

**ANALISIS SENTIMEN PADA JUDUL BERITA ONLINE EKONOMI DENGAN  
MENGUNAKAN *SUPPORT VECTOR MACHINE***

**SKRIPSI**

Oleh :  
**ANDHIKA MAULANA EFFENDI**  
NIM. 18650078



**PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS SAINS DAN TEKNOLOGI  
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM  
MALANG  
2023**

**ANALISIS SENTIMEN PADA JUDUL BERITA ONLINE EKONOMI  
DENGAN MENGGUNAKAN *SUPPORT VECTOR MACHINE***

**SKRIPSI**

Oleh :  
**ANDHIKA MAULANA EFFENDI**  
**NIM. 18650078**

**Diajukan Kepada:  
Fakultas Sains dan Teknologi Universitas Islam Negeri (UIN) Malang  
untuk Memenuhi Salah Satu Persyaratan Dalam  
Memperoleh Gelar Sarjana Komputer (S.Kom)**

**PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS SAINS DAN TEKNOLOGI  
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM  
MALANG  
2023**

**HALAMAN PERSETUJUAN**

**ANALISIS SENTIMEN PADA JUDUL BERITA ONLINE EKONOMI  
DENGAN MENGGUNAKAN *SUPPORT VECTOR MACHINE***

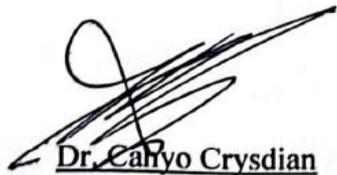
**SKRIPSI**

**Oleh :**

**ANDHIKA MAULANA EFFENDI**  
**NIM. 18650078**

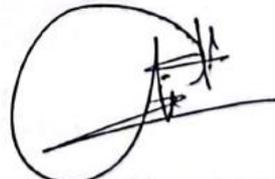
Telah Diperiksa dan Disetujui untuk Diuji  
Tanggal : 25 Mei 2023

**Pembimbing I**



**Dr. Cahyo Crysdian**  
NIP. 19740424 200901 1 008

**Pembimbing II**



**Ajib Hanani, M.T**  
NIDT. 19840731 20160801 1 076

**Mengetahui,**  
**Ketua Program Studi Teknik Informatika**  
**Fakultas Sains dan Teknologi**  
**Universitas Islam Negeri Maulana Malik Ibrahim Malang**



**Dr. Fachrul Kurniawan, M.MT**  
NIP. 1971020 200912 1 001

HALAMAN PENGESAHAN

ANALISIS SENTIMEN PADA JUDUL BERITA ONLINE EKONOMI  
DENGAN MENGGUNAKAN *SUPPORT VECTOR MACHINE*

SKRIPSI

Oleh:

**ANDHIKA MAULANA EFFENDI**  
**NIM. 18650078**

Telah Dipertahankan di Depan Dewan Penguji  
dan Dinyatakan Diterima Sebagai Salah Satu Persyaratan  
untuk Memperoleh Gelar Sarjana Komputer (S.Kom)  
Pada Tanggal: 09 Juni 2023

**Susunan Dewan Penguji**

Ketua Penguji	:	<u>Dr. Irwan Budi Santoso, M.Kom</u>	(		)
		NIP. 19770103 201101 1 004			
Anggota Penguji I	:	<u>Fajar Rohman Hariri, M.Kom</u>	(		)
		NIP. 19890515 201801 1 001			
Anggota Penguji II	:	<u>Dr. Cahyo Crysdian</u>	(		)
		NIP. 19740424 200901 1 008			
Anggota Penguji III	:	<u>Ajib Hanani, M.T</u>	(		)
		NIDT. 19840731 20160801 1 076			

Mengetahui,  
Ketua Program Studi Teknik Informatika  
Fakultas Sains dan Teknologi  
Universitas Islam Negeri Maulana Malik Ibrahim Malang

  
  
**Dr. Fachrul Kurniawan, M.MT**  
**NIP. 19771020 200912 1 001**

## PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan di bawah ini :

Nama : Andhika Maulana Effendi  
NIM : 18650078  
Fakultas : Sains dan Teknologi  
Jurusan : Teknik Informatika  
Judul Skripsi : Analisis Sentimen Pada Judul Berita Online Ekonomi Dengan Menggunakan *Support Vector Machine*

Menyatakan dengan sebenarnya bahwa Skripsi yang saya tulis ini benar-benar hasil karya saya sendiri, bukan merupakan pengambilalihan data, tulisan atau pikiran orang lain yang saya akui hasil tulisan atau pikiran saya sendiri, kecuali dengan mencantumkan sumber cuplikan pada daftar pustaka.

Apabila di kemudian hari terbukti atau dapat dibuktikan Skripsi ini hasil jiplakan, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Malang, 26 Juni 2023  
Yang Membuat pernyataan,



Andhika Maulana Effendi  
18650078

## **HALAMAN MOTTO**

**“Jangan takut berbuat salah, takutlah jika tidak bisa belajar dari kesalahan”**

**“Jangan tenggelam dalam rasa sesal atas kejadian masa lalu. Andai waktu dapat diputar, tetaplah melangkah maju, terima konsekuensi, dan perbaiki”**

**“Hargai selagi ada, usahakan selagi bisa”**

## HALAMAN PERSEMBAHAN

Alhamdulillah, Puji syukur kehadirat Allah Subhanahu Wa Ta'ala, sholawat serta salam bagi Rasulullah Muhammad Shalallahu Alaihi Wassallam.

Dengan segenap hati penulis mempersembahkan karya ini kepada :

Keluarga yang sangat disayangi dan dicintai, Bapak Suefendi, Ibu Julinarwati, Adik Zaki Adyatma Effendi, dan Adik Keysa Tsaqila Ayundhiza Effendi yang senantiasa memberi dukungan baik dalam bentuk semangat, materi, do'a, serta rasa kasih sayang dan pengertian kepada penulis yang tiada hentinya, sehingga penulis tetap kuat dalam menjalani kehidupan hingga titik ini.

Bapak Dr. Cahyo Crysidian selaku dosen pembimbing I yang telah membimbing dan memberikan jalan keluar dari keluh kesah serta permasalahan yang dihadapi oleh penulis selama menjalani tahap skripsi. Bapak Ajib Hanani, M.T selaku dosen pembimbing II yang telah memberi dukungan dan arahan secara agamis, serta kiat-kiat sukses dalam menjalani tahap skripsi.

Bapak Irwan Budi Santoso, M.Kom dan Bapak Fajar Rohman Hariri, M.Kom selaku dosen penguji I dan dosen penguji II yang telah membimbing dan memberi masukan sehingga penulis memperoleh hasil skripsi yang lebih baik. Seluruh Dosen Program Studi Teknik Informatika yang telah memberikan ilmunya selama masa perkuliahan sehingga peneliti mendapatkan pengetahuan yang lebih luas dan In sya Allah menjadikan manfaat bagi orang sekitar.

Dimas Agung Surya Mahendra selaku teman, sahabat dan tetangga penulis sekaligus Bapak Taufik Hendrasno dan Ibu Iin Priharyuni selaku orang tua dari Dimas yang juga ikut andil dalam keberlangsungan skripsi ini.

Saudara seperjuangan mahasiswa Teknik Informatika angkatan 2018 (*Unity of Informatics Force*) yang telah menemani penulis dari fase mahasiswa baru hingga tahap ini.

Pihak-pihak yang secara langsung maupun tidak langsung membantu dalam proses penyelesaian karya ini.

## KATA PENGANTAR

Puji syukur kehadirat Allah Subhanahu Wa Ta'ala atas ahmat dan hidayah-Nya sehingga penulis dapat mencapai pada tahap kehidupan saat ini. Sholawat serta salam semoga tercurahkan kepada Rasulullah Muhammad Shalallahu Alaihi Wassallam yang diharapkan syafaatnya di hari akhir kelak.

Penulis mengucapkan banyak terima kasih kepada semua pihak yang telah berkontribusi dalam keberlangsungan skripsi ini dengan bentuk dukungan maupun perbuatan. Ucapan terima kasih ini penulis tujukan kepada :

1. Prof. Dr. M. Zainuddin, MA selaku rektor Universitas Islam Negeri Maulana Malik Ibrahim Malang.
2. Dr. Sri Hariani, M.Si., selaku Dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang
3. Dr. Fachrul Kurniawan ST., M.MT, selaku ketua Program Studi Teknik Informatika.
4. Dr. Cahyo Crysidian selaku dosen pembimbing I yang telah membimbing dan selalu memberikan waktunya untuk mendengarkan keluh kesah dari penulis.
5. Bapak Ajib Hanani, M.T selaku dosen pembimbing II yang telah memberi dukungan dan arahan dari segi agama dan memberikan tips dilancarkannya skripsi.
6. Bapak Irwan Budi Santoso, M.Kom dan Bapak Fajar Rohman Hariri, M.Kom selaku dosen penguji I dan dosen penguji II yang telah membimbing

dan memberi masukan sehingga penulis memperoleh hasil skripsi yang lebih baik.

7. Kedua orang tua penulis yang sangat disayangi dan dicintai, Bapak Suefendi, Ibu Julinarwati yang senantiasa memberi dukungan baik dalam bentuk semangat, materi, do'a, serta rasa kasih sayang dan pengertian kepada penulis yang tiada hentinya, sehingga penulis tetap kuat dalam menjalani kehidupan hingga titik ini.
8. Adik Zaki Adyatma Effendi, dan Adik Keysa Tsaqila Ayundhiza Effendi yang selalu sabar menghadapi sifat usil dari penulis walaupun tidak jarang ada yang menangis di akhir sesi.
9. Yuliana Romadhoni selaku teman, sahabat, dan orang penting milik penulis yang selalu menemani dan menerima penulis dalam keadaan apapun serta sabar dalam menghadapi sifat dan perilaku penulis.
10. Muhammad Farid Muhtarom selaku teman yang sudah bisa dianggap sebagai saudara yang mana pernah selalu menemani kemanapun penulis pergi dan telah mewarnai kehidupan perkuliahan penulis dengan candaan yang selalu seru sejak masih berstatus mahasiswa baru hingga akhirnya sibuk dengan kehidupan masing-masing.
11. Miftahul Dewanto dan Fatma Indika Sari selaku teman, sahabat, sekaligus sobat ngopi penulis sejak masih menjadi mahasiswa baru hingga sekarang.
12. Ghani Mutaqin, Aditya Prasetyo, Wildan Arif Riuditama, beserta teman lain yang berjuang bersama-sama dengan penulis dalam mengerjakan skripsi.

13. Dimas Agung Surya Mahendra selaku teman, sahabat dan tetangga penulis sekaligus Bapak Taufik Hendrasno dan Ibu Iin Priharyuni selaku orang tua dari Dimas yang juga ikut berkontribusi dalam keberlangsungan skripsi ini.
14. Seluruh Dosen Program Studi Teknik Informatika yang telah memberikan ilmunya selama masa perkuliahan sehingga peneliti mendapatkan pengetahuan yang lebih luas dan In sya Allah menjadikan manfaat bagi orang sekitar.
15. Nia Faricha, S.Si selaku Admin Program Studi Teknik Informatika yang selalu sabar dalam menghadapi mahasiswa yang sedang menjalankan skripsi dengan pertanyaan yang sangat banyak dan berulang.

