Analisis Sentimen *Review* Film Menggunakan Naive Bayes Classifier Dengan Fitur TF-IDF

1st Muhammad Thaariq Razaq
Fakultas Informatika
Universitas Telkom
Bandung, Indonesia
thaariqrazaq@student.telkomuniversity.ac.id

2nd Dade Nurjanah
Fakultas Informatika
Universitas Telkom
Bandung, Indonesia
dadenurjanah@telkomuniversity.ac.id

3rd Hani Nurrahmi
Fakultas Informatika
Universitas Telkom
Bandung, Indonesia
haninurrahmii@telkomuniversity.ac.id

Abstrak-Penilaian mengenai isi dari review film dapat disebut dengan sentiment analysis. Sentiment analysis pada review film terbagi menjadi 2 yaitu berupa review positif dan review negatif. Salah satu algoritma data mining yang paling sering digunakan dalam penelitian adalah Naïve Bayes karena bekerja dengan cepat dan efisien sebagai metode pengklasifikasian teks tetapi memiliki kekurangan yang sangat sensitif dalam pemilihan fitur. Pada umumnya, data review film memuat isi yang sangat panjang sehingga diperlukan feature selection atau pemangkasan fitur yang berguna untuk mengurangi dimensi pada saat proses klasifikasi. Pada penelitian ini menggunakan fitur Tf-Idf sebagai salah satu solusi untuk mempermudah dan mempercepat pencarian informasi yang sesuai adalah dengan meringkas konten tersebut. Tf-Idf (Term Frequency Inverse Document Frequency) merupakan metode pembobotan dalam bentuk integrasi antar term frequency dengan inverse document frequency. Metode Tf-Idf digunakan pada penelitian ini untuk memilih fitur sebagai hasil ringkasan, dengan penerapannya pada seleksi fitur bobot kata. Sebelum proses klasifikasi, dilakukan tahapan preprocessing yang meliputi data cleaning dan case folding, stop words removal, stemming, dan tokenization. Pada penelitian ini menghasilkan nilai akurasi mencapai 86.48%. Sehingga Naïve Bayes dengan fitur Tf-Idf pada masalah analisis klasifikasi sentimen review film terbukti memberikan akurasi yang akurat

Kata kunci- sentiment analysis, film, Naïve Bayes, TF-IDF

I. PENDAHULUAN

A. Latar Belakang

Review tentang film merupakan kebutuhan bagi semua orang untuk mendapatkan informasi mengenai sebuah film sehingga dapat digunakan untuk membantu mendapatkan informasi tentang isi film yang akan ditonton. Informasi yang bisa didapat melalui sebuah review film adalah mengenai jalan cerita, aktor sampai dengan konflik yang terjadi di dalamnnya serta kelebihan dan kekurangan sebuah film. Informasi-informasi hasil review yang dibuat kemudian digunakan sebagai bahan pertimbangan dalam menentukan kualitas dari sebuah film sehingga pecinta film dapat mengetahui sejauh mana film tersebut layak atau tidak layak di tonton. Penilaian mengenai isi dari review film dapat disebut dengan sentiment analysis. Sentiment analysis adalah proses penerapan natural language processing (NLP) dan analisis teks untuk mengidentifikasi dan melakukan ektrak informasi subjektif dari sebuah teks [1]. Sentiment analysis dapat diaplikasikan menggunakan sebuah metode klasifikasi untuk mempermudah dalam pengelompokan data berupa data positif atau data negatif yaitu dengan menggunakan metode Naïve Bayes. Metode Naïve Bayes digunakan pada proses klasifikasi dalam sebuah penelitian karena bekerja dengan cepat dan efisien sebagai metode pengklasifikasian teks. Penerapan klasifikasi sentiment analysis menjadi kalimat positif maupun negatif dapat dilakukan setelah pemangkasan pada data subjek yang digunakan untuk mengurangi fitur sehingga menghindari banyaknya dimensi yang digunakan pada saat proses klasifikasi [2]. Review film dapat mempunyai ukuran dataset yang cukup besar baik itu pada data training maupun data testing. Dimensi dan fitur yang berlebihan akan meningkatkan ruang pencarian semakin tinggi sehingga akan menyebabkan kesulitan dalam memproses data dan akan menurunkan kinerja serta membuat data tidak konsisten. Analisis dan mining dalam data juga membutuhkan waktu yang lama dalam pemrosesan data. Pengurangan dimensi dapat diterapkan untuk mengurangi dimensi dari data, dimana nantinya akan meningkatkan kinerja dari tehnik machine learning dengan menghilangkan fitur yang tidak perlu digunakan. Penyelesaian dalam permasalahan penelitian ini menggunakan metode Tf-Idf yang akan mengukur pembobotan berbasis kata sebagaifitur hasil ringkasan

