Analisis Sentimen Terhadap Pengaruh Metaverse Kepada Masyarakat Menggunakan Metode Naïve Bayes dan K-NN Pada Aplikasi *Twitter*

**Igrid Adelin1, Yusran Timur Samuel2**

Universitas Advent Indonesia; Jl. Kolonel Masturi No.288, Cihanjuang Rahayu,

Kec.Parongpong, Kab. Bandung Barat, (022) 2700163

Fakultas Teknologi Informasi, Universitas Advent Indonesia

e-mail: 11981042@unai.edu, 2y.tarihoran@unai.edu

**Abstrak**

Seiring berkembangnya zaman di dunia, perkembangan teknologi yang terjadi semakin pesat sehingga memberikan kemudahan kepada masyarakat melalui *platform* yang ada saat ini. Salah satunya adalah sebuah *platform* yang dapat membawa dunia nyata ke dalam bentuk digital. Bermula dari sebuah novel berjudul *Snow Crash* yang menceritakan virtual 3D yang dihuni oleh para *Avatar*. Metaverse memungkinkan untuk melakukan banyak hal seperti konser virtual, melakukan perjalanan *online* dan lain sebagainya. Sejak datangnya pandemi *Covid-19* tentunya pengimplementasian metaverse mengundang masyarakat Indonesia beropini. Ribuan opini disampaikan oleh banyak orang setiap harinya di berbagai media sosial, salah satunya adalah *Twitter*. Penelitian ini bertujuan untuk mengetahui akurasi paling tinggi dari dua algoritma yaitu Naïve Bayes dan K-NN menggunakan pembobotan TF-IDF melalui opini masyarakat lebih mengarah ke positif atau negatif dengan jarak waktu tiga bulan semenjak 10 Agustus 2022 sampai dengan 10 November 2022. Dikarenakan hasil penelitian ini mendapat ketidakseimbangan data (*imbalance* data), maka peneliti menggunakan algoritma *Smote* yang menghasilkan akurasi paling tinggi yaitu pada algoritma Naïve Bayes dengan nilai k=5 optimalnya adalah 85,91% dimana 135 kalimat negatif dan 431 kalimat positif yang berarti sentimen masyarakat terhadap Metaverse lebih mengarah ke sppositif.

**Kata Kunci:** Analisis Sentimen, Naïve Bayes, K-NN, *Twitter*, Metaverse, TF-IDF

*Analysis of Sentiment Towards Metaverse to Society on Twitter Application Using Naïve Bayes and K-NN*

***Abstract***

*Along with the development of the times in the world, technological developments are occurring more rapidly so as to provide convenience to the public through the platforms that exist today. One of them is a platform that can bring the real world into digital form. Starting from a novel called Snow Crash which tells a 3D virtual world inhabited by avatars. Metaverse makes it possible to do many things such as virtual concerts, travel online and so on. Since the arrival of the Covid-19 pandemic, of course the implementation of the metaverse has invited the Indonesian people to share their opinions. Thousands of opinions are conveyed by many people every day on various social media, one of which is Twitter. This study aims to determine the highest accuracy of the two algorithms, namely Naïve Bayes and K-NN using TF-IDF weighting through public opinion which is more positive or negative with an interval of three months from 10 August 2022 to 10 November 2022. Due to the results of this research this gets an imbalance of data (data imbalance), so researchers use the Smote algorithm which produces the highest accuracy, namely the Naïve Bayes algorithm with an optimal k = 5 value of 85.91% where 135 sentences are negative and 431 sentences are positive which means that people's sentiment towards Metaverse is more pointing to the positive.*

***Keywords:*** *Sentiment Analysis, Naïve Bayes, K-NN, Twitter, Metaverse, TF-IDF*

# Pendahuluan

Perkembangan teknologi banyak berdampak pada berbagai hal, diantaranya pengembangan aplikasi, cara bertransaksi dan pola hidup masyarakat Indonesia. Tidak sampai disitu saja perkembangan teknologi perlahan-lahan memiliki kemajuan. Dalam sebuah *platform* digital, manusia dapat saling berinteraksi bahkan dapat memiliki penghasilan. Hal tersebut dapat digabungkan dalam sebuah *platform* digital yang dikenal dengan metaverse. Konsep metaverse pertama kali muncul pada tahun 1992 dalam novel fiksi ilmiah *Snow Crash* oleh novelis Amerika Neal Stephenson. Karakter dalam *Snow Crash* yang mempresentasikan Avatar dan bekerja dalam realitas virtual tiga dimensi (3D), dan dunia virtual di mana orang berinteraksi satu sama lain dan lingkungannya tanpa fisik, keterbatasan dunia nyata yang disebut metaverse. Metaverse terdiri dari kata “meta”, berarti transendensi atau keberadaan dan “*verse*”, yang berarti dunia dan alam semesta. Metaverse mengacu pada realitas virtual yang ada diluar realitas. Ribuan opini disampaikan oleh banyak orang setiap harinya di setiap sosial media. Setiap orang dapat memberikan pendapatnya secara bebas, opini-opini tersebut dapat bersifat positif ataupun negatif mengenai suatu topik tertentu [3].