Penulis menyadari dalam penyusunan skripsi ini masih terdapat kekurangan dan penulis berharap semoga skripsi ini dapat memberikan manfaat kepada pembaca khususnya bagi penulis secara pribadi. *Aamiin Ya Rabbal Alamin.*

Malang, 29 Mei 2023

Penulis

## DAFTAR ISI

<b>HALAMAN JUDUL</b> .....	Error! Bookmark not defined.
<b>HALAMAN PERSETUJUAN</b> .....	<b>iii</b>
<b>HALAMAN PENGESAHAN</b> .....	Error! Bookmark not defined.
<b>PERNYATAAN KEASLIAN TULISAN</b> .....	Error! Bookmark not defined.
<b>HALAMAN MOTTO</b> .....	<b>vi</b>
<b>HALAMAN PERSEMBAHAN</b> .....	<b>vii</b>
<b>KATA PENGANTAR</b> .....	<b>ix</b>
<b>DAFTAR ISI</b> .....	<b>xii</b>
<b>DAFTAR GAMBAR</b> .....	<b>xiv</b>
<b>DAFTAR TABEL</b> .....	<b>xv</b>
<b>ABSTRAK</b> .....	<b>xvi</b>
<b>ABSTRACT</b> .....	<b>xvii</b>
<b>خلاصة</b> .....	<b>xviii</b>
<b>BAB 1 PENDAHULUAN</b> .....	<b>1</b>
1.1 Latar Belakang .....	1
1.2 Pernyataan Masalah .....	3
1.3 Tujuan Penelitian .....	3
1.4 Batasan Masalah.....	4
1.5 Manfaat Penelitian .....	4
<b>BAB II STUDI PUSTAKA</b> .....	<b>5</b>
2.1 Analisis Sentimen .....	5
2.2 <i>Support Vector Machine</i> .....	7
<b>BAB III DESAIN DAN IMPLEMENTASI PENELITIAN</b> .....	<b>9</b>
3.1 Pengumpulan Data .....	9
3.2 Desain Sistem.....	10
3.3 <i>Preprocessing</i> .....	12
3.3.1 Cleaning data .....	12
3.3.2 Case Folding .....	14
3.3.3 Tokenizing .....	15
3.3.4 Stemming.....	16
3.3.5 Stopword Removal .....	17
3.4 TF-IDF .....	18
3.5 Support Vector Machine .....	21
<b>BAB IV UJI COBA DAN PEMBAHASAN</b> .....	<b>27</b>
4.1 Skenario Uji Coba.....	27
4.2 Hasil Uji Coba.....	28
4.2.1 Hasil Uji Coba Ke-1 .....	33
4.2.2 Hasil Uji Coba Ke-2.....	33
4.2.3 Hasil Uji Coba Ke-3 .....	34
4.2.4 Hasil Uji Coba Ke-4.....	35
4.3 Pembahasan.....	35
<b>BAB V KESIMPULAN DAN SARAN</b> .....	<b>39</b>
5.1 Kesimpulan .....	39

5.2	Saran .....	39
<b>DAFTAR PUSTAKA</b>		
<b>LAMPIRAN</b>		

## DAFTAR GAMBAR

Gambar 3. 1 Desain Sistem.....	10
Gambar 3. 2 Alur Preprocessing .....	11
Gambar 3. 3 Flowchart Cleansing.....	12
Gambar 3. 4 Flowchart Case Folding .....	13
Gambar 3. 5 Flowchart Tokenizing .....	14
Gambar 3. 6 Flowchart Stemming .....	16
Gambar 3. 7 Flowchart Stopwords Removal .....	17
Gambar 3. 8 Flowchart Pembentukan Hyperplane .....	23

## DAFTAR TABEL

Tabel 3. 1 Data Sampel .....	10
Tabel 3. 2 Contoh Cleaning .....	13
Tabel 3. 3 Contoh Case Folding.....	14
Tabel 3. 4 Contoh Tokenizing.....	15
Tabel 3. 5 Contoh Stemming .....	16
Tabel 3. 6 Contoh Stopword Removal.....	17
Tabel 3. 7 Perhitungan TF.....	18
Tabel 3. 8 Perhitungan IDF.....	19
Tabel 3. 9 Perhitungan TF-IDF.....	19
Tabel 3. 10 Hasil Perhitungan TF-IDF .....	20
Tabel 4. 1 Confusion Matrix .....	28
Tabel 4. 2 Sampel Dataset Penelitian.....	29
Tabel 4. 3 Hasil Cleaning Data .....	30
Tabel 4. 4 Hasil Case Folding .....	30
Tabel 4. 5 Hasil Tokenizing .....	31
Tabel 4. 6 Hasil Stemming.....	31
Tabel 4. 7 Hasil Stopword Removal .....	32
Tabel 4. 8 Hasil Preprocessing.....	32
Tabel 4. 9 Confusion Matrix Uji Coba Ke-1 .....	33
Tabel 4. 10 Confusion Matrix Uji Coba Ke-2.....	34
Tabel 4. 11 Confusion Matrix Uji Coba Ke-3.....	34
Tabel 4. 12 Confusion Matrix Uji Coba Ke-4.....	35
Tabel 4. 13 Hasil Akurasi Uji Coba.....	36

## ABSTRAK

Effendi, Andhika Maulana, 2023. **Analisis Sentimen Pada Judul Berita Online Ekonomi Dengan Menggunakan *Support Vector Machine***. Skripsi. Program Studi Teknik Informatika, Fakultas Sains Dan Teknologi, Universitas Islam Negeri Maulana Malik Ibrahim Malang. Pembimbing: (I) Dr. Cahyo Crysdiyan. (II) Ajib Hanani, M.T.

---

Kata Kunci: *Analisis Sentimen, Support Vector Machine, Berita Online, TF-IDF, COVID-19*.

Media massa merupakan sarana komunikasi yang digunakan untuk menyampaikan pesan dari suatu sumber kepada khalayak ramai menggunakan alat- alat komunikasi. Kemajuan teknologi mengakibatkan media massa ini tidak hanya berupa media cetak namun juga media elektronik. Hal ini juga memudahkan masyarakat untuk mengkritik atau hanya beropini mengenai berita yang tersebar. Sejak kemunculan kasus COVID-19 pemberitaan media massa mulai ramai membahas tentang COVID-19 beserta akibatnya terhadap sektor masyarakat terutama di bidang ekonomi. Oleh sebab itu perlu dilakukan analisis sentimen terhadap judul berita online ekonomi dengan mengimplementasikan metode *support vector machine*. Tujuan dari penelitian ini adalah mengukur performa dari sistem dimana data yang digunakan berupa judul berita dengan jumlah 1000 data. Data tersebut dilabeli dengan label positif dan negatif yang kemudian diolah dengan beberapa rasio pembagian *data training* dan *data testing*. Nilai akurasi terbaik didapatkan dari hasil uji coba dengan rasio 9:1 dengan akurasi 76%, presisi sebesar 80,48%, recall sebesar 89,18%, dan f-measure sebesar 84,61%.

## ABSTRACT

Effendi, Andhika Maulana, 2023. **Sentiment Analysis in Online Economic News Titles Using Support Vector Machine**. Thesis. Department of Informatics Engineering, Faculty of Science and Technology, Maulana Malik Ibrahim State Islamic University Malang. Supervisors: (I) Dr. Cahyo Crysdiyan. (II) Ajib Hanani, M.T.

---

Mass media is a means of communication used to convey messages from a source to the general public using communication tools. Advances in technology have resulted in this mass media not only in the form of print media but also electronic media. This also makes it easier for people to criticize or just have opinions about the news that is being spread. Since the emergence of the COVID-19 case, mass media reports have begun to be busy discussing COVID-19 and its consequences for the community sector, especially in the economic sector. Therefore it is necessary to do sentiment analysis of online economic news titles by implementing the support vector machine method. The purpose of this study is to measure the performance of the system where the data used is in the form of news titles with a total of 1000 data. The data is labeled with positive and negative labels which are then processed with several distribution ratios of training data and data testing. The best accuracy value was obtained from the test results with a 9:1 ratio with 76% accuracy, 80.48% precision, 89.18% recall and 84.61% f-measure.

Keywords: *Sentiment Analysis, Support Vector Machine, Online News, TF-IDF, COVID-19.*

## خلاصة

أفندي ، أنديكا مولانا ، 2023. تحليل المشاعر على عناوين الأخبار الاقتصادية على الإنترنت باستخدام آلة المتجه الداعمة أطروحة. برنامج دراسة هندسة المعلوماتية ، كلية العلوم والتكنولوجيا ، جامعة الولاية الإسلامية مولانا مالك إبراهيم مالانج. المشرفون: (١) دكتور كاهو كريسيديان. (٢) عجيب حناني

الكلمات المفتاحية: تحليل المشاعر ، آلة المتجهات الداعمة ، الأخبار عبر الإنترنت، TF-IDF، COVID-19 .

.وسائل الإعلام الجماهيرية هي وسيلة اتصال تستخدم لنقل الرسائل من مصدر إلى عامة الناس باستخدام أدوات الاتصال. أدى التقدم التكنولوجي إلى ظهور هذه الوسائط ليس فقط في شكل وسائل الإعلام المطبوعة ولكن أيضًا في الوسائط الإلكترونية بدأت ، COVID-19 هذا أيضًا يسهل على الناس انتقاد أو مجرد تكوين آراء حول الأخبار التي يتم نشرها. منذ ظهور حالة وعواقبه على قطاع المجتمع ، وخاصة في القطاع الاقتصادي. لذلك من COVID-19 تقارير وسائل الإعلام مشغولة بمناقشة الضروري إجراء تحليل المشاعر للأخبار الاقتصادية عبر الإنترنت من خلال تنفيذ طريقة آلة المتجه الداعمة. الغرض من هذه الدراسة هو قياس أداء النظام حيث تكون البيانات المستخدمة على شكل عناوين أخبار إجمالي 1000 بيانات. يتم تصنيف البيانات بعلامات موجبة وسالبة يتم معالجتها بعد ذلك بنسب توزيع متعددة لبيانات التدريب واختبار البيانات. يتم الحصول على أفضل قيمة .84,61% بنسبة f ومقياس ، 89,18% واسترجاع ، 80,84% تليها دقة ، 76% بدقة 1: 9دقة من نتائج الاختبار بنسبة

# **BAB 1**

## **PENDAHULUAN**

### **1.1 Latar Belakang**

Media massa merupakan sarana komunikasi yang digunakan untuk menyampaikan pesan dari suatu sumber kepada khalayak ramai menggunakan alat-alat komunikasi (Solten Rajagukguk & Olilia, 2020). Media massa bergerak secara kontinu sesuai dengan jadwal edar masing masing. Media massa ini berisikan informasi perihal suatu kondisi, peristiwa, perubahan, bahkan opini – opini yang berisikan tanggapan perihal permasalahan atau isu aktual yang sedang terjadi di masyarakat. Tak sedikit pemberitaan ini berisikan berita yang menyangkut keadaan sektor kehidupan masyarakat dalam suatu wilayah. Selain itu masyarakat juga mampu mengkritik berita yang tersebar melalui media massa.

Seiring perkembangan zaman, media massa ini tidak hanya berupa media cetak namun juga media elektronik. Di era gobalisasi ini media massa elektronik merupakan media yang ramai digunakan. Media massa elektronik ini dapat berupa website dan aplikasi. Media massa ini diakses secara mudah dan gratis dengan melalui smartphone. Hal ini juga memudahkan masyarakat untuk mengkritik atau hanya beropini mengenai berita yang tersebar. Di sisi lain, hal ini juga mampu memudahkan pihak yang diberitakan untuk mengetahui opini masyarakat yang mampu membantu membangun maupun memperbaiki dalam rencana untuk menjalankan tujuannya.

Rasulullah sebagai utusan Allah SWT selama hidupnya telah mencontohkan kaumnya dengan sifat tabligh. Sehingga sebagai seorang muslim hendaklah untuk

berhati - hati dalam menyampaikan berita supaya berita yang disebarkan nanti merupakan berita yang benar. Maka dari itu, alangkah baiknya bila mendapatkan sebuah berita untuk dikaji ulang kebenarannya sebelum disebarkan kepada orang lain. Sebagaimana telah dijelaskan dalam Surah Al-Hujurat ayat 6 tentang pentingnya berhati-hati dalam menyampaikan berita :

يَا أَيُّهَا الَّذِينَ آمَنُوا إِن جَاءَكُمْ فَاسِقٌ بِنَبَأٍ فَتَبَيَّنُوا أَن تُصِيبُوا قَوْمًا بِجَهَالَةٍ فَتُصْبِحُوا عَلَىٰ مَا فَعَلْتُمْ  
نَادِمِينَ

*“Wahai orang-orang yang beriman! Jika seseorang yang fasik datang kepadamu membawa suatu berita, maka telitilah kebenarannya, agar kamu tidak mencelakakan suatu kaum karena kebodohan (kecerobohan), yang akhirnya kamu menyesali perbuatanmu itu..”* (Q.S. Al-Hujurat : 6).

Menurut tafsir Ibnu Katsir, ayat ini Allah SWT memrintahkan (kaum mukmin) untuk memeriksa secara teliti terhadap berita yang asalnya dari orang fasik, dan hendaklah kaum mukmin untuk bersikap hati-hati dalam menerima berita. Menerima pesan begitu saja sama dengan mengikuti orang jahat. Allah SWT telah melarang orang beriman untuk mengikuti jalan orang yang rusak (Ibnu Katsir Online, 2022).

Di Indonesia sudah terdapat banyak sekali media berita online yang bisa dibilang terpercaya. Menurut Ketua Dewan Pers 2016-2019 Yosep Adi Prasetyo memperkirakan bahwa jumlah media massa di Indonesia mencapai 47.000 media dan media online bisa mencapai 43.300 (Dewan Pers, 2022). Pada tahun 2021, 5 website teratas dengan kategori media massa online yaitu Kompas.com, TribunNews.com, Detik.com, IdnTimes.com, dan Jpnn.com (BeritaManado, 2021).

Terhitung sejak Maret 2020 yaitu kasus pertama COVID-19 masuk di

Indonesia, pemberitaan media massa mulai ramai membahas tentang COVID-19 beserta akibatnya terhadap sektor masyarakat. Dari berbagai macam berita tersebut dapat dianalisis dan diklasifikasikan untuk mengetahui berita yang bersifat positif dan negatif. Sedangkan dalam melakukan klasifikasi terdapat berbagai macam metode seperti contohnya adalah *Naïve Bayes*, *KNearest Neighbour*, *K-Means*, *Support Vector Machine* (SVM), dan lain sebagainya.