Ada beberapa algoritma klasifikasi yang banyak digunakan untuk analisis review sentimen antara lain Support Vector Machine, Na¨ıve Bayes, dan K-Nearest Neighbor [3]. Beberapa penelitian yang telah dilakukan dalam klasifikasi sentimen ulasan online diantaranya, Comparative machine learning untuk klasifikasi sentimen analisis ulasan film [4]. Analisis sentimen pada opini review film menggunakan algoritma NB, KNN, dan Random Forest [5]. Analisis sentimen pada opini review film menggunakan Features and Opinion Words Extraction [6]. Analisis sentimen pada opini review film menggunakan algoritma NB, KNN, dan Decision Tree [7]. Analisis sentimen pada opini review film menggunakan algoritma NB dan fitur Gini Index [8]. Analisis sentimen pada opini review film menggunakan algoritma NB dan fitur Chi Square [9]. Analisis sentimen pada You tube movie trailer comments menggunakan algoritma NB dan fitur Tf-Idf [10].

B. Topik dan Batasannya

Tugas akhir ini difokuskan pada analisis sentimen dari review film yang berpotensi menghasilkan review postif dan review negatif. Banyak situs memberikan review suatu produk yang dapat mencerminkan opini pengguna. Salah satu contohnya adalah situs Internet Movie Database (IMDb). IMDb adalah situs web yang berhubungan dengan produksi film dan film. Penelitian ini mengusulkan untuk menggunakan metode Naïve Bayes. Pemilihan metode didasarkan pada beberapa penelitian terdahulu sebagai referensi.

Pada tugas akhir ini, topik yang akan dibahas adalah bagaimana performansi metode Naïve Bayes dan Fitur Tf-Idf dalam analisis sentimen dari review film yang berasal dari data IMDb. Batasan masalah pada tugas akhir ini adalah sebagai berikut: Pertama, jumlah sample atau data sebanyak 50000. Kedua, data yang berupa Bahasa Inggris. Ketiga, penelitian analisis sentimen menggunakan metode Naïve Bayes Classifier dan fitur Tf-Idf sebagai pengklasifikasi data review film.

C. Tujuan

Tujuan dari penelitian ini adalah untuk mengidentifikasi akurasi dari Naïve Bayes dengan adanya indikasi review postif dan review negatif pada review film IMDb. Dengan hasil Naïve Bayes dibantu dengan fitur Tf-Idf ini diharapkan dapat membantu bagi semua orang untuk mendapatkan informasi mengenai sebuah film sehingga dapat digunakan untuk membantu mendapatkan informasi tentang isi film yang akan ditonton.

Organisasi Tulisan

Struktur penulisan dari tugas akhir ini disusun sebagai berikut: Bagian pertama berisi pendahuluan terkait tugas akhir ini. Bagian kedua menjelaskan studi yang terkait dengan tugas akhir ini. Bagian ketiga akan menjelaskan pemodelan dan performansi dari sistem yang dibangun. Bagian keempat menjelaskan hasil dan evaluasi hasil pengujian yang telah dilakukan pada bagian ketiga. Kemudian, pada bagian terakhir menjelaskan kesimpulan dan saran berdasarkan hasil pengujian yang dilakukan pada tugas akhir ini.

