**SENTIMENT ANALYSIS WSBK MANDALIKA MENGGUNAKAN NAMED ENTITY RECOGNITION**

**Dimas Agung Gumelar1, M. Hafidh Ilmi Nafian2, Lutfi Mitra Tanjung3**

1Universitas Brawijaya

2Universitas Brawijaya

3Universitas Brawijaya

Email: 1dagumelar27@student.ub.ac.id, 2@student.ub.ac.id, 3lutfimt@student.ub.ac.id

\*Penulis Korespondensi

(Naskah masuk: dd mmm yyyy, diterima untuk diterbitkan: dd mmm yyyy)

**Abstrak**

WSBK (World Superbike) merupakan ajang balap motor yang menggunakan mesin motor komersil. Ajang ini berlangsung di Mandalika, Lombok Tengah, Nusa Tenggara Barat (NTB). Dataset WSBK Mandalika didapatkan dari proses stream langsung menggunakan API Twitter. Proses ini memanfaatkan library Kafka. Dataset yang didapatkan memuat beberapa informasi terkait *tweet* dari pengguna Twitter yang terdiri dari tanggal, topik, id pengguna, *username*, *name*, isi *tweet*, jumlah *like*, jumlah *reply*, jumlah *retweet*, dan status *retweet*. Dataset yang diperoleh disimpan ke dalam database yang selanjutnya dilakukan proses machine learning. Dataset ini akan dianalisis menggunakan algoritma NER *(Named Entity Recognition)*. NER adalah algoritma yang dapat mengenali orang dan tempat serta dapat digunakan untuk melakukan *sentiment analysis*. Isi tweet pengguna akan diklasifikasikan ke dalam salah satu kategori yaitu kategori positif atau negatif. Hasil dari proses analisis menunjukkan bahwa tweet yang memiliki sentimen positif lebih banyak daripada sentimen negatif.

**Kata kunci**: *WSBK Mandalika, API Twitter, Kafka, Sentiment Analysis, NER (Named Entity Recognition)*

***SENTIMENT ANALYSIS OF WSBK MANDALIKA USING NAMED ENTITY RECOGNITION***

***Abstract***

*WSBK (World Superbike) is a motorcycle racing event that uses commercial motorcycle engines. This event took place in Mandalika, Central Lombok, West Nusa Tenggara (NTB). The WSBK Mandalika dataset is obtained from a live stream process using the Twitter API. This process makes use of the Kafka library. The dataset obtained contains some information related to tweets from Twitter users consisting of date, topic, user id, username, name, tweet content, number of likes, number of replies, number of retweets, and retweet status. The dataset obtained is stored in a database which is then carried out by machine learning processes. This dataset will be analyzed using the NER (Named Entity Recognition) algorithm. NER is an algorithm that can recognize people and places and can be used to perform sentiment analysis. The contents of the user's tweet will be classified into one of the categories, namely the positive or negative category. The results of the analysis process show that tweets that have more positive sentiments than negative sentiments.*

**Keywords**: *WSBK Mandalika, API Twitter, Kafka, Sentiment Analysis, NER (Named Entity Recognition)*

# PENDAHULUAN

WSBK (World Superbike) merupakan ajang balap motor yang menggunakan mesin motor komersil. Indonesia sukses menyelenggarakan ajang ini yang mana berlangsung di Mandalika, Lombok Tengah, Nusa Tenggara Barat (NTB) pada tanggal 19-21 November 2021. Hal tersebut ramai menjadi perbincangan masyarakat Indonesia hingga di media sosial khususnya Twitter.

Dataset WSBK Mandalika didapatkan dari proses stream langsung menggunakan API Twitter. Proses ini memanfaatkan library Kafka. Dataset yang didapatkan memuat beberapa informasi terkait *tweet* dari pengguna Twitter yang terdiri dari tanggal, topik, id pengguna, *username*, *name*, isi *tweet*, jumlah *like*, jumlah *reply*, jumlah *retweet*, dan status *retweet*. Dataset yang diperoleh disimpan ke dalam database yang selanjutnya dilakukan proses machine learning.