Penelitian ini menerapkan metode *Support Vector Machine* (SVM) dan dilakukan pembobotan kata dengan menggunakan *Term Frequency-Inverse Document Frequency* TF-IDF. *Support Vector Machine* merupakan salah satu metode yang tepat dengan tujuan membuat pemisah antar kelas klasifikasi atau *hyperplane*. *Support Vector Machine* ini berusaha membuat *hyperplane* dengan cara mencari jarak maksimal antar kelas. Metode *Support Vector Machine* sering digunakan karena sangat cepat dan akurat pada klasifikasi data (Husna *et al.*, 2020).

## 1.2 Pernyataan Masalah

1. Berapa besar jumlah nilai akurasi, presisi, *recall*, dan *f-measure* dalam analisis sentimen pada judul berita online ekonomi dengan menggunakan *Support Vector Machine*.
2. Faktor apa yang berpengaruh terhadap performa *Support Vector Machine* pada judul berita online ekonomi.

## 1.3 Tujuan Penelitian

1. Mengukur seberapa besar nilai akurasi, presisi, *recall*, dan *f-measure* dalam analisis sentimen pada judul berita online ekonomi dengan menggunakan

*Support Vector Machine*.

2. Menganalisis faktor-faktor yang mempengaruhi performa *Support Vector Machine* pada judul berita online ekonomi.

#### **1.4 Batasan Masalah**

1. Data penelitian yang diambil hanya dari media berita online berbahasa Indonesia.
2. Data penelitian yang diambil hanya dari media berita online selama 2 tahun terakhir.
3. Data penelitian yang diambil hanya berita yang menyangkut sektor ekonomi.

#### **1.5 Manfaat Penelitian**

Manfaat bagi Kementerian Koordinator Bidang Perekonomian, dengan adanya analisis sentimen terhadap sektor kehidupan masyarakat ini dapat digunakan untuk mengetahui sentimen pada berita massa yang diharapkan pemerintah mampu menyiapkan strategi yang paling baik untuk masa mendatang.

## BAB II

### STUDI PUSTAKA

#### 2.1 Analisis Sentimen

Analisis sentimen merupakan sebuah metode untuk memperoleh data dari berbagai macam media yang ada di internet. Analisis sentimen bertujuan untuk menganalisis suasana publik, suasana hati, serta gambaran perasaan secara otomatis (Que *et al.*, 2020). Dalam melakukan analisis sentimen diperlukan data sentimen yang bersih dari berbagai macam *noise*. Maka dari itu dalam proses analisis perlu ditambahkan tahap *preprocessing* untuk mendapatkan data yang bersih dari *noise*.

*Text preprocessing* merupakan merupakan tahap pertama dari *text mining*. *Text preprocessing* adalah sebuah proses melibatkan transformasi teks sebelum melakukan analisis yang mana tranformasi teks ini meliputi *tokenizing*, *removing stopwords and non alphabetic words*, *stemming*, dan lain sebagainya (Hickman *et al.*, 2022).

Dalam penelitian sebelumnya (Alam & Yao, 2019) melakukan penelitian mengenai dampak *text preprocessing* dalam analisis sentimen. Penelitian ini menggunakan 3 metode klasifikasi, yaitu Naïve Bayes, *Maximum Entropy* (MaxE), dan *Support Vector Machine* (SVM). Tahapan *text preprocessing* yang dilakukan pada penelitian ini merupakan *emoticons removal*, *stopwords removal*, *stemming*, dan *word vectorization*. Hasil dari penelitian ini didapatkan perbandingan hasil akurasi data sebelum dilakukan *text preprocessing* dengan hasil penambahan proses *text preprocessing*. Pada metode SVM didapatkan peningkatan akurasi dari

81.09% menjadi 81.63%. Pada metode Naïve Bayes didapatkan peningkatan yang cukup signifikan yaitu dari 83.69% menjadi 91.81%. Berbeda dari kedua metode sebelumnya, metode *Maximum Entropy* tidak membuahkan hasil yang beda, yaitu tetap dengan hasil akurasi 88.27%.

Selain *preprocessing*, dalam analisis sentimen perlu juga menambahkan tahapan pembobotan kata. Untuk melakukan pembobotan kata dalam penelitian ini adalah menggunakan *Term Frequency Inverse Document Frequency* (TF-IDF). Algoritma TF-IDF merupakan metode statistik yang digunakan untuk menilai pentingnya sebuah kata. TF-IDF terdiri dari dua bagian, yaitu *frequency of words* dan *frequency of inverse text*. *Frequency of words* mengacu pada jumlah kemunculan kata dalam satu dokumen. Sedangkan *Frequency of inverse text* mengacu pada ukuran kepentingan sebuah kata (Zhou, 2022).

Dalam penelitian sebelumnya (Haddi *et al.*, 2013) membandingkan tiga metode yang populer untuk pembobotan kata, yaitu *Feature Frequency* (FF), *Term Frequency Inverse Document Frequency* (TF-IDF), dan *Feature Presence* (FP). Penelitian ini menggunakan metode yang sama yaitu metode *Support Vector Machine* dengan jumlah 2000 data yang terdiri atas 1000 data positif dan juga 1000 data negatif. Hasil dari penelitian ini didapatkan akurasi paling besar dengan TF-IDF sebesar 93.5% sedangkan FF hanya 90.5% dan FP sebesar 93%.

Sedangkan pada penelitian lain (Cahyani & Patasik, 2021) melakukan perbandingan antara TF-IDF dengan Word2Vec menggunakan data tweet tentang commuter line dan transjakarta. Metode yang diaplikasikan pada penelitian ini adalah *Support Vector Machine* (SVM) dan *Multinomial Naïve Bayes* (MNB).

Penelitian ini membandingkan tiga skenario, yaitu SVM menggunakan TF-IDF, SVM menggunakan Word2Vec, dan MNB menggunakan TF-IDF. Hasil dari penelitian ini didapatkan kesimpulan bahwa metode SVM menggunakan TF-IDF lebih unggul dari kedua skenario yang lain.

## 2.2 *Support Vector Machine*

*Support Vector Machine* (SVM) merupakan sebuah metode yang baru namun memiliki performa yang lebih baik dalam berbagai bidang aplikasi seperti klasifikasi teks, pengenalan tulisan tangan, bioinformatika, dan lain sebagainya. Metode ini digunakan untuk menganalisa data dan mengenali pola klasifikasi. Dalam penggunaannya, metode ini mengharuskan menggunakan pelabelan, yaitu pelabelan kalimat positif dan negatif (Suhardjono *et al.*, 2019).

Dalam penelitian sebelumnya (Permana & Sahara, 2019) menggunakan metode *Support Vector Machine* (SVM) sebagai penentu kelulusan mahasiswa pada pembelajaran elektronik. Hasil penelitian ini didapatkan besar akurasi 81.25%. Kemudian hasil tersebut dites untuk mendapatkan akurasi dan nilai *Area Under Curve* (AUC) setiap algoritma sehingga hasil tes tertinggi dapat diperoleh menggunakan *support vector* dengan akurasi 85.02% dan nilai AUC 0.710.

Sedangkan pada penelitian lain (Kurniawan *et al.*, 2019) melakukan perbandingan metode klasifikasi analisis sentimen tokoh politik pada komentar media berita online. Penelitian ini membandingkan metode *Support Vector Machine* (SVM) dan Naïve Bayes dengan tambahan *Particle Swarm Optimization* (PSO). Hasil dari penelitian ini didapatkan bahwa besar akurasi metode SVM lebih

besar dari Naïve Bayes. Metode SVM menghasilkan akurasi sebesar 76,09%, sedangkan metode Naïve Bayes menghasilkan akurasi sebesar 68,21%. Dijelaskan juga pada hasil bahwa pengimplementasian PSO mampu meningkatkan akurasi dengan hasil peningkatan menjadi 78,40% pada metode SVM dan 74,98% pada metode Naïve Bayes.

Pada penelitian (Ratino *et al.*, 2020) dilakukan perbandingan dengan metode yang serupa yaitu Naïve Bayes dengan *Support Vector Machine* (SVM) dengan objek berupa komentar pada media social Instagram. Penelitian ini dilengkapi dengan tahap *preprocessing*. Dari hasil didapatkan kesimpulan bahwa kedua metode tersebut mampu digunakan dalam analisis sentiment dengan objek komentar pada media social Instagram. Hasil perbandingan akurasi didapatkan bahwa akurasi metode *Support Vector Machine* lebih unggul dari metode Naïve Bayes dengan perbandingan 81,16% oleh SVM dan 79,07% milik Naïve Bayes.

Selain itu dalam penelitian lain yang memiliki objek dan metode yang serupa (Ulfah & Anam, 2020) melakukan analisis sentimen *hate speech* pada berita online menggunakan Support Vector Machine. Penelitian ini menggunakan jumlah total 33 berita dengan 207 komentar digunakan sebagai data. Penelitian ini dilengkapi dengan tahapan *preprocessing* dan pembobotan dengan TF-IDF. Dari hasil klasifikasi didapatkan hasil uji performa dengan akurasi 53,88%, *Recall* 49,69%, *Precision* 48,77%, dan *fmeasure* 49,23%.

## BAB III

### DESAIN DAN IMPLEMENTASI PENELITIAN

#### 3.1 Pengumpulan Data

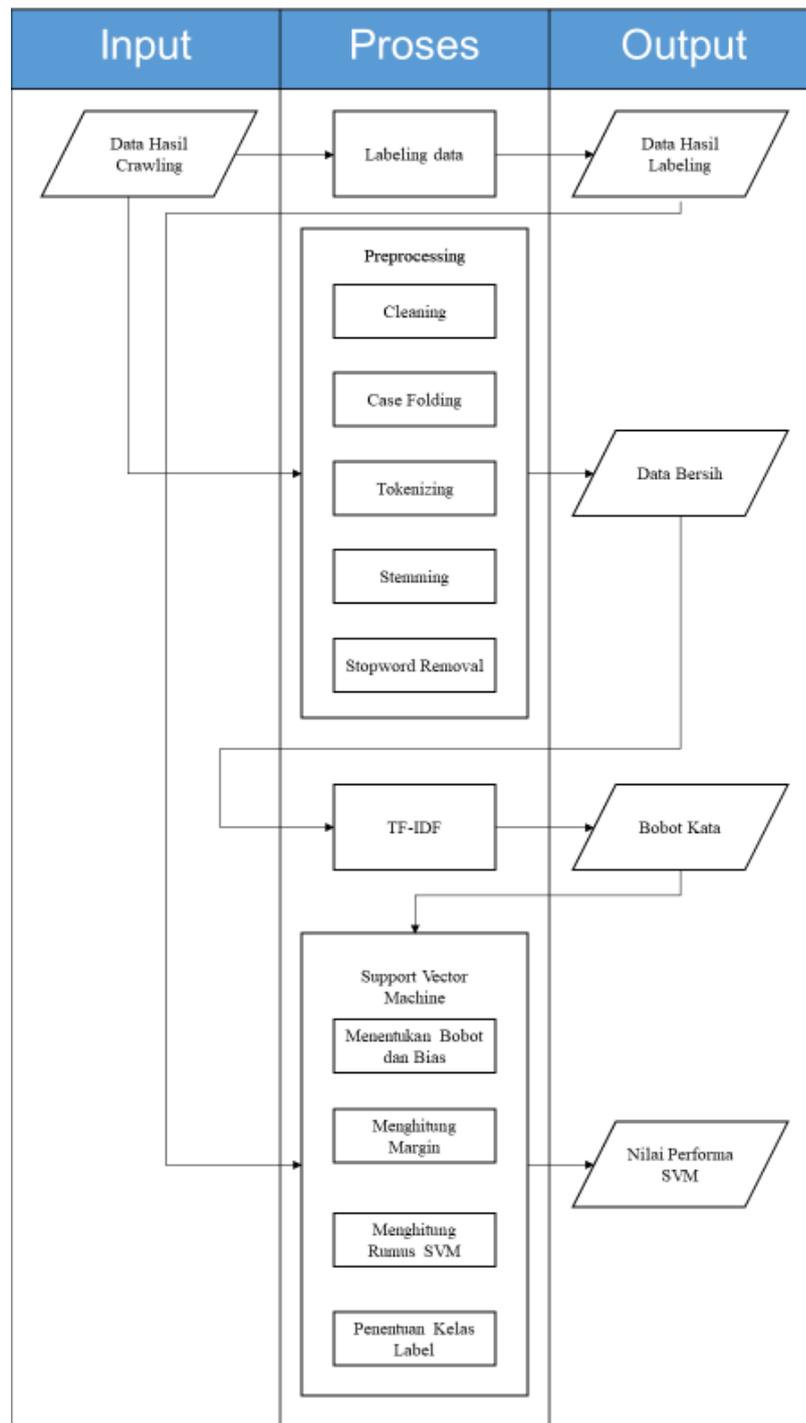
Data yang digunakan dalam penelitian ini adalah data hasil dari *website* berita online *cnnindonesia.com*, *jpnn.com*, *detik.com*, *bisnis.com*, dan *kompas.com* dengan teknik *web scrapping*. Data diambil pada tanggal 18 November 2021 berjumlah 1000 data judul berita terkait ekonomi, Data dari web yang diambil merupakan judul dari berita yang mana setiap berita dilabeli dengan label positif atau negatif. Pelabelan dalam penelitian ini dilakukan dengan bantuan ahli yang sedang bekerja dalam bidang ekonomi yaitu pada kantor pajak, dan juga menggunakan metode *Crowdsourcing*. Data yang dilabeli oleh ahli adalah sebanyak 100 data sedangkan 900 data dilabeli dengan metode *Crowdsourcing* dari berbagai macam latar belakang keahlian dengan pelabel berjumlah ganjil sehingga dapat dijadikan perbandingan . Latar belakang profesi orang yang melabeli 900 data ini terdiri atas 8 mahasiswa program studi Teknik Informatika, 1 mahasiswa program studi Manajemen, 3 mahasiswa Politeknik Negeri Malang program studi Teknik Kimia, 1 pegawai bank. Data-data tersebut kemudian disimpan dalam file excel dengan kolom berisikan teks dan label. Contoh sampel data yang akan digunakan pada penelitian ditunjukkan pada Tabel 3.1.

