



**ANALISIS SENTIMEN RESPONS MASYARAKAT TERHADAP KARTU
PRAKERJA MENGGUNAKAN ALGORITMA K-NN, NAÏVE BAYES
DAN SVM**

TUGAS AKHIR

Tri Putra Adimas Nugraha
41518010057

UNIVERSITAS
MERCU BUANA
**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS MERCU BUANA
JAKARTA
2022**



**ANALISIS SENTIMEN RESPONS MASYARAKAT
TERHADAP KARTU PRAKERJA MENGGUNAKAN
ALGORITMA K-NN, NAÏVE BAYES DAN SVM**

Tugas Akhir

Diajukan Untuk Melengkapi Salah Satu Syarat
Memperoleh Gelar Sarjana Komputer

Oleh:

Tri Putra Adimas Nugraha

41518010057

UNIVERSITAS
MERCU BUANA
PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS MERCU BUANA
JAKARTA
2022

LEMBAR PERNYATAAN ORISINALITAS

LEMBAR PERNYATAAN ORISINALITAS

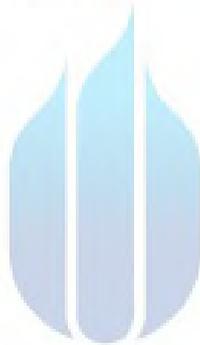
Yang bertanda tangan dibawah ini:

NIM : 41518010057

Nama : Tri Putra Adimas Nugraha

Judul Tugas Akhir : Analisis Sentimen Respons Masyarakat Terhadap Kartu Prakerja Menggunakan Algoritma K-NN, Naïve Bayes dan SVM.

Menyatakan bahwa Laporan Tugas Akhir saya adalah hasil karya sendiri dan bukan plagiat. Apabila ternyata ditemukan didalam laporan Tugas Akhir saya terdapat unsur plagiat, maka saya siap untuk mendapatkan sanksi akademik yang terkait dengan hal tersebut.



Jakarta, 15 Agustus 2022



Tri Putra Adimas Nugraha

UNIVERSITAS
MERCU BUANA

SURAT PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR

SURAT PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR

Sebagai mahasiswa Universitas Mercu Buana, saya yang bertanda tangan di bawah ini:

Nama Mahasiswa : Tri Putra Adimas Nugraha
NIM : 41518010057
Judul Tugas Akhir : Analisis Sentimen Respons Masyarakat Terhadap Kartu Prakerja Menggunakan Algoritma K-NN, Naive Bayes dan SVM.

Dengan ini memberikan izin dan menyetujui untuk memberikan kepada Universitas Mercu Buana **Hak Bebas Royalti Noneksklusif** (*None-exclusive Royalty Free Right*) atas karya ilmiah saya yang berjudul diatas beserta perangkat yang ada (jika diperlukan).

Dengan Hak Bebas Royalti/Noneksklusif ini Universitas Mercu Buana berhak menyimpan, mengalihmedia/formatkan, mengelola dalam bentuk pangkalan data (*database*), merawat dan mempublikasikan tugas akhir saya.

Selain itu, demi pengembangan ilmu pengetahuan di lingkungan Universitas Mercu Buana, saya memberikan izin kepada Peneliti di Lab Riset Fakultas Ilmu Komputer, Universitas Mercu Buana untuk menggunakan dan mengembangkan hasil riset yang ada dalam tugas akhir untuk kepentingan riset dan publikasi selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Jakarta, 15 Agustus 2022

UNIVERSITAS
MERCU BUANA



Tri Putra Adimas Nugraha

SURAT PERNYATAAN LUARAN TUGAS AKHIR

SURAT PERNYATAAN LUARAN TUGAS AKHIR

Sebagai mahasiswa Universitas Mercu Buana, saya yang bertanda tangan di bawah ini:

Nama Mahasiswa : Tri Putra Adimas Nugraha
NIM : 41518010057
Judul Tugas Akhir : Analisis Sentimen Respons Masyarakat Terhadap Kartu Prakerja Menggunakan Algoritma K-NN, Naive Bayes dan SVM.

Menyatakan bahwa:

1. Luaran Tugas Akhir saya adalah sebagai berikut:

No	Luaran	Jenis	Status
1	Publikasi Ilmiah	Jurnal Nasional Tidak Terakreditasi	Diajukan ✓
		Jurnal Nasional Terakreditasi	✓
		Jurnal International Tidak Bereputasi	Diterima
		Jurnal International Bereputasi	
Disubmit/dipublikasikan di:	Nama Jurnal	: Rekayasa Sistem dan Teknologi Informasi (RESTI)	
	ISSN	: 2580-0760	
	Link Jurnal	: http://jurnal.iaii.or.id/index.php/RESTI	
	Link File Jurnal Jika Sudah di Publish	:	

2. Bersedia untuk menyelesaikan seluruh proses publikasi artikel mulai dari submit revisi artikel sampai dengan dinyatakan dapat diterbitkan pada jurnal yang dituju.
3. Diminta untuk melampirkan scan KTP dan Surat Pernyataan (Lihat Lampiran Dokumen HKI), untuk kepentingan pendaftaran HKI apabila diperlukan

Demikian pernyataan ini saya buat dengan sebenarnya.

Jakarta, 15 Agustus 2022



Tri Putra Adimas Nugraha

LEMBAR PERSETUJUAN PENGUJI

LEMBAR PERSETUJUAN PENGUJI

NIM : 41518010057
Nama : Tri Putra Adimas Nugraha
Judul Tugas Akhir : Analisis Sentimen Respons Masyarakat Terhadap
Kartu Prakerja Menggunakan Algoritma K-NN,
Naïve Bayes dan SVM

Tugas Akhir ini telah diperiksa dan disidangkan sebagai salah satu persyaratan untuk memperoleh gelar Sarjana pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana.

Jakarta, 29 Agustus 2022



(Wawan Gunawan, S.Kom, MT)

UNIVERSITAS
MERCU BUANA

LEMBAR PERSETUJUAN PENGUJI

LEMBAR PERSETUJUAN PENGUJI

NIM : 41518010057
Nama : Tri Putra Adimas Nugraha
Judul Tugas Akhir : Analisis Sentimen Respons Masyarakat Terhadap
Kartu Prakerja Menggunakan Algoritma K-NN,
Naïve Bayes dan SVM

Tugas Akhir ini telah diperiksa dan disidangkan sebagai salah satu persyaratan untuk memperoleh gelar Sarjana pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana.