Metaverse menjadi salah topik yang dibahas dalam platform digital Twitter, bagaimana pengaruh dari penerapan metaverse tersebut pada kehidupan sehari-hari serta bagaimana masyarakat menghadapinya dengan melakukan analisis sentimen berdasarkan opini dari *Twitter*. Analisis sentimen ialah suatu metode untuk melihat nilai atau pendapat publik pada objek tertentu bisa dikelompokkan menjadi kalimat yang positif atau negatif [7]. Jadi analisa sentimen merupakan aktivitas dimana peneliti dapat mengindentifikasi pendapat yang ada dari masyarakat terhadap suatu hal yang akan diteliti [2]. *Platform* metaverse menjadi objek penelitian karena kedatangannya membuat banyak pendapat berbagai kalangan masyarakat. Pada akhir bulan Juni 2021, Mark Zuckerberg sebagai pendiri aplikasi *Facebook* mengubah nama aplikasi tersebut dengan meta secara legal [1].

Berdasarkan penelitian dengan mengaplikasikan algoritma Naïve Bayesian *Classification* dalam aplikasi *Twitter* dimana memiliki akurasi sebesar 81.3559% pada aplikasi Peduli Lindungi [4]. Penelitian selanjutnya dengan menerapkan tiga algoritma yaitu K-NN, *Random Forest*, dan Naïve Bayes dalam *Event Flash*. Naïve Bayes menjadi rekomendasi algoritma untuk mengelompokkan data sentimen dengan akurasi lebih besar dan stabil [5]. Penelitian selanjutnya menggunakan Naïve Bayes *Classifier*, K-NN dan *Decision Tree* untuk menelaah sentimen antar hubungan netizen dan pemerintah dimana Naïve Bayes memiliki akurasi yang lebih besar dibandingkan dengan algoritma yang lain [13]. Menurut penelitian dengan memakai algoritma K-NN, *Decision Tree* serta Naïve Bayes untuk implementasi *orange* data *mining* untuk klasifikasi kelulusan mahasiswa dimana menggunakan Naïve Bayes yang memiliki akurasi tertinggi yaitu 89% [14]. Bersumber pada penelitian pada penerapan tiga algoritma K-NN, Naïve Bayes dan *Decision Tree* untuk menganalisa pendapat terhadap PT PAL Indonesia, Naïve Bayes mendapat akurasi lebih tinggi dari yang lain sebesar 84,08% [15]

Minimnya informasi mengenai metaverse menjadi alasan bagi penulis melakukan penelitian ini, yang bertujuan untuk membuat model yang dapat memprediksi sentimen masyarakat terhadap metaverse. Dengan membandingkan kedua algoritma yaitu Naïve Bayes dan K-NN serta pembobotan data menggunakan *Term Frequency-Inverse Document Frequency (TF-IDF)* pada aplikasi *Twitter* diharapkan dapat melihat akurasi sentimen mana yang paling tinggi mengenai metaverse.

# Metode Penelitian

Metode penelitian memiliki langkah-langkah terstruktur yang harus dilakukan untuk menjadi tujuan penelitian. Agar memperoleh hasil yang diinginkan maka harus memiliki langkah-langkah pada Gambar 1 sebagai berikut.

**Gambar 1** Metode Penelitian

## Data *Twitter*

Pada media sosial *Twitter* kita dapat memberikan ulasan atau menyampaikan pesan kepada pemakai aplikasi. Di Twitter kita dapat membuat *tweet* maksimum sebanyak 140 karakter. *Twitter* juga mempersiapkan *Application Programming Interface* (API) untuk menarik data [4]. Elemen yang dapat digunakan sebagai sumber data.

1. *Hashtag* : Tagar yang berkaitan dengan topik tertentu
2. *Replies* : Balasan dari teks pesan
3. *Tweet text* : Isi dari sebuah *tweet*
4. *Retweet* : Membagikan sebuah *tweet*

## *Crawling* Data

Pada penelitian ini, terdapat beberapa tahapan untuk menganalisis data *tweet* *Twitter* [3]. Data yang diperoleh tidak dilakukan secara manual melainkan menggunakan *software* yang bernama *Rapid Miner* dengan tambahan API yang dihubungkan dengan *Twitter* yang terlihat pada Gambar 2. Tahapan *crawling* data dengan pengambilan data *tweet* sejak 10 Agustus 2022 sampai dengan 10 November 2022 data yang diperoleh berjumlah 2.110 data. Pencarian data menggunakan kata kunci metaverse dengan *hashtag* #metaverse.



**Gambar 2** Proses Crawling Data

## *Filtering* Data

Pada tahap ini, proses *filtering* data yaitu menghapus link, *hashtag*, *retweet, tweet* dan lain sebagainya pada sebuah kalimat yang terlihat pada Tabel 1.

**Tabel 1** Proses *Filtering* Data

|  |  |  |
| --- | --- | --- |
| **No** | ***Tweet*** | **Hasil** |
| 1 | Indonesia siap metaverse https://t.co/jiGckXnpCt | Indonesia siap metaverse |
| 2 | @Metacostindo: Penasaran ga, gimana rasanya kitab isa ngampus dan belajar di Metaverse? Bertemu dengan banyak teman baru untuk belajar.#metaverse #foryou | Penasaran ga, gimana rasanya kitab isa ngampus dan belajar di Metaverse? Bertemu dengan banyak teman baru untuk belajar. |

