

# ANALISIS SENTIMEN KOMENTAR PADA KANAL YOUTUBE THE LAZY MONDAY MENGGUNAKAN ALGORITMA NAÏVE BAYES

Egi Fransisco Saputra<sup>1</sup>, Muhammad Rizky Pribadi<sup>2</sup>

<sup>1,2</sup> Program Studi, Informatika, Fakultas Ilmu Komputer & Rekayasa, Universitas Multi Data Palembang.  
<sup>1</sup>siegik@mhs.mdp.ac.id, <sup>2</sup>Rizky@mdp.ac.id.

---

## Kata kunci:

naïve bayes; oversampling; sentimen; tf-idf ; youtube.

---

---

**Abstract:** Youtube is one of the video platforms and is the largest source of information worldwide, especially in Indonesia, many videos about product reviews have comments that can cause sentiment in certain classes, based on this, a study was conducted using the Naïve Bayes method to analyze public sentiment on YouTube Game Review video comment data on The lazy monday channel. data is taken through crawling and labeling and preprocessing is carried out and then ektarki using TF-IDF and K-fold cross validation is carried out to get the best validation and naive bayes modeling. From the tests conducted, it shows that imbalanced data only gets 45% accuracy performance on three data labels. While in the case of imbalance data, by oversampling SMOTE, an accuracy of 67% is obtained in three data labels.

**Abstrak:** Youtube adalah salah satu platform video dan menjadi sumber informasi terbesar di seluruh dunia terutama di Indonesia, banyak video yang mengenai review produk memiliki komentar yang dapat menimbulkan sentimen pada kelas tertentu, Berdasarkan hal tersebut maka dilakukan penelitian menggunakan metode Naïve Bayes untuk menganalisis sentimen masyarakat terhadap data komentar video Review Game youtube di kanal The lazy monday. data diambil melalui crawling dan dilakukan labelisasi dan preprocessing lalu di ektarki menggunakan TF-IDF dan dilakukan K-fold cross validation guna mendapatkan best validation serta pemodelan naive bayes. Dari pengujian yang dilakukan, menunjukkan bahwa data yang imbalance hanya mendapatkan performa akurasi 45% pada tiga label data. Sedangkan pada kasus data imbalance, dengan melakukan oversampling SMOTE didapatkan akurasi sebesar 67 % dalam pada tiga label data.

---

Saputra dan Pribadi. (2023). Analisis Sentimen Komentar Pada Kanal Youtube The Lazy Monday Menggunakan Algoritma Naïve Bayes. *MDP Student Conference 2023*

---

## PENDAHULUAN

Youtube adalah platform video berbagi yang populer yang mana para penggunanya bisa berinteraksi, menonton, kementar hingga berbagi video secara gratis. Banyak jenis video yang tentunya mampu pengguna upload di dalam platform Youtube. Salah satu contoh jenis video tersebut ialah video yang membahas suatu produk tertentu. Banyak nya produk yang di Bahasa atau di review menimbulkan komentar yang tidak relevan yakni sentimen publik yang berbeda beda[1].

The lazy monday merupakan jurnalis media yang aktif dan bergerak dalam game analisis di indonesia, The lazy monday mampu mengenalkan game melalui sudut pandang berbeda dari narasi yang

dibuat melalui video di Youtube. Hal yang paling penting ialah *point of view* dari the lazy Monday sehingga video yang mereka unggah dapat membuat masyarakat berkomentar dengan komentar yang berbeda beda.[2]