II. KAJIAN TEORI

Menurut informasi yang telah dijelaskan, akan dilakukan penelitian tentang sentiment analisis review film di IMDb menggunakan algoritma Naïve Bayes. Data yang diperoleh akan diproses dengan memerlukan text mining, kemudian akan dilanjutkan mengklasifikasikan komentar IMDb menjadi dua kelas, yaitu positif dan negatif. Klasifikasi yang dilakukan menerapkan algoritma Naïve Bayes. Klasifikasi berfungsi memberikan kemudahan kepada pengguna untuk mengetahui opini positif atau negatif. Nilai akurasi yang dihasilkan dengan menggunakan algoritma akan memberikan pengaruh pada hasil klasifikasi.

Ada banyak penelitian tentang analisis sentimen pada topik seperti komentar media sosial, produk, politik, dan banyak lagi. Ada begitu banyak teknik untuk mengklasifikasikan teks. Banyak peneliti mencoba melakukan kombinasi teknik untuk mencapai kinerja yang lebih baik.

Palak Baid dkk. dalam makalah ini [5] ulasan film terklasifikasi untuk analisis sentimen menggunakan WEKA Tool. Data yang didapat dari IMDb mereka menganalisis 2000 data review yang mengekspresikan sentimen positif atau negatif. Dalam makalah ini, mereka juga membandingkan teknik klasifikasi Naïve Bayes, Random Forest, dan K-Nearest Neighbour. Mereka melakukan eksperimen mereka di WEKA dan menyimpulkan bahwa hasil terbaik diberikan oleh pengklasifikasi Naïve Bayes. Pengklasifikasi Naïve Bayes mencapai akurasi 81,45%, pengklasifikasi Random Forest mencapai akurasi 78,65%, pengklasifikasi K-Nearest Neighbor mencapaiakurasi 55,30%.

Penelitian Gurshobit Singh [6] dilakukan dengan menggunakan data film dari TMDB ID. Penelitian dilakukan dengan fitur Based Opinion Mining. Sistem mengekstrak semua kata benda, frasa kata benda, kata kerja, dan kata sifat dari review film dan membandingkan dengan daftar kata yang ada. Kata-kata ini diklasifikasikan berdasarkan polaritasnya. Polaritas kalimat mengikuti aturan yang sama seperti ekspresi aritmatika. Sentimen negatif mengandung semua kata opini negatif dan sentimen positif mengandung semua kata opini positif. Lalu menjumlah total kalimat positif atau negatif yang ditemukan dalam ulasan. Jika jumlah total kalimat positif lebih besar dari jumlah total kalimat negatif maka polaritas review akan menjadi positif. Demikian pula sebaliknya, lalu penelitian ini menghasilkan akurasi sebesar

81.22%

Dalam penelitian ini [7] penelitian dilakukan dengan 100 opini review film.. Data yang didapat akan diekspresikan menjadi sentimen positif atau negatif. Dalam makalah ini, mereka juga membandingkan teknik klasifikasi Naïve Bayes, Decision Tree, dan K-Nearest Neighbour. Mereka melakukan eksperimen dengan menghitung frekuensi word lalu melakukan klasifikasi menggunakan Naïve Bayes, Decision Tree, dan K-Nearest Neighbour. Kemudian mereka menyimpulkan bahwa hasil terbaik diberikan oleh pengklasifikasi Naïve Bayes.

Dalam penelitian Riko dkk. [8] melakukan penelitian analisis sentimen review film IMDb, Rotten Tomatoes, and Metacritic. Riko dkk menggunakan klasifikasi MNNB dan fitur Gini Index. GIT digunakan untuk memisahkan atribut. Untuk setiap fitur dalam film, review akan dihitung berdasarkan GIT A, GIT B, dan GIT C pada kelas positif dan negatif. Kemudian dari hasil penelitian menggunakan MNNB dan GIT menghasilkan performansi yang berbeda seperti pada GIT A memiliki performansi yang lebih baik dibandingkan dengan GIT B dan GIT C yaitu 59,54%, sedangkan GIT B dan GIT C sebesar 59,29% dengan selisih 0,25%.

