</i>	<i>True Negative (TN)</i>

Dengan empat kemungkinan yaitu *True Positive (TP)* menunjukkan data positif yang menghasilkan prediksi positif, *False Positive (FP)* menunjukkan data positif yang menghasilkan prediksi negatif, *True Negative (TN)* menunjukkan data negatif yang menghasilkan prediksi negatif, dan *False Negative (FN)* menunjukkan data negatif yang menghasilkan prediksi positif.

2.7.2 *Classification Report*

Classification report merupakan visualisasi untuk mengukur kualitas performa prediksi dari metode klasifikasi dengan menampilkan sejumlah indikator pengukuran antara lain sebagai berikut:

1. *Precision*

Precision atau presisi digunakan untuk menghitung berapa rasio atau keakuratan prediksi yang dilakukan oleh model adalah benar.

$$precision = \frac{TP}{TP + FP} \quad (2. 25)$$

2. *Recall*

Recall merupakan kemampuan metode klasifikasi untuk menemukan semua kasus positif.

$$recall = \frac{TP}{TP + FN} \quad (2. 26)$$

3. *F1-Score*

F1-score merupakan pembobotan dari rata-rata harmonik nilai *precision* dan *recall*.

$$F1-score = \frac{2 \times precision \times recall}{precision + recall} \quad (2. 5)$$

4. *Accuracy*

Accuracy digunakan untuk menghitung rasio dari prediksi yang benar, baik positif maupun negatif dengan keseluruhan data yang ada.

$$accuracy = \frac{(TP + TN)}{TP + FP + TN + FN} \quad (2. 28)$$

2.8 Berita Hoaks

Menurut Kamus Besar Bahasa Indonesia (KBBI), hoaks berarti informasi bohong atau informasi yang lalu lalang. Pengertian tersendiri oleh masyarakat terkait hoaks berdasarkan survei yang dilakukan oleh Masyarakat telematika Indonesia (Mastel) yaitu 88% memilih pengertian hoaks sebagai berita bohong yang sengaja dibuat, 49% menjawab berita yang menghasut, 15% menyebutkan hoaks merupakan berita ramalan atau yang berkaitan dengan fiksi ilmiah, 14% menyebutkan sebagai berita yang membuat pemerintah merasa tersudutkan, 31%

menganggap hoaks sebagai berita yang dibuat untuk menjelekan orang lain dan 1% lainnya menjawab tidak tahu (Survei Mastel, 2019).

Menurut Hadi (2023), penyebab munculnya berita hoaks antara lain sebagai berikut:

1. Keinginan masyarakat untuk mencari sensasi dan perhatian orang lain.
2. Minimnya penindakan hukum untuk menanggulangi masalah penyebaran berita hoaks.
3. Merupakan lahan bisnis sebagai dampak samping tujuan sosial politik.
4. Merupakan alat kampanye hitam yang dibuat untuk mencera pihak lain dengan tujuan memberikan kesan buruk atas personal seseorang.
5. Untuk memunculkan teror secara psikologis sehingga publik resah dan cemas.

Saat ini berita hoaks banyak beredar di berbagai platform media sosial seperti *twitter, facebook, instagram, whatsapp* dan blog-blog tertentu, dan dalam berbagai segi kehidupan masyarakat yaitu sosial, ekonomi, hukum, dan politik. Dampak buruk penyebaran berita hoaks sangat mengkhawatirkan. Berita hoaks tak jarang memuat isu-isu sensitif dalam berbagai media yang berpotensi menyinggung suku, agama, ras dan antar golongan (SARA) dan dapat menimbulkan keresahan bahkan perpecahan dalam masyarakat (Hadi, 2023).

III. METODOLOGI PENELITIAN

3.1 Waktu dan Tempat Penelitian

Penelitian ini dilaksanakan pada perkuliahan semester ganjil tahun akademik 2023/2024 bertempat di Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.

3.2 Data Penelitian

Data yang digunakan dalam penelitian ini adalah data sekunder judul berita hoaks Indonesia yang berjumlah 1000 data yang bersumber dari situs resmi Kaggle dengan link <https://www.kaggle.com/datasets/linkgish/indonesian-fact-and-hoax-political-news>. Terdapat 4 informasi pada data yaitu *url* berita, tanggal terbit berita, judul berita, dan label yang menunjukkan kelas pada data. Angka 0 menunjukkan bahwa berita tersebut merupakan berita bukan hoaks, sedangkan angka 1 menunjukkan berita tersebut merupakan berita hoaks. Tabel 2 menyajikan data yang digunakan pada penelitian ini.