Jakarta, 29 Agustus 2022



UNIVERSITAS
(Harni Kusniyati, M.Kom)
MERCU BUANA

LEMBAR PERSETUJUAN PENGUJI

LEMBAR PERSETUJUAN PENGUJI

NIM : 41518010057
Nama : Tri Putra Adimas Nugraha
Judul Tugas Akhir : Analisis Sentimen Respons Masyarakat Terhadap
Kartu Prakerja Menggunakan Algoritma K-NN,
Naïve Bayes dan SVM

Tugas Akhir ini telah diperiksa dan disidangkan sebagai salah satu persyaratan untuk memperoleh gelar Sarjana pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana.

Jakarta, 10 Agustus 2022



(Umniy Salamah, ST., MMSI)

UNIVERSITAS
MERCU BUANA

LEMBAR PENGESAHAN

LEMBAR PENGESAHAN

NIM : 41518010057
Nama : Tri Putra Adimas Nugraha
Judul Tugas Akhir : Analisis Sentimen Respons Masyarakat Terhadap Kartu Prakerja Menggunakan Algoritma K-NN, Naïve Bayes dan SVM

Tugas Akhir ini telah diperiksa dan disidangkan sebagai salah satu persyaratan untuk memperoleh gelar Sarjana pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana.

Jakarta, 29 Juli 2022

Menyetujui,


(Eugenius Kau Suni ST.MT)
Dosen Pembimbing

Mengetahui,


UNIVERSITAS
MERCU BUANA

(Wawan Gunawan, S.Kom, MT)
Koord. Tugas Akhir Teknik Informatika

(Ir. Emil R. Kaburuan, Ph.D., IPM.)
Ka. Prodi Teknik Informatika

ABSTRAK

Nama : Tri Putra Adimas Nugraha
NIM : 41518010057
Pembimbing TA : Eugenius Kau Suni ST. MT
Judul : Analisis Sentimen Respons Masyarakat Terhadap Kartu Prakerja Menggunakan Algoritma K-NN, Naïve Bayes dan SVM

Virus Covid-19 yang memasuki Indonesia berdampak pada penurunan akibat pemberlakuan peraturan *social distancing* oleh Pemerintah. Banyak perusahaan yang terpaksa mengurangi jumlah pekerja yang artinya semakin banyak masyarakat yang kehilangan mata pencahariannya. Banyak upaya yang dilakukan pemerintah dalam menekan angka pengangguran yang terjadi di Indonesia, salah satunya yaitu meluncurkan program Kartu Prakerja. Twitter sebagai media sosial yang kerap kali digunakan oleh masyarakat Indonesia untuk menyatakan opini mereka terhadap Program Kartu Prakerja tersebut. Analisis sentimen biasa digunakan untuk mengetahui nilai dari sebuah opini dari subjek penelitian, apakah bernilai negatif atau positif. Penggunaan analisa sentimen dalam mengetahui apakah mayoritas opini dari masyarakat di twitter terhadap Program Kartu Prakerja mengarah ke positif atau negatif. Algoritma KNN adalah sebuah metode untuk melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek yang diuji. Metode *Naïve Bayes* menerapkan pendekatan pada perhitungan statistik dengan melakukan asumsi antar kelas. *Support Vector Machine* (SVM) adalah salah satu metode klasifikasi dengan konsep dasar yang memaksimalkan batas *hyperplane* yang memisahkan suatu set data. Untuk itu algoritma KNN, SVM dan *Naïve Bayes* dirasa tepat untuk diterapkan dalam melakukan proses analisis sentimen terhadap opini masyarakat di twitter tentang Kartu Prakerja.

Kata kunci:

Machine Learning, Kartu Prakerja, Klasifikasi, *K-Nearest Neighbor*, *Support Vector Machine*, *Naïve Bayes*, TF-IDF

ABSTRACT

Name : Tri Putra Adimas Nugraha
Student Number : 41518010057
Counsellor : Eugenius Kau Suni ST. MT
Title : Sentiment Analysis Citizens Response to Pre-Employment Card Using K-NN, Naïve Bayes and SVM Algorithms

The Covid-19 virus that entered Indonesia had an impact on the decline due to the implementation of social distancing regulations by the Government. Many companies impacted to reduce the number of workers, which means more people lose their jobs. The government try a many way to resolved and reduce the unemployment rate in Indonesia, one of solution is launching the Pre-Employment Card program. Twitter as a social media with a much users by Indonesian people to explain their opinion about Pre-Employment Card Program. Sentiment analysis is usually used to determine the value of an opinion from the research subject, whether it is negative or positive. The use of sentiment analysis to find out whether the majority of public opinion on Twitter towards the Pre-Employment Card Program is positive or negative. The KNN algorithm is a method for classifying objects based on learning data that is closest to the object by tested. The Naïve Bayes method applies an approach to statistical calculations by making assumptions between classes. Support Vector Machine (SVM) is a classification method with a basic concept that maximizes the hyperplane boundary that separates a data set. For this reason, the KNN, SVM and Naïve Bayes algorithms are proper to use the process of sentiment analysis on public opinion on Twitter about Pre-Employment Cards.

Key words:

Machine Learning, Pre-employment card, Classification, K-Nearest Neighbor, Support Vector Machine, Naïve Bayes, TF-IDF.