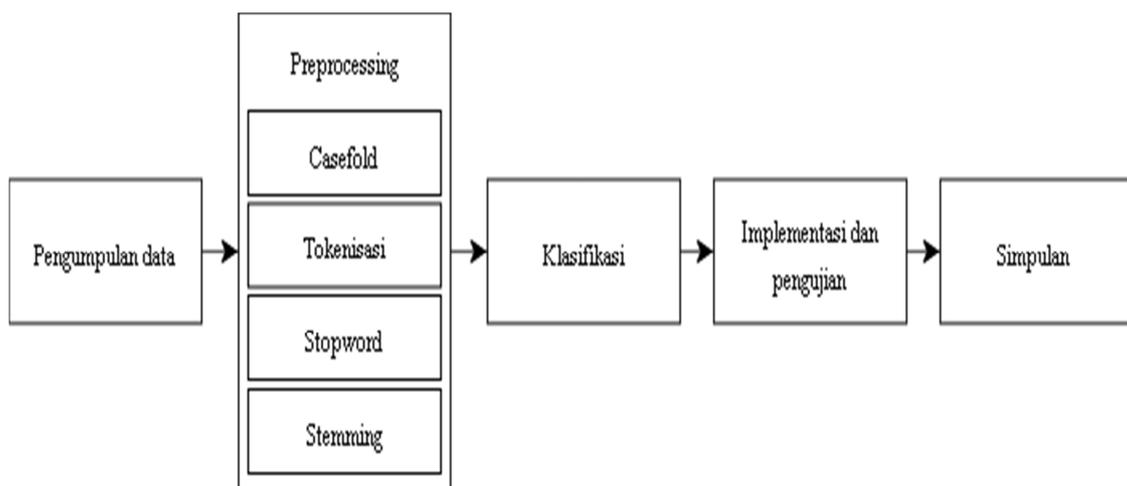
Tentunya dapat menimbulkan banyak sentimen yang berbeda beda juga. Salah satu cara melakukan pengalisan sentiment adalah dapat dilakukan dengan menggunakan data dari media sosial seperti youtube[3]. dengan mengumpulkan pendapat atau sentimen dari data dari media sosial tentunya kita dapat membedakan sentimen mana yang bersifat positif yang secara definisi pendapat yang baik dari suatu konteks, sentiment negatif yang mengungkapkan pendapat buruk dari suatu konteks dan netral yang menyatakan hal hal yang tidak mendukung baik atau buruk[4].Berhubungan Dalam Melakukan analisis, data komentar yang dianalisis selanjutnya dilakukan klasifikasi sehingga data dilihat yang menjadi keunggulan dan kelemahan dari suatu produk. Kegiatan Analisis sentimen juga tak luput dari sebuah istilah *text mining*. *Text mining* merupakan sesuatu yang mencoba menggali hingga memecahkan informasi masalah kelebihan beban dengan menggunakan beberapa Teknik dari *data mining*, *machine learning*, *Natural Preprocessing* dan sebagainya[5].

Pada penelitian[3] Klasifikasi Opini Masyarakat Terhadap Pertanian di Indonesia dengan Naïve Bayes pada Twitter menyimpulkan bahwa dalam penerapan naïve bayes dapat menghasilkan akurasi dengan 0.523% dan dalam Penelitian[6] yang berjudul Sistem Analisis Sentimen pada Ulasan Produk Menggunakan Metode Naïve Bayes menyimpulkan jika naïve bayes dapat dengan mudah mengklasifikasi sentiment pada ulasa produk sesuai dengan yang telah di sesuaikan dengan mendapatkan akurasi sebesar 77.78 %. Algoritma *naïve bayes* adalah algoritma yang mampu membuat tingkat akurasi dalam penganalisisasn jauh lebih baik serta *naïve bayes* juga dapat dikenal dengan baiknya dalam mengklasifikasi teks dengan jumlah data kecil atau sebuah dokumen[7]

Fokus dalam penelitian ini adalah untuk mengklasifikasikan serta menganalisis komentar pada kanal youtube dengan menggunakan algoritma *naïve bayes*.

## METODE

Metodelogi adalah tahapan proses yang sudah tersusun secara sistematis pada melakukan suatu penelitian yang berguna untuk menggapai tujuan yang di harapkan. Tahapan yang akan dilakukan pada penelitian ini yaitu pada gambar 1.