Ahmad Zuili dkk. [9] melakukan penelitian sentimen review film IMDb dimana 1400 data opini, terdiri dari 700 buah opini positif dan 700 buah opini negative. Penelitian dilakukan dalam beberapa fase. Pertama, dilakukan preprocessing kemudian fase penyeleksian menggunakan fitur Chi Square. Terakhir fase klasifikasi menggunakan Naïve Bayes menghasilkan 64.40%

Risky Novendri dkk. [10] melakukan penelitian analisi sentimen YouTube movie trailer menggunakan klasikasi Naïve Bayes dan fitur Tf-Idf. Dalam penelitian ini, penulis menerapkan teknik crawling di Youtube tentang film Money Heist menggunakan otomatisasi pengujian web dengan Selenium WebDriver dan Python. Lalu penulis melakukan proses preprocessing kemudian dilanjutkan dengan penyeleksian menggunakan fitur Tf-Idf. Terakhir untuk proses klasifikasi penulis menggunakan klasifikasi Naïve Bayes dan mendapatkan hasil akurasi sebesar 81%

A. Sentiment Analysis (SA)

Sentiment Analysis (SA) atau biasa di sebut juga sebagai opinion mining adalah suatu riset komputasional dari emosi yang diungkapkan atau diekspresikan berupa tulisan (tekstual) dan opini sentiment [11]. Sentiment Analysis (SA) merupakan suatu proses untuk memahami data, mengolah data dan mengekstrak data tekstual secara otomatis dengan tujuan mendapatkan informasi sentimen atau intisari dari data yang terdapat di dalam suatu kalimat opini. Sentiment Analysis (SA) ini sendiri untuk melihat pendapat atau kecenderungan opini terhadap suatu masalah atau objek oleh seseorang, apakah kecenderungan tersebut mengarah ke hal positif atau negatif [12].

B. Text Mining

Text mining merupakan konsep terapan dalam teknik data mining untuk mencari pola inti suatu teks, dengan tujuan mendapatkan informasi yang terkandung dalam suatu teks yang dapat di manfaatkan dengan tujuan tertentu. Tahapan proses yang harus di lewati text mining di bagi menjadi :

1. Text preprocessing

Tahapan awal dalam text mining adalah text prepocessing dengan tujuan mempersiapkan data teks yang nantinya akan mengalami pengolahan data teks berikutnya. Dalam tahap ini hasil yang di dapatkan dari proses text preprocessing akan dilakukan proses transformasi. Proses transformasi ini dilakukan dengan mengurangi jumlah dari setiap kata dalam data teks stop word removal dan mengubah kata-kata menjadi kata dasar dalam data teks stemming. Kemudian melakukan pemecahan kalimat menjadikan data dalam bentuk token.

2. Feature selection

Dalam tahapan feature selection adalah tahapan penting dalam text mining. Karena dalam tahap ini dilakukan proses pembuangan beberapa term atau kata yang tidak terkait sehingga memperoleh term atau kata penting sebagai wakil kumpulan dokumen yang di analisis.

Dalam feature selection terdapat beberapa metode yang digunakan yaitu Term Frequency-Inverse Document Frequency (**Tf-Idf**). **Tf-Idf** itu sendiri terdiri dari Term Frequency dan Inverse Document Frequency Merupakan transformasi data yang memberikan nilai numerik dan menghitung setiap term yang berada dalam text review ini. Hal ini perlu dilakukan dikarenakan term dapat berbentuk kata atau frase dan agar dokumen atau text dapat diketahui konteksnya oleh sistem maka harus diberi indikator berupa pembobotan berupa nilai biner. Penelitian ini menggunakan metode Tf-Idf dengan formula:

$$W_{dt} = tf_d \times idf_t = tf_d \times \log(\frac{N}{df_t})$$

GAMBAR 2.1
RUMUS TF-IDF

Keterangan :