## *Labelling* Data

Pada tahap ini, proses pembagian data menjadi kalimat yang termasuk menjadi negatif atau positif menggunakan metode *crowdsourcing*. Proses pelabelan dilakukan untuk menentukan apakah sebuah *tweet* masuk ke dalam kelas positif yang termasuk pujian, saran, masukan dari emosi positif seperti kepuasan dan kebahagiaan. Sedangkan kelas negatif termasuk ketidakpuasan, sindiran, kritikan dan kekecewaan. Proses *labelling* akan dilakukan oleh empat orang dimana dua orang bertugas mengklasifikasikan secara manual dan dua akan melakukan *cross check* hasil klasifikasi yang dapat dilihat pada Tabel 2. Setelah proses *filtering* data berjumlah sebanyak 566 data dimana 135 kalimat negatif dan 431 kalimat positif. Dikarenakan data tidak seimbang, maka mempengaruhi nilai prediksi akurasi yang baik terhadap kelas data mayoritas dan nilai prediksi akurasi yang buruk pada kelas minoritas, maka penulis menggunakan algoritma *upsampling Smote* dengan mencari k (tetangga) terdekat pada setiap data di kelas minoritas [15].

**Tabel 2** Proses *Labelling data*

|  |  |
| --- | --- |
| **Label** | ***Tweet*** |
| Negatif | Sesungguhnya proyek Metaverse tidak bermanfaat untuk masyarakat |
| Kayaknya manusia nggk butuh metaverse. Dunia nyata aja nggk habis2 di garap. |
| Positif | Metaverse akan membawa kita ke dalam era baru ekonomi |
| ”Metaverse” dan Web3 adalah angin perubahan kuat yang akan mendorong kita ke dunia masa depan yang baru |

## *Pre-processing* Data

Suatu teknik untuk memproses data melalui dimana melalui beberapa tahapan agar data yang dikumpulkan dapat diolah menjadi data yang bersih dan mudah untuk dipahami dan dianalisis.

## Transform Cases

Dalam tahap ini dilakukan proses kesamaan huruf dimulai dari menyamakan huruf yang kapital menjadi huruf kecil yang terlihat pada Tabel 3.

**Tabel 3** Proses *Transform Cases*

|  |  |  |
| --- | --- | --- |
| **No** | ***Tweet*** | **Hasil** |
| 1 | GUE MAU PINDAH KE METAVERSE AJA | gue mau pindah ke metaverse aja |
| 2 | Salah satunya dengan bekerja sama dengan perusahaan asal Korea, Creature Hunters Inc. Kerja sama ini meliputi Metaverse, Mobile Game Production, NFT Blockchain, Creative Entertainment, Animation Production. | salah satunya dengan bekerja sama dengan perusahaan asal korea, creature hunters inc. kerja sama ini meliputi metaverse, mobile game production, nft blockchain, creative entertainment, animation production. |

## Tokenisasi (*Tokenize*)

Dalam tahap ini dilakukan memecah kalimat menjadi beberapa bagian dalam bentuk kata-kata untuk menentukan sekumpulan kata terstruktur pada setiap kata yang terlihat pada Tabel 4.

|  |  |  |
| --- | --- | --- |
| **No** | ***Tweet*** | **Hasil** |
| 1 | gue mau pindah ke metaverse aja | [ gue, mau, pindah, ke, metaverse, aja] |
| 2 | salah satunya dengan bekerja sama dengan perusahaan asal korea, creature hunters inc. kerja sama ini meliputi metaverse, mobile game production, nft blockchain, creative entertainment, animation production. | [ salah, satunya, dengan, bekerja, sama, dengan, perusahaan, asal, korea, creature, hunters, inc, kerja, sama, ini, meliputi, metaverse, mobile, game, production, nft, blockchain, creative, entertainment, animation, production ] |

 **Tabel 4** Proses Tokenisasi

## Stopword

Dalam tahap ini dilakukan proses untuk membuang kata yang tidak memiliki arti atau tidak relevan seperti ‘untuk’, ‘dengan’, ‘lainnya’ , ‘masa’, ‘telah’, ‘dalam’, ‘adalah’, dan lain sebagainya yang terlihat pada Tabel 5. Tahap ini menggunakan suatu file *stopword* bahasa Indonesia, cara kerjanya yaitu jikalau suatu kata tercantum pada kumpulan daftar *stopword* yang digunakan maka kata tersebut akan dihapus dan sebaliknya jikalau suatu kata tidak tercantum pada kumpulan daftar *stopword* maka *rapid miner* akan memproses.

**Tabel 5** Proses *Stopword*

|  |  |  |
| --- | --- | --- |
| **No** | ***Tweet*** | **Hasil** |
| 1 | masa depan khas web 3.0 telah menawarkan terobosan dalam ranah digital yang terkait dengan teknologi blockchain metaverse dan konsep relevan lainnya seperti identitas digital | [ khas, web, 3.0, menawarkan, terobosan, ranah, digital, terkait, teknologi, blockchain, metaverse, konsep, relevan, identitas, digital] |
| 2 | meta, microsoft, adobe, dan epic games termasuk di antara 30+ perusahaan yang bergabung dengan metaverse standards forum untuk menetapkan standar teknologi metaverse | [ meta, microsoft, adobe, epic, games, 30+, perusahaan, bergabung, metaverse, standards, forum, menetapkan, standar, teknologi, metaverse]  |

## Stemming

Dalam tahap ini dilakukan proses untuk menghapus konjungsi dan mengubah setiap kata menjadi imbuhan kata dasar untuk mendapatkan lebih banyak data atau informasi yang relevan dan dapat dilihat pada Tabel 6 dan menggunakan file kata dasar bahasa Indonesia.

