

Penelitian

A

Jurnal Penelitian SENAFTI 1

Surat Keputusan Dekan

 UNIVERSITAS BUDI LUHUR FAKULTAS TEKNOLOGI INFORMASI KAMPUS PUSAT : Jl. Raya Gading, Pelakangan Utara, Jakarta Selatan 12280 Telp : (021) 5853753 (Pusat), (021) 7371994, Fax : (021) 5853753 Website : http://www.budiluhur.ac.id		 UNIVERSITAS BUDI LUHUR FAKULTAS TEKNOLOGI INFORMASI KAMPUS PUSAT : Jl. Raya Gading, Pelakangan Utara, Jakarta Selatan 12280 Telp : (021) 5853753 (Pusat), (021) 7371994, Fax : (021) 5853753 Website : http://www.budiluhur.ac.id	
KEPUTUSAN DEKAN FAKULTAS TEKNOLOGI INFORMASI UNIVERSITAS BUDI LUHUR NOMOR : K/UBL/FTI/000/004/00/22 TENTANG: PENUGASAN KEGIATAN TRI DHARMA & PENUNJANG BAGI DOSEN FAKULTAS TEKNOLOGI INFORMASI UNIVERSITAS BUDI LUHUR SEMESTER GASAL TAHUN AKADEMIK 2022/2023 DEKAN FAKULTAS TEKNOLOGI INFORMASI UNIVERSITAS BUDI LUHUR		d. Pengabdian Kepada Masyarakat (PPM) , dalam kegiatan terprogram, terjadwal atau insidental;	
Menimbang :		KEDUA :	Dosen-dosen yang melaksanakan penugasan wajib membuat Laporan Kegiatan, dengan mengikuti pedoman dari Fakultas/Program Studi, sebagai pertanggungjawaban atas kegiatan yang diikuti;
		KETIGA :	Kegiatan Tri Dharma yang tidak termasuk dalam surat keputusan ini akan memiliki penugasan tersendiri;
		KEEMPAT :	Keputusan ini berlaku sejak tanggal ditetapkan dan akan diubah sebagaimana mestinya apabila di kemudian hari terdapat kekeliruan.
Mengingat :			
1) Bahwa Dosen adalah pendidik profesional dan ilmu dengan tugas utama mentransformasikan, mengembangkan, dan menyebarkan ilmu pengetahuan, teknologi, dan seni melalui pendidikan/pengajaran penelitian & karya ilmiah, dan Pengabdian pada masyarakat yang dikenal dengan istilah Tri Dharma Perguruan Tinggi;			
2) Bahwa untuk meningkatkan profesionalitas dan kompetensi sebagai pendidik profesional maka dipandang perlu untuk memberikan tugas-tugas tambahan/penunjang dalam lingkup kegiatan penunjang Tri Dharma;			
1) Undang – undang Republik Indonesia Nomor 12 Tahun 2012 tentang Pendidikan Tinggi;			
2) Peraturan Pemerintah Republik Indonesia Nomor 17 Tahun 2010 tentang Pengelolaan dan Penyelenggaraan Pendidikan;			
3) Peraturan Menteri Riset, Teknologi, dan Pendidikan Tinggi Republik Indonesia Nomor 44 Tahun 2015 tentang Standar Nasional Pendidikan Tinggi;			
4) Peraturan Menteri Riset, Teknologi dan Pendidikan Tinggi Republik Indonesia Nomor 15 Tahun 2017 tentang Penamaan Program Studi Pada Perguruan Tinggi;			
5) Akta Yayasan Pendidikan Budi Luhur Tanggal 23 Desember 1991;			
6) Peraturan Pengurus Yayasan Pendidikan Budi Luhur Cakdi Nomor: K/YBL/C/KEP/000/000/00/17 tanggal 24 Agustus 2017 tentang Statuta Universitas Budi Luhur;			
MEMUTUSKAN			
Menetapkan :			
PERTAMA :			
Menugaskan dosen-dosen Fakultas Teknologi Informasi Universitas Budi Luhur untuk melaksanakan kegiatan Tri Dharma Perguruan Tinggi dan penunjangnya pada Semester Gasal Tahun Akademik 2022/2023 yang meliputi:			
a. Kegiatan partisipasi aktif dalam Pertemuan Ilmiah sebagai Ketua/Anggota/Peserta/Pembicara/Penulis/Narasumber pada kegiatan Seminar, Workshop, Konferensi, Pelatihan, Simposium, Lokakarya, Forum Diskusi, Sosialisasi dan sejenisnya;			
b. Publikasi Ilmiah pada Prosiding, Jurnal/majalah/jurnal luar dan sejenisnya;			
c. Partisipasi dalam organisasi profesi, organisasi keilmuan dan/atau organisasi lain yang menunjang kegiatan Tri Dharma Pendidikan Tinggi;			
		Ditetapkan di : Jakarta Pada Tanggal : 1 September 2022 Dekan Fakultas Teknologi Informasi  Dr. Ir. Deni Mahdiana, S.Kom, M.M., M.Kom	