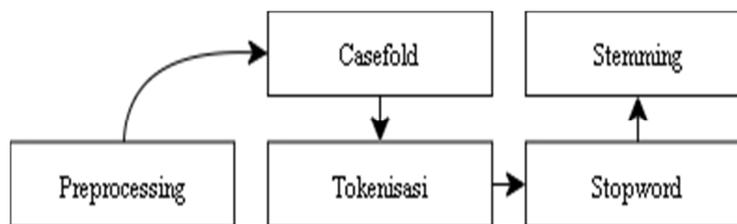
Gambar 1. Tahapan Metode

**Pengumpulan Data**

Pada tahap ini, akan dilakukan khususnya pada penelitian ini yakni dengan dilakukannya sebuah Teknik pengumpulan data dengan dua tahap, yaitu *crawling* dan *labelisasi*. Tahapan *crawling* yaitu sebuah proses dilakukannya pengambilan data komentar dari youtube[8]. lebih spesifiknya data komentar pada video pada kanal youtube the lazy monday dengan judul “Review Apex Legends – Battle Royale harusnya begini, sempurna | lazy review”. Pengumpulan atau pengambilan data menggunakan sebuah metode *crawling* pada Bahasa *python* dan akan di dapatkan 1000 komentar yang telah *import* dalam bentuk *csv*. Setelah berhasil mendapatkan dataset yang diinginkan yaitu 1000 data komentar, maka data akan dilakukan proses tahap kedua yaitu *labelisasi*. Proses ini memberikan data yang telah didapatkan sebelumnya dengan memberikan label pada data yang telah di dapatkan sebelumnya. Dengan memberikan label apakah data didalamnya itu mempunyai *value* positif, negatif dan netral berdasarkan makna kata atau kalimat pada dataset tersebut. Proses *labelisasi* data menggunakan pakar Bahasa Indonesia. Agar data tersebut dapat mendapatkan nilai yang lebih baik dan demi mencapai tujuan penelitian.

**Preprocessing**

Pada tahap ini *preprocessing* atau *text preprocessing* yakni proses memberikan serta membuat data gar lebih terstruktur dan bersih sebelum diolah nantinya[9]. Pada tahapan ini terbagi mendai 4 bagian pada gambar 2.



**Gambar 2. Preprocessing**

**Casefold**

Pada tahap ini mempunyai tujuan untuk membuat data terstruktur dan menyeragamkan teks ke dalam bentuk *lowercase*. Proses pada tabel 1.

**Tabel 1. Proses Casefold**

Sebelum <i>casefold</i>	Real Combat Bukan Keliling Map Gak Jelas
Sesudah <i>casefold</i>	real combat bukan keliling map gak jelas

**Tokenisasi**

Pada tahap ini mempunyai tujuan untuk memecahkan kalimat. Singkatnya dilakukan pemecahan kalimat menjadi kata. Proses pada tabel 2.

**Tabel 2. Proses Tokenisasi**

Sebelum Tokenisasi	real combat bukan keliling map gak jelas
Sesudah Tokenisasi	[real, combat, bukan, keliling, luas, jelas]

### Stopword

Pada tahap ini dilakukan sebuah proses penghilangan kata. singkatnya *stopword* ialah kata mempunyai kemunculan yang lebih tentunya dianggap tidak ada arti. Layaknya kata hubung. Proses pada tabel 3.

**Tabel 3. Proses Stopword**

Sebelum <i>Stopword</i>	[real, combat, bukan, keliling, luas, jelas]
Sesudah <i>Stopword</i>	[real, combat, keliling, luas]

### Stemming

Pada tahap ini dilakukan dan mempunyai tujuan untuk mencari kata dasar, singkatnya dengan menghilangkan semua imbuhan yang telah menyatu kata. Proses pada tabel 4

**Tabel 4. Proses Stemming**

Sebelum <i>Stemming</i>	[real, combat, keliling, luas]
Sesudah <i>Stemming</i>	[real, combat, keliling, luas]