**Tabel 6** Proses *Stemming*

|  |  |  |
| --- | --- | --- |
| **No** | ***Tweet*** | **Hasil** |
| 1 | metaverse era dimana setiap orang dapat mengekspresikan diri dengan cara menarik | [metaverse, era, dimana, orang, ekspresi, diri, cara, tarik]  |
| 2 | teknologi ini mampu mewujudkan interaksi di dunia virtual yang unik selayaknya kehidupan nyata | [teknologi, wujud, interaksi, dunia, virtual, unik, layak, hidup, nyata] |

## Text Mining

Cara mendapatkan data yang tidak sistematis, tidak tertata atau tidak teratur tetapi memiliki peluang informasi yang baik seperti data teks dapat dikenal dengan sebutan *text mining* [6]. Berdasarkan penelitian [9], *text mining* serta data *mining* yaitu hal yang sama. Oleh karena itu, metode yang dimiliki sama untuk dipakai pada *mining* [9].

## Naïve Bayes

Naïve Bayes adalah tingkatan kategori individual yang menggunakan teorema bayes dan memperkirakan tingkat independensi yang tinggi dari semua keadaan dan kejadian. Pada metode ini terdapat dua tahapan pengklasifikasian teks yaitu tahapan klasifikasi dan *training*, adapun rumusnya sebagai berikut:

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

Dimana:

X = Data asing

H = Kelas Khusus atau khas (hipotesa data X)

P(H|X) = Prediksi hipotesis H pada X

P(H) = Prediksi hipotesis H

P(X|H) = Prediksi hipotesis X pada H

P(X) = Prediksi X

## *K-Nearest Neighbor* (K-NN)

K-NN adalah algoritma berlandaskan *instance* *learning* atau *lazy learning*. Dari seluruh algoritma *machine learning* algoritma *K-Nearest Neighbor* yaitu algoritma yang paling sederhana. K-NN akan menemukan jarak atau tetangga yang terdekat diantara Adapun rumusnya sebagai berikut:

$d\_{i}= \sqrt{(\sum\_{i=1}^{n}(x\_{ij}-y\_{j})}$ (2)

Dimana:

di = Jarak sampel

xij = Data sampel pengetahuan

yj = Data *input* var ke-j

n = Jumlah sampel

***TF-IDF***

Dalam tahap ini terbagi menjadi 2 dimana TF (*Term* Frequency) dan IDF (*Inverse Document* Frequency). TF adalah banyaknya kata pada data dan *inverse* frekuensi adalah dokumen yang menyimpan kata. Jadi di dalam proses ini yaitu teknik menghitung banyaknya kata yang muncul di dalam sebuah dokumen data yang dapat kita lihat pada Gamb 9, nilai yang bernilai 0 artinya sebuah kata yang tidak muncul dalam dokumen tersebut dan yang bernilai 1 artinya sebuah kata yang muncul dalam dokumen tersebut. Berikut ini adalah rumus TF-IDF [12]:

$TF-IDF=tf.idf$(3)

Dimana:

TF = *Term Frequency*

Idf = *Inverse document frequency*

Contoh penerapan TF-IDF

Dokumen 1: metaverse maju

Dokumen 2: bali via metaverse kak

**Tabel 7** *Document Frequency*

|  |  |  |  |
| --- | --- | --- | --- |
| **Term (t)** | **Dokumen 1** | **Dokumen 2** | **df** |
| metaverse | 1 | 1 | 2 |
| maju | 1 | 0 | 1 |
| bali | 0 | 1 | 1 |
| via | 0 | 1 | 1 |
| kak | 0 | 1 | 1 |

Menghitung *Inverse Document Frequency* (IDf) dengan rumus yang terlihat pada Tabel 8.

$Idf =\frac{1}{df} atau Idf= log( \frac{N}{df}) $ (4)

Dimana:

Idf = *Inverse document frequency*

df = *document frequency*

N = Jumlah keseluruhan dokumen

**Tabel 8** *Inverse Document Frequency*

|  |  |  |
| --- | --- | --- |
| **Term (t)** | **df** | **Idf** |
| metaverse | 2 | 0,5 | log 4 |
| maju | 1 | 1 | log 1 |
| bali | 1 | 1 | log 1 |
| via | 1 | 1 | log 1 |
| kak | 1 | 1 | log 1 |

Menghitung TF-IDF (*Term Frequency Inverse Document Frequency*) dengan rumus yang terlihat pada rumus ke tiga di Tabel 9.

**Tabel 9** Perhitungan *Term Frequency Inverse Document Frequency*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Term (t)** | **Dokumen 1** | **Dokumen 2** | **Idf** | **Tf.idf** |
| **Dokumen 1** | **Dokumen 2** |
| metaverse | 1 | 1 | 0,5 | 0,5 | 0,5 |
| maju | 1 | 0 | 1 | 1 | 0 |
| bali | 0 | 1 | 1 | 0 | 1 |
| via | 0 | 1 | 1 | 0 | 1 |
| kak | 0 | 1 | 1 | 0 | 1 |

## Evaluasi

Evaluasi adalah tingkat analisis yang dipergunakan untuk memberikan evaluasi terhadap hasil telah selesai dilakukan. Perhitungan dilakukan dengan parameter *Accuracy, Precision, Recall* dan *Specificity.* Di bawah ini adalah table *Confusion Matrix* yang terlihat pada Tabel 10 beserta rumus perhitungannya.