KATA PENGANTAR

Puji syukur kita panjatkan atas nikmat, rahmat serta karunia yang senantiasa diberikan oleh Allah SWT sehingga dapat menyelesaikan Tugas Akhir dengan judul “Analisis Sentimen Respons Masyarakat Terhadap Kartu Prakerja Menggunakan Algoritma K-NN, Naïve Bayes dan SVM” sebagai salah satu syarat untuk menyelesaikan Program Sarjana (S1) pada Program Sarjana Fakultas Ilmu Komputer Jurusan Teknik Informatika

Dalam penyusunan Tugas Akhir ini Penulis menyadari bahwa laporan Tugas Akhir ini tak akan bisa selesai tanpa bantuan dari banyak pihak. Oleh karena itu, penulis mengucapkan terima kasih kepada:

1. Ayah dan Ibu yang senantiasa membesarkan saya dengan penuh kasih, cinta dan kehangatan yang tidak ternilai harganya, serta tak luput pula doa yang selalu dipanjatkan kepada saya agar dapat meraih hasil yang terbaik.
2. Bapak Dr. Harwikarya, M.T., selaku Dosen Pembimbing Akademik atas arahan dan masukan yang telah diberikan kepada saya selama menjalani masa studi.
3. Bapak Eugenius Kau Suni S.T., M.T., selaku Dosen Pembimbing Tugas Akhir saya yang telah rela meluangkan sebagian besar waktunya untuk turut serta memberi saya saran selama proses pengerjaan Tugas Akhir.
4. Seluruh Dosen Program Studi Teknik Informatika yang sudah memberikan ilmu yang bermanfaat selama proses perkuliahan berlangsung.
5. Sahabat dan rekan-rekan yang telah membantu dan memberi semangat kepada saya selama proses pengerjaan Tugas Akhir.
6. Seluruh Staff Administrasi dan Tata Usaha yang telah membantu dan memberikan kemudahan, terimakasih atas semua pelayanan dan arahannya.
7. Kedua Kakak saya yang telah memberikan segala bentuk bantuan selama proses menjalani studi Universitas Mercu Buana selama ini.
8. Semua pihak dan personal yang tidak dapat disebut satu per satu yang terlibat dalam proses pembuatan Tugas Akhir ini sehingga dapat terselesaikan dengan baik

Akhir kata, penulis menyadari bahwa Tugas Akhir ini masih jauh dari kata sempurna. Masih banyak terdapat kekurangan dalam penelitian maupun penulisan Tugas Akhir ini. Untuk itu, kritik dan saran pembaca sangat dihargai dan diharapkan. Semoga Tugas Akhir ini dapat memberikan manfaat bagi para pembaca.

Jakarta, 06 Juli 2022

Penulis



DAFTAR ISI

HALAMAN SAMPUL.....	i
HALAMAN JUDUL	i
LEMBAR PERNYATAAN ORISINALITAS	ii
SURAT PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR... iii	
SURAT PERNYATAAN LUARAN TUGAS AKHIR.....	iv
LEMBAR PERSETUJUAN PENGUJI	v
LEMBAR PENGESAHAN	viii
ABSTRAK	ix
ABSTRACT.....	x
KATA PENGANTAR.....	xi
DAFTAR ISI.....	xiii
NASKAH JURNAL	1
KERTAS KERJA.....	10
BAB 1. LITERATUR REVIEW.....	11
BAB 2. ANALISIS DAN PERANCANGAN.....	37
BAB 3. SOURCE CODE.....	41
BAB 4. DATASET.....	52
BAB 5. TAHAPAN EKSPERIMEN.....	54
BAB 6. HASIL SEMUA EKSPERIMEN.....	61
DAFTAR PUSTAKA	65
LAMPIRAN DOKUMEN HAKI.....	68
LAMPIRAN KORESPONDENSI	70
CURRICULUM VITAE.....	71

NASKAH JURNAL

Terbit online pada laman web jurnal: <http://jurnal.iaii.or.id>



JURNAL RESTI

(Rekayasa Sistem dan Teknologi Informasi)

Vol. 6 No. x (2022) x - x

ISSN Media Elektronik:
2580-0760

Analisis Sentimen Respons Masyarakat Terhadap Kartu Prakerja Menggunakan Algoritma K-NN, Naïve Bayes dan SVM

Sentiment Analysis Citizens Response to Pre-Employment Card Using K-NN, Naïve Bayes and SVM Algorithms

Tri Putra Adimas Nugraha¹, Eugenius Kau Suni²

Teknik Informatika, Ilmu Komputer, Universitas Mercu Buana

141518010057@student.mercubuana.ac.id, 2eugenius@mercubuana.ac.id

Abstract

The Covid-19 virus that entered Indonesia had an impact on the decline due to the implementation of social distancing regulations by the Government. Many companies are forced to reduce the number of workers, which means more people lose their livelihoods. Sentiment analysis is usually used to determine the value of an opinion from the research subject, whether it is negative or positive. The use of sentiment analysis to find out whether the majority of public opinion on Twitter towards the Pre-Employment Card Program is positive or negative. The KNN algorithm is a method for classifying objects based on learning data that is closest to the object being tested. The Naïve Bayes method applies an approach to statistical calculations by making assumptions between classes. Support Vector Machine (SVM) is a classification method with a basic concept that maximizes the hyperplane boundary that separates a data set. For this reason, the KNN, SVM and Naive Bayes algorithms are deemed appropriate to be applied in the process of sentiment analysis on public opinion on Twitter about Pre-Employment Cards.

Keywords: Machine Learning, Kartu Prakerja, Klasifikasi, K-Nearest Neighbor, Support Vector Machine, Naïve Bayes, TF-IDF.

Abstrak

Virus Covid-19 yang memasuki Indonesia berdampak pada penurunan akibat pemberlakuan peraturan *social distancing* oleh Pemerintah. Banyak perusahaan yang terpaksa mengurangi jumlah pekerja yang artinya semakin banyak masyarakat yang kehilangan mata pencahariannya. Twitter sebagai media sosial yang kerap kali digunakan oleh masyarakat Indonesia untuk menyatakan opini mereka terhadap Program Kartu Prakerja tersebut. Analisis sentimen biasa digunakan untuk mengetahui nilai dari sebuah opini dari subjek penelitian, apakah bernilai negatif atau positif. Penggunaan analisa sentimen dalam mengetahui apakah mayoritas opini dari masyarakat di twitter terhadap Program Kartu Prakerja mengarah ke positif atau negatif. Algoritma KNN adalah sebuah metode untuk melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek yang diuji. Metode *Naïve Bayes* menerapkan pendekatan pada perhitungan statistik dengan melakukan asumsi antar kelas. *Support Vector Machine* (SVM) adalah salah satu metode klasifikasi dengan konsep dasar yang memaksimalkan batas *hyperplane* yang memisahkan suatu set data. Untuk itu algoritma KNN, SVM dan *Naïve Bayes* dirasa tepat untuk diterapkan dalam melakukan proses analisis sentimen terhadap opini masyarakat di twitter tentang Kartu Prakerja.