Tabel 3. 1 Data Sampel

<b>Dokumen</b>	<b>Judul berita</b>	<b>Kelas</b>
D1	Pemerintah Wajibkan Pengusaha Sediakan Minyak Goreng Curah	P
D2	Asosiasi Pedagang: Harga Minyak Goreng, Kedelai, hingga Daging Sapi Naik Tidak Wajar	N
D3	Awal Pekan, Harga Minyak Dunia Naik Lagi ke Level 110 Dollar AS Per Barrel	N
D4	Harga Cabai Merah Keriting hingga Bawang Putih Naik di Akhir Pekan	N
D5	Harga Minyak Goreng Curah Menurun, Tak Lebih dari Rp14.000 per Liter	P

### 3.2 Desain Sistem

Alur jalannya sistem pada penelitian ini terdiri dari beberapa tahap sebagaimana telah ditunjukkan pada Gambar 3.1.



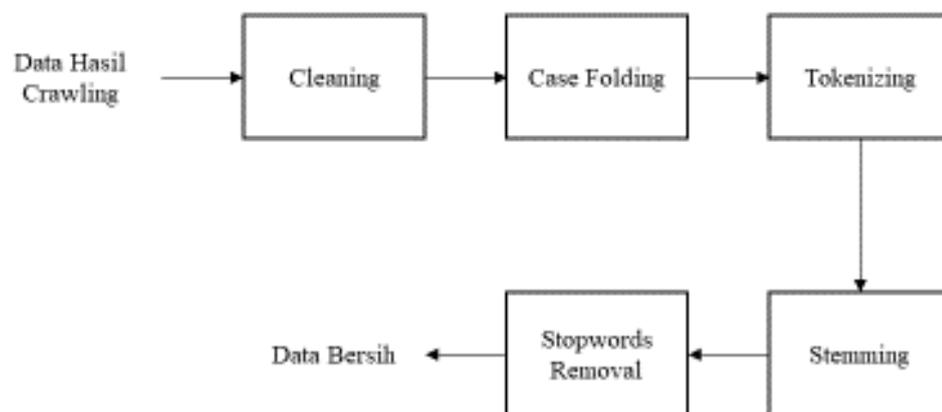
Gambar 3. 1 Desain Sistem

Sistem dalam penelitian ini dikembangkan dengan menggunakan bahasa pemrograman python. Berdasarkan desain sistem diatas dapat diketahui input dari

penelitian ini merupakan data hasil *web scrapping* yang kemudian diproses melalui tahap *preprocessing*, *feature extraction* dengan TF-IDF, dan klasifikasi dengan metode *Support Vector Machine* (SVM). Setelah itu dilakukan pengujian dengan *confusion matrix* untuk mendapatkan hasil akurasi, presisi, recall, dan *f-measure*.

### 3.3 *Preprocessing*

Data hasil perolehan dari *web scrapping* harus melalui tahapan *preprocessing*. Data dengan format file excel ini harus diproses melalui beberapa tahap supaya dihasilkan data yang bersih. Alur tahap *preprocessing* ditunjukkan oleh Gambar 3.2.



Gambar 3. 2 Alur *Preprocessing*

#### 3.3.1 *Cleaning data*

Tahap *cleaning* merupakan tahap awal dalam *preprocessing* data. Pada tahap ini data akan dibersihkan dari berbagai macam *noise* seperti symbol, angka, tanda baca, emotikon, dll. *Flowchart cleaning data* ditunjukkan pada Gambar 3.3.



Gambar 3. 3 *Flowchart Cleaning*

Tahap *cleaning* memiliki input judul teks berita. Kemudian jika terdeteksi angka dan tanda baca maka perlu dihilangkan. Hasil tahap *cleaning* ini ditunjukkan pada tabel 3.2.

Tabel 3. 2 Contoh *Cleaning*

Sebelum <i>Cleaning</i>	Harga Minyak Goreng Curah Menurun, Tak Lebih dari Rp14.000 per Liter
Sesudah <i>Cleaning</i>	Harga Minyak Goreng Curah Menurun, Tak Lebih dari Rp per Liter

### 3.3.2 Case Folding

Tahap *case folding* merupakan tahap untuk menyeragamkan teks dengan cara mengubah semua teks menjadi bentuk huruf kecil (*lowercase*). *Flowchart case folding* ditunjukkan pada Gambar 3.4.



Gambar 3. 4 *Flowchart Case Folding*

Tahap *case folding* memiliki input teks hasil *cleaning*. Teks tersebut kemudian diubah menjadi format *lowercase*. Hasil dari tahap *case folding* terdapat pada tabel 3.3.

Tabel 3. 3 Contoh *Case Folding*

Sebelum <i>Case Folding</i>	Harga Minyak Goreng Curah Menurun, Tak Lebih dari Rp per Liter
Sesudah <i>Case Folding</i>	harga minyak goreng curah menurun, tak lebih dari rp per liter

### 3.3.3 Tokenizing

Tahap *tokenizing* merupakan tahapan dimana teks akan dipecah tiap katanya. Tiap potongan kata ini biasa disebut *token* atau *term*. *Flowchart tokenizing* ditunjukkan oleh Gambar 3.5.



Gambar 3. 5 Flowchart Tokenizing

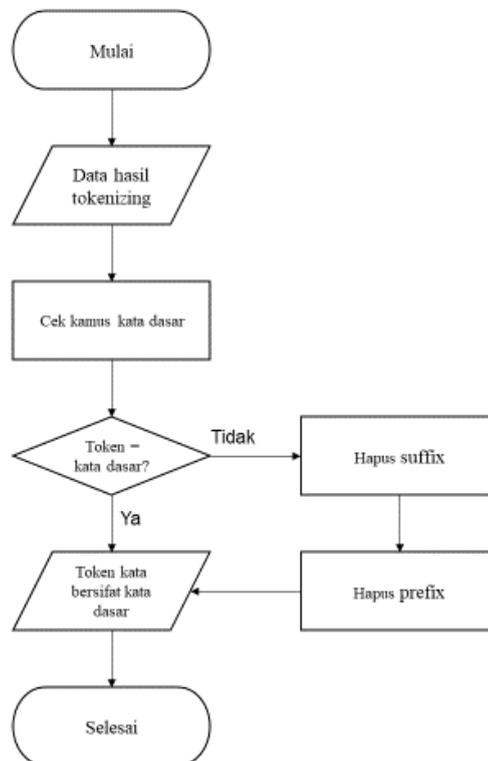
Tahap *tokenizing* memiliki input data hasil proses *case folding*. Teks tersebut kemudian diubah menjadi potongan kata (*token*). Hasil proses *tokenizing* terdapat pada tabel 3.4.

Tabel 3. 4 Contoh *Tokenizing*

Sebelum <i>Tokenizing</i>	harga minyak goreng curah menurun, tak lebih dari rp per liter
Sesudah <i>Tokenizing</i>	“harga”, “minyak”, “goreng”, “curah”, “menurun”, “tak”, “lebih”, “dari”, “rp”, “per”, “liter”

### 3.3.4 Stemming

Tahap *stemming* merupakan tahap untuk merubah kata dalam dokumen menjadi kata dasar. Dengan kata lain tahap ini akan menghilangkan berbagai imbuhan pada kata. Dalam penerapannya dilakukan dengan bantuan library Sastrawi. *Flowchart stemming* ditunjukkan pada Gambar 3.6.



Gambar 3. 6 Flowchart Stemming

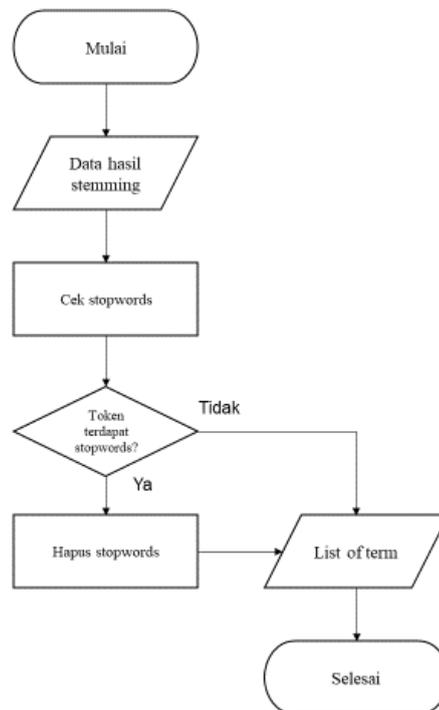
Tahap *stemming* memiliki input berupa data hasil proses *tokenizing*. Jika teks tersebut mengandung *prefix* dan *suffix*, maka *prefix* dan *suffix* perlu dihapus. Hasil tahap *stemming* dapat dilihat oleh tabel 3.5.

Tabel 3. 5 Contoh Stemming

Sebelum <i>Stemming</i>	“harga”, “minyak”, “goreng”, “curah”, “menurun”, “tak”, “lebih”, “dari”, “rp”, “per”, “liter”
Sesudah <i>Stemming</i>	“harga”, “minyak”, “goreng”, “curah”, “turun”, “tak”, “lebih”, “dari”, “rp”, “per”, “liter”

### 3.3.5 Stopword Removal

Tahap ini merupakan tahap terakhir dalam *preprocessing*. *Stopword Removal* berfungsi untuk menghilangkan kata yang tidak mengandung sentimen. Jenis *stopword* yang digunakan pada penelitian ini adalah *stopword* Indonesia. *Flowchart stopwords removal* ditunjukkan pada Gambar 3.7.



Gambar 3. 7 Flowchart Stopwords Removal

Tahap *stopwords removal* memiliki input berupa data hasil proses *stemming*. Jika teks tersebut mengandung *stopwords* Indonesia, maka *stopwords* perlu dihapus. Hasil tahap *stopwords removal* dapat ditunjukkan oleh tabel 3.5.

Tabel 3. 6 Contoh *Stopword Removal*

Sebelum <i>Stopword Removal</i>	“harga”, “minyak”, “goreng”, “curah”, “turun”, “tak”, “lebih”, “dari”, “rp”, “per”, “liter”
Sesudah <i>Stopword Removal</i>	“harga”, “minyak”, “goreng”, “curah”, “turun”, “rp”, “liter”

### 3.4 TF-IDF

Data bersih yang telah melalui tahapan *preprocessing* akan dihitung bobot katanya dengan TF-IDF. TF merupakan perhitungan nilai frekuensi kemunculan kata sedangkan IDF merupakan bobot setiap kata. Semakin jarang munculnya kata dalam suatu dokumen, maka semakin besar bobot kata tersebut (Madasu & Elango, 2020). Perhitungan TF ditunjukkan pada persamaan 3.1, sedangkan contoh perhitungannya ditunjukkan pada Tabel 3.7.

$$TF_{(w)} = \frac{\text{Jumlah kemunculan sebuah kata dalam dokumen}(d)}{\text{Jumlah total kata dalam satu dokumen}} \quad (3.1)$$

Tabel 3. 7 Perhitungan TF

Dokumen	Kata	Frekuensi	TF
D1	harga	0	0/5
D2		1	1/5
D3		1	1/5
D4		1	1/5
D5		1	1/5

Perhitungan IDF ditunjukkan pada persamaan 3.2, sedangkan contoh perhitungannya ditunjukkan pada Tabel 3.8.

$$IDF_{(w)} = \log_e \frac{\text{Jumlah Total Dokumen}}{\text{Jumlah total dokumen dengan kata } (w)} \quad (3.2)$$

Tabel 3. 8 Perhitungan IDF

Dokumen	Frekuensi	Kata	IDF
D1	0	harga	$\log 5/4 = 0,09691$
D2	1		
D3	1		
D4	1		
D5	1		

Perhitungan TF-IDF ditunjukkan pada persamaan 3.3, sedangkan contoh perhitungannya ditunjukkan pada Tabel 3.9.

$$TF - IDF(w) = TF(w) \times IDF(w) \quad (3.3)$$

Tabel 3. 9 Perhitungan TF-IDF

Dokumen	Kata	TF	IDF	TFIDF
D1	harga	0	0,09691	0
D2		0,2		0,019382
D3		0,2		0,019382
D4		0,2		0,019382
D5		0,2		0,019382

Hasil perhitungan bobot dari masing – masing *term* ditunjukkan oleh Tabel 3.10.