Tabel 2. Data Penelitian

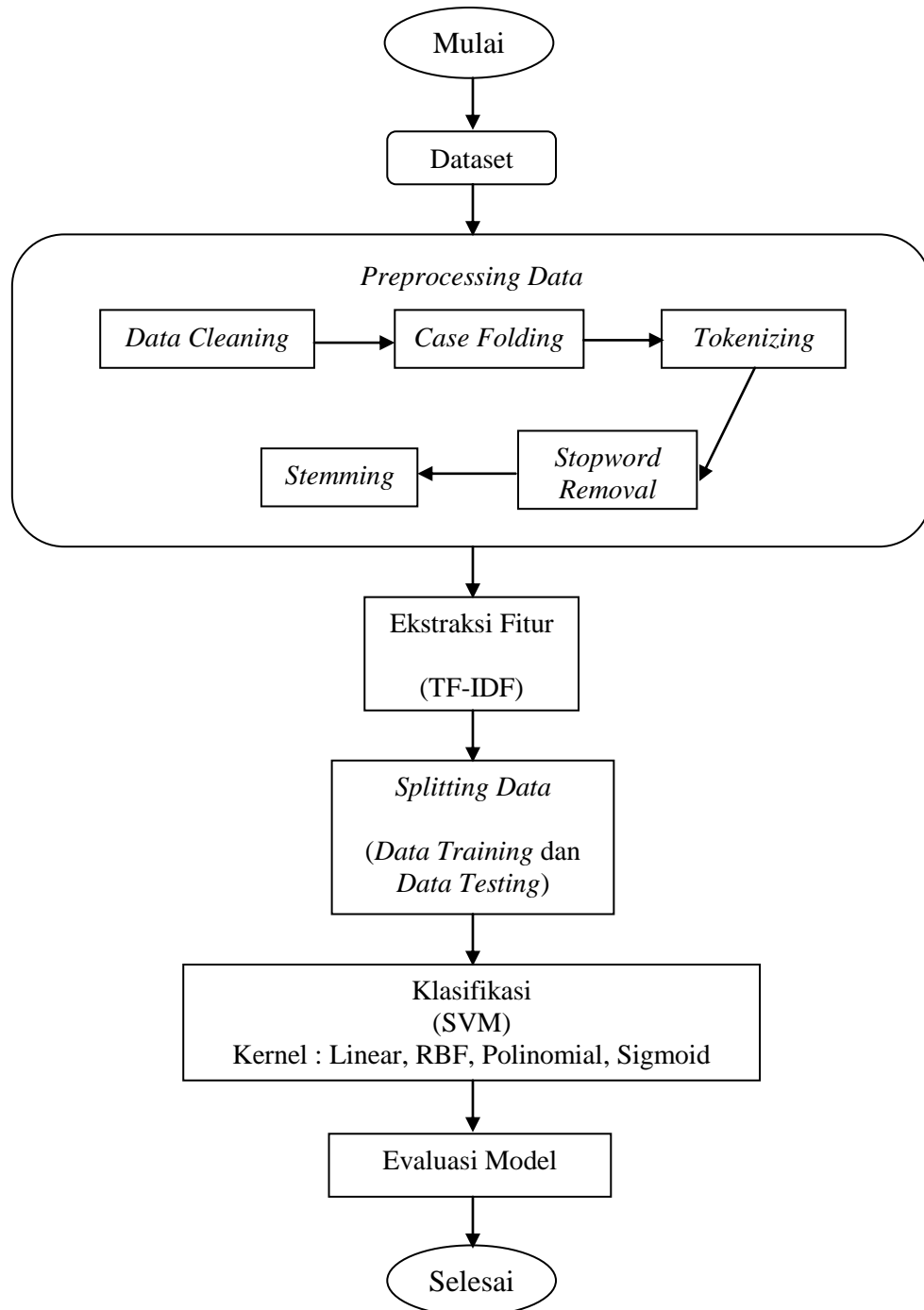
	Judul Berita	Tanggal	Url	Label
0	“Australia keok tidak mampu hadapi Indonesia, minta bantuan ke iran”	2023-01-03	https://turnbackhoax.id/2023/01/03/salah-video-australia-keok-gak-mampu-hadapi-indonesia-minta-bantuan-ke-iran/	1
1	“Tahun Baru Mencekam Malaysia !!Serangan Darat Besar Besaran Pasukan Indonesia Terhadap Malaysia”	2023-01-03	https://turnbackhoax.id/2023/01/03/salah-foto-tahun-baru-mencekam-malaysia-serangan-darat-besar-besaran-pasukan-ri-terhadap-malaysia/	1
2	Daftar Kegentingan Memaksa yang membuat Jokowi Terbitkan Perppu Ciptaker	2023-01-03	https://www.cnnindonesia.com/nasional/20230103122823-32-895465/daftar-kegentingan-memaksa-yang-bikin-jokowi-terbitkan-perppu-ciptaker	0
3	Menilik Janji Kampanye Ganjar untuk Atasi Banjir Semarang dan Jateng	2023-01-03	https://nasional.tempo.co/read/1675404/menilik-janji-kampanye-ganjar-pranowo-untuk-atasi-banjir-semarang-dan-jawa-tengah	0
...
999	Luhut Akhirnya Ditangkap! Ternyata Hasil Kekayaannya Adalah Hasil Merugikan Negara	2022-12-31	https://turnbackhoax.id/2022/12/31/salah-luhut-akhirnya-ditangkap-ternyata-hasil-kekayaannya-adalah-hasil-merugikan-negara/	1
1000	Kas Negara Menipis Maruf Minta Masyarakat Bantu Pemerintah	2022-12-31	https://turnbackhoax.id/2022/12/31/salah-kas-negara-menipis-maruf-amin-minta-masyarakat-bantu-pemerintah/	1