Kata kunci: Machine Learning, Kartu Prakerja, Klasifikasi, K-Nearest Neighbor, Support Vector Machine, Naïve Bayes, TF-IDF.

Universitas Mercu Buana

1. Pendahuluan

Virus Covid-19 yang memasuki Indonesia di awal tahun 2020 mengakibatkan nilai perekonomian Indonesia menurun serta melemahnya nilai tukar rupiah dan perlambatan pertumbuhan ekonomi. Pertumbuhan ekonomi Indonesia mengalami penurunan akibat pemberlakuan peraturan *social distancing* oleh Pemerintah yang berdampak pula pada sektor ketenagakerjaan di Indonesia [1]. Banyak perusahaan yang terpaksa mengurangi jumlah pekerja yang artinya semakin banyak masyarakat yang kehilangan mata pencahariannya. Menurut Data Badan Pusat Statistik, angka pengangguran tertinggi terjadi pada bulan Agustus 2020 yang menyentuh angka 9,77 juta atau sekitar 7,07% masyarakat Indonesia yang tidak memiliki pekerjaan. Banyak upaya yang dilakukan pemerintah dalam menekan angka pengangguran yang terjadi di Indonesia, salah satunya yaitu meluncurkan program Kartu Prakerja.

Pada akhir Maret 2020 Pemerintah Negara Republik Indonesia resmi meluncurkan program bantuan melalui Kartu Prakerja. Berlandaskan pada Peraturan Presiden Nomor 36 tahun 2020 tentang Pengembangan Kompetensi Kerja Melalui Program Kartu Prakerja, memiliki tujuan untuk perluasan kesempatan kerja, peningkatan produktivitas dan daya saing dengan pengembangan kompetensi kerja. Program ini menargetkan para pencari kerja, pekerja atau buruh yang terkena pemutusan hubungan kerja hingga pekerja atau buruh yang membutuhkan peningkatan kompetensi [2].

Asosiasi Penyelenggara Jasa Internet Indonesia (APJII) mencatat pada periode 2019-kuartal II/2020, total pengguna jasa internet menyentuh angka 196,7 juta jiwa di Indonesia [3]. Jumlah pengguna yang tergolong banyak, menjadikan Twitter sebagai media sosial yang kerap kali digunakan oleh masyarakat Indonesia untuk menyatakan opini mereka terhadap suatu hal, bahkan tak jarang terdapat pro dan kontra pada *tweet* yang dilontarkan oleh masyarakat tersebut [4]. Hal ini pula yang dimanfaatkan masyarakat Indonesia untuk melontarkan opini mereka terkait Kartu Prakerja yang

merupakan program pemerintah Indonesia.

Analisis sentimen atau *opinion mining* merupakan proses memahami, mengesktrak dan mengolah data tekstual secara otomatis untuk mendapatkan informasi sentimen yang terkandung dalam suatu kalimat opini [5]. Analisis sentimen biasa digunakan untuk mengetahui nilai dari sebuah opini dari subjek penelitian, apakah bernilai negatif atau positif [6]. Penggunaan analisa sentimen dalam mengetahui apakah mayoritas opini dari masyarakat di twitter terhadap Program Kartu Prakerja mengarah ke positif atau negatif.

Menurut Pramudiono, *data mining* merupakan analisis otomatis dari data yang berjumlah besar atau kompleks dengan tujuan untuk menemukan pola atau kecenderungan yang penting yang biasanya tidak disadari keberadaannya. Sehingga *data mining* dapat diartikan sebagai proses pengambilan sebuah data dari suatu sumber dengan tujuan mendapatkan sebuah informasi untuk kemudian dianalisis dengan menggunakan metode algoritma *data mining* [7]. Mengacu pada fungsinya, *data mining* sendiri terbagi menjadi *classification*, *clustering*, *association*, *regression*, dan *forecasting*.

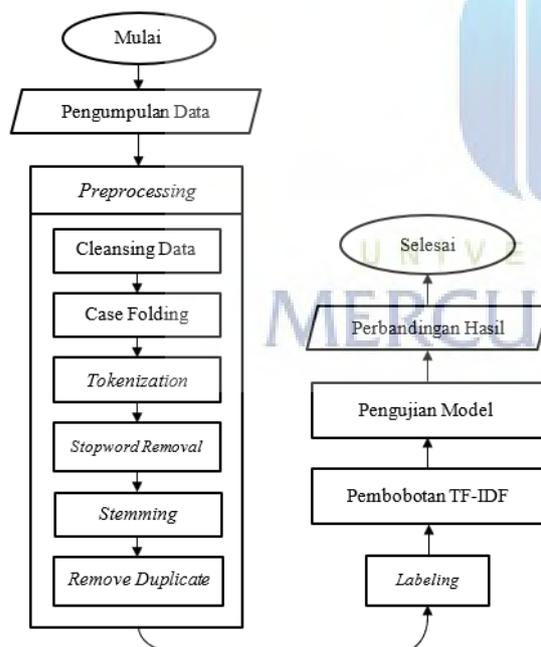
Algoritma K-Nearest Neighbor (KNN) merupakan suatu metode *data mining classification* kepada sekumpulan data yang telah diklasifikasikan sebelumnya [8]. Lebih tepatnya algoritma KNN adalah sebuah metode untuk melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek yang diuji. Metode *Naïve bayes* merupakan algoritma data mining *classification* dengan menerapkan konsep prediksi probabilitas pada sebuah kasus. Metode ini menerapkan pendekatan pada perhitungan statistik dengan melakukan asumsi antar kelas [9]. *Support Vector Machine* (SVM) adalah salah satu metode klasifikasi dengan konsep dasar yang memaksimalkan batas *hyperplane* yang memisahkan suatu set data [10].

Ketiga algoritma yang telah disebutkan diatas memiliki kesamaan, yaitu *classification*. Berlandaskan hal tersebut, algoritma KNN, SVM dan *Naïve Bayes* dirasa tepat untuk diterapkan dalam

melakukan proses analisis sentimen terhadap opini masyarakat di twitter tentang Kartu Prakerja. Selain itu dilakukannya penelitian ini diharapkan mampu untuk mengetahui kecenderungan opini masyarakat terhadap program yang dijalankan oleh pemerintah khususnya program Kartu Prakerja agar dapat dijadikan bahan evaluasi kedepannya oleh pemerintah.