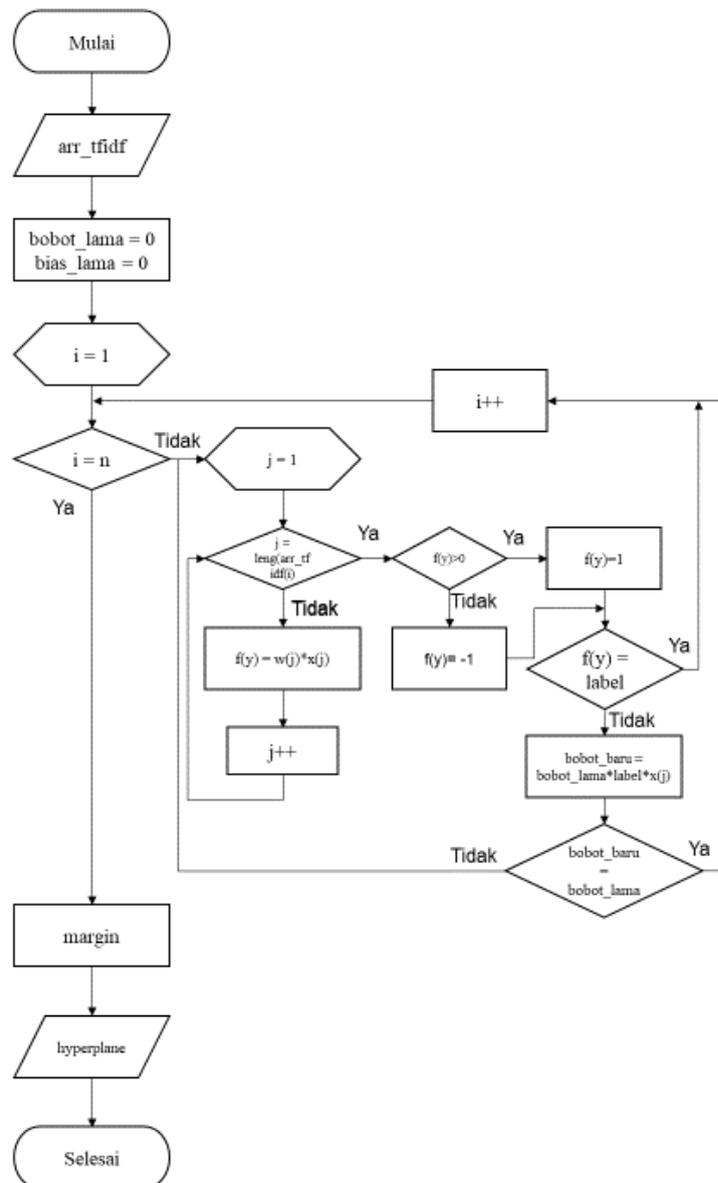
Tabel 3. 10 Hasil Perhitungan TF-IDF

Token	TF-IDF				
	D1	D2	D3	D4	D5
pemerintah	0,139794	0	0	0	0
wajib	0,139794	0	0	0	0
usaha	0,139794	0	0	0	0
sedia	0,139794	0	0	0	0
minyak	0,019382	0,019382	0,019382	0	0,019382
goreng	0,04437	0,04437	0	0	0,04437
curah	0,079588	0	0	0	0,079588
asosiasi	0	0,139794	0	0	0
dagang	0	0,139794	0	0	0
harga	0	0,019382	0,019382	0,019382	0,019382
kedelai	0	0,139794	0	0	0
daging	0	0,139794	0	0	0
sapi	0	0,139794	0	0	0
wajar	0	0,139794	0	0	0
pekan	0	0	0,079588	0,079588	0
dunia	0	0	0,139794	0	0
level	0	0	0,139794	0	0
dollar	0	0	0,139794	0	0
as	0	0	0,139794	0	0
barrel	0	0	0,139794	0	0

cabai	0	0	0	0,139794	0
merah	0	0	0	0,139794	0
keriting	0	0	0	0,139794	0
bawang	0	0	0	0,139794	0
putih	0	0	0	0,139794	0
turun	0	0	0	0	0,139794
rp	0	0	0	0	0,139794
liter	0	0	0	0	0,139794

### 3.5 Support Vector Machine

Penelitian ini menggunakan metode *Support Vector Machine* (SVM) dengan model linear, hal ini dikarenakan penelitian ini hanya memisahkan antara sentimen positif dan negatif dengan bantuan *hyperplane*. Komponen awal yang dibutuhkan dalam metode ini adalah pengumpulan data dari kemunculan kata dalam satu kelas atau lebih dikenal dengan *term frequency*. *Flowchart* metode *Support Vector Machine* ditunjukkan pada Gambar 3.8.



Gambar 3. 8 Flowchart Pembentukan Hyperplane

Klasifikasi dengan metode SVM bertujuan untuk mencari posisi *hyperplane* yang berfungsi sebagai batas pemisah kelas positif dan negatif. Dari visualisasi pada Gambar 3.8 dapat disimpulkan bahwa dalam penggunaan rumus SVM berasal dari rumus matematika yaitu Sistem Persamaan Linear Dua Variabel yang mana persamaan umumnya dapat dilihat pada persamaan 3.4.

$$y = ax + b \quad (3.4)$$

Sehingga jika diterapkan ke rumus SVM maka konstanta  $a$  diubah menjadi  $w$ , dimana  $w$  merupakan bobot pada metode SVM. Persamaan *hyperplane* SVM dapat dilihat pada persamaan 3.5.

$$y = w \cdot x + b \quad (3.5)$$

$w$  merupakan bobot dari setiap fitur,  $x$  merupakan nilai dari setiap fitur dari tahap TF-IDF, sedangkan  $b$  adalah bias. Dalam analisis sentimen terdapat lebih dari satu bobot dan  $x$  yang merupakan nilai bobot dari setiap fitur yang telah didapatkan dari perhitungan TF-IDF juga berjumlah banyak, sehingga rumus persamaan SVM diubah seperti yang dituliskan pada persamaan 3.6.

$$y = \sum_{i=1}^n w_i \cdot x_i + b \quad (3.6)$$

Dari persamaan 3.6 yang diterapkan dalam perhitungan rumus *Support Vector Machine*, maka perlu dicari nilai bobot awal. Dalam kasus penelitian serupa sebelumnya belum ditemukan rumus untuk penentuan nilai bobot dan bias, sehingga peneliti menetapkan dengan nilai bobot yang telah diasumsikan dengan  $w_i = 0$  dan  $b = 0$ . Selain itu perlu diinisialisasikan nilai  $C$  yang mana variabel  $C$  ini dikenal sebagai *error term*. Berikut adalah implementasinya pada program *python* :

```
def __init__(self, C = 0.0001):
    self.C = C
    self.bobot_lama = 0
    self.bias_lama = 0
```

Bobot awal dan bias sementara ini diasumsikan sedemikian rupa untuk keperluan mencari bobot dan bias yang optimal. Cara ini disebut dengan aturan perceptron, dimana aturan ini akan melakukan training melalui beberapa tahap iterasi yang biasa disebut dengan istilah *epoch*. *Epoch* sendiri merupakan istilah dalam *machine learning* yang mengacu pada satu kali iterasi melalui seluruh set data latih yang mana persamaan 3.6 adalah persamaan yang diacu untuk *epoch* tersebut. Pemberian nilai *epoch* ini dibebaskan, namun semakin besar nilai *epoch* maka semakin baik pula kinerja modelnya.

Sebelum melakukan *training* data, perlu diinisialisasikan rumus untuk mencari *margin*. *Hinge Loss Calculation* merupakan cara yang digunakan untuk memaksimalkan *margin* dan meminimalkan *loss*. Rumus untuk *Hinge Loss Calculation* sendiri dapat dilihat pada persamaan 3.7.

$$\min \frac{1}{2} \|w\|^2 + C_i \sum_{i=1}^N \max\{0, 1 - y_n(w \cdot x + b)\} \quad (3.7)$$

Berikut adalah implementasinya pada program *python* :

```
def hingeloss(self, bobot_lama, bias_lama, x, y):
    # Regularizer term sesuai rumus
    regularz = 0.5 * (bobot_lama * bobot_lama)

    for i in range(x.shape[0]):
        # Optimization term sesuai rumus
        opt_term = y[i] * ((np.dot(bobot_lama, x[i])) +
bias_lama)

        # menghitung loss
        nilai_loss = regularz + self.C * max(0, 1-opt_term)
    return nilai_loss[0][0]
```

Untuk melakukan iterasi dan menentukan bobot dan bias semmentaranya, dalam aturan perceptron juga perlu menentukan *learning rate*. Dalam aturan perceptron *learning rate* disimbolkan dengan  $\alpha$ . Nilai dari *learning rate* ditentukan

oleh peneliti, sebagian besar nilai *learning rate* diasumsikan dengan nilai  $\alpha = 0,001$  untuk sederhananya. Selain itu peneliti menginisialisasikan *batch size* untuk mengatur jumlah data yang masuk dalam satuan waktu. Berikut adalah implementasinya pada program *python* :

```
def fit(self, X, Y, batch_size=100, learning_rate=0.1,
epochs=500):

    jumlah_fitur = X.shape[1]
    jumlah_sampel = X.shape[0]
    c = self.C
    sampel = np.arange(jumlah_sampel)
    np.random.shuffle(sampel)

    # membuat array dengan jumlah feature yang ada
    bobot_lama = np.zeros((1, jumlah_fitur))
    bias_lama = 0
    losses = []

    while(self.C < 0.0001):
        l = self.hingeloss(bobot_lama, bias_lama, X, Y)
        losses.append(l)

        for batch_initial in range(0, jumlah_sampel,
batch_size):
            gradbobot = 0
            gradbias = 0
            for j in range(batch_initial, batch_initial+
batch_size):
                if j < jumlah_sampel:
                    x = sampel[j]
                    ti = Y[x] * (np.dot(bobot_lama, X[x].T)
+ bias_lama)

                    if ti > 1:
                        gradbobot += 0
                        gradbias += 0
                    else:
                        # Menghitung gradient bobot dan
bias

                        gradbobot += c * Y[x] * X[x]
                        gradbias += c * Y[x]

                        # Update nilai bobot and bias
                        bobot = bobot - learning_rate * gradbobot
                        bias = bias + learning_rate * gradbias
                    delta_w = abs(self.bobot_baru - bobot_lama)
                    delta_b = abs(self.bias_baru - bias_lama)
                    self.C = max(delta_w, delta_b)
                    self.bobot_lama = self.bobot_baru
                    self.bias_lama = self.bias_baru
```

```

self.bobot = bobot
self.bias = bias

return self.bobot, self.bias, losses

```

Dari perhitungan *training* ini, sistem mencoba mengurangi *loss* dengan melakukan iterasi berturut-turut dan mencari bobot serta bias yang optimal. Dengan cara *Gradient Descent* ini bobot dan bias diperbarui pada setiap iterasi dengan melibatkan gradien dan *learning rate* untuk meminimalisir *loss*. Rumus untuk membarui bobot dapat dilihat pada persamaan 3.8

$$w_{baru} = w_{lama} - \alpha * w_{lama} + \alpha * \text{gradien } w \quad (3.8)$$

Sehingga persamaan *hyperplane* dapat ditunjukkan pada persamaan 3.11.

$$y(\sum_{i=1}^n w_i \cdot x_i) + b \geq \text{margin} \quad (3.9)$$

Tahap selanjutnya adalah tahap implementasi sistem yang telah dibuat, yang mana hasil pengujian sistem berupa penentuan kelas berdasarkan nilai yang dihasilkan dari persamaan 3.9. Berikut adalah implementasinya pada program *python* :

```

def predict(self, X):
    prediction = np.dot(X, self.bobot[0]) + self.bias #
    bobot.x + bias
    return np.sign(prediction)

```

Fungsi *predict* ini merupakan fungsi untuk *testing* data dengan jumlah data yang ditentukan sesuai skenario uji coba. Dengan fungsi ini, sistem akan mampu memprediksi data uji dengan perbandingan dari hasil pembelajaran sistem terhadap pola penentuan kelas.