Dataset ini akan dianalisis menggunakan algoritma NER *(Named Entity Recognition)*. NER adalah algoritma yang dapat mengenali orang dan tempat serta dapat digunakan untuk melakukan *sentiment analysis*. Sentiment Analysis adalah sebuah proses yang menggunakan text analysis untuk mendapatkan data dari berbagai sumber yang biasanya data tersebut merupakan opini dari berbagai pengguna. Isi tweet pengguna akan diklasifikasikan ke dalam salah satu kategori yaitu kategori positif atau negatif. Hasil dari proses analisis akan menunjukkan perbandingan jumlah tweet yang memiliki sentimen positif dan sentimen negatif.

# DATASET WSBK MANDALIKA

Dataset WSBK Mandalika didapatkan dari proses stream langsung menggunakan API Twitter. Proses ini memanfaatkan library Kafka dimasukkan ke dalam producer dengan topik 'topik-wsbk-mandalika'. Pada penelitian ini didapatkan 80 baris data berbeda. Dataset ini memuat beberapa informasi terkait *tweet* dari pengguna Twitter yang terdiri dari tanggal, topik, id pengguna, *username*, *name*, isi *tweet*, jumlah *like*, jumlah *reply*, jumlah *retweet*, dan status *retweet*. Dataset yang diperoleh diambil dari consumer Kafka dan disimpan ke dalam database menggunakan library sqlite3 ke dalam tabel ‘mandalika’ di dalam database ‘wsbk.db’. Data ini selanjutnya dilakukan proses machine learning.

# DATA VALIDATION

Sebelum dilakukan proses machine learning, dilakukan validasi terhadap dataset WSBK. Proses validasi ini menggunakan metode *Great Expectations*. Dataset diubah ke dalam bentuk file berformat csv. Kemudian dilakukan proses instalasi dan inisialisasi Great Expectations. Hasil kesuksesan dari proses validasi data adalah 100%. Rincian hasil validasi terdapat pada tabel 1, tabel 2, tabel 3, dan tabel 4.

Tabel 1. Statistik dataset WSBK

| Evaluated Expectations | | | 11 | |
| --- | --- | --- | --- | --- |
| Successful Expectations | | | 11 | |
| Unsuccessful Expectations | | | 0 | |
| Success Percent | | | 100% | |

Tabel 2. Table-Level Expectations

| **Status** | **Expectation** | | | **Observed Value** | |
| --- | --- | --- | --- | --- | --- |
| v | Must have greater than or equal to 64 and less than or equal to 79 rows. | | | 72 | |
| v | Must have exactly 11 columns | | | 11 | |
| v | Must have these columns in this order: Unnamed: 0, tanggal, topic, id, username, name, tweet, like\_count, reply\_count, retweet\_count, retweeted | | | ['Unnamed: 0', 'tanggal', 'topic', 'id', 'username', 'name', 'tweet', 'like\_count', 'reply\_count', 'retweet\_count', 'retweeted'] | |

Tabel 3. Unnamed: 0 Expectations

| **Status** | **Expectation** | | | **Observed Value** | |
| --- | --- | --- | --- | --- | --- |
| v | values must never be null. | | | 100% not null | |
| v | minimum value must be greater than or equal to -1 and less than or equal to 1. | | | 0 | |
| v | maximum value must be greater than or equal to 70 and less than or equal to 72. | | | 71 | |
| v | mean must be greater than or equal to 34.5 and less than or equal to 36.5. | | | 35.5 | |
| v | median must be greater than or equal to 34.5 and less than or equal to 36.5. | | | 35.5 | |
| v | quantiles must be within the following value ranges.   | Quantile | Min Value | Max Value | | --- | --- | --- | | 0.05 | 3 | 5 | | Q1 | 17 | 19 | | Median | 35 | 37 | | Q3 | 52 | 54 | | 0.95 | 66 | 68 | | | | | Quantile | Value | | --- | --- | | 0.05 | 4 | | Q1 | 18 | | Median | 36 | | Q3 | 53 | | 0.95 | 67 | | |