### Klasifikasi

Pada Tahap ini dilanjutkan dengan proses kalsifikasi menggunakan metode yang akan digunakan. Adapun metode yang digunakan yaitu metode *naïve bayes*. Tahap ini data yang sebelumnya telah melewati proses *crawling* dan labelisasi akan diproses menggunakan metode *naïve bayes*. *Naïve bayes* dikembangkan oleh Reverend Thomas Bayes pada abad 18-an. *Naïve bayes* juga dikenal sebagai pengklasifikasian yang dapat diandalkan dan digunakan untuk memprediksi suatu *class*. khususnya pada tahap ini diharapkan *machine* akan dapat mengenali pola yang selanjutnya dilakukan pengklasifikasi ke dalam tiga sentimen atau kelas tertentu[10]. Layaknya positif, negatif dan netral. bentuk dari *naïve bayes* ditunjukkan pada nomor 1.

$$P(X) = \frac{P(H|X) \cdot P(H)}{P(H)} \quad (1)$$

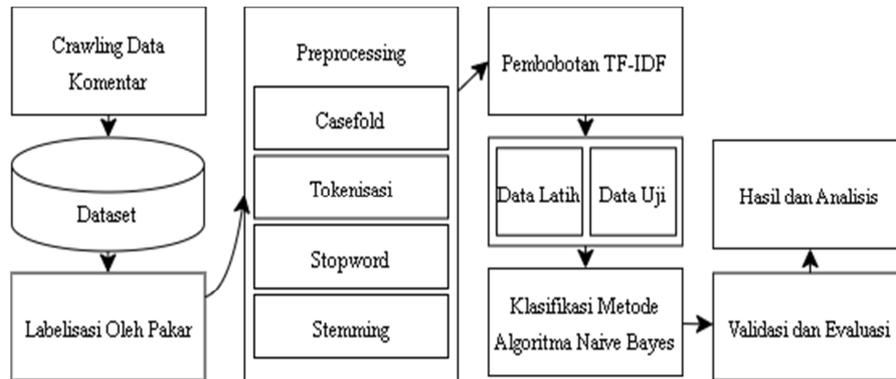
Dimana:

- $X$  : Data dengan *class* yang belum diketahui
- $H$  : Hipotesis data  $X$  merupakan suatu *class* spesifik
- $P(H|X)$  : Probabilitas hipotesis  $H$  berdasarkan kondisi  $X$
- $P(H)$  : Probabilitas hipotesis  $H$
- $P(X|H)$  : Probabilitas  $X$  berdasarkan kondisi pada hipotesis  $H$
- $P(X)$  : Probabilitas  $X$

Tahapan untuk perhitungan *naïve bayes* ialah pertama *naïve bayes classifier* merupakan menghitung probabilitas setiap *class* seluruh data latih dan pada data uji, proses ini dilakukan untuk mengetahui keakuratan model yang dibangun pada proses latih.

### Perancangan Sistem

Pada tahap ini, perancangan sistem akan dibangun berdasarkan pengamatan sebelumnya. Beberapa rancangan ini dilakukan dengan membangun serta merancang dan tentunya menghasilkan *output* yang baik. Sehingga metode yang nantinya dapat dengan mudah dalam menerapkan metode tersebut. Alur rancangan pada gambar 3.



Gambar 3. Rancangan Sistem

**Implementasi dan Pengujian**

Pada tahap ini, penelitian menggunakan Bahasa pemrograman *python* yang dijalankan pada platform *Google Collab*. Proses ini akan dilakukan untuk proses *crawling*, *preprocessing*, Pembobotan TF-IDF, menghitung nilai pada *K-fold Cross validation* dan mengindetifikasi metode *naive bayes* terhadap *machine* yang telah berhasil di bangun menggunakan [11] *Confusion matrix* dengan menghitung *accuracy*, *precision*, dan *recall*. Pada nomor 2, 3 dan 4.