2. Metode

Penelitian ini tergolong sebagai penelitian Kuantitatif, yaitu penelitian yang berlandaskan data numerik serta penyelesaian dengan teknik analisis atau statistika [24]. Penelitian ini diawali dengan proses pengumpulan data yang dilanjutkan dengan tahapan *pre-processing* data. Setelah itu, data yang sudah siap olah melalui tahapan analisis sentimen secara otomatis dan kemudian dicari nilai presisi dari ketiga metode algoritma yang digunakan.



GRAFIK PERBANDINGAN ALGORITMA



Gambar 1. Tahapan Penelitian

2.1. Metode Pengumpulan Data

Pengumpulan data pada penelitian ini mengambil dari data *tweet* berdasarkan kata kunci tertentu yang opininya ingin kita ambil atau biasa disebut dengan istilah *scrapping* [11]. Pada penelitian ini menggunakan kata kunci “kartu prakerja” dan penelitian ini menggunakan data sekunder yang bersumber dari data Twitter. Data yang dikumpulkan melalui proses autentifikasi antara twitter *Application Programming*



Interface (API). Data *tweet* yang akan melalui proses *scrapping* ini merupakan opini-opini yang diutarakan masyarakat Indonesia mengenai kartu prakerja, untuk kemudian opini tersebut dikelompokkan apakah positif, negatif dan netral. Setelah mendapatkan data berbentuk CSV maka dataset tersebut siap diolah pada tahapan penelitian.

Gambar 2. Metode Pengumpulan Data

2.2. Metode Analisis Data

2.2.1 *Pre-processing*

Tahapan ini bertujuan untuk membersihkan data mentah yang baru saja didapatkan melalui proses *crawling* dari aplikasi Rapid Miner agar menjadi data yang siap olah [12]. Berikut adalah tahapan *pre-processing* pada penelitian ini:

- Menghilangkan *noise* yang ada di dataset, seperti HTML, enter, *mention*, RT, *username*, *hashtag*, angka dan URL [13].
- Merubah bentuk kata menjadi bentuk dasarnya agar sebuah karakter dapat seragam.
- Memisahkan kalimat menjadi kata-kata.
- Menghapus kata-kata yang tidak mendiskripsikan sesuatu atau yang tidak memiliki makna [14].

- e. Mengubah kata jamak menjadi kata dasar.
- f. Menghapus dataset yang memiliki duplikasi.

2.2.2 Pelabelan Data

Data yang sudah melewati tahapan *pre-processing* akan dilakukan pelabelan kelas data. Pelabelan kelas data dibagi menjadi dua, kelas positif dan kelas negatif. Proses pelabelan dilakukan dengan menggunakan bahasa R dengan menggunakan *tools* Google Colab. Proses pelabelan dengan menggunakan Bahasa R diperlukan kamus Bahasa Indonesia yang terintegrasi dengan Google Drive. Kamus berbahasa Indonesia dengan kata positif dan kamus berbahasa Indonesia dengan kata Negatif.

Tabel 1. Hasil Pelabelan Data

Positif	Negatif
692	510

Pada tahapan labeling, dilakukan skoring pada tiap kalimat. Setiap kata yang terdeteksi diberikan skoring untuk menilai kelas sentimen. Untuk kata positif, diberikan nilai 1 sedangkan untuk kata negatif diberikan nilai -1. Jika terdapat kata yang tidak ada dalam kamus positif ataupun negatif, diberikan nilai 0. Pemberian skoring dilakukan dengan cara menghitung jumlah nilai pada tiap kata dalam satu kalimat. Jika nilai ≥ 0 maka dilabeli sebagai sentimen tweet positif, sebaliknya jika nilai < 0 maka dilabeli sebagai tweet sentimen negatif. Jika hasil skoring bernilai 0 maka dilabeli sebagai tweet positif. Kelas data positif akan diberi nilai 1, sedangkan kelas data Negatif akan diberi nilai 0. Berikut adalah contoh hasil labeling otomatis tersebut.

Tabel 2. Contoh Tweet Hasil Labeling

Score	Tweet	Sentimen
1	alam abis daftar prakerja mesti lulus kursus	Positif
-2	kadrun ga syukur sih	Negatif
-1	arditho pramono nganggur ardhito prakerja	Negatif
2	arti status prakerja evaluasi tanda lolos	Positif

-1 awal maret mu gagal dapet prakerja Negatif

2.2.3 Proses TF-IDF

Pembobotan kata dengan TF-IDF adalah proses pemberian bobot untuk setiap kata yang terdapat dalam dokumen [15]. Tahapan ini dilakukan untuk mengetahui seberapa penting sebuah kata mewakili sebuah kalimat, dilakukan pembobotan atau perhitungan. Pemberian skor dalam TF-IDF berdasarkan frekuensi munculnya kata dalam dokumen [16]. TF-IDF dilakukan dengan maksud untuk melakukan proses memberikan bobot pada setiap kata untuk menghitung bobot pada masing-masing kata yang akan digunakan sebagai fitur dalam menilai sentimen kata [17].

2.2.4 Klasifikasi Algoritma

Penelitian ini menggunakan tiga algoritma klasifikasi sebagai pembandingan, yaitu *Support Vector Machine* (SVM), *Naïve Bayes* dan *K-Nearest Neighbor* (KNN). Proses melakukan penilaian objek data untuk memasukkannya ke dalam kelas yang tersedia biasa disebut klasifikasi [18]. SVM merupakan teknik yang relatif baru sebagai metode algoritma prediksi, baik dalam klasifikasi maupun regresi [19]. Konsep klasifikasi dengan SVM adalah mencari *hyperlane* terbaik yang berfungsi sebagai pemisah dua kelas data. Algoritma SVM bekerja dengan memisahkan data dengan menggunakan metode pengukuran jarak dari *hyperlane* dari titik terdekat daripada mencari titik maksimum untuk memaksimalkan jarak antar label kelas berdasarkan pengukuran batas keanggotaan kelas [20]. Metode ini terbilang handal dalam menyelesaikan permasalahan klasifikasi data, permasalahan tersebut diselesaikan dengan persamaan *Lagrangian equation* yang merupakan bentuk ganda dari SVM melalui pemrograman kuadrat [21]. SVM juga dapat bekerja pada dataset berdimensi tinggi menggunakan trik kernel [16]. SVM sendiri memiliki prinsip dasar pengklasifikasian linear, yaitu kasus klasifikasi yang dapat dipisahkan secara linear. *Support Vector Machine* juga dapat bekerja pada dataset berdimensi tinggi menggunakan trik kernel.