3.3 Metode Penelitian

Penelitian ini menggunakan metode SVM untuk mengklasifikasikan judul berita hoaks, selanjutnya mencari nilai terbaik pada beberapa fungsi kernel yang digunakan, berdasarkan nilai akurasi, *presicion*, *recall*, *f1-score*, dan *confusion matrix*, dengan bantuan *software Python*. Adapun tahapan-tahapan dalam penelitian ini antara lain sebagai berikut:

1. Melakukan input data teks judul berita Indonesia yang sudah dilabeli untuk selanjutnya dilakukan klasifikasi dengan metode SVM.
2. Melakukan prapemrosesan data, bertujuan mengubah data yang semula tidak terstruktur menjadi terstruktur. Adapun tahapannya sebagai berikut:
 - a. *Data cleaning*, yaitu proses penghapusan kata dan karakter yang tidak diperlukan seperti *url*, *tag*, angka, tanda baca, dan simbol dalam data akan dihilangkan.
 - b. *Case folding*, yaitu mengubah seluruh data teks menjadi huruf kecil (*lowercase*).
 - c. *Stopword removal*, yaitu menghapus kata umum yang tidak bermakna.
 - d. *Tokenizing*, yaitu proses pemecahan kalimat menjadi kata.
 - e. *Stemming*, yaitu proses mencari kata dasar yang dihasilkan dari proses *stopword removal*
3. Melakukan ekstraksi fitur, bertujuan untuk mengubah data dalam bentuk teks menjadi data numerik. Proses ekstraksi fitur dilakukan dengan menggunakan metode TF-IDF.
4. *Splitting data*, yaitu membagi data pelatihan (*training*) dan data pengujian (*testing*). Dengan persentase:
 - a. 80% *training*, dengan 20% *testing*
 - b. 90% *training*, dengan 10% *testing*
 - c. 95% *training*, dengan 5% *testing*
5. Membuat model klasifikasi dengan metode SVM menggunakan kernel linear, RBF, sigmoid, dan polinomial.
6. Mengevaluasi model berdasarkan *confusion matrix*, dan *classification report*

Adapun tahapan dari penelitian ini digambarkan dalam diagram alir sebagai berikut:



Gambar 2. Diagram Alir Penelitian

V. KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan, dapat diambil kesimpulan dalam klasifikasi judul berita hoaks politik menggunakan *Support Vector Machine* (SVM), antara lain sebagai berikut:

1. Pada penelitian ini dilakukan klasifikasi judul berita hoaks politik menggunakan metode SVM dengan menggunakan 4 fungsi kernel, yaitu kernel linear, kernel RBF, kernel polinomial, dan kernel sigmoid yang terdiri dari dua kelas yaitu kelas berita bukan hoaks dan kelas berita hoaks. Data yang digunakan adalah data sekunder judul berita politik yang bersumber dari *platform* Kaggle. Banyak data yang digunakan adalah 1000 data yang terdiri dari 540 data judul berita politik yang masuk ke dalam kelas berita bukan hoaks dan 460 data, judul berita politik yang masuk ke dalam kelas berita hoaks. Rasio perbandingan data *training* dan data *testing* yang digunakan adalah 80%:20%, 90%:10%, dan 95%:5%.
2. Nilai akurasi terbaik diperoleh pada klasifikasi menggunakan fungsi kernel RBF, dengan perbandingan data *training* dan data *testing* sebesar 80%:90%, dan pada perbandingan 95%:5% dengan nilai akurasi sebesar 92%. Selain itu melalui *confusion matrix* dapat dilihat jika hasil model klasifikasi menunjukkan sebanyak 184 judul berita dapat diklasifikasikan dengan benar sedangkan 16 judul berita lainnya termasuk ke dalam kesalahan proses klasifikasi.

5.2 Saran

Adapun saran yang dapat penulis harapkan untuk penelitian selanjutnya antara lain sebagai berikut:

1. Menggunakan perbandingan parameter lain dalam melakukan perhitungan manual pada proses *training* dan membandingkan kinerja dari setiap parameter.
2. Melakukan perbandingan fungsi kernel pada tahap perhitungan manual, sehingga dapat dilihat perbedaan hasil dari beberapa fungsi kernel yang digunakan.
3. Menjelaskan bagaimana pengaruh dalam penggunaan parameter dan fungsi kernel yang berbeda dalam perhitungan manual pada proses *training*.

DAFTAR PUSTAKA

- Alpaydin, E. 2020. *Adaptive Computation and Machine Learning: Introduction to Machine Learning*, Second Edition. London, England: The MIT press.
- Apriani, I., Sibaroni, Y., dan Palupi, I. 2023. Perbandingan Pembobotan Fitur TF-IDF dan TF-ABS Dalam Klasifikasi Berita Online Menggunakan Support Vector Machine (SVM). *e-Proceeding of Engineering*. **10**(3): 3652-3663.
- Asiyah, S. N., dan Fithriasari, K. 2016. Klasifikasi Berita Online Menggunakan Metode Support Vector Machine dan K-Nearest Neighbor. *Jurnal Sains dan Seni ITS*. **5**(2): 317-322.
- Budiman, I., Faisal, M. R., dan Nugrahadi, D. T. 2020. Studi Ekstraksi Fitur Berbasis Vektor Word2Vec pada Pembentukan Fitur Berdimensi Rendah. *Jurnal Komputasi*. **8**(1): 62-69.
- Cahyo, P. W., dan Aesy, U. S. 2023. Perbandingan LSTM dengan Support Vector Machine dan Multinomial Naive Bayes pada Klasifikasi Kategori Hoax. *Jurnal Transformatika*. **20**(2): 23-29.
- Darujati, C., dan Gumelar, A. B. 2012. Pemanfaatan Teknik Supervised untuk Klasifikasi Teks Bahasa Indonesia. *Jurnal Bandung Text Mining*. **16**(1): 1-8.