SEMINAR NASIONAL MAHASISWA FAKULTAS TEKNOLOGI INFORMASI

Call For Paper

**PERANAN ARTIFICIAL INTELLIGENCE
YANG CERDAS BERBUDI LUHUR
DALAM MENGHADAPI ERA SOCIETY 5.0**

TOPIK

Cyber Security :

- Cryptography
- Steganography
- Wireless Sensor Network
- Network Management and Maintenance

Artificial Intelligence :

- Expert Systems/Artificial Intelligence
- Semantics Web and Ontologies
- Data Mining
- Text Mining
- Natural Language Processing

SELASA

**06 SEPTEMBER
2022**



UNIVERSITAS BUDI LUHUR
Jl. Ciledug Raya, Petukangan Utara
Jakarta Selatan 12260 - Indonesia

VIA



ZOOM

Terbitan Terkini

Vol 1 No 1 (2022): Prosiding SENAFTI (Buku 2 - Artificial Intelligence) September 2022

Buku Prosiding Seminar Nasional Mahasiswa Fakultas Teknologi Informasi (SENAFTI)

Buku 2: Artificial Intelligence

Diterbitkan: 30-09-2022

Artificial Intelligence

Penerapan Algoritma KNN Pada Twitter Untuk Analisis Sentimen Masyarakat Terhadap Event MotoGP Di Sirkuit Mandalika

Erlangga Dwi Nugraha, Grace Gata

1-7



Penerapan K-Means Clustering Untuk Menentukan Peminatan Barang Pada Jayyid Shop

Fikri Dermawan, Gunawan Pria Utama

8-17



Implementasi Algoritma Apriori Untuk Analisis Penjualan Produk Pada Toko Perjuangan Collection

I Komang Ardika Viantama, Painem Painem

18-24



<https://senafiti.budiluhur.ac.id/index.php/senafiti/issue/view/2>

Penerapan Algoritma KNN Pada Twitter Untuk Analisis Sentimen Masyarakat Terhadap Event MotoGP Di Sirkuit Mandalika

Erlangga Dwi Nugraha
Universitas Budi Luhur

Grace Gata
Universitas Budi Luhur

Kata Kunci: data mining, k-nearest neighbours, analisis sentimen, event motogp di sirkuit mandalika

Abstrak

Setelah 25 tahun menunggu, Indonesia akan kembali menjadi tuan rumah ajang balap MotoGP. ajang balap MotoGP 2022 di Indonesia digelar pada tanggal 18-20 Maret 2022, Nusa Tenggara Barat (NTB). Pada tahun 1996 sampai dengan 1997, Indonesia pernah juga berkesempatan menjadi tuan

 PDF

Diterbitkan
30-09-2022

Cara Mengutip

Nugraha, E. D., & Gata, G. (2022). Penerapan Algoritma KNN Pada Twitter Untuk Analisis Sentimen Masyarakat Terhadap Event MotoGP Di Sirkuit Mandalika. *Seminar Nasional Mahasiswa Fakultas Teknologi Informasi (SENAFTI) Universitas Budi Luhur*, 1(1), 1-7. Diambil dari <https://senafiti.budiluhur.ac.id/index.php/senafiti/article/view/38>

Format Sitasi Lainnya ▼

<https://senafiti.budiluhur.ac.id/index.php/senafiti/article/view/38>

<https://senafiti.budiluhur.ac.id/index.php/senafiti/article/view/38/1>



Seminar Nasional Mahasiswa Fakultas Teknologi Informasi (SENAFTI)
Jakarta – Indonesia, 6 September 2022

Penerapan Algoritma KNN Pada Twitter Untuk Analisis Sentimen Masyarakat Terhadap Event MotoGP Di Sirkuit Mandalika

Erlangga Dwi Nugraha^{1*}, Grace Gata²

^{1,2}Fakultas Teknologi Informasi, Sistem Informasi, Universitas Budi Luhur, Jakarta Selatan, Indonesia

Email: ¹erlanggadwi@ugraha454@gmail.com, ²grace.gata@budiluhur.ac.id
(* corresponding author)

