

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

Erlangga Dwi Nugraha<sup>1\*</sup>, Grace Gata<sup>2</sup>

<sup>1\*2</sup>Fakultas Teknologi Informasi, Sistem Informasi, Universitas Budi Luhur, Jakarta Selatan, Indonesia

Email: <sup>1</sup>erlanggadwinugraha454@gmail.com, <sup>2</sup>grace.gata@budiluhur.ac.id  
(\* : corresponding author)

**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 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 objek tersebut. Maka tujuan dari penelitian ini untuk melakukan analisis sentimen dengan menggunakan algoritma *K-Nearest-Neighbors*, untuk mengetahui berapa besar akurasi yang dihasilkan dari algoritma *K-Nearest Neighbors* dalam mengklasifikasi sentimen dari sebuah tweet yang berkaitan dengan Sirkuit Mandalika. Penelitian ini akan menerapkan metodelogi *Cross-Industry Standard Process for Data Mining (CRISP-DM)*. Kemudian data yang telah dikumpulkan tersebut harus melewati beberapa tahapan seperti tahapan *preprocessing* dan *modelling*. Sehingga didapat 5000 tweet yang terdiri dari 3022 tweet sentimen positif dan 566 tweet sentimen negatif. Maka bisa disimpulkan bahwa terdapat 85.91% tweet dengan sentimen positif terhadap Event MotoGP di Sirkuit Mandalika. hasil yang diperoleh dari tahap *modelling* dengan menggunakan algoritma *k-nearest neighbors* dan perbandingan 80:20 untuk data *training* dan *testing*, maka nilai *accuracy* yang dihasilkan sebesar 96.51%, *precision* 99.15%, *recall* 96.69%.

**Kata Kunci:** *data mining*, *k-nearest neighbors*, analisis 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 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 object. 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 566 negative sentiment tweets. So it can be concluded that there are 85.91% 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 99.15%, recall is 96.69%.

**Keywords:** *data mining*, *k-nearest neighbors*, *sentiment analysis*, *motogp event at mandalika circuit*

---

## 1. PENDAHULUAN

Selepas 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 Mandalika, Nusa Tenggara Barat (NTB). Balapan GP500 digelar pada 7 April 1996 dan 28 Maret 1996. Di Sirkuit Sentul Bogor, Jawa Barat, September 1997 [1].

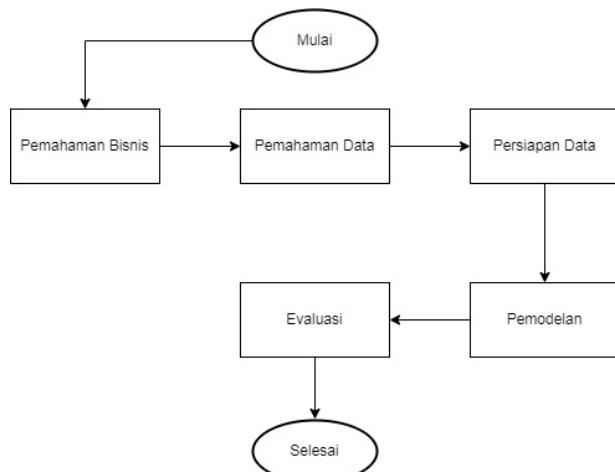
MotoGP di sirkuit mandalika 2022 disambut bersemangat oleh masyarakat Indonesia. Banyak warga Indonesia menyambut kedatangan para pembalap MotoGP itu tiba di Bandara Internasional Soekarno-Hatta. Presiden Joko Widodo yang bersemangat menyambut acara perlombaan ini dengan cara menyambut sekelompok para pembalap di Istana Negara. Semangat dari masyarakat Indonesia juga dapat diamati dari tiket nonton 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 pertunjukan WSBK 2021, petunjukan 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 akibat terjadinya bencana gempa bumi pada akhir tahun 2018. Terlebih lagi pariwisata Nusa Tenggara Barat kembali terguncang dengan adanya wabah Covid-19 pada tahun 2020, Namun setelah adanya Pergelaran ajang balap MotoGP Mandalika 2022 berdampak positif membuat angka kunjungan wisatawan kabupaten Lombok terus membaik [3].

Seiring berkembangnya teknologi, masyarakat menjadi sering membicarakannya. 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 mudah diimplementasikan dan paling banyak digunakan untuk menyelesaikan masalah tersebut[4]. *K-Nearest Neighbors* dengan tingkat ke efektifan 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 [5].