$$Accuracy = (TP + TN) / (TP+FP+FN+TN) \tag{2}$$

$$Precision = (TP) / (TP + FP) \tag{3}$$

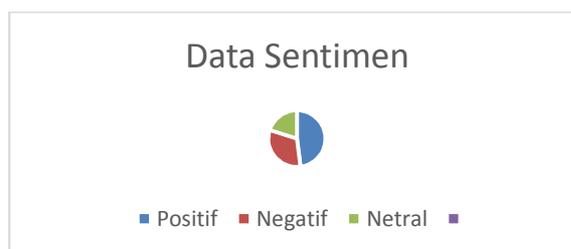
$$Recall = (TP) / (TP + FN) \tag{4}$$

Dimana:

- TP = *True Positif Count*.
- TN = *True Negative Count*.
- FP = *False Positif Count*.
- FN = *False Negative Count*.

**HASIL DAN PEMBAHASAN**

Pada tahap *preprocessing* didalam data didapatkan komentar yang ditulis oleh para penonton di video youtube. Pada komentar ini terdapat tiga pembagian komentar yang telah dilakukan pada proses labelisasi oleh pakar Bahasa Indonesia dengan kategori positif, negatif dan netral. data yang terdiri dari 1000 komentardari video dengan judul “*Review Apex Legends – Battle Royale harusnya begini, sempurna | lazy review*” men dapatkan komentar positif sebanyak 48 % positif, 32 % negatif dan 20 % netral, pada gambar 4.



Gambar 4. Perbandingan Sentimen

Pada analisis sentiment ini dilakukan 2 skenario. Skenario pertama yaitu menggunakan *K-fold cross validation* dengan iterasi 10 dengan alokasi data imbalance. Skenario kedua menggunakan *K-fold cross validation* dengan iterasi 10 dengan alokasi data menggunakan oversampling SMOTE pada data imbalance. Tabel 5 skenario pertama dan tabel 6 skenario kedua.

**Tabel 5. Skenario Pertama**

No	Data	Accuracy	Precession	Recall
1	Positif	45 %	42%	50%
2	Negatif		36%	57%
3	Netral		58%	35%

**Tabel 6. Skenario Kedua**

No	Data	Accuracy	Precession	Recall
1	Positif	67 %	62%	71%
2	Negatif		67%	93%
3	Netral		76%	35%

Berdasarkan kedua skenario yang telah dilakukan sebelumnya, kedua skenario tersebut menghasilkan *accuracy* yang paling besar yaitu 67 % dibandingkan dengan skenario pertama.

## SIMPULAN

Berdasarkan hasil pengujian dan penelitian ini maka analisis sentiment komentar pada kanal youtube The lazy monday dengan melakukan pengujian atau skenario sebanyak dua kali yaitu pada data imbalance dan pada metode oversampling SMOTE. mendapatkan sebuah konklusi jika data yang mulanya mempunyai ketidakseimbangan dapat mempengaruhi akurasi yang di dapatkan. melalui serangkaian pengujian dengan menggunakan *K-fold cross validation* agar mendapatkan *best validation* yang lebih baik, data *imbalance* pada pengujian pertama hanya mendapatkan tingkat akurasi sebesar 45% dengan precession positif 42%, negatif 36% dan 58%, Recall positif 50%, negatif 57% dan netral 35%. Pada pengujian atau skenario kedua dengan menggunakan data *imbalance* yang sama namun di iringin dengan metode *oversampling SMOTE* dengan melakukan *K-fold cross validation* untuk mendapatkan *best validation* dapat dihasilkan rata rata akurasi yang di dapatkan lebih baik yakni sebesar 67 % precession positif 62%, negatif 67 % dan 76%, Recall positif 71%, negatif 93% dan netral 35%.

Berdasarkan penelitian ini diharapkan dilakukannya penelitian lainnya pada komentar di kanal youtube menggunakan metode lainnya sehingga diharapkan dapat memberikan nilai akurasi yang lebih optimal. Dengan adanya analisis sentiment ini dapat memberikan informasi mengenai hasil sentiment masyarakat mengenai video pada kanal the lazy monday yang dapat digunakan sebagai masukan agar selalu memberikan video yang terbaik untuk masyarakat.