Abstrak—Setelah 25 tahun menunggu, Indonesia akan kembali menjadi tuan rumah ajang balap MotoGP 2022 di Indonesia digelar pada tanggal 18-20 Maret 2022, Nusa Tenggara Barat (NTB). Pada tahun 1996 sampai dengan 1997, Indonesia pernah juga berkesempatan menjadi tuan rumah festival balap motor paling bergengsi, kala itu bernama GP500 yang digelar di Sirkuit Sentul, Bogor, Jawa Barat pada tanggal 7 April 1996 dan 28 September 1997. Event balap MotoGP Mandalika 2022 di Indonesia ini ada beberapa kemungkinan respon yang muncul dari masyarakat yaitu respon positif maupun respon negatif, tergantung pada pandangan publik terhadap event tersebut. Maka tujuan dari penelitian ini untuk melakukan analisa sentimen dengan menggunakan algoritma *K-Nearest Neighbors*, untuk mengetahui berapa besar skorasi yang dihasilkan dari algoritma *K-Nearest Neighbors* dalam mengklasifikasi sentimen dari sebuah *tweet* yang berkaitan dengan Sirkuit Mandalika. Penelitian ini akan menerapkan metode *Cross-Industry Standard Process for Data Mining (CRISP-DM)*. Kemudian data yang telah dikumpulkan tersebut harus melalui beberapa tahapan seperti tahapan *preprocessing* dan *modeling*. Sehingga didapat 5000 *tweet* yang terdiri dari 3022 *tweet* sentimen positif dan 1978 *tweet* sentimen negatif. Maka bisa disimpulkan bahwa terdapat 55.91% *tweet* dengan sentimen positif terhadap Event MotoGP di Sirkuit Mandalika. Hasil yang diperoleh dari tahap *modeling* dengan menggunakan algoritma *K-Nearest Neighbors* dan perbandingan 80:20 untuk data *training* dan *testing*, maka nilai *accuracy* yang dihasilkan sebesar 96.51%, *precision* 99.13%, *recall* 96.69%.

Kata Kunci: *data mining, K-Nearest Neighbors, analisa sentimen, event MotoGP di sirkuit mandalika*

Application of KNN Algorithm on Twitter for Analysis of Community Sentiment on Events MotoGP in Mandalika Circuit

Abstract—After 25 years of waiting, Indonesia will again host the MotoGP racing event. The 2022 MotoGP racing event in Indonesia will be held on March 18-20 2022, West Nusa Tenggara (NTB). From 1996 to 1997, Indonesia also had the opportunity to host the most prestigious motorcycle racing festival, at that time named the GP500 which was held at the Sentul Circuit, Bogor, West Java on April 7, 1996 and September 28, 1997. The MotoGP Mandalika 2022 racing event in Indonesia, there are several possible responses that arise from the community, namely positive responses or negative responses, depending on the public's view of the event. So the purpose of this study is to conduct sentiment analysis using the K-Nearest Neighbors algorithm, to find out how much accuracy is generated from the K-Nearest Neighbors algorithm in classifying the sentiment of a *tweet* related to the Mandalika Circuit. This research will apply the Cross-Industry Standard Process For Data Mining (CRISP-DM) methodology. Then the data that has been collected must go through several stages such as *preprocessing* and *modeling* stages. So that we get 5000 *tweets* consisting of 3022 positive sentiment *tweets* and 1978 negative sentiment *tweets*. So it can be concluded that there are 55.91% of *tweets* with positive sentiment towards the MotoGP Event at the Mandalika Circuit. The results obtained from the modeling stage using the K-Nearest Neighbors algorithm and a comparison of 80:20 for *training* and *testing* data, the resulting *accuracy* value is 96.51%, *precision* is 99.13%, *recall* is 96.69%.

Keywords: *data mining, K-Nearest Neighbors, sentiment analysis, MotoGP event at mandalika circuit*

1. PENDAHULUAN

Selapas menunggu sangat lama Indonesia kembali menjadi tuan rumah ajang balap MotoGP (sebelumnya dikenal sebagai Grand Prix 500 atau GP500), balapan MotoGP 2022 di Indonesia akan berlangsung pada 18-20 Maret 2022, 1996 dan 18-20 Maret 1997 di Sirkuit Sentul Bogor, Jawa Barat, September 1997 [1].

MotoGP di sirkuit mandalika 2022 disambut bersemangat oleh masyarakat Indonesia. Banyak warga Indonesia menyambut kedatangan para pembalap MotoGP itu tiba di Bandara Internasional Soekarno-Hatta, Presiden Joko Widodo yang bersemangat menyambut secara langsung ini dengan cara menyambut selebritas/pemirsa para pembalap di Istana Negara. Semangat dari masyarakat Indonesia juga dapat diamati dari tiket stadion langsung GP mandalika yang habis terjual. Sirkuit Mandalika dapat memberikan dampak positif yang besar dalam pengembangan pariwisata di Indonesia, serupa dengan dampak finansial kinerja dari pernyjukan WSBK 2021, pernyjukan balap MotoGP Mandalika 2022 juga akan berdampak ekonomi jangka pendek, dengan kata lain, akan meningkatkan perekonomian daerah, khususnya perekonomian NTB[2]. Berdasarkan statistik pariwisata Kabupaten Lombok tahun 2018 angka kunjungan wisatawan kabupaten Lombok terus mengalami penurunan



Seminar Nasional Mahasiswa Fakultas Teknologi Informasi (SENAFITI)
Jakarta - Indonesia, 6 September 2022

akibat terjadinya bencana gempa bumi pada akhir tahun 2018. Terlebih lagi pariwisata Nusa Tenggara Barat kembali terganggu dengan adanya wabah Covid-19 pada tahun 2020. Namun setelah adanya Pergelaran yang digelar MotoGP Mandalika 2021 berdampak positif membuat angka kunjungan wisatawan ke kabupaten Lombok terus meningkat [3].