## 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 ditunjukkan pada Gambar 1, pada penelitian menggunakan metodologi *Cross-Industry Standard Prosess 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 dikumpulkan 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 27 Maret 2021, Dengan target data *tweet* yang dikumpulkan sebanyak 5000 *tweet*. Setelah mengumpulkan data kemudian data digabung 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:](https://), dan @username.
  3. *Case folding* adalah proses mengubah semua data *tweets* 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 *Invers Document Frequency (IDF)*, dan menghitung bobot[9].
- d. Pemodelan
- Pada tahapan ini penelitian akan melakukan pemodelan terhadap *dataset* yang sudah dilakukan *preprocessing* dan *dataset* yang belum dilakukan *preprocessing*. 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 60:40, 70:30 dan 80:20. Dan menggunakan *cross validation* dengan pembagian secara acak ke dalam 10 bagian. Pada tahap pemodelan peneliti menggunakan algoritma *k-nearest neighbors* namun sebelumnya telah dilakukan perbandingan model klasifikasi dengan menggunakan *algoritma decision tree* dan *naïve bayes*[10].
- e. Evaluasi
- Pada tahapan ini peneliti mengevaluasi metode klasifikasi dengan mengukur kinerjanya 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 diangkat selama penelitian. Masalah yang di angkat oleh peneliti adalah seberapa besar akurasi yang dihasilkan algoritma *k-nearest neighbors* dalam mengklasifikasikan 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 suatu berita kepada para pengguna yang lain. Pada media sosial *twitter* juga memiliki satu lelebihan yaitu jika terdapat banyak *tweet* dengan topik yang sama, maka *twitter* akan otomatis menjadikan topik tersebut menjadi sebuah trending topik. Sebuah trending topik dapat dijadikan sebagai *keywords* 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 *keywords* seperti Sirkuit Mandalika untuk metode *crawling* data *tweet*. Peneliti juga melakukan pengambilan data *tweet* yang bersangkutan dengan Sirkuit Mandalika pada tanggal 17 Maret 2022 hingga 27 Maret 2022. Data yang dikumpulkan pada proses *Crawling* ini dengan menggunakan *keywords* 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 *modelling*

### 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 kamus *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 notepad* atau .txt yang kemudian dimasukan 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*

<b>Preprocessing</b>	<b>Sebelum</b>	<b>Sesudah</b>
<i>Transform Cases</i>	Kalo Sirkuit Mandalika Bisa Dipake F1 Ga Si	Kalo sirkuit mandalika bisa dipake F1 ga si
<i>Tokenize</i>	Kalo Sirkuit Mandalika Bisa Dipake F1 Ga Si	Kalo Sirkuit Mandalika Bisa Dipake F1 Ga Si
<i>Filter Tokens (by Length)</i>	Ini baru berita Moto GP Finlandia 2022 batal digelar IMI ajukan sirkuit Mandalika sebagai pengganti	Baru berita Moto Finlandia 2022 batal digelar ajukan sirkuit Mandalika sebagai pengganti
<i>Filter Stopwords (Dictionary)</i>	Ini baru berita Moto GP Finlandia 2022 batal digelar IMI ajukan sirkuit Mandalika sebagai pengganti	Ini berita Moto GP Finlandia 2022 batal digelar IMI ajukan sirkuit Mandalika pengganti
<i>Stem (Dictionary)</i>	Sebelumnya 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 Sukseskan Event MXGP	Sebelum 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 Sukses Event MXG

### 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)

<b>Perbandingan 60:40</b>			
<i>Sentimen</i>	<i>Training</i>	<i>Testing</i>	<i>Total</i>
<i>Positive</i>	1813	1209	3022
<i>Negative</i>	340	226	566
<i>Total</i>	2153	1435	3588
<b>Perbandingan 70:30</b>			
<i>Sentimen</i>	<i>Training</i>	<i>Testing</i>	<i>Total</i>
<i>Positive</i>	2115	907	3022
<i>Negative</i>	396	170	566
<i>Total</i>	2511	1077	3588
<b>Perbandingan 80:20</b>			
<i>Sentimen</i>	<i>Training</i>	<i>Testing</i>	<i>Total</i>
<i>Positive</i>	2418	604	3022
<i>Negative</i>	453	113	566
<i>Total</i>	2871	737	3588