**Tabel 10** *Confusion Matrix*

|  |  |
| --- | --- |
| **Kelas Prediksi** | **Kelas Aktual** |
| Positif | Negatif |
| Positif | TP | FP |
| Negatif | FN | TN |

Dimana :

TP = *True Positive*, yang artinya data diprediksi positif dan nilai sebenarnya positif

FP = *False Positive,* yang artinya data diprediksi positif, tetapi sebenarnya negatif

TN = *True Negative*, yang artinya data diprediksi negatif dan nilai sebenarnya negatif

FN = *False Negative,* yang artinya data diprediksi negatif, tetapi sebenarnya positif

*Accuracy* merupakan perhitungan secara akurat dimana akan diklasifikasikan data positif dan juga data negatif. Adapun cara menghitungnya yaitu sebagai berikut:

$Accuracy= \frac{(TP+TN)}{(TP+TN+FN+FP)} x 100\%$ (6)

*Precision* merupakan prediksi dari rasio yang membandingkan hasil perhitungan prediksi *true* positifdengan data keseluruhan yang diprediksi positif. Adapun cara menghitungnya yaitu sebagai berikut:

$Precision= \frac{TP}{(TP+FP)} x 100\%$ (7)

*Recall* merupakan prediksi dari rasio yang membandingkan hasil perhitungan prediksi *true* positif dengan keseluruhan data yang *true* positif. Adapun cara menghitungnya:

$Recall= \frac{TP}{(TP+FN)} x 100\%$ (8)

*Specificity* merupakan kebenaran memprediksi data negatif dengan jumlah keseluruhan data negatif. Adapun cara menghitungnya yaitu sebagai berikut:

$Specificity= \frac{TN}{(TN+FP)} x 100\%$ (9)

## *Synthetic Minority Over-samping Technique (SMOTE)*

Data yang tidak seimbang adalah di saat total kelas data yang satu lebih minim dibandingkan dengan total kelas data lainnya. Metode *Smote* adalah cara yang dipakai untuk mengatasi ketidakseimbangan data kelas minoritas. Metode *Smote* juga merupakan metode turunan dari *oversampling*. Cara kerjanya yaitu dengan mencari nilai tetangga (k *nearest neighbors*) secara acak dengan membuat data sintesis sebanyak duplikasi yang diinginkan oleh data minor [15].

## Penarikan Kesimpulan

Tahapan yang paling akhir yang merupakan langkah untuk menentukan hasil dari pengujian yang sudah dilakukan.

# Hasil dan Pembahasan

# *Pre-Processing*

Pada tahap ini yang dapat dilihat dari Gambar 3, peneliti menggunakan operator *Transform Cases* pada tahap pertama *preprocessing* untuk menyamakan kalimat menjadi huruf kecil pada data yang sudah peneliti kumpulkan, lalu menggunakan operator *Tokenize* untuk memisahkan kata per kata dalam sebuah kalimat. Tahap selanjutnya yang peneliti lakukan yaitu menggunakan operator *Filter Stopwords (Dictionary)* dimana operator ini perlu ditambahkan dokumen *stopword* bahasa Indonesia sebagai acuan untuk kata mana yang akan dihilangkan dan kata mana yang akan diproses, kemudian operator terakhir yaitu *Stem (Dictionary)* dimana operator ini juga perlu ditambahkan dokumen kata dasar sebagai acuan untuk memproses kata menjadi kata dasar dan menghapus konjungsi. Jumlah data yang diolah menjadi 566 data setelah melalui tahapan *Pre-Processing* data.



**Gambar 3** Isi Operator *Process Document*

# Pengujian Algoritma Naïve Bayes tanpa *Smote*

# Dengan menggunakan *software Rapid Miner*, pada Gambar 4 tahap pertama yang peneliti lakukan adalah menggunakan operator *Read Excel* untuk membaca data sentimen yang sudah diproses pada tahapan pre-processing, lalu menggunakan operator *Split Data* untuk memisahkan *training* dataset dan *testing* dataset dengan ratio 70%:30%, kemudian menggunakan operator *Set Role* untuk memilih *role* label pada dataset. Operator selanjutnya yang digunakan yaitu *Nominal to Text* dimana operator ini berguna untuk mengubah data *nominal* menjadi *text*, selanjutnya menggunakan operator *Process Documents from data* dengan pembobotan TF-IDF dimana terdapat operator yang ada di Gambar 3 pada tahap *pre-processing*. Tahap selanjutnya menggunakan operator *Cross Validation* yang didalamnya terdapat beberapa operator untuk memasukkan algoritma.

#

**Gambar 4** Pengujian Algoritma Naïve Bayes tanpa *Smote*

# Evaluasi Kedua Algoritma

Pada tahapan ini yang dapat dilihat pada Gambar 5 dan Gambar 6, peneliti menggunakan operator Naïve Bayes dan K-NN untuk pengklasifikasian algoritma pada *Cross Validation*, lalu selanjutnya menggunakan operator *Apply Model* untuk membentuk model dan *Performance* untuk memberikan hasil dari klasifikasi.