Sering berkembangnya teknologi, masyarakat menjadi sering membicarakan. Masyarakat memberikan pendapat melalui berbagai media sosial *Twitter*. Pendapat tersebut bisa berupa pendapat *negative* atau *positif* tergantung dari pandangan publik. Oleh karena itu dibutuhkan suatu analisis terhadap opini-opini tersebut dalam penelitian ini menggunakan metode *K-Nearest Neighbors*. *K-Nearest Neighbors* merupakan salah satu algoritma yang sudah diimplementasikan dan paling banyak digunakan untuk menyelesaikan masalah tersebut [4]. *K-Nearest Neighbors* dengan tingkat ke efektifitas yang tinggi, serta cocok untuk berbagai masalah yang berhubungan dengan klasifikasi. Maka tujuan dari penelitian ini adalah melakukan analisis sentimen dengan menggunakan algoritma *K-Nearest Neighbors* dalam mengklasifikasikan sentimen dari sebuah *tweet* yang berkaitan dengan Sirkuit Mandalika [4].

2. METODE PENELITIAN

a. Tahapan Penelitian



Gambar 1. Tahapan Penelitian Analisis Sentimen Masyarakat Terhadap Sirkuit Mandalika Menggunakan Algoritma KNN

Pada Penelitian ini terdiri dari beberapa tahapan seperti yang di tunjukkan pada Gambar 1, pada penelitian menggunakan metodologi *Cross-Industry Standard Process for Data Mining* (CRISP-DM), CRISP-DM merupakan metode yang menggunakan model proses pengembangan data yang di terapkan ke dalam strategi untuk memecahkan suatu masalah [6]. Berikut ini adalah penjelasan mengenai tahapan-tahapan yang dilakukan selama penelitian:

a. Pemahaman Bisnis

Pada tahap ini penelitian mencoba memahami masalah yang diangkat yaitu analisis sentimen terhadap Sirkuit Mandalika, penelitian ini dilakukan dengan mempelajari artikel atau berita di internet, kemudian penelitian ini mencari *tweet* atau trending topik terkait Sirkuit Mandalika di media sosial *twitter*.

b. Pemahaman Data

Pada tahapan penelitian ini mencoba untuk memahami data yang akan digunakan sekaligus menentukan *tweet* yang sesuai dengan topik permasalahan dengan menggunakan beberapa *keyword* seperti Sirkuit Mandalika. Data penelitian diumpulkan menggunakan *software* *rapidminer*. Sebelum melakukan proses pengambilan *tweet*, peneliti membuat koneksi yang menghubungkan antara *rapidminer* dengan *twitter* agar proses pengumpulan data *tweet* dapat dilakukan. Maka selanjutnya peneliti akan menentukan lama periode data *tweet* yang akan digunakan pada penelitian ini melakukan pengambilan data *tweet* dari media sosial *twitter* pada tanggal 18 Maret 2022 hingga 17 Maret 2021. Dengan target data *tweet* yang di kumpulkan sebanyak 5000 *tweet*. Setelah mengumpulkan data kemudian data di grupung menjadi satu, lalu peneliti melakukan *remove duplicate* menggunakan *rapid miner* untuk menghapus data yang muncul secara berulang. Setelah melakukan *remove duplicate* maka sebelum melakukan *preprocessing* data, data perlu diberi label sentimen terlebih dahulu, pelabelan data dilakukan dengan menggunakan *azure machine* yang terapat dalam *get add ins* di *Microsoft excel* [7].

c. Persiapan Data

Pada tahapan ini data yang telah diberikan pelabelan sentiment, maka data *tweet* akan dilakukan *data preprocessing*. Berikut ini adalah langkah-langkah yang dilakukan selama *data preprocessing* [8]:

1. *Remove duplicates* adalah proses untuk menghapus data yang muncul secara berulang.

2. *Replace* adalah proses untuk menghilangkan kata yang terdapat pada tweets seperti RT (*Retweet*), <https://twitter.com/username>.
3. *Case folding* adalah proses mengubah semua data *tweet* menjadi huruf kecil.
4. *Tokenizing* adalah proses pemecahan *tweet* dari kalimat menjadi kata individual, menghilangkan karakter yang bersifat pembatas, dan menghapus angka atau tanda baca.
5. *Stopwords* adalah proses yang dilakukan untuk menghilangkan kata sambungan atau kata yang tidak penting.
6. *Stemming* adalah proses untuk mencari akar kata. pada proses *stemming* ini peneliti membuat *file stem* yang berfungsi menghilangkan imbuhan atau mencari akar kata.
7. *Term Weighting* (TF-IDF) adalah proses untuk melakukan pembobotan kata, dengan cara mencari nilai *Term Frequency* (TF), lalu mencari nilai *Document Frequency* (DF), kemudian mencari nilai *Inverse Document Frequency* (IDF), dan menghitung bobot[9].