### 3.4 Pemodelan

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

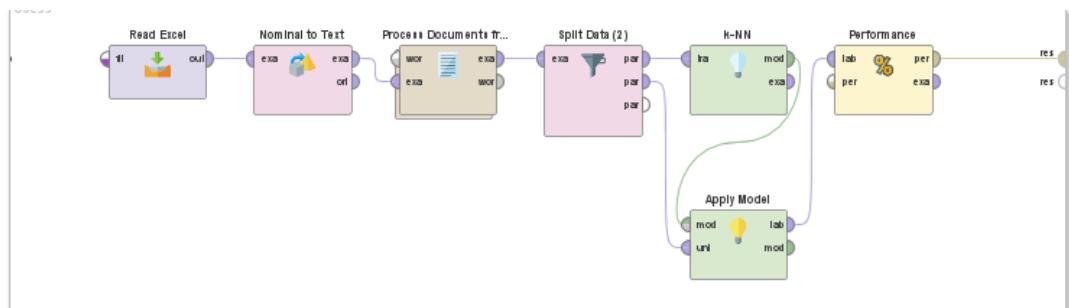
- a. Berikut *Modelling* menggunakan *split* data Pada tahapan ini maka data akan di *split* menjadi 3 model yaitu 60:40, 70:30, dan 80:20.

- b. *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 memasukan 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.

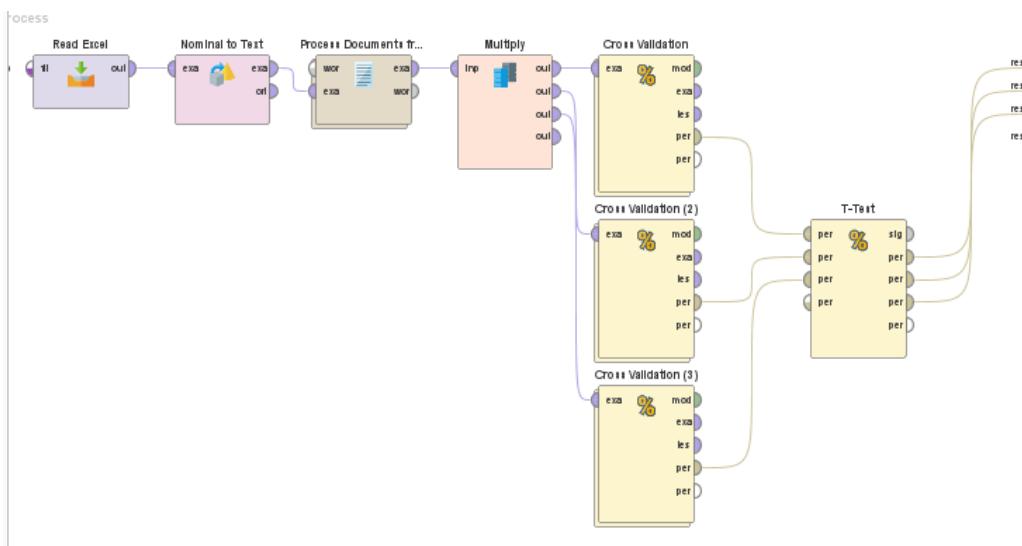


Gambar 2. Penerapan Algoritma Dengan Model *Split Data*

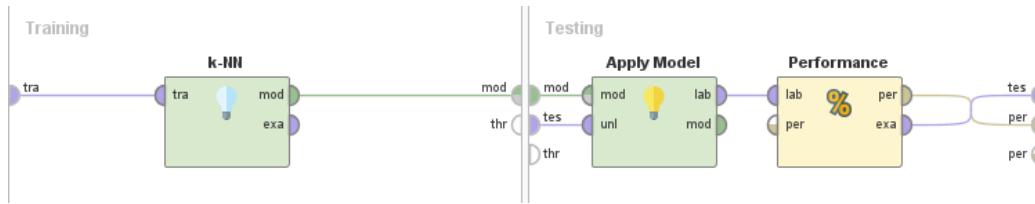
### 3.4.2 Pemodelan Menggunakan Cross Validation

Berikut ini adalah penjelasan algoritma K-Nearest Neighbors menggunakan Cross Validation (Gambar3.) dan (Gambar 4.):

- 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 memasukan 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 bersarang yang memiliki dua subproses yaitu subproses training dan subproses testing. Subproses training digunakan untuk melatih model. model yang dilatih kemudian diterapkan dalam subproses testing, tahap akhir 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 yaitu 'sarana 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.