## **BAB IV**

### **UJI COBA DAN PEMBAHASAN**

Bab ini berisi hasil pengujian sistem yang terdiri dari *text preprocessing*, *Term Frequency Inverse Document Frequency*, proses klasifikasi dengan metode *Support Vector Machine*, dan yang terakhir adalah menghitung hasil uji coba dengan perhitungan nilai akurasi, presisi, *recall*, dan *f-measure* dengan menggunakan *confusion matrix*.

#### **4.1 Skenario Uji Coba**

Tahapan ini membahas tentang berbagai skenario uji coba data dalam program. Data pada sistem ini dipisah menjadi dua belah bagian yaitu data *training* dan data *testing*. Perbandingan data training dan data testing yang diterapkan dalam penelitian ini adalah 6:4, 7:3, 8:2, dan 9:1. *Confusion matrix* adalah tabel yang menyatakan klasifikasi jumlah data uji yang benar dan salah. Dalam *confusion matrix* terdapat beberapa istilah yaitu TP, TN, FP, dan FN. True Positive (TP) merupakan jumlah dokumen dari kelas positif yang diklasifikasikan benar sebagai kelas positif. True Negative (TN) merupakan jumlah dokumen dari kelas negatif yang diklasifikasikan benar sebagai kelas negatif. False Positive (FP) merupakan jumlah dokumen dari kelas negatif yang diklasifikasikan salah sebagai kelas positif. Sedangkan False Negative (FN) merupakan jumlah dokumen dari kelas positif yang diklasifikasikan salah sebagai kelas negatif. Representasi dari *confusion matrix* ditunjukkan pada Tabel 4.1. (Normawati & Prayogi, 2021).

Tabel 4. 1 *Confusion Matrix*

Kelas Sebenarnya	Kelas Prediksi	
	Positif	Negatif
Positif	True Positive (TP)	False Positive (FP)
Negatif	False Negative (FN)	True Negative (TN)

Persamaan *Confusion Matrix* yang dapat digunakan sebagai perhitungan performa akurasi ditunjukkan pada persamaan 4.1, performa presisi pada persamaan 4.2, performa recall pada persamaan 4.3, dan *f-measure* pada persamaan 4.4 (Rohman & Arifianto, 2021).

$$Akurasi(\%) = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \quad (4.1)$$

$$Presisi(\%) = \frac{TP}{TP+FP} \times 100\% \quad (4.2)$$

$$Recall(\%) = \frac{TP}{TP+FN} \times 100\% \quad (4.3)$$

$$F - measure = \frac{2 \times Presisi \times Recall}{Presisi + Recall} \quad (4.4)$$

## 4.2 Hasil Uji Coba

Sebagaimana dijelaskan pada bab sebelumnya, data yang digunakan dalam penelitian ini diperoleh dari data proses crawling di media massa online dengan kategori berita ekonomi dengan jumlah 1000 data. Data tersebut dilabeli dengan label positif dan negatif dengan jumlah data positif sebanyak 721 data dan data negatif sebanyak 279 data. Beberapa sampel data pada penelitian ini ditunjukkan pada tabel 4.2.

Tabel 4. 2 Sampel *Dataset* Penelitian

No.	Text	Label
1.	PPKM Longgar, Hunian Hotel di Makassar Naik Jadi 65 Persen	P
2.	Pemerintah Sebut Ekraf Bakal Jadi Lokomotif Penggerak di Kala Pandemi	P
3.	Lapor ke Jokowi, Perpadi Jamin Stok Beras Cukup hingga Akhir Tahun	P
4.	Ekonom Wanti-wanti Bea Masuk Ekstra Baju Impor Bikin Ogah Belanja	N
5.	Perusahaan Beri Gaji di Bawah UMP Terancam Denda Rp400 Juta	P
6.	Sri Mulyani Senang Banyak Anak Muda Paham APBN	P
7.	Tagihan Klaim Pasien Covid-19 di RI Nyaris Rp100 T pada 2021	N
8.	Kadin Kalbar: UMKM Paling Terdampak Banjir Sintang	N
9.	Garuda Tambah Penerbangan ke Lombok Jelang WSBK 2021	P
10.	Dirut BPJS Kesehatan Ungkap Pentingnya Tunjangan Sakit di Masa Pandemi	P
11.	Sri Mulyani Waspada Fenomena Pekerja Lebih Gemar WFH	P
12.	Erick Thohir Bantah Ambil Untung dari Aturan Wajib PCR	P
13.	BPDPKS Ungkap Peran Penting Riset dalam Pengembangan Industri Sawit	P
14.	Pengusaha Angkat Suara soal UMP 2022 Naik 1,09 Persen	P
15.	BI: Penyaluran KPR Naik 8,87 Persen per Oktober 2021	N
16.	KAI Luncurkan KAIPay, Bayar Tiket Bisa Pakai Uang Elektronik	P
17.	Rupiah Perkasa ke Rp14.220 per Dolar AS pada Kamis Sore	P
18.	IHSG Loyo ke 6.636 Pada Hari Ini, 18 November 2021	N
19.	Jokowi Sebut RI Bakal Ekspor Jarum Suntik	P
20.	Karyawan Laporkan Dugaan Korupsi di Garuda ke BPK	N

Data pada penelitian ini kemudian diolah melalui tahap *preprocessing*.

Tahap ini terdiri dari beberapa tahap di dalamnya yaitu *cleaning data*, *case folding*, *tokenizing*, *stemming*, dan *stopword removal*. Hasil dari penerapan *cleaning data* terdapat pada tabel 4.3.

Tabel 4. 3 Hasil *Cleaning Data*

No.	Sebelum <i>cleaning</i>	Sesudah <i>cleaning</i>
1.	PPKM Longgar, Hunian Hotel di Makassar Naik Jadi 65 Persen " -	PPKM Longgar Hunian Hotel di Makassar Naik Jadi Persen
2.	Pemerintah Sebut Ekraf Bakal Jadi Lokomotif Penggerak di Kala Pandemi " -	Pemerintah Sebut Ekraf Bakal Jadi Lokomotif Penggerak di Kala Pandemi
3.	Lapor ke Jokowi, Perpadi Jamin Stok Beras Cukup hingga Akhir Tahun " -	Lapor ke Jokowi Perpadi Jamin Stok Beras Cukup hingga Akhir Tahun
4.	Ekonom Wanti-wanti Bea Masuk Ekstra Baju Impor Bikin Ogah Belanja " -	Ekonom Wantiwanti Bea Masuk Ekstra Baju Impor Bikin Ogah Belanja
5.	Perusahaan Beri Gaji di Bawah UMP Terancam Denda Rp400 Juta " -	Perusahaan Beri Gaji di Bawah UMP Terancam Denda Rp Juta

Pada tabel 4.3 ditunjukkan perbedaan antara data pada kolom sebelum dilakukan *cleaning data* dan sesudah *cleaning data*. Pada tahap ini data dibersihkan dari berbagai macam *noise data* seperti simbol, angka, dan tanda baca. Tahapan selanjutnya adalah tahap *case folding* sebagaimana hasil uji coba tahap *case folding* ditunjukkan pada tabel 4.4.

Tabel 4. 4 Hasil *Case Folding*

No.	Sebelum <i>case folding</i>	Sesudah <i>case folding</i>
1.	PPKM Longgar Hunian Hotel di Makassar Naik Jadi Persen	ppkm longgar hunian hotel di makassar naik jadi persen
2.	Pemerintah Sebut Ekraf Bakal Jadi Lokomotif Penggerak di Kala Pandemi	pemerintah sebut ekraf bakal jadi lokomotif penggerak di kala pandemi
3.	Lapor ke Jokowi Perpadi Jamin Stok Beras Cukup hingga Akhir Tahun	lapor ke jokowi perpadi jamin stok beras cukup hingga akhir tahun
4.	Ekonom Wantiwanti Bea Masuk Ekstra Baju Impor Bikin Ogah Belanja	ekonom wantiwanti bea masuk ekstra baju impor bikin ogah belanja
5.	Perusahaan Beri Gaji di Bawah UMP Terancam Denda Rp Juta	perusahaan beri gaji di bawah ump terancam denda rp juta

Pada tabel 4.4 ditunjukkan perbedaan data sebelum *case folding* dan sesudah *case folding*. Tahap ini mengolah data dengan cara menyeragamkan teks menjadi huruf kecil (*lowecase*). Tahap selanjutnya dari proses *preprocessing* adalah *tokenizing* sebagaimana ditunjukkan pada tabel 4.5.

Tabel 4. 5 Hasil *Tokenizing*

No.	Sebelum <i>tokenizing</i>	Sesudah <i>tokenizing</i>
1.	ppkm longgar hunian hotel di makassar naik jadi persen	ppkm,longgar,hunian,hotel,di,makassar,naik,jadi,persen
2.	pemerintah sebut ekraf bakal jadi lokomotif penggerak di kala pandemi	pemerintah,sebut,ekraf,bakal,jadi,lokomotif,penggerak,di,kala,pandemi
3.	lapor ke jokowi perpadi jamin stok beras cukup hingga akhir tahun	lapor,ke,jokowi,perpadi,jamin,stok,beras,cukup,hingga,akhir,tahun
4.	ekonom wantiwanti bea masuk ekstra baju impor bikin ogah belanja	ekonom,wantiwanti,bea,masuk,ekstra,baju,impor,bikin,ogah,belanja
5.	perusahaan beri gaji di bawah ump terancam denda rp juta	perusahaan,beri,gaji,di,bawah,ump,terancam,denda,rp,juta

Pada tabel 4.5 ditunjukkan perbedaan hasil *tokenizing*, dimana perbedaannya dapat dilihat bahwa data yang awalnya berupa kalimat dirubah menjadi dipisahkan per kata/token. Tahap selanjutnya pada proses *preprocessing* data adalah *stemming*, sebagaimana ditunjukkan pada tabel 4.6.

Tabel 4. 6 Hasil *Stemming*

No.	Sebelum <i>stemming</i>	Sesudah <i>stemming</i>
1.	ppkm,longgar,hunian,hotel,di,makassar,naik,jadi,persen	ppkm,longgar,huni,hotel,di,makassar,naik,jadi,persen
2.	pemerintah,sebut,ekraf,bakal,jadi,lokomotif,penggerak,di,kala,pandemi	perintah,sebut,ekraf,bakal,jadi,lokomotif,gerak,di,kala,pandemi
3.	lapor,ke,jokowi,perpadi,jamin,stok,beras,cukup,hingga,akhir,tahun	lapor,ke,jokowi,padi,jamin,stok,beras,cukup,hingga,akhir,tahun
4.	ekonom,wantiwanti,bea,masuk,ekstra,baju,impor,bikin,ogah,belanja	ekonom,wantiwanti,bea,masuk,ekstra,baju,impor,bikin,ogah,belanja
5.	perusahaan,beri,gaji,di,bawah,ump,terancam,denda,rp,juta	usaha,beri,gaji,di,bawah,ump,ancam,denda,rp,juta

Pada tabel 4.6 ditunjukkan perbedaan hasil *stemming*, dimana tahap ini berfungsi untuk menghilangkan imbuhan pada kata sehingga menghasilkan kata dasar. Tahap terakhir pada proses *preprocessing* yaitu *stopword removal* ditunjukkan pada tabel 4.7.

Tabel 4. 7 Hasil *Stopword Removal*

No.	Sebelum <i>stopword removal</i>	Sesudah <i>stopword removal</i>
1.	ppkm,longgar,huni,hotel,di,makassar,naik,jadi,persen	ppkm longgar huni hotel makassar persen
2.	perintah,sebut,ekraf,bakal,jadi,lokomotif,gerak,di,kala,pandemi	perintah ekraf lokomotif gerak pandemi
3.	lapor,ke,jokowi,padi,jamin,stok,beras,cukup,hingga,akhir,tahun	lapor jokowi padi jamin stok beras
4.	ekonom,wantiwanti,bea,masuk,ekstra,baju,impor,bikin,ogah,belanja	ekonom wantiwanti bea masuk ekstra baju impor bikin ogah belanja
5.	usaha,beri,gaji,di,bawah,ump,ancam,denda,rp,juta	usaha gaji ump ancam denda rp juta

Tabel 4.7 menunjukkan hasil dari proses *stopword removal*, dimana tahap ini berfungsi untuk menghilangkan kata yang tidak mengandung sentimen. Setelah melalui proses *preprocessing* data, maka data yang digunakan hanyalah data pada kolom sesudah *stopword removal* seperti yang ditunjukkan pada tabel 4.7 dan labelnya saja. Beberapa hasil *preprocessing* data dapat dilihat pada tabel 4.8.