**Gambar 5** Desain Algoritma Naïve Bayes

**Gambar 6** Desain Algoritma K-NN

# Hasil Algoritma Naïve Bayes tanpa menggunakan *Smote*

Hasil *Accuracy* pada Gambar 7 menggunakan algoritma Naïve Bayes memperoleh akurasi yang artinya bahwa model ini bisa secara akurat mengklasifikasikan data positif dan data negatif sebesar 78,06% dengan *Precision* yang artinya prediksi dari rasio benar positif dengan keseluruhan hasil yang diprediksi positif sebesar 88,56%, *Recall* yang artinya prediksi dari rasio yang membandingkan hasil perhitungan prediksi benar positif dengan keseluruhan data yang benar positif sebesar 81,08%, dan *Specificity* yang artinya prediksi data negatif dengan keseluruhan data negatif sebesar 69,00%.

****

**Gambar 7** Hasil *Accuracy, Precision,* Recall dan *Specificity* pada algoritma Naïve Bayes

Dihitung secara manual:

$Accuracy = \frac{(TP+TN)}{(TP+TN+FN+FP)}x100\%= \frac{240+69}{240+69+56+31}x 100\%=\frac{309}{396}x 100\%=78,06\%$

$Precision = \frac{TP}{(TP+FP)}x100\%= \frac{240}{240+31} x 100\%= \frac{240}{271}x 100\%=88,56\%$

$Recall = \frac{TP}{(TP+FN)}x100\%= \frac{240}{240+56}x 100\%= \frac{240}{296} x 100\%=81,08\%$

$Specificity = \frac{TN}{(TN+FP)}x 100\%=\frac{69}{69+31} x 100\%= \frac{69}{100} x 100\%=69,00\%$

Dapat disimpulkan bahwa melalui *software Rapid Miner* dan juga dihitung secara manual, hasil dari semua parameter (*Accuracy, Precision, Recall* dan *Specificity*) sama.

# Hasil Algoritma K-NN tanpa menggunakan *Smote*

Hasil *Accuracy* pada Gambar 8 menggunakan algoritma K-NN memperoleh akurasi yang artinya bahwa model ini bisa secara akurat mengklasifikasikan data positif dan data negatif sebesar 78,55% dengan *Precision* yang artinya prediksi dari rasio benar positif dengan keseluruhan hasil yang diprediksi positif sebesar 81,12%, *Recall* yang artinya prediksi dari rasio yang membandingkan hasil perhitungan prediksi benar positif dengan keseluruhan data yang benar positif sebesar 92,91%, dan *Specificity* yang artinya prediksi data negatif dengan keseluruhan data negatif sebesar 36,00%.



**Gambar 8** Hasil algoritma K-NN

Dihitung secara manual:

$Accuracy = \frac{(TP+TN)}{(TP+TN+FN+FP)}x100\%= \frac{275+36}{275+36+21+64}x 100\%=\frac{311}{399}x 100\%=78,55\%$

$Precision = \frac{TP}{(TP+FP)}x100\%= \frac{275}{275+64} x 100\%= \frac{275}{339}x 100\%=81,12\%$

$Recall = \frac{TP}{(TP+FN)}x100\%= \frac{275}{275+21}x 100\%= \frac{275}{296} x 100\%=92,91\%$

$Specificity = \frac{TN}{(TN+FP)}x 100\%=\frac{36}{36+64} x 100\%= \frac{36}{100} x 100\%=36,00\%$

Dapat disimpulkan bahwa melalui *software* Rapid Miner dan juga dihitung secara manual, hasil dari semua parameter (*Accuracy, Precision, Recall* dan *Specificity*) sama.

# Pengujian Kedua Algoritma dengan *Smote*

Pada tahapan ini menggunakan penambahan operator *Smote* *(Synthetic Minority Oversampling Technique)* yaitu untuk mengatasi ketidakseimbangan kelas pada data yang dapat dilihat pada Gambar 9. Operator yang digunakan adalah “*SMOTE* Upsampling” dengan percobaan nilai k dari satu sampai kelima.



**Gambar 9** Desain Model Naïve Bayes dengan *Smote*

# Hasil Membandingkan Algoritma Naïve Bayes tanpa *Smote*  dan Algoritma Naïve Bayes dengan *Smote*

Pada tahap ini merupakan hasil *Accuracy* yang menggunakan operator *SMOTE (Synthetic Minority Oversampling Technique)* dengan membandingkan k=1 hingga k=5 serta hasil yang tidak menggunakan operator *Smote*. Hasil pada algoritma Naïve Bayes tanpa *Smote* diperoleh *Accuracy* dimana dapat diartikan bahwa model ini bisa secara akurat mengklasifikasikan data positif dan data negatif sebesar 78,06%. Dengan menggunakan *Smote* diperoleh nilai *Accuracy* yang paling tinggi pada percobaan k=5 artinya bahwa model ini bisa secara akurat mengklasifikasikan data positif dan data negatif sebesar 85,91% yang dapat dilihat pada Tabel 11.