d. **Penodekan**
Pada tahapan ini penelitian akan melakukan penodekan terhadap *dataset* yang sudah dilakukan *pre-processing* dan *dataset* yang belum dilakukan *pre-processing*. Kemudian *dataset* tersebut akan dipecah menjadi data *training* dan data *testing*, pada tahapan ini akan menjadi dua tahapan yaitu *split data* dan *cross fold validation*. Pada *split data* peneliti akan membagi *dataset* tersebut dengan perbandingan data *training* dan data *testing* yaitu 80:40, 70:30 dan 50:50. Dan menggunakan *cross validation* dengan pembagian secara acak ke dalam 10 bagian. Pada tahap penodekan peneliti menggunakan algoritma *k-nearest neighbors* namun sebelumnya telah dilakukan perbandingan model klasifikasi dengan menggunakan algoritma *decision tree* dan *naive bayes*[10].

e. **Evaluasi**
Pada tahapan ini peneliti mengevaluasi metode klasifikasi dengan mengukur kinerja yang menggunakan *confusion matrix* terhadap algoritma *k-nearest neighbors*.

3. HASIL DAN PEMBAHASAN

3.1 Pemahaman Terhadap Bisnis (*Business Understanding*)

Pada tahapan *business understanding* ini peneliti mencoba untuk memahami objek atau masalah yang di angkat selama penelitian. Masalah yang di angkat oleh peneliti adalah seberapa besar akurat yang dihasilkan algoritma *k-nearest neighbors* dalam mengklasifikasi sentimen dari sebuah *tweet* yang berkaitan dengan penerapan Sirkuit Mandalika. Pada penelitian ini menggunakan data yang didapat melalui media sosial *twitter*, karena masih banyak sekali warga Indonesia yang menggunakan media sosial *twitter* sebagai media untuk memberitahu atau info tentang apa yang mereka lakukan atau rasakan, membagikan informasi, dan melaporkan segala cerita berita kepada para pengguna yang lain. Pada media sosial *twitter* juga memiliki satu kelebihan yaitu jika terdapat banyak *tweet* dengan topik yang sama, maka *twitter* akan otomatis menjadikan topik tersebut menjadi sebuah *trending topic*. Sebuah *trending topic* dapat dijadikan sebagai *keyword* pada saat proses *crawling* data, karena dapat memudahkan peneliti untuk mendapatkan data *tweet* sesuai dengan topik yang di butuhkan.

3.2 Pemahaman Terhadap Data (*Data Understanding*)

Pada tahap data *understanding* maka peneliti mencoba untuk memahami data yang akan digunakan. Pada tahap ini peneliti juga melakukan pengumpulan data yang akan diambil dari media sosial *twitter* dengan menggunakan *rapidminer* serta menentukan *keyword* seperti Sirkuit Mandalika untuk metode *crawling* data *tweet*. Peneliti juga melakukan pengumpulan data *tweet* yang berangkutan dengan Sirkuit Mandalika pada tanggal 17 Maret 2021 hingga 27 Maret 2021. Data yang dikumpulkan pada proses *Crawling* ini dengan menggunakan *keyword* Sirkuit Mandalika adalah 5000 data *tweet*.

3.3 Persiapan Data (*Data Preparation*)

Pada tahapan data *preparation* ini maka data yang sudah terkumpul dan sudah diberi label sentimen maka akan dilakukan data *preprocessing* oleh peneliti. Pada tahapan data *preparation* ini peneliti akan membagi menjadi 2 tahapan sebagai berikut :

- a. Tahapan *preprocessing*, tahapan ini merupakan proses yang akan dilakukan oleh peneliti setelah data *tweet* sudah memiliki sentimen positif atau negatif.
- b. Penentuan data *training* dan data *testing*, tahapan ini merupakan proses untuk menentukan perbandingan data *training* dan data *testing* terlebih dahulu sebelum masuk ke dalam tahapan *modeling*.

3.3.1 Tahapan Preprocessing

Membersihkan data dilakukan dengan menggunakan tahapan preprocessing yang terdiri dari *transform cases*, *tokenize*, *filter tokens (by Length)*, *filter stopwords (Dictionary)* dan *stem (Dictionary)*. Dimana dalam penggunaan *filter stopwords* peneliti menggunakan *lemmas stopwords* milik *Kaggle.com* dan penggunaan *stem* peneliti membuat manual dengan bantuan kamus besar bahasa Indonesia. Simpan kata-kata yang ada dalam dataset dalam bentuk *file stopwords.txt* yang kemudian dimasukkan ke dalam *rapid miner* yang selanjutnya membuat kalimat menjadi kata dasar. Berikut ini adalah contoh hasil dari tahapan *preprocessing* Tabel 1.