Tabel 4. 8 Hasil *Preprocessing*

No.	Text	Label
1.	ppkm longgar huni hotel makassar persen	P
2.	perintah ekraf lokomotif gerak pandemi	P
3.	lapor jokowi padi jamin stok beras	P
4.	ekonom wantiwanti bea masuk ekstra baju impor bikin ogah belanja	N
5.	usaha gaji ump ancam denda rp juta	P
6.	sri mulyani senang anak muda paham apbn	P
7.	tagih klaim pasien covid ri rp t	N
8.	kadin kalbar umkm dampak banjir sintang	N
9.	garuda terbang lombok jelang wsbk	P
10.	dirut bpjs sehat tunjang sakit pandemi	P

Dari skenario uji coba yang sudah dijelaskan sebelumnya dalam bab 4.1 didapatkan hasil performa yang berbeda. Hal ini dikarenakan oleh perbedaan pembagian data yang mempengaruhi jumlah data *training* dan data *testing*.

#### 4.2.1 Hasil Uji Coba Ke-1

Dari uji coba skenario pembagian data menggunakan rasio 6:4 menghasilkan pembagian data yang terdiri dari 600 data *training* dan juga 400 data *testing*. Jumlah data dengan sentimen positif pada data *training* sebanyak 422 dan data negatif sebanyak 178. Sedangkan pada data *testing* jumlah data dengan sentimen positif adalah 299 dan data dengan sentimen negatif sebanyak 101. *Confusion matrix* dari uji coba ke-1 dapat dilihat pada tabel 4.9.

Tabel 4. 9 *Confusion Matrix* Uji Coba Ke-1

Kelas Sebenarnya	Kelas Prediksi	
	Positif	Negatif
Positif	263(TP)	74(FP)
Negatif	30(FN)	33(TN)

Berdasarkan hasil perhitungan dengan skenario perbandingan data 6:4 didapatkan akurasi sebesar 74%, presisi sebesar 78,04%, *recall* sebesar 89,76%, dan *f-measure* sebesar 83,49%

#### 4.2.2 Hasil Uji Coba Ke-2

Dari uji coba skenario pembagian data menggunakan rasio 7:3 menghasilkan pembagian data yang terdiri dari 700 data *training* dan 300 data *testing*. Jumlah data dengan sentimen positif pada data *training* sebanyak 502 dan

data negatif sebanyak 198. Sedangkan pada data *testing* jumlah data dengan sentimen positif adalah 219 dan data dengan sentimen negatif sebanyak 81. *Confusion matrix* dari uji coba ke-1 dapat dilihat pada tabel 4.10.

Tabel 4. 10 *Confusion Matrix* Uji Coba Ke-2

Kelas Sebenarnya	Kelas Prediksi	
	Positif	Negatif
Positif	204(TP)	54(FP)
Negatif	23(FN)	19(TN)

Berdasarkan hasil perhitungan dengan skenario perbandingan data 7:3 didapatkan akurasi sebesar 74,33%, presisi sebesar 79,06%, *recall* sebesar 89,86%, dan *f-measure* sebesar 84,12%.

#### 4.2.3 Hasil Uji Coba Ke-3

Dari uji coba skenario pembagian data menggunakan rasio 8:2 menghasilkan pembagian data yang terdiri dari 800 data *training* dan 200 data *testing*. Jumlah data dengan sentimen positif pada data *training* sebanyak 577 dan data negatif sebanyak 223. Sedangkan pada data *testing* jumlah data dengan sentimen positif adalah 144 dan data sentimen negatif sebanyak 56. *Confusion matrix* dari uji coba ke-1 dapat dilihat pada tabel 4.11.

Tabel 4. 11 *Confusion Matrix* Uji Coba Ke-3

Kelas Sebenarnya	Kelas Prediksi	
	Positif	Negatif
Positif	136(TP)	41(FP)
Negatif	14(FN)	9(TN)

Berdasarkan hasil perhitungan dengan skenario perbandingan data 8:2 didapatkan akurasi sebesar 72,5 %, presisi sebesar 76,8%, *recall* sebesar 90,6%, dan *f-measure* sebesar 83,1%.

#### 4.2.4 Hasil Uji Coba Ke-4

Dari uji coba skenario pembagian data menggunakan rasio 9:1 menghasilkan pembagian data yang terdiri dari 900 data *training* dan 100 data *testing*. Jumlah data dengan sentimen positif pada data *training* sebanyak 647 dan data negatif sebanyak 253. Sedangkan pada data *testing* jumlah data dengan sentimen positif adalah 74 dan data dengan sentimen negatif sebanyak 26. *Confusion matrix* dari uji coba ke-1 dapat dilihat pada tabel 4.12.

Tabel 4. 12 *Confusion Matrix* Uji Coba Ke-4

Kelas Sebenarnya	Kelas Prediksi	
	Positif	Negatif
Positif	66(TP)	16(FP)
Negatif	8(FN)	10(TN)

Berdasarkan hasil perhitungan dengan skenario perbandingan data 9:1 didapatkan akurasi sebesar 76%, presisi sebesar 80,48%, *recall* sebesar 89,18%, dan *f-measure* sebesar 84,61%.

### 4.3 Pembahasan

Berdasarkan dari hasil dari uji coba yang sudah dilakukan terhadap 1000 data berita yang telah diberi label dapat diketahui bahwa data positif memiliki jumlah yang lebih banyak dari data sentimen negatif. Data hasil *web crawling*

tersebut perlu diolah terlebih dahulu dengan proses *preprocessing* dengan urutan tahapan *cleaning data*, *case folding*, *tokenizing*, *stemming*, dan *stopword removal*. Selain membantu mempermudah dalam mengolah data, tahap *preprocessing* data ini merupakan salah satu faktor yang mempengaruhi performa klasifikasi metode *Support Vector Machine*. Selain itu TF-IDF juga merupakan faktor yang membantu dalam pengolahan data sentimen karena TF-IDF mampu merubah kata menjadi angka sehingga mampu diolah dengan metode *Support Vector Machine*.

Pengukuran hasil evaluasi model dengan beberapa skenario uji coba juga memberikan hasil performa yang berbeda. Sebagaimana telah diejelaskan sebelumnya pada bab 4.2 hasil perbandingan akurasi dari beberapa skenario uji coba dapat dilihat pada tabel 4.13.

Tabel 4. 13 Hasil Akurasi Uji Coba

No.	Rasio Uji Coba	Akurasi
1.	6:4	74%
2.	7:3	74,33%
3.	8:2	72,5%
4.	9:1	76%

Dari tabel 4.13 diketahui perbandingan dari masing masing rasio dan diketahui dari beberapa skenario uji coba diperoleh hasil akurasi terbesar pada rasio 9:1. Selain itu tabel 4.13 juga menunjukkan pengaruh besar pembagian rasio dapat mempengaruhi akurasi.

Sistem ini dibuat dengan harapan mampu melakukan klasifikasi sentimen dari berita ekonomi pasca pandemi COVID-19 dengan baik dan akurat. Sesuai dengan manfaat penelitian, adanya sistem ini diharapkan mampu mengurangi resiko kesalahan informasi yang menyebabkan kesalahan dalam menyusun strategi

untuk masa mendatang. Sistem ini diharapkan mampu menjadi alat bantu dengan tujuan mencari kebenaran data. Mengingat dalam Islam Allah *Subhanahu Wa Ta'ala* telah berfirman dalam surah Al – Baqarah ayat 42, yang berbunyi :

وَلَا تَلْبِسُوا الْحَقَّ بِالْبَاطِلِ وَتَكْتُمُوا الْحَقَّ وَأَنْتُمْ تَعْلَمُونَ

“Jangan kalian mencampur kebenaran dan kebatilan. Jangan juga kalian menyembunyikan kebenaran. Padahal kalian menyadarinya.” (Qur'an Surah Al – Baqarah ayat 42).

Dari ayat tersebut dapat dilihat bahwa Allah *Subhanahu Wa Ta'ala* memerintahkan manusia untuk menyampaikan kebenaran tanpa menutupinya. Dalam hubungannya pada sistem ini yaitu adanya kebenaran yang berupa data ril dan juga kebatilan yang berupa *error*.

Menurut tafsir Ibnu Katsir, dalam ayat ini Allah *Subhanahu Wa Ta'ala* melarang orang – orang Yahudi melakukan hal yang biasa mereka kerjakan di masa lalu, misalnya mencampuradukkan antara perkara yang benar dan yang salah, menyembunyikan perkara yang benar dan salah, serta memulas perkara yang batil dengan perkara yang hak. Hal serupa diriwayatkan dari Sa'id ibnu Jubair dan Ar-Rabi' ibnu Anas. Qatadah mengatakan bahwa dalam firman-Nya, "Janganlah kalian campur adukkan yang hak dengan yang batil" (Al-Baqarah: 42), yakni janganlah kalian campur adukkan antara Yahudi dan Nasrani dengan Islam, sedangkan kalian mengetahui bahwasanya agama Allah itu adalah agama Islam; agama Yahudi dan agama Nasrani itu adalah bid'ah, bukan dari Allah. Hal yang semisal telah diriwayatkan dari Al-Hasan Al-Basri. (*Tafsir Surat Al - Baqarah, ayat 42, n.d.*)

Selain itu Allah *Subhanahu Wa Ta'ala* juga berfirman dalam surah Al – Imran ayat 71 yang berbunyi :

يَا أَهْلَ الْكِتَابِ لِمَ تَلْبِسُونَ الْحَقَّ بِالْبَاطِلِ وَتَكْتُمُونَ الْحَقَّ وَأَنْتُمْ تَعْلَمُونَ

*"Wahai Ahli Kitab! Mengapa kamu mencampuradukkan kebenaran dengan kebatilan, dan kamu menyembunyikan kebenaran, padahal kamu mengetahui?"*  
(QS Ali Imran: 71).

Menurut tafsir Ibnu Katsir dalam ayat ini Allah *Subhanahu Wa Ta'ala* memberitakan perihal kedengkian para orang Yahudi kepada kaum mukmin dan mereka selalu menginginkan agar kaum mukmin menjadi sesat. Allah memberitakan juga bahwa perbuatan mereka itu justru menjadi senjata makan tuan, sedangkan mereka tidak sedikitpun merasakan bahwa tipu daya diri mereka justru akibat buruknya menimpa diri mereka sendiri. (*Tafsir Surat Al - Imran, ayat 71, n.d.*)

## BAB V

### KESIMPULAN DAN SARAN

#### 5.1 Kesimpulan

*Dataset* yang digunakan pada penelitian ini didapatkan dari proses *crawling data* pada berita online dari 5 *website* yang berbeda yaitu *cnnindonesia.com*, *jppn.com*, *detik.com*, *bisnis.com*, dan *kompas.com* dengan teknik *web scrapping*. Jumlah data yang digunakan sebanyak 1000 data dengan masing-masing *website* sebanyak 200 data. Data-data tersebut dilabeli dengan label positif dan negatif dengan jumlah data positif sebanyak 721 data dan data negatif sebanyak 279 data. Uji coba dalam penelitian ini dikombinasikan dalam hal pembagian *data training* dan *data testing*, rasio pembagiannya adalah 6:4, 7:3, 8:2, dan 9:1. Pengujian dilakukan menggunakan variabel dengan jumlah yang sama yaitu, bobot = 0, bias = 0, *learning rate* = 0.001, dan *epoch* = 1000. Dari hasil uji coba dengan berbagai macam pembagian *dataset* didapatkan hasil terbaik pada perbandingan *dataset* 9:1 dengan akurasi 76%, presisi sebesar 80,48%, *recall* sebesar 89,18%, dan *f-measure* sebesar 84,61%. Oleh karena itu dapat disimpulkan bahwa besar rasio pembagian *dataset* mempengaruhi besar performa.

#### 5.2 Saran

Berdasarkan hasil uji coba dari penelitian ini penulis menyadari bahwa hasil dari pengolahan terhadap *dataset* masih bisa ditingkatkan lebih lagi pada hasil performanya. Oleh karena itu penulis menyarankan beberapa hal bagi penelitian selanjutnya, yaitu :

1. Menggunakan metode yang berbeda untuk pembobotan kata seperti *Word2Vec*, *FastText*, dan metode yang lain.
2. Menggunakan metode pengolahan data selain *support vector machine* agar dapat dibandingkan hasil performanya antar metode sehingga didapatkan nilai performa yang terbaik.
3. Menggunakan *K-fold Cross Validation* untuk pembagian rasio antara *data training* dan *data testing* sehingga mampu didapatkan kombinasi rasio terbaik dalam mengukur performa.