**Tabel 11**  Perbandingan Algoritma Naïve Bayes dengan *Smote* dan Naïve Bayes tanpa *Smote*

|  |  |  |
| --- | --- | --- |
| **Performa** | **Naïve Bayes dengan *Smote*** | **Naïve Bayes** |
| **k=1** | **k=2** | **k=3** | **k=4** | **k=5** |
| *Accuracy* | 84,57% | 85,74% | 85,74% | 85,74% | 85,91% | 78,06% |

# Hasil Membandingkan Algoritma K-NN tanpa *Smote* danAlgoritma K-NN dengan *Smote*

Pada tahap ini merupakan hasil *Accuracy* yang menggunakan operator *SMOTE* *(Synthetic Minority Oversampling Technique)* dengan membandingkan k=1 hingga k=5 serta hasil yang tidak menggunakan operator *Smote*. Hasil pada algoritma K-NN tanpa *Smote* diperoleh *Accuracy* dimana dapat diartikan bahwa model ini bisa secara akurat mengklasifikasikan data positif dan data negatif sebesar 78,55%. Dengan menggunakan *Smote* diperoleh nilai *Accuracy* yang paling tinggi pada percobaan k=5 artinya bahwa model ini bisa secara akurat mengklasifikasikan data positif dan data negatif sebesar 80,44% yang dapat dilihat pada Tabel 12.

**Tabel 12** Perbandingan Algoritma K-NN dengan *Smote* dan K-NN tanpa *Smote*

|  |  |  |
| --- | --- | --- |
| **Performa** | **K-NN dengan *Smote*** | **K-NN** |
| **k=1** | **k=2** | **k=3** | **k=4** | **k=5** |
| *Accuracy* | 80,11% | 78,95% | 79,44% | 79,94% | 80,44% | 78,55% |

# Perbandingan *Confusion Matrix* pada Algoritma Naïve Bayes dan K-NN

Hasil *Confusion Matrix* dengan nilai k=1 sampai k=5 untuk melihat perbandingan *True Positive, False Positive, True Negative* dan *False Negative* melalui metode Naïve Bayes dan K-NN. Dapat dilihat dari Tabel 13 bahwa *True Positive* dan *True Negative* yang paling baik ada pada k=5 yaitu *True Positive* sebesar 232 dan *True Negative* sebesar 65 pada algoritma Naïve Bayes. *True Positive* sebesar 218 dan *True Negative* sebesar 79 pada algoritma K-NN.

**Tabel 13** Hasil Perbandingan *Confusion Matrix*

|  |  |  |  |
| --- | --- | --- | --- |
| Metode | K=1 | K=2 | K=3 |
| TP | FP | TN | FN | TP | FP | FN | TN | TP | FP | FN | TN |
| Naïve Bayes | 231 | 21 | 285 | 66 | 231 | 20 | 286 | 66 | 231 | 20 | 286 | 66 |
| K-NN | 229 | 52 | 254 | 68 | 220 | 50 | 256 | 77 | 219 | 46 | 260 | 78 |
| Metode | K=4 | K=5 |
| TP | FP | TN | FN | TP | FP | FN | TN |
| Naïve Bayes | 231 | 20 | 286 | 66 | 232 | 20 | 286 | 65 |
| K-NN | 219 | 43 | 263 | 78 | 218 | 39 | 267 | 79 |

# Visualisasi

Pada tahap ini yaitu menggunakan *word cloud* pada *Rapid Miner* dimana dapat memvisualisasikan hasil analis sentimen. Dengan menggunakan word cloud terlihat secara keseluruhan kata-kata yang sering kelihatan dapat dilihat pada gambar 10 yaitu teknologi, belajar, virtual, bisnis, dan lain sebagainya. Pada Gambar 11 kata-kata positif yaitu keren, *industry*, bisnis, layanan, *best*, luncurkan, ekonomi dan lain sebagainya. Pada Gambar 12 kata-kata negatif yang sering muncul yaitu gaya, kayak, aslinya, gimana, ngomongin dan lain sebagainya .

**Gambar 10** *world cloud* keseluruhan **Gambar 11** *world cloud* positif **Gambar 12** *world cloud* negatif

# Kesimpulan

Digunakan dua algoritma pada penelitian ini yaitu Naïve Bayes dan K-NN dalam menentukan sentimen masyarakat terhadap metaverse. Dikarenakan penelitian ini diperoleh ketidakseimbangan data atau *imbalance* data, maka menggunakan penambahan algoritma *Smote* dengan melakukan perbandingan dari nilai k=1 sampai k=5 pada algoritma Naïve Bayes secara berturut-turut adalah 84,57%, 85,74%, 85,74%, 85,74% dan 85,91%. Pada algoritma K-NN secara berturut-turut adalah 80,11%, 78,95%, 79,44%, 79,44%, 80,44%. Dari hasil yang diberikan algoritma Naïve Bayes dengan menggunakan *Smote* *upsampling* memberikan *performance* terbaik pada nilai k=5 sebesar 85,91% hal ini menerangkan bahwa untuk menganalisis sentimen pada metaverse kepada masyarakat lebih mengarah ke positif. Diharapkan pada penelitian berikutnya dapat mengaplikasikan algoritma lain serta pembobotan berbeda untuk memperoleh kesimpulan yang lebih baik lagi.

# Daftar Pustaka

[1] C. Newton, “MARK IN THE METAVERSE,” www.theverge.com, Jul 22, 2021. [Online]. Available: https://www.theverge.com/22588022/mark-zuckerberg-facebook-ceo-metaverse-interview.