Wdt = Nilai bobot term ke-t pada dokumen d

tfd = Jumlah munculnya term t pada dokumen d

N = Jumlah dokumen secara keseluruhan d

ft = Jumlah dokumen yang mengandung term t

C. Naïve Bayes Classifier

Naïve Bayes Classifier adalah konsep probabilitas penentuan kelompok. Algoritma klasifikasi ini dapat mengolah data dalam jumlah besar dengan hasil akurasi yang tinggi [13]. Naive Bayes Classifier memiliki beberapa keunggulan, antara lain yaitu proses komputasi yang cepat, mudah diterapkan dengan struktur yang sederhana, dan efektif. Naive Bayes telah digunakan dengan cukup sukses dalam konteks beragam aplikasi, dan sangat populer dalam konteks klasifikasi teks. Dalam melakukan klasifikasi teks, algoritme Naive Bayes mampu mendapatkan nilai akurasi yang tinggi dan kompleksitas run time yang baik dengan jumlah data yang besar.

Dengan Persamaan Teorema Bayes :

$$P(H|E) = \frac{P(E|H).P(H)}{P(E)}$$

Keterangan :

E = Data yang belum diketahui classnya

H = Suatu class spesifikasi hipotesis data E

P(H|E) = probabilitas posterior, probabilitas maka akan muncul H jika diketahui E

P(E|H) = probabilitas posterior, probabilitas maka akan muncul E jika diketahui H

P(H) = probabilitas prior, probabilitas kejadian H

P(E) = probabilitas prior, probabilitas kejadian E

Di mana E adalah kata, H adalah kelas, P(E|H) adalah peluang kata di kelas H, P(c) adalah peluang dari kelas H dan P(E) adalah peluang dari kata E.

Dalam proses pengklasifikasian Naïve Bayes memerlukan Multinomial Naïve Bayes. Multinomial Naïve Bayes adalah model yang dikembangkan dari algoritma Bayes yang cocok dalam hal pengklasifikasian teks atau dokumen. Model dari multinomial memperhitungkan frekuensi dari setiap kata yang muncul pada dokumen tertentu. Maksud dari multinomial adalah suatu keadaan dimana value fitur memiliki lebih dari dua kejadian [14].

Untuk library scikit learn disini yang digunkan adalah Pipeline, CountVectorizer, MultinomialNB, Confusion Matrix, TfidfTransformer, dan f1 Score.

D. Evaluasi

Hasil klasifikasi dapat diuji dengan menggunakan metode pengujian dimana akan diukur tingkat akurasi sistem yang dibuat. Pengujian yang dapat dilakukan terdiri dari beberapa cara yaitu seperti accuracy, precision, recall dan f-measure. Accuracy adalah sebuah tingkat kedekatan antara nilai prediksi dengan nilai aktual. Precision merupakan jumlah jumlah dokumen relevan yang ditemukan dibagi dengan jumlah semua dokumen yang ditemukan. Recall merupakan jumlah dokumen relevan yang ditemukan dibagi dengan jumlah semua dokumen relevan di dalam koleksi [16].

TABEL 2.1
CONFUSION MATRIX

	Kelas Klasifikasi (True)	Kelas Klasifikasi (False)
Kelas Sebenarnya (True)	True Positive (TP)	False Negative (FN)
Kelas Sebenarnya (False)	False Positive (FP)	True Negative (TN)

Dari tabel tersebut terdapat 4 kategori, yaitu :

1. True Positive (TP) adalah kondisi saat suatu kelas true dan berhasil diklasifikasikan sebagai kelas true.
2. True Negative (TN) adalah kondisi saat suatu kelas false dan berhasil diklasifikasikan sebagai kelas false.
3. False Negative (FN) adalah kondisi saat suatu kelas true diklasifikasikan sebagai kelas false.
4. False Positive (FP) adalah kondisi saat suatu kelas false diklasifikasikan sebagai kelas true

Performansi yang akan diuji pada penelitian ini sebagai berikut :

1. Precision adalah tingkat ketepatan antara informasi yang diminta oleh pengguna dengan jawaban yang diberikan oleh sistem. Untuk menghitung nilai precision dapat dilakukan dengan menggunakan persamaan

$$Precision = \frac{TP}{TP+FP}$$

2. Recall adalah tingkat keberhasilan sistem dalam menemukan kembali sebuah informasi. Untuk menghitung nilai recall menggunakan persamaan