Tabel 1. Perbandingan Sebelum dan Sesudah Preprocessing

Preprocessing	Sebelum	Sesudah
<i>Transform Cases</i>	Kata Sirkuit Mandalika Bisa Dipukul F1 On Si	Kata sirkuit mandalika bisa dipukul F1 ga si
<i>Tokenize</i>	Kata Sirkuit Mandalika Bisa Dipukul F1 On Si	Kata Sirkuit Mandalika Bisa Dipukul F1 On Si
<i>Filter Tokens (by Length)</i>	Ini baru berita MotoGP Finlandia 2022 bakal digelar DMI sirkuit Mandalika sebagai pengganti	Baru berita MotoGP Finlandia 2022 bakal digelar sirkuit Mandalika sebagai pengganti
<i>Filter Stopwords (Dictionary)</i>	Ini baru berita MotoGP Finlandia 2022 bakal digelar DMI sirkuit Mandalika sebagai pengganti	Ini berita MotoGP Finlandia 2022 bakal digelar DMI sirkuit Mandalika sebagai pengganti
<i>Stem (Dictionary)</i>	Sekelompok Indonesia sukses menggelar World Superbike tahun lalu untuk musim ini akan berlangsung pada November 2022 dan MotoGP di Sirkuit Mandalika Lombok Nusa Tenggara Barat NTB Sukanan Event MotoGP	Sekelompok Indonesia sukses menggelar World Superbike tahun lalu untuk musim ini akan langsung pada November 2022 dan MotoGP di Sirkuit Mandalika Lombok Nusa Tenggara Barat NTB Sukanan Event MotoGP

3.3.2 Pembagian Data Training dan Data Testing

Data *training* adalah data latih yang akan digunakan untuk proses latihan bagi model klasifikasi. Data *testing* adalah data yang akan digunakan untuk uji *rule* klasifikasi. Berikut ini adalah perbandingan data *training* dan data *testing* 60:40, 70:30, dan 80:20 yang dapat dilihat pada Tabel 2. Metode pembagiannya menggunakan metode *stratified sampling*.

Tabel 2. Sebaran Data Training dan Data Testing Pada dataset (Ulasan Masyarakat terhadap Sirkuit Mandalika)

Perbandingan 60:40				
Sentimen	Training	Testing	Total	
Positive	1615	1009	3022	
Negative	340	226	566	
Total	2153	1435	3588	
Perbandingan 70:30				
Sentimen	Training	Testing	Total	
Positive	2115	907	3022	
Negative	296	170	566	
Total	2511	1077	3588	
Perbandingan 80:20				
Sentimen	Training	Testing	Total	
Positive	2418	604	3022	
Negative	463	113	566	
Total	2881	717	3588	

3.4 Pemodelan

Pada tahapan *modeling* ini, maka peneliti akan melakukan pengukuran performa klasifikasi dalam 2 tahap :

- Berikut *Modeling* menggunakan *split* data Pada tahapan ini maka data akan di *split* menjadi 3 model yaitu 60:40, 70:30, dan 80:20.
- Modeling* menggunakan *cross validation* Pada tahapan ini maka data *training* dan data *testing* di *split* secara acak menjadi 10 bagian (*number of folds* = 10).

3.4.1 Modeling Menggunakan Split Data

Berikut ini adalah penjelasan algoritma K-Nearest Neighbors menggunakan *split* data (Gambar 2):

- Read Excel adalah operator untuk memilih file excel yang akan digunakan.

- Nominal to Text berfungsi untuk mengubah jenis atribut nominal yang dipilih menjadi teks. Data teks diproses untuk mengekstrak kata-kata dan mengirimkan kata menjadi vektor.
- Process Documents from Data adalah operator untuk preprocessing data yang digunakan untuk memasukkan operator transform cases, tokenize, filter tokens (By Length), filter stem (Dictionary) dan stopwords (Dictionary) didalamnya.
- Split data adalah operator yang digunakan untuk membagi data training dan data testing.
- K-NN adalah operator yang menghasilkan algoritma yang digunakan untuk klasifikasi.
- Apply Model digunakan untuk menerapkan model yang telah dilatih sebelumnya menggunakan data training pada data testing.
- Performance digunakan untuk mengevaluasi kinerja model yang memberikan daftar nilai kinerja secara otomatis, misalnya accuracy, precision, dan recall.