[2] Ali Ahmad, Windu Gata, “Sentimen Analisis Masyarakat Indonesia di Twitter Terkait Metaverse dengan Algoritma Support Vector Machine,” *J.JTIK (Jurnal Teknologi Informasi Dan Komunikasi),* vol. 6, no. 4, pp. 548-555, Oktober. 2022, Available doi: https://doi.org/10.35870/jtik.v6i4.569

[3] T. W. Putra, A. Triayudi, and Andrianingsih, “Analisis Sentimen Pembelajaran Daring menggunakan Metode Naïve Bayes, K-NN, dan Decision Tree,” *J. JTIK (Jurnal Teknologi Informasi Dan Komunikasi)*, vol. 6, no. 1, pp. 157–163, Oktober. 2021, Available doi: https://doi.org/10.35870/jtik.v6i1.368.

[4] Y. T. Samuel, K. J. Manurip, and H. Sitorus, “Analisis Sentimen Tokoh Publik Menggunakan Metode Naïve Bayesian Classification Pada Aplikasi *Twitter*,” *J. TeIKa (Jurnal Teknologi Informasi dan Komunikasi)*, vol. 7, no. 1, pp. 81-88, April. 2017, Available doi: https://doi.org/10.36342/teika.v7il.2218.

[5] A. Wandani, Fauziah, and Andrianingsih, “Sentimen Analisis Pengguna Twitter pada Event Flash Sale Menggunakan Algoritma K-NN, Random Forest, dan Naive Bayes,” *Jurnal Sains Komputer Dan Informatika J-SAKTI*, vol. 5, no. 2, pp. 651–665, September. 2021, Available doi: http://dx.doi.org/10.30645/j-sakti.v5i2.365.

[6] H. Hassani, C. Beneki, S. Unger, M.T. Mazinani, and M.R Yeganegi, “Text mining in big data analytics,*” Big Data Cogn. Comput.,* vol. 4, no.1, pp. 1-34, Januari. 2020, Available doi: 10.3390/bdcc401001.

[7] M. A. Kausar, A. Soosaimanickam, and M. Nasar, “Public Sentiment Analysis on Twitter Data during COVID-19 Outbreak,” Int. J. Adv. Comput. Sci. Appl., vol. 12, no. 2, pp. 415–422, Januari. 2021, Available doi: 10.14569/IJACSA.2021.0120252.

[8] R. Puspita and A. Widodo, “Perbandingan Metode K-NN, Decision Tree, dan Naïve Bayes Terhadap Analisis Sentimen Pengguna Layanan BPJS,” *J. Inform. Univ. Pamulang*, vol. 5, no. 4, pp. 646–654, Desember. 2020, Available doi: 10.32493/informatika.v5i4.7622.

[9] A. K. Fauziyyah; “Analisis Sentimen Pandemi Covid19 Pada Streaming Twitter Dengan Text Mining Python,” *Jurnal Imiah* *SINUS (STMIK Sinar Nusantara Surakarta)*, vol. 18, no.2, pp. 31-42, Juli. 2020, Available doi: 10.30646/sinus.v18i2.491.

[10] H. P. Doloksaribu, Y. T. Samuel, “Komparasi Algoritma Data *Mining* Untuk Analisis Sentimen Aplikasi Pedulilindungi,” *Jurnal Teknologi Informasi (JTI) Jurnal Keilmuan dan Aplikasi Bidang Teknik Informatika*, vol. 16, no. 1, pp. 1-11, Januari. 2022, Available doi: https://doi.org/10.47111/JTI.

[11] Rizki Tri Wahyuni, Dhidik Prastiyanto, dan Eko Supraptono, “Penerapan Algoritma Cosine Similarity dan Pembobotan TF-IDF pada Sistem Klasifikasi Dokumen Skripsi”, *Jurnal Teknik Elektro*, vol. 9, no. 1, Januari – Juni. 2017, pp. 18-23, Available doi: https://doi.org/10.15294/jte.v9i1.10955.

[12] M.Khairul Anam, Bunga Nanti Pikir, Muhammad Bambang Firdaus, Susi Erlinda dan Agustin, “Penerapan Naïve Bayes Classifier, K-Nearest Neighbor (K-NN) dan *Decision Tree* untuk Menganalisis Sentimen pada Interaksi Netizen dan Pemerintah”, *Jurnal Manajemen, Teknik Informatika, dan Rekayasa Komputer*, vol. 21, no. 1, pp. 139-150, November. 2021, Available doi: 10.30812/matrik.v2li1.1092.

[13] Hozairi, Anwari, dan Syariful Alim, “Implementasi Orange Data Mining Klasifikasi Kelulusan Mahasiswa dengan Model K-Nearest Neighbor, Decision Tree serta Naïve Bayes”, *Jurnal Ilmiah Nero*, vol. 6, no. 2, 2021, pp. 133-144, Available doi: http://dx.doi.org/10.21107/nero.v6i2.237.

[14] Franly Salmon Pattiiha dan Hendry, “Perbandingan Metode K-NN, Naïve Bayes, Decision Tree untuk Analisis Sentimen Tweet Twitter Terkait Opini Terhadap PT PAL Indonesia”, *Jurnal Riset Komputer* *(JURIKOM)*, vol. 9, no. 2, April. 2022, pp. 506-514, Available doi: 10.30865/jurikom.v9i2.4016.

[15] Rimbun Siringoringo, “Klasifikasi Data Tidak Seimbang Menggunakan Algoritma Smote dan k-Nearest Neighbor”, Jurnal *Information System Development* (ISD), vol. 3, no. 1, pp. 44-49, Februari. 2018, Available: https://ejournal-medan.uph.edu/index.php/isd/article/view/177/63.