## DAFTAR PUSTAKA

- [1] E. Yosef, S. Sihombing, and B. T. Hanggara, “Pemanfaatan API Youtube Dalam Pengembangan Aplikasi Portal Video Penangkaran Kenari Untuk Peternak Kenari Berbasis Android,” Vol. 4, No. 7, pp. 2067–2074, 2020.
- [2] W. M. Adiputra, “Jurnalisme Game dan Membaurnya Karakter Konten Media Studi Kasus The Lazy Monday, Gamebrott dan Playverse,” *J. Commun. Spectr.*, Vol. 8, No. 2, pp. 166–189, 2019, doi: 10.36782/jcs.v8i2.1853.
- [3] H. Irsyad and M. R. Pribadi, “Klasifikasi Opini Masyarakat Terhadap Pertanian di Indonesia Dengan Naive Bayes pada Twitter,” *J. Tek. Inform. Unika St. Thomas*, Vol. 4, No. 1, pp. 89–98, 2019.
- [4] K. A. B. Permana, M. Sudarma, and W. G. Ariastina, “Analisis Rating Sentimen pada Video di Media Sosial Youtube Menggunakan STRUCT-SVM,” *Maj. Ilm. Teknol. Elektro*, Vol. 18, No. 1, p. 113, 2019, doi: 10.24843/mite.2019.v18i01.p17.
- [5] A. S. Hilda Rachmi, “Classification of Text Mining Review Oil Disfusser Products Using Naive Bayes Classification,” *J. Mantik*, Vol. 4, No. 1, pp. 187–192, 2020, [Online]. Available: <https://iocscience.org/ejournal/index.php/mantik/index>.
- [6] B. Gunawan, H. S. Pratiwi, and E. E. Pratama, “Sistem Analisis Sentimen pada Ulasan Produk Menggunakan Metode Naive Bayes,” *J. Edukasi dan Penelit. Inform.*, Vol. 4, No. 2, p. 113, 2018, doi: 10.26418/jp.v4i2.27526.
- [7] K. Chai, H. T. Hn, and H. L. Cheiu, “Naive-Bayes Classification Algorithm,” *Proc. 25th Annu. Int. ACM SIGIR Conf. Res. Dev. Inf. Retr.*, pp. 97–104, 2002.
- [8] D. Lab, “Crawling Sebagai Teknik Pengumpulan Data Sekunder,” *Dqlab.Com*. 2021, [Online]. Available: <https://www.dqlab.id/crawling-sebagai-teknik-pengumpulan-data-sekunder#:~:text=Ya%2C web crawling adalah teknik,penambahan dataset yang lebih besar>.
- [9] V. A. Fitri, R. Andreswari, and M. A. Hasibuan, “Sentiment Analysis of Social Media Twitter with Case of Anti-LGBT Campaign In Indonesia Using Naïve Bayes, Decision Tree, and Random Forest Algorithm,” *Procedia Comput. Sci.*, Vol. 161, pp. 765–772, 2019, doi: 10.1016/j.procs.2019.11.181.
- [10] L. Ardiani, H. Sujaini, and T. Tursina, “Implementasi Sentiment Analysis Tanggapan Masyarakat Terhadap Pembangunan di Kota Pontianak,” *J. Sist. dan Teknol. Inf.*, Vol. 8, No. 2, p. 183, 2020, doi: 10.26418/justin.v8i2.36776.
- [11] Burhanudin, Y. Musa’adah, and Y. Wihardi, “Klasifikasi Komentar Spam pada Youtube Menggunakan Metode Naïve Bayes, Support Vector Machine, Dan K-Nearest Neighbors,” *J. Inform. dan Komput.*, Vol. 3, No. 2, pp. 54–59, 2018.