Naïve Bayes Classifier merupakan teknik pembelajaran algoritma data mining yang memanfaatkan metode probabilitas dan statistik. *Naïve Bayes Classifier* dalam mengklasifikasi ada dua proses penting yaitu pembelajaran (*training*) dan pengujian [22]. Algoritma klasifikasi *Naïve Bayes* merupakan algoritma yang digunakan untuk mencari nilai probabilitas tertinggi untuk mengklasifikasikan data uji pada kategori yang paling tepat, *classifier Bayesian* memiliki akurasi dan kecepatan yang lebih tinggi terutama bila diterapkan pada dataset yang banyak [23]. *Naïve Bayes* merupakan salah satu algoritma data mining yang mudah digunakan dan memiliki waktu pemrosesan yang cepat, mudah diimplementasikan dengan struktur yang cukup sederhana dan memiliki tingkat efektivitas yang tinggi. Algoritma KNN merupakan suatu metode *data mining classification* kepada sekumpulan data yang telah diklasifikasikan sebelumnya. Lebih tepatnya algoritma KNN adalah sebuah metode untuk melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek yang diuji. Metode *Naïve bayes* merupakan algoritma data mining *classification* dengan menerapkan konsep prediksi probabilitas pada sebuah kasus. Metode ini menerapkan pendekatan pada perhitungan statistik dengan melakukan asumsi antar kelas.

3. Hasil dan Pembahasan

Pengujian yang dilakukan pada penelitian ini menggunakan tiga algoritma yaitu SVM, *Naïve Bayes* dan KNN. Pada penelitian ini dilakukan skenario pengujian pada komposisi data yang sudah melalui tahap data *cleansing* dan *Pre-processing*, percobaan dilakukan pada ukuran *test_size*. Evaluasi hasil pada penelitian ini menggunakan *confusion matrix* yang akan diambil hasil dari nilai akurasi pada masing-masing skenario pengujian algoritma.

3.1. Support Vector Machine

Skenario pengujian pada algoritma SVM ditunjukkan pada table 4, dimana pengujian pada komposisi data

yaitu *test_size* 0.1, 0.2 dan 0.3. Hasil Pengujian pada *test_size* = 0.1 memiliki nilai akurasi sebesar 93%, pada *test_size* 0.2 mendapat nilai akurasi sebesar 94%, sedangkan pada *test_size* = 0.3 akurasi yang didapat sebesar 91%. Dari hasil tersebut dapat diketahui bahwa nilai akurasi tertinggi didapatkan pada skenario pengujian dengan *test_size* sebesar 0.2. Berdasarkan hasil dari penelitian ini dapat dibuktikan bahwa penerapan ukuran *test_size* = 0.2 pada algoritma SVM bisa menghasilkan akurasi yang baik. Oleh karenanya, algoritma SVM dapat digunakan dan diterapkan dengan baik dalam proses analisis sentiment.

Tabel 3. Hasil Pengujian Algoritma SVM

Test_size = 0.1	Test_size = 0.2	Test_size = 0.3
93%	94%	91%

3.2. Naïve Bayes

Skenario pengujian pada algoritma *Naïve Bayes* ditunjukkan pada tabel 3, dimana pengujian pada komposisi data yaitu *test_size* 0.1, 0.2 dan 0.3. Hasil Pengujian pada *test_size* = 0.1 memiliki nilai akurasi sebesar 85%, pada *test_size* 0.2 mendapat nilai akurasi sebesar 87%, sedangkan pada *test_size* = 0.3 akurasi yang didapat sebesar 88%. Dari hasil tersebut dapat diketahui bahwa nilai akurasi tertinggi didapatkan pada skenario pengujian dengan *test_size* sebesar 0.3. Berdasarkan hasil dari penelitian ini dapat dibuktikan bahwa penerapan ukuran *test_size* = 0.3 pada algoritma *Naïve Bayes* bisa menghasilkan akurasi yang cukup baik. Meskipun dalam penelitian ini SVM memiliki hasil jauh lebih baik dibandingkan dengan *Naïve Bayes*. Tetapi, algoritma *Naïve Bayes* juga dapat menjadi alternatif penerapan algoritma dalam proses analisis sentiment selain algoritma SVM karena memiliki nilai akurasi yang sudah masuk dalam kategori cukup baik.

Tabel 4. Hasil Pengujian Algoritma Naïve Bayes

Test_size = 0.1	Test_size = 0.2	Test_size = 0.3
-----------------	-----------------	-----------------

85% 87% 88%

3.3. K-NN

Skenario pengujian pada algoritma KNN ditunjukkan pada table 4, dimana pengujian pada komposisi data yaitu test_size 0.1, 0.2 dan 0.3. Hasil Pengujian pada test_size = 0.1 memiliki nilai akurasi sebesar 96%, pada test_size 0.2 mendapat nilai akurasi sebesar 93%, sedangkan pada test_size = 0.3 akurasi yang didapat sebesar 94%. Dari hasil tersebut dapat diketahui bahwa nilai akurasi tertinggi didapatkan pada skenario pengujian dengan test_size sebesar 0.1. Berdasarkan hasil dari penelitian ini dapat dibuktikan bahwa penerapan ukuran test_size = 0.1 pada algoritma KNN bisa menghasilkan akurasi yang sangat baik. Oleh karenanya, algoritma SVM dapat digunakan dan diterapkan dengan sangat baik dalam proses analisis sentiment dibandingkan algoritma SVM dan Naïve Bayes.