Gambar 2. Penerapan Algoritma Dengan Model Split Data

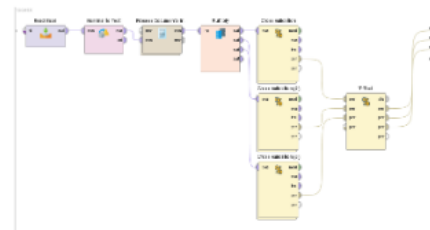
3.4.2 Pemodelan Menggunakan Cross Validation

Berikut ini adalah penjelasan algoritma K-Nearest Neighbors menggunakan Cross Validation (Gambar 3) dan (Gambar 4):

- Read File adalah operator untuk memilih file excel yang akan digunakan.
- Nominal to Text berfungsi untuk mengubah jenis atribut nominal yang dipilih menjadi teks. Data teks diproses untuk mengekstrak kata-kata dan mengirimkan kata menjadi vektor.
- Process Documents from Data adalah operator untuk preprocessing data yang digunakan untuk memasukkan operator transform cases, tokenize, filter tokens (By Length), filter stem (Dictionary) dan stopwords (Dictionary) didalamnya.
- Split data adalah operator yang digunakan untuk membagi data training dan data testing.
- Multiply adalah operator port input dan mengirimkan salinannya ke port output. Setiap port yang terhubung membuat salinan independen. Jadi mengubah satu salinan tidak berpengaruh pada salinan lainnya.
- Cross Validation adalah operator berulang yang memiliki dua subproses yaitu subproses training dan subproses testing. Subproses training digunakan untuk melatih model, model yang dilatih kemudian diterapkan dalam subproses testing, tetapi tidak kinerja model diukur selama fase testing.
- T-Test adalah Operator yang digunakan untuk perbandingan vektor kinerja. Operator ini melakukan uji-t untuk menentukan probabilitas hipotesis nol yang benar sebenarnya adalah sama.
- K-NN adalah operator yang menghasilkan algoritma yang digunakan untuk klasifikasi.
- Apply Model digunakan untuk menerapkan model yang telah dilatih sebelumnya menggunakan data training pada data testing.
- Performance digunakan untuk mengevaluasi kinerja model yang memberikan daftar nilai kinerja secara otomatis, misalnya accuracy, precision, dan recall.



Seminar Nasional Mahasiswa Fakultas Teknologi Informasi (SENAFITI)
Jakarta - Indonesia, 6 September 2022



Gambar 3. Penerapan Algoritma K-NN Dengan Model Cross Validation



Gambar 4. Proses Cross Validation

3.4.3 Hasil Pemodelan Menggunakan Split Data

Setelah melakukan tahapan pemodelan dengan menggunakan perbandingan data *training* dan data *testing* terhadap dataset yang sudah dilakukan *preprocessing*. Maka nilai *accuracy* yang dihasilkan dapat dilihat pada Tabel 3.

Tabel 3. Perbandingan Accuracy algoritma KNN, Naive Bayes dan SVM Menggunakan Metode Split Data

Perbandingan	Algoritma		
	<i>K-Nearest Neighbors</i>	<i>Naive Bayes</i>	<i>Support Vector Machine</i>
60:40	94.07%	84.16%	84.38%
70:30	94.24%	84.68%	84.29%
80:20	93.86%	82.34%	83.96%

3.4.4 Hasil Proses Pemodelan Menggunakan Cross Validation

Setelah melakukan tahapan pemodelan terhadap dataset yang sudah dilakukan *preprocessing*. Maka nilai *accuracy* yang dihasilkan dapat dilihat pada Tabel 4.

Tabel 4. Hasil Accuracy algoritma KNN, Naive Bayes dan SVM Menggunakan Metode Cross Validation

Algoritma	Accuracy
<i>K-Nearest Neighbors</i>	91.30%
<i>Naive Bayes</i>	85.28%
<i>Support Vector Machine</i>	84.36%

Berdasarkan Tabel 3 dan Tabel 4 yang berisikan hasil pengukuran performa model klasifikasi *k-nearest neighbors* dengan 3 cara yaitu pembagian dataset dengan *split data* dan *cross validation*, maka dapat disimpulkan bahwa pemodelan dengan *split data* 70:30 memiliki akurasi sebesar 94.24% sedangkan dengan pemodelan *cross validation* memiliki akurasi sebesar 91.30% sehingga akurasi tertinggi yang dihasilkan pada penelitian ini adalah dengan pemodelan dengan *split data* 70:30.



SENAFTI
SEMINAR NASIONAL MAHASISWA
FAKULTAS TEKNOLOGI INFORMASI

Seminar Nasional Mahasiswa Fakultas Teknologi Informasi (SENAFTI)
Jakarta - Indonesia, 6 September 2022

3.5 Pengujian

Penelitian ini akan melakukan pengujian dengan menggunakan *confusion matrix* yang diperoleh dari tahapan *modelling* dengan menggunakan algoritma *K-Nearest Neighbors* dimana pada tabel 5 akurasi dengan *split data* 70:30 lebih tinggi dibandingkan *cross validation*. Sehingga *confusion matrix* yang disajikan pada Tabel 5 adalah hasil dari waktuu pengujian Klasifikasi *K-Nearest Neighbors* dengan *dataset* yang berjumlah 5558, terdapat 2871 data *training* dan 757 data *testing*.