## DAFTAR PUSTAKA

- Alam, S., & Yao, N. (2019). The impact of preprocessing steps on the accuracy of machine learning algorithms in sentiment analysis. *Computational and Mathematical Organization Theory*, 25(3), 319–335. <https://doi.org/10.1007/s10588-018-9266-8>
- Cahyani, D. E., & Patasik, I. (2021). Performance comparison of tf-idf and word2vec models for emotion text classification. *Bulletin of Electrical Engineering and Informatics*, 10(5), 2780–2788. <https://doi.org/10.11591/eei.v10i5.3157>
- Dewan Pers.* (n.d.). Retrieved March 24, 2022, from [https://dewanpers.or.id/publikasi/opini\\_detail/173/Media\\_Online\\_Perlu\\_Berbenah\\_Diri](https://dewanpers.or.id/publikasi/opini_detail/173/Media_Online_Perlu_Berbenah_Diri)
- Haddi, E., Liu, X., & Shi, Y. (2013). The role of text pre-processing in sentiment analysis. *Procedia Computer Science*, 17, 26–32. <https://doi.org/10.1016/j.procs.2013.05.005>
- Hickman, L., Thapa, S., Tay, L., Cao, M., & Srinivasan, P. (2022). Text Preprocessing for Text Mining in Organizational Research: Review and Recommendations. *Organizational Research Methods*, 25(1), 114–146. <https://doi.org/10.1177/1094428120971683>
- Husna, R. El, Wasono, R., Haris, M. Al, Semarang, U. M., & Sentimen, A. (2020). *ANALISIS SENTIMEN PADA TWITTER MENGENAI NETFLIX DIBLOKIR TELKOM MENGGUNAKAN SUPPORT VECTOR* Menonton merupakan salah satu hiburan yang sangat digemari oleh masyarakat . Di zaman digital yang serba cepat dan serba ada , banyak masyarakat yang beralih dari m. 214–222.
- Ini Daftar Top 5 Ranking Website Media Online di Indonesia 2021 - BeritaManado.com: Berita Terkini Manado, Sulawesi Utara.* (n.d.). Retrieved March 17, 2022, from <https://beritamanado.com/ini-daftar-top-5-ranking-website-media-online-di-indonesia-2021/>
- Kurniawan, S., Gata, W., Puspitawati, D. A., -, N., Tabrani, M., & Novel, K. (2019). Perbandingan Metode Klasifikasi Analisis Sentimen Tokoh Politik Pada Komentar Media Berita Online. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 3(2), 176–183. <https://doi.org/10.29207/resti.v3i2.935>
- Madasu, A., & Elango, S. (2020). Efficient feature selection techniques for sentiment analysis. *Multimedia Tools and Applications*, 79(9–10), 6313– 6335. <https://doi.org/10.1007/s11042-019-08409-z>
- Normawati, D., & Prayogi, S. A. (2021). Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter. *Jurnal Sains Komputer & Informatika (J-Sakti)*, 5(2), 697–711.

- Permana, R. A., & Sahara, S. (2019). Metode Support Vector Machine Sebagai Penentu Kelulusan Mahasiswa pada Pembelajaran Elektronik. *Jurnal Khatulistiwa Informatika*, 7(1), 50–58. <https://doi.org/10.31294/jki.v7i1.5743>
- Que, V. K. S., Iriani, A., & Purnomo, H. D. (2020). Analisis Sentimen Transportasi Online Menggunakan Support Vector Machine Berbasis Particle Swarm Optimization. *Jurnal Nasional Teknik Elektro Dan Teknologi Informasi*, 9(2), 162–170. <https://doi.org/10.22146/jnteti.v9i2.102>
- Ratino, Hafidz, N., Anggraeni, S., & Gata, W. (2020). Sentimen Analisis Informasi Covid-19 menggunakan Support Vector Machine dan Naïve Bayes. *Jurnal JUPITER*, 12(2), 1–11.
- Rohman, M. A., & Arifianto, D. (2021). Penerapan Metode Euclidean Probability dan Confusion Matrix dalam Diagnosa Penyakit Koi. *Jurnal Smart Teknologi*, 2(2), 122–130.
- Solten Rajagukguk, S., & Olilia, S. (2020). PERAN MEDIA MASSA TERHADAP MASYARAKAT DI SAAT PANDEMI COVID 19 Solten Rajagukguk 1 Sandra Olifia 2. *Peran Media Massa Terhadap Masyarakat Di Saat Pandemi Covid-19*, 3(1), 42–52. <http://isip.usni.ac.id/jurnal/Solten.pdf>
- Suhardjono, Ganda, W., & Abdul, H. (2019). Prediksi Waktu Kelulusan Mahasiswa Menggunakan Svm Berbasis Pso. *Bianglala Informatika*, 7(2), 97–101.
- Tafsir Surat Al-Hujurat, ayat 6-8.* (n.d.). Retrieved March 24, 2022, from <http://www.ibnukatsironline.com/2015/10/tafsir-surat-al-hujurat-ayat-6-8.html>
- Tafsir Surat Al - Baqarah, ayat 42.* (n.d.). Retrieved May 8, 2023, from <http://www.ibnukatsironline.com/2014/08/tafsir-surat-al-baqarah-ayat-42-43.html>
- Tafsir Surat Al - Imran, ayat 71.* (n.d.). Retrieved May 8, 2023, from <https://tafsir.learn-quran.co/id/surat-3-al-'imran/ayat-71>
- Ulfah, A. N., & Anam, M. K. (2020). Analisis Sentimen Hate Speech Pada Portal Berita Online Menggunakan Support Vector Machine (SVM). *JATISI (Jurnal Teknik Informatika Dan Sistem Informasi)*, 7(1), 1–10. <https://doi.org/10.35957/jatisi.v7i1.196>
- Zhou, H. (2022). Research of Text Classification Based on TF-IDF and CNN- LSTM. *Journal of Physics: Conference Series*, 2171(1), 218–222. <https://doi.org/10.1088/1742-6596/2171/1/012021>

# LAMPIRAN

## Lampiran 1

Tabel Dataset dan Label

No	teks	label
1	PPKM Longgar, Hunian Hotel di Makassar Naik Jadi 65 Persen " -	P
2	Pemerintah Sebut Ekraf Bakal Jadi Lokomotif Penggerak di Kala Pandemi " -	P
3	Lapor ke Jokowi, Perpadi Jamin Stok Beras Cukup hingga Akhir Tahun " -	P
4	Ekonom Wanti-wanti Bea Masuk Ekstra Baju Impor Bikin Ogah Belanja " -	N
5	Perusahaan Beri Gaji di Bawah UMP Terancam Denda Rp400 Juta " -	P
6	Sri Mulyani Senang Banyak Anak Muda Paham APBN " -	P
7	Tagihan Klaim Pasien Covid-19 di RI Nyaris Rp100 T pada 2021 " -	N
8	Kadin Kalbar: UMKM Paling Terdampak Banjir Sintang " -	N
9	Garuda Tambah Penerbangan ke Lombok Jelang WSBK 2021 " -	P
10	BI Beli SBN di Pasar Perdana Rp143,32 T per 16 November 2021 " -	P
11	Dirut BPJS Kesehatan Ungkap Pentingnya Tunjangan Sakit di	P

	Masa Pandemi " -	
12	Pengusaha soal PPKM Level 3 saat Nataru: Otomatis Berdampak ke Ekonomi " -	N
13	Sri Mulyani Waspada Fenomena Pekerja Lebih Gemar WFH " -	P
14	Jokowi Bidik Ekspor Produk Turunan Nikel Tembus Rp284 T Tahun Ini " -	P
15	Erick Thohir Bantah Ambil Untung dari Aturan Wajib PCR " -	P
16	BPDPKS Ungkap Peran Penting Riset dalam Pengembangan Industri Sawit " -	P
17	Pengusaha Angkat Suara soal UMP 2022 Naik 1,09 Persen " -	P
18	BI: Penyaluran KPR Naik 8,87 Persen per Oktober 2021 " -	N
19	Airlangga: Top Up Kartu Sembako Cair Awal atau Akhir Desember " -	P
20	Daftar Negara yang Menggunakan Mata Uang Euro " -	P
21	Jokowi Bakal Larang Ekspor Bauksit 2022 dan Tembaga 2023 " -	P
22	KAI Luncurkan KAIPay, Bayar Tiket Bisa Pakai Uang Elektronik	P

	" -	
23	Jokowi Batasi Waktu Pengembangan Digital: 2 Tahun " -	P
24	Gojek Perusahaan Luhut Kucurkan Rp17 T Buat Ekosistem Motor Listrik " -	P
25	SKK Migas Kembali Gelar Konvensi Migas Terbesar di Indonesia " -	P
...		
976	Didenda Rp 1,3 Triliun, Indosat GIG Berhenti Beroperasi 25 November	N
977	Tips agar Omzet Usaha Tetap Stabil Tanpa Obral Diskon	P
978	Daftar Harga Sembako Hari Ini di Jakarta	P
979	BI Catat Sepekan Aliran Modal Asing Rp 2,32 Triliun Keluar dari Indonesia	N
980	IHSG Sepekan Naik 1,04 Persen, Kapitalisasi Pasar Jadi Rp 8.245 Triliun	P
981	Baru Diangkat Jadi Manajer? Segeralah Tancap Gas Lakukan Ini	P
982	Turun Rp 6.000, Simak Rincian Harga Emas Batangan Antam Hari Ini	P
983	Sri Mulyani Minta Pengusaha Tak Lupa Bayar Pajak meski Sibuk Ekspansi Bisnis	P
984	SKB CPNS Kemenkumham 2021: Tahapan, Jadwal, Lokasi, dan Materi Ujian	P
985	Intellectual Humility	P
986	Wall Street Ditutup Bervariasi, Indeks Nasdaq Cetak Rekor Baru	P
987	Mengenal Lebih Dekat dengan Upah Minimum	P
988	Banyak Diskon, Ini Katalog Promo Alfamart 19-21 November 2021	P

989	Gojek dan Perusahaan Luhut Siapkan Rp 17 Triliun untuk Kembangkan Kendaraan Listrik	P
990	Airlangga: Presidensi G20 Indonesia Jadi Momentum Branding di Dunia Internasional	P
991	[POPULER MONEY] Akan Banyak Warga Punya Solar Panel   Gaji Jenderal Polisi	P
992	Ada Diskon Minyak Goreng, Ini Katalog Promo Indomaret 19-21 November 2021	P
993	Mulai Besok, Akses Keluar/Masuk di Stasiun Bekasi akan Mengalami Perubahan	P
994	Kenaikan Upah 1,09 Persen, KSPSI: Ini Tidak Adil, Kami Tak Akan Diam	N
995	Bos KCIC Pamer Kereta Cepat Bisa Prediksi Gempa dan Tangkal Petir	P
996	Luhut Bangga: WSBK di Sirkuit Mandalika Ditonton 1,6 Miliar Penduduk Dunia	P
997	Rincian UMP 2022 Se-Pulau Jawa, dari Tertinggi hingga Terendah	P
998	Usai WSBK, Menhub Siapkan Prasarana Transportasi untuk MotoGP dan F1	P
999	Bahagiaanya Luhut, Sirkuit Mandalika Jadi Buah Bibir Seantero Negeri	P
1000	Naik Rp 31.135 dan Rp 13.956, Segini UMP Jabar dan Jateng 2022	P

## Lampiran 2

Tabel Perbandingan Label Ahli dan Label *Crowdsourcing*

No	label	label ahli	No	label	label ahli
1	P	P	26	P	P
2	P	P	27	P	P
3	P	P	28	N	N
4	N	N	29	P	P
5	P	P	30	N	N
6	P	P	31	P	P

7	N	N	32	N	N
8	N	N	33	P	P
9	P	P	34	N	N
10	P	P	35	N	N
11	P	P	36	N	N
12	N	N	37	P	P
13	P	P	38	P	P
14	P	P	39	P	P
15	P	P	40	N	N
16	P	P	41	N	N
17	P	P	42	P	P
18	N	N	43	P	P
19	P	P	44	N	N
20	P	P	45	N	N
21	P	P	46	P	P
22	P	P	47	P	P
23	P	P	48	N	N
24	P	P	49	P	P
25	P	P	50	N	N

No	label	label ahli	No	label	label ahli
51	P	P	76	P	P
52	P	P	77	P	P
53	P	P	78	P	P
54	P	P	79	P	P
55	P	P	80	P	P
56	P	P	81	P	P
57	N	N	82	P	P
58	N	N	83	P	P
59	P	P	84	P	P
60	P	P	85	P	P
61	P	P	86	N	N
62	P	P	87	N	N
63	P	P	88	P	P
64	P	P	89	P	P
65	P	P	90	P	P
66	N	N	91	N	N
67	P	P	92	N	N
68	P	P	93	P	P
69	N	N	94	P	P
70	P	P	95	N	N