Tabel 5. Hasil Pengujian Algoritma KNN

Test_size = 0.1	Test_size = 0.2	Test_size = 0.3
96%	93%	96%

3.4 Perbandingan Algoritma

Gambar 3. Perbandingan Algoritma

Berdasarkan hasil perbandingan pengujian yang tertera pada tabel 5, dapat diketahui bahwa nilai akurasi tertinggi dimiliki oleh model KNN yaitu sebesar 96%, selanjutnya posisi kedua untuk nilai akurasi yang masih masuk dalam kategori baik dimiliki oleh model SVM dengan nilai sebesar 94%. Sedangkan nilai akurasi paling rendah didapat pada algoritma *Naïve Bayes* yaitu sebesar 88%. Pada Algoritma KNN dan SVM dapat dilihat bahwa selisih nilai akurasi dari kedua model tersebut tidak terlalu signifikan. Dimana hanya memiliki selisih sebesar 2%. Sedangkan selisih untuk algoritma *Naïve Bayes* dengan algoritma KNN sebesar 8% dan dengan algoritma SVM sebesar 6%. Berdasarkan

hal tersebut dari hasil penelitian ini untuk penerapan pada proses analisis sentiment, algoritma yang paling unggul dan memiliki nilai akurasi tertinggi dimiliki oleh Model KNN. Berikut merupakan grafik perbandingan algoritma yang dapat dilihat pada gambar tiga.

Tabel 6. Perbandingan Hasil Pengujian

SVM	Naïve Bayes	KNN
94%	88%	96%

4. Kesimpulan & Saran

4.1 Kesimpulan

Berdasarkan hasil dari beberapa pengujian yang telah dilakukan dalam mencari model yang paling tepat untuk penerapan proses analisis sentiment. Dimana dilakukan skenario pengujian pada ukuran test_size masing-masing algoritma. Komposisi data yang digunakan yaitu 0.1, 0.2, dan 0.3. Algoritma yang digunakan dalam penelitian ini yaitu *Support Vector Machine*, *Naïve Bayes* dan *K-Nearest Neighbor*. Dari hasil pengujian, menunjukkan bahwa KNN berada pada posisi paling unggul dengan nilai akurasi yang didapat yaitu 96%. Dilanjut dengan SVM dengan hasil akurasi sebesar 94% dan yang mendapat nilai akurasi terendah yaitu model *Naïve Bayes* dengan hasil sebesar 88%. Berdasarkan hasil tersebut dapat disimpulkan bahwa ketiga algoritma tersebut cocok untuk penerapan dalam proses analisis sentiment. Meskipun KNN memiliki nilai terbaik tetapi hasil dari kedua algoritma lainnya pun masuk dalam kategori baik. Dengan adanya penelitian ini diharapkan dapat membantu pemerintah dan masyarakat dalam mengetahui informasi mengenai respons terhadap kartu prakerja, serta dapat membantu pemerintah untuk memperbaiki kinerja dalam proses penyelenggaraan proses prakerja. Pada penelitian selanjutnya, dapat dilakukan penerapan *cross validation* pada proses validasi dalam menghindari dan mencegah terjadinya kasus *overfitting*, dan untuk skenario pengujian terhadap parameter yang diuji lebih diperbanyak

lagi agar pengoptimalan kinerja algoritma bisa lebih baik.

4.2 Saran

Melihat hasil dari perhitungan jumlah sentimen positif dan negatif didapatkan hasil positif yang lebih besar yaitu dengan angka persentase 52% jika dibandingkan dengan sentimen negatif. Hal ini menunjukkan bahwa respons masyarakat terhadap program pemerintah Kartu Prakerja dapat dinilai sudah baik, meski begitu angka 48% pada opini negatif menunjukkan bahwa tak sedikit pula masyarakat yang mengemukakan kekecewaannya. Hal ini diharapkan agar menjadi perhatian bagi pemerintah dalam mengevaluasi program Kartu Prakerja ini untuk dapat lebih baik lagi kedepannya.

Daftar Rujukan

- [1] V. R. Krisnandika, D. Aulia, L. Jannah, S. T. Media, and N. Citra, "Dampak Pandemi Covid-19 Terhadap Pengangguran Di Indonesia," *Jurnal Ilmu Sosial dan Pendidikan*, vol. 5, no. 3, pp. 2598–9944, 2021, doi: 10.36312/jisip.v5i3.2227/http.
- [2] PRESIDEN REPUBLIK INDONESIA, "PERATURAN PRESIDEN NOMOR 36 TAHUN 2020 TENTANG PENGEMBANGAN KOMPETENSI KERJA MELALUI PROGRAM KARTU PRAKERJA," 2020
- [3] D. Yordan Sihombing and Y. Nataliani, "SISTEMASI: Jurnal Sistem Informasi Analisis Interaksi Pengguna Twitter pada Strategi Pengadaan Barang Menggunakan Social Network Analysis." [Online]. Available: <http://sistemasi.ftik.unisi.ac.id>
- [4] Winda Putri Anggraini, Manda Syari Utami, Juliafatin Malinda Berlianty, Elvira Sellya Hutagalung, Yogi Juniarto, and Rani Nooraeni, "Klasifikasi Sentimen Masyarakat Terhadap Kebijakan Kartu Prakerja di Indonesia," *ISSN*, vol. Vol. 13, No. 4, Dec. 2020.
- [5] W. E. Nurjanah, R. Setya Perdana, and M. A. Fauzi, "Analisis Sentimen Terhadap Tayangan Televisi Berdasarkan Opini Masyarakat pada Media Sosial Twitter menggunakan Metode K-Nearest Neighbor dan Pembobotan Jumlah Retweet," 2017. [Online]. Available: <http://j-ptiik.ub.ac.id>
- [6] M. Nur Habibi and Sunjana, "Analysis of Indonesia Politics Polarization before 2019 President Election Using Sentiment Analysis and Social Network Analysis," *International Journal of Modern Education and Computer Science*, vol. 11, no. 11, pp. 22–30, Nov. 2019, doi: 10.5815/ijmecs.2019.11.04.
- [7] M. Rangga, A. Nasution, and M. Hayaty, "Perbandingan Akurasi dan Waktu Proses Algoritma K-NN dan SVM dalam Analisis Sentimen Twitter," *JURNAL INFORMATIKA*, vol. 6, no. 2, pp. 212–218, 2019, [Online]. Available: <http://ejournal.bsi.ac.id/ejurnal/index.php/ji>
- [8] D. Cahyanti, A. Rahmayani, and S. Ainy Husniar, "Indonesian Journal of Data and Science Analisis performa metode Knn pada Dataset pasien pengidap Kanker Payudara," vol. 1, no. 2, pp. 39–43, 2020.
- [9] R. Fajar, S. Program, P. Rekayasa, N. Lunak, and R. Bengkalis, "Implementasi Algoritma Naive Bayes Terhadap Analisis Sentimen Opini Film Pada Twitter," vol. 3, no. 1.
- [10] S. Hota and S. Pathak, "KNN classifier based approach for multi-class sentiment analysis of twitter data," *International Journal of Engineering and Technology(UAE)*, vol. 7, no. 3, pp. 1372–1375, 2018, doi: 10.14419/ijet.v7i3.12656.
- [11] V. Kevin, S. Que, : Analisis, S. Transportasi, A. Iriani, and H. D. Purnomo, "Analisis Sentimen Transportasi Online Menggunakan Support Vector Machine Berbasis Particle Swarm Optimization (Online Transportation Sentiment Analysis Using Support Vector Machine Based on Particle Swarm Optimization)," 2020. [Online]. Available: www.tripadvisor.com,
- [12] Vamshi Krishna. B, Dr. Ajeet Kumar Pandey, and Dr. Siva Kumar A. P, "Topic