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,Naïve Bayes dan SVM Menggunakan Metode Split Data

Perbandingan	Algoritma		
	<i>K-Nearest Neighbors</i>	<i>Naïve Bayes</i>	<i>Support Vector Machine</i>
60:40	94.07%	64.16%	84.38%
70:30	94.24%	64.68%	84.29%
80:20	93.86%	62.34%	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,Naïve Bayes dan SVM Menggunakan Metode *Cross Validation*

Algoritma	Accuracy
<i>K-Nearest Neighbors</i>	91.30%
<i>Naïve Bayes</i>	63.28%
<i>Support Vector Machine</i>	84.36%

Berdasarkan Tabel 3 dan Tabel 4 yang berisikan hasil pengukuran performa model klasifikasi *k-nearest neighbors* dengan 2 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.

### 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 evaluasi pengukuran klasifikasi *K-Nearest Neighbors* dengan *dataset* yang berjumlah 3588, terdapat 2871 data training, dan 737 data testing.

**Tabel 5. Confusion Matrix**

	<b>True positive</b>	<b>True negative</b>	<b>Class precision</b>
<b>Pred. positive</b>	874	7	99.21%
<b>Pred. negative</b>	32	163	83.59%
<b>Class recall</b>	96.47%	95.88%	

Perhitungan dari Tabel 5 adalah sebagai berikut :

- $Precision = \frac{TP}{TP+FP} = \frac{874}{874+7} = 0,9921 = 99.21\%$
- $Recall = \frac{TP}{TP+FN} = \frac{874}{874+32} = 0,9647 = 96.47\%$
- $Accuracy = \frac{TP+TN}{TP+TN+FP+FN} = \frac{874+163}{874+163+7+32} = 0,9638 = 96.38\%$

## 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 social *twitter* pada tanggal 17-27 Maret yang sudah dilakukan *preprocessing* dan pemberian lebel sentiment dengan jumlah dataset yang digunakan 3588 tweet menghasilkan data tweet dengan sentimen positif sebanyak 3022, lalu data tweet dengan sentimen 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 91.30%.

## DAFTAR PUSTAKA

- [1] Rasbin, “Dampak Ekonomi Dari Pergelaran MotoGP Mandalika 2022 : Jangka Pendek dan Jangka Panjang,” *Info Singkat, Pus. Penelit. Badan Keahlian DPR RI*, vol. 14, no. 6, pp. 13–18, 2022.
- [2] N. Tri Romadloni, I. Santoso, and S. Budilaksono, “Perbandingan Metode Naive Bayes, KNN, dan Decision Tree Terhadap Analisis Sentimen Transportasi KRL Commuter Line,” *J. IKRA-ITH Inform.*, vol. 3, no. 2, pp. 1–9, 2019.
- [3] K. Kunci, “Analisis Pemberdayaan Masyarakat Dalam Pengembangan Pariwisata Kuta Mandalika 1 Fakultas Ekonomi dan Bisnis Universitas Mataram \* Corresponding Author : diswandi@unram.ac.id Info Artikel Lombok Tengah merupakan salah satu daerah tujuan wisata yang cukup s,” vol. 4, no. 1, 2022.
- [4] R. H. Satrio and M. A. Fauzi, “Klasifikasi Tweets Pada Twitter Menggunakan Metode K-Nearest Neighbour ( K-NN ) Dengan Pembobotan TF-IDF,” vol. 3, no. 8, pp. 8293–8300, 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 Jasa Ekspedisi Barang J&T Ekpress Indonesia),” *Pros. SINTAK*, pp. 480–486, 2018.
- [6] A. D. Adhi Putra, “Analisis Sentimen pada Ulasan pengguna Aplikasi Bibit Dan Bareksa dengan Algoritma KNN,” *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 8, no. 2, pp. 636–646, 2021, doi: 10.35957/jatisi.v8i2.962.
- [7] S. Ernawati and R. Wati, “Penerapan Algoritma K-Nearest Neighbors Pada Analisis Sentimen Review Agen Travel,” *J. Khatulistiwa Inform.*, vol. 6, no. 1, pp. 64–69, 2018.
- [8] M. S. Alrajak, I. Ernawati, and I. Nurlaili, “Analisis Sentimen Terhadap Pelayanan PT PLN di Jakarta pada Twitter dengan Algoritma K- Nearest Neighbor (K-NN),” *Semin. Nas. Mhs. Ilmu Komput. dan Apl.*, pp. 110–122, 2020.
- [9] A. Supriatman, “Pembobotan TF-IDF pada Judul Penelitian Dosen Sebagai Dasar Klasifikasi Menggunakan Algoritma K-NN (Studi Kasus: Universitas Siliwangi),” *J. Serambi Eng.*, vol. 6, no. 1, pp. 1573–1579, 2021, doi:

10.32672/jse.v6i1.2645.

- [10] F. Sodik and I. Kharisudin, “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.