Tabel 4. Tanggal Expectations

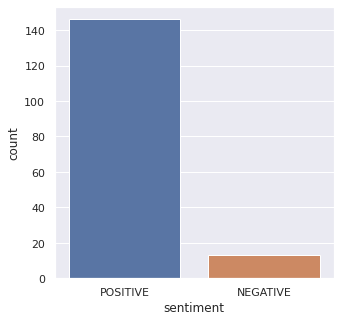
| **Status** | **Expectation** | | | **Observed Value** | |
| --- | --- | --- | --- | --- | --- |
| v | values must never be null. | | | 100% not null | |
| v | values must always be greater than or equal to 1 characters long. | | | 0% unexpected | |

# ALGORITMA NER (Named Entity Recognition)

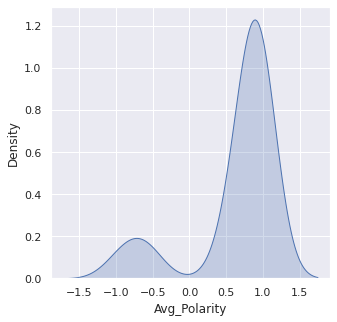
NER

# HASIL MODELING

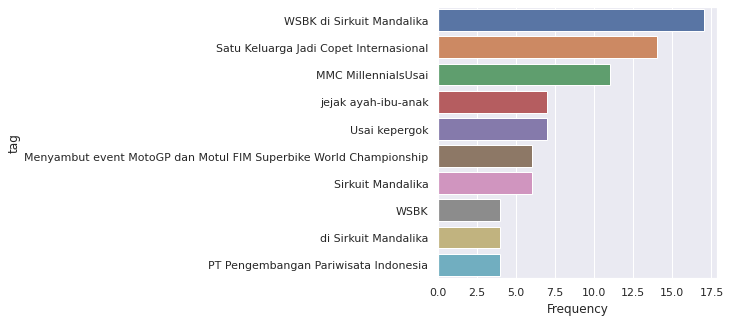
NER



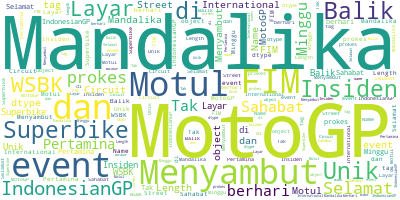
Gambar 1. Grafik jumlah sentimen positif dan negatif



Gambar 2. Grafik area polarity



Gambar 3. Grafik peringkat 10 tag terbanyak



Gambar 4. Word cloud dataset WSBK

# DAFTAR PUSTAKA

Ismail, A. M., 2018. Medium. [Online]

Available at: https://medium.com/bee-solution-partners/cara-kerja-algoritma-k-nearest-neighbor-k-nn-389297de543e

[Accessed 17 12 2021].

BAKTI KOMINFO, 2021. INFORMASI TENTANG TEKNOLOGI GYROSCOPE. [Online]

Available at: https://www.baktikominfo.id/id/informasi/pengetahuan/informasi\_tentang\_teknologi\_gyroscope\_fungsi\_dan\_cara\_kerjanya\_yang\_wajib\_dibaca-780

[Accessed 17 December 2021].

BAKTI KOMINFO, 2021. PENGERTIAN, FUNGSI DAN KELEBIHAN ACCELEROMETER. [Online]

Available at: https://www.baktikominfo.id/id/informasi/pengetahuan/pengertian\_fungsi\_dan\_kelebihan\_accelerometer\_yang\_tak\_banyak\_orang\_ketahui-785

[Accessed 17 December 2021].

Yanuar, A., 2018. Random Forest. [Online]

Available at: https://machinelearning.mipa.ugm.ac.id/2018/07/28/random-forest/

[Accessed 17 December 2021].