- Model Based Opinion Mining and Sentiment Analysis,” 2018.
- [13] Anu J Nair, Aadithya Vinayak, and Veena G, “Comparative study of Twitter Sentiment On COVID - 19 Tweets ,” 2021.
- [14] Kusriani and Mochamad Mashuri, “Sentiment Analysis In Twitter Using Lexicon Based and Polarity Multiplication,” 2019.
- [15] I. Made and A. Agastya, “PENGARUH STEMMER BAHASA INDONESIA TERHADAP PEFORMA ANALISIS SENTIMEN TERJEMAHAN ULASAN FILM,” 2018. [Online]. Available: <https://github.com/arthaAgastya/dataset/tree/master/Mov>
- [16] S. Fahmi, L. Purnamawati, G. F. Shidik, M. Muljono, and A. Z. Fanani, “Sentiment analysis of student review in learning management system based on sastrawi stemmer and SVM-PSO,” in *Proceedings - 2020 International Seminar on Application for Technology of Information and Communication: IT Challenges for Sustainability, Scalability, and Security in the Age of Digital Disruption, iSemantic 2020*, Sep. 2020, pp. 643–648. doi: 10.1109/iSemantic50169.2020.9234291.
- [17] Reddy, “Analisis Sentimen Pengguna Twitter Terhadap Polemik Persepakbolaan Indonesia Menggunakan Pembobotan TF-IDF dan K-Nearest Neighbor.” [Online]. Available: <https://t.co/9Wl0aWpfd5>
- [18] F. A. Wenando, R. Hayami, Bakaruddin, and A. Y. Novermahakim, “Tweet Sentiment Analysis for 2019 Indonesia Presidential Election Results using Various Classification Algorithms,” in *Proceeding - 1st International Conference on Information Technology, Advanced Mechanical and Electrical Engineering, ICITAMEE 2020*, Oct. 2020, pp. 279–282. doi: 10.1109/ICITAMEE50454.2020.9398513.
- [19] Kai-xu Han, Chien-Ching Chiu, and Wei Chien, “The Application of Support Vector Machine (SVM) on the Sentiment Analysis of Internet Posts,” 2019.
- [20] S. E. Saad and J. Yang, “Twitter Sentiment Analysis Based on Ordinal Regression,” *IEEE Access*, vol. 7, pp. 163677–163685, 2019, doi: 10.1109/ACCESS.2019.2952127.
- [21] R. Ren, D. D. Wu, and D. D. Wu, “Forecasting stock market movement direction using sentiment analysis and support vector machine,” *IEEE Systems Journal*, vol. 13, no. 1, pp. 760–770, Mar. 2019, doi: 10.1109/JSYST.2018.2794462.
- [22] W. Khofifah, D. N. Rahayu, and A. M. Yusuf, “Analisis Sentimen Menggunakan Naive Bayes Untuk Melihat Review Masyarakat Terhadap Tempat Wisata Pantai Di Kabupaten Karawang Pada Ulasan Google Maps,” *Jurnal Interkom: Jurnal Publikasi Ilmiah Bidang Teknologi Informasi dan Komunikasi*, vol. 16, no. 4, pp. 28–38, Jan. 2022, doi: 10.35969/interkom.v16i4.192.
- [23] S. Thomas, T. Informasi, and I. Shanti Bhuana, “Studi Analisis Metode Analisis Sentimen pada YouTube,” *JIFOTECH (JOURNAL OF INFORMATION TECHNOLOGY)*, vol. 1, no. 1, 2021.
- [24] M. Yusuf, M. Valensyah and W. Gunawan, “PENERAPAN ALGORITMA K-NEAREST NEIGHBOR (KNN) DALAM MEMREDIKSI DAN MENGHITUNG TINGKAT AKURASI DATA CUACA DI INDONESIA,” vol. 2, no. 2, 2021.



KERTAS KERJA

Ringkasan

Kertas kerja ini berisi tentang kelengkapan material dari artikel jurnal dengan judul “Analisis Sentimen Respons Masyarakat Terhadap Kartu Prakerja Menggunakan Algoritma K-NN, Naïve Bayes dan SVM”. Seluruh hasil penelitian Tugas Akhir yang tidak dimasukkan kedalam artikel jurnal. Pada kertas kerja ini disajikan: *literature review*, analisis perancangan, *source code*, dataset yang digunakan, tahapan eksperimen dan seluruh hasil eksperimen.

- Bagian 1: *Literature review* menjabarkan mengenai beberapa jurnal yang terkait dalam penelitian.
- Bagian 2: Analisis dan Perancangan menjabarkan analisis masalah dan analisis model.
- Bagian 3: *Source code* menjabarkan kumpulan kode di setiap prosesnya mulai dari Pengambilan Data, *preprocessing*, Implementasi model, sampai pengujian dan evaluasi.
- Bagian 4: Tahapan eksperimen berisi pengumpulan data, *preprocessing*, Implementasi model dan Evaluasi model.
- Bagian 5: Hasil implementasi eksperimen secara keseluruhan yang mencakup jawaban dari scenario percobaan pengujian yang sudah disebutkan pada bagian 4. Bagian ini berguna untuk menemukan model prediksi terbaik yang dilihat dari perbandingan algoritma.