	True positive	True negative	Class precision
Pred. positive	874	7	99.21%
Pred. negative	32	163	83.39%
Class recall	96.47%	95.58%	

Perhitungan dari Tabel 5 adalah sebagai berikut :

$$a. Precision = \frac{TP}{TP+FP} = \frac{874}{874+7} = 0.9921 = 99.21\%$$

$$b. Recall = \frac{TP}{TP+FN} = \frac{874}{874+32} = 0.9647 = 96.47\%$$

$$c. Accuracy = \frac{TP+TN}{TP+FP+FN} = \frac{874+163}{874+163+7+32} = 0.9638 = 96.38\%$$

4. KESIMPULAN

Penelitian ini melakukan Analisis Sentimen Masyarakat Terhadap Sirkuit Mandalika Dengan Algoritma *K-Nearest Neighbors*. Berdasarkan hasil pengujian dan analisis yang telah dilakukan, Menggunakan data dari media sosial *twitter* pada tanggal 17-27 Maret yang sudah dilakukan *pre-processing* dan pemberian label sentiment dengan jumlah *dataset* yang digunakan 5558 *tweet* menghasilkan data *tweet* dengan sentiment positif sebanyak 3023, lalu data *tweet* dengan sentiment negatif sebanyak 566 data maka dapat disimpulkan bahwa metode algoritma *K-Nearest Neighbors* mendapatkan nilai akurasi sebesar 96.38% saat menggunakan model pengujian *split data* dengan perbandingan 70:30 dan saat model pengujian *Cross Validation* mendapatkan nilai akurasi sebesar 91.50%.

DAFTAR PUSTAKA

- [1] Khabib, "Dampak Ekonomi Dari Pergelaran MotoGP Mandalika 2022 : Jangka Pendek dan Jangka Panjang," *Jaya Singkat, Pus. Penelit. Badan Keahlian DPR RI*, vol. 14, no. 6, pp. 13-18, 2022.
- [2] N. Tri Ramadani, I. Samsan, and S. Budilaksana, "Perbandingan Metode Naive Bayes, KNN, dan Decision Tree Terhadap Analisis Sentimen Transparansi KRL Commuter Line," *J. IDRA-JTW Inform.*, vol. 3, no. 2, pp. 1-9, 2019.
- [3] K. Kusni, "Analisis Pemberdayaan Masyarakat Dalam Pengembangan Pariwisata Kota Mandalika 1 Fakultas Ekonomi dan Bisnis Universitas Mataram * Corresponding Author : diawandi@unram.ac.id Info Artikel Lombok Tengah merupakan salah satu daerah tujuan wisata yang cukup," vol. 4, no. 1, 2022.
- [4] R. H. Samsi and M. A. Fandi, "Klasifikasi *Tweets* Pada *Twitter* Menggunakan Metode *K-Nearest Neighbour (K-NN)* Dengan Pembebasan TF-IDF," vol. 3, no. 8, pp. 8299-8300, 2019.
- [5] A. Salam, J. Zeniarja, and R. S. U. Khasanah, "Analisis Sentimen Data Komentar Sosial Media Facebook Dengan *K-Nearest Neighbor* (Studi Kasus Pada Akun Jass Elaypedin Burang J&T Elaypedi Indonesia)," *Pros. IJNTAS*, pp. 480-486, 2018.
- [6] A. D. Adhi Putra, "Analisis Sentimen pada Ulasan pengguna Aplikasi Bibit Dan Berkes dengan Algoritma KNN," *JATIS (Jurnal Tek. Inform. dan Sasi Informan)*, vol. 3, no. 2, pp. 636-646, 2021, doi: 10.33937/jatin.v3i2.362.
- [7] S. Esmarati and S. Wan, "Penerapan Algoritma *K-Nearest Neighbors* Pada Analisis Sentimen Review Agen Travel," *J. Khasanah Inform.*, vol. 6, no. 1, pp. 64-69, 2018.
- [8] M. S. Almagh, I. Esmarati, and I. Nurilili, "Analisis Sentimen Terhadap Pelayanan PT PLN di Jakarta pada *Twitter* dengan Algoritma *K-Nearest Neighbors (K-NN)*," *Seminar Nas. Mahasiswa Kampus dan Apm*, pp. 110-122, 2020.
- [9] A. Supriatman, "Pembebasan TF-IDF pada Judul Penelitian Dosen Sebagai Dasar Klasifikasi Menggunakan Algoritma *K-NN* (Studi Kasus: Universitas Biliwangi)," *J. Serambi Eng.*, vol. 6, no. 1, pp. 1573-1579, 2021, doi: 10.33267/jse.v6i1.2643.
- [10] F. Jodit and D. Khairudin, "Analisis Sentimen dengan SVM, Naive Bayes dan KNN untuk Studi Tanggapan Masyarakat Indonesia Terhadap Pandemi Covid-19 pada Media Sosial *Twitter*," *Prisma*, vol. 4, pp. 628-634, 2021.