- Firmansyah, M. R., Ilyas, R., dan Kasyidi, F. 2020. Klasifikasi Kalimat Ilmiah Menggunakan Recurrent Neural Network. *Prosiding Industrial Research Workshop and National Seminar*. **11**(1): 488-495. Bandung, 26-27 Agustus 2020.
- Guo, X., Yin, Y., Dong, C., Yang, G., and Zhou, G. 2008. On the Class Imbalance Problem. *Fourth international conference on natural computation*, China: School of Computer Science and Technology, Shandong University, Jinan.
- HaCohen-Kerner, Y., Miller, D., and Yigal, Y. 2020. The influence of preprocessing on text classification using a bag-of-words representation. *PloS one*. **15**(5): 1-22.
- Hadi, S. P. 2023. Pemahaman Literasi Digital Penyebab Munculnya Hoaks. *Jurnal Karya untuk Masyarakat*. **4**(2): 150-160.
- Harjanta, A. T. J. 2015. Preprocessing Text untuk Meminimalisir Kata yang Tidak Berarti dalam Proses Text Mining. *Jurnal Informatika UPGRI*. **1**(1): 1-9.
- Hasil Survey Wabah Hoax Nasional. 2019. (<https://mastel.id/wp-content/uploads/2019/04/Survey-Hoax-Mastel-2019-10-April-2019.pdf>, diakses 5 November 2023).
- Hotho, A., Nürnberger, A., and Paass, G. 2005. A Brief Survey of Text Mining. *Journal for Language Technology and Computational Linguistics*. **20**(1): 19-62.
- Ikonomakis, M., Kotsiantis, S., and Tampakas, V. 2005. Text Classification Using Machine Learning Techniques. *WSEAS Transactions on Computers*. **4**(8): 966-974.

- Kim, S. W., and Gil, J. M. 2019. Research paper classification systems based on TF-IDF and LDA schemes. *Human-centric Computing and Information Sciences*. **9**(30): 1-21.
- Luqyana, W. A., Cholissodin, I., dan Perdana, R. S. 2018. Analisis Sentimen Cyberbullying Pada Komentar Instagram dengan Metode Klasifikasi Support Vector Machine. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*. **2**(11): 4704-4713.
- Nidhi, and Gupta, V. 2011. Recent Trends in Text Classification Techniques . *International Journal of Computer Application*. **35**(6): 45-51.
- Nugroho, A. S., Witarto, A. B., dan Handoko W. 2003. Support Vector Machine. *Proceeding Indones. Sci. Meeting*. Jepang, 20 Desember 2003.
- Putra, P. R. B., Indriati, dan Perdana, R. S. 2023. Klasifikasi Judul Berita Online menggunakan Metode Support Vector Machine (SVM) dengan Seleksi Fitur Chi-square. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*. **7**(5): 2132-2141.
- Prihatini, P. M. 2016. Implementasi Ekstraksi Fitur Pada Pengolahan Dokumen Berbahasa Indonesia. *Jurnal Matrix*. **6**(3): 174-177.
- Ropikoh, I. A., Abdulhakim, R., Enri, U., dan Sulistiyowati, N. 2021. Penerapan Algoritma Support Vector Machine (SVM) untuk Klasifikasi Berita Hoax Covid-19. *Journal of Applied Informatics and Computing*. **5**(1): 64-73.
- Sani, R. R., Pratiwi, Y. A., Winarno, S., Udayanti, E. D., dan Alzami, F. 2022. Analisis Perbandingan Algoritma Naive Bayes Classifier dan Support Vector Machine untuk Klasifikasi Berita Hoax pada Berita Online Indonesia. *Jurnal Masyarakat Informatika*. **13**(2): 85-98.

- Sebastiani, F. 2001. Machine Learning in Automated Text Categorization. *ACM Computing Surveys (CSUR)*. **34**(1): 1-55.
- Triwulan Pertama 2023, Kominfo Identifikasi 425 Isu Hoaks. 2023. (https://www.kominfo.go.id/content/detail/48363/siaran-pers-no-50hmkominfo042023-tentang-triwulan-pertama-2023-kominfo-identifikasi-425-isu-hoaks/0/siaran_pers, diakses 13 Oktober 2023).
- Vijayakumar, Sethu, and Si Wu. 1999. Sequential Support Vector Classifiers and Regression. *IIA/SOCO*.
- Wang, Z. Q., Sun, X., Zhang, D. X., and Li, X. 2006. An optimal SVM-based text classification algorithm. In *2006 International Conference on Machine Learning and Cybernetics*. 1378-1381.
- Zaki, M. J., and Meira, W. 2014. *Data mining and analysis: fundamental concepts and algorithms*. Cambridge University Press.
- Zhang, W., Yoshida, T., and Tang X. 2008. Text Classification Based On Multi-Word With Support Vector Machine. *Knowledge-Based Systems*, **21**(8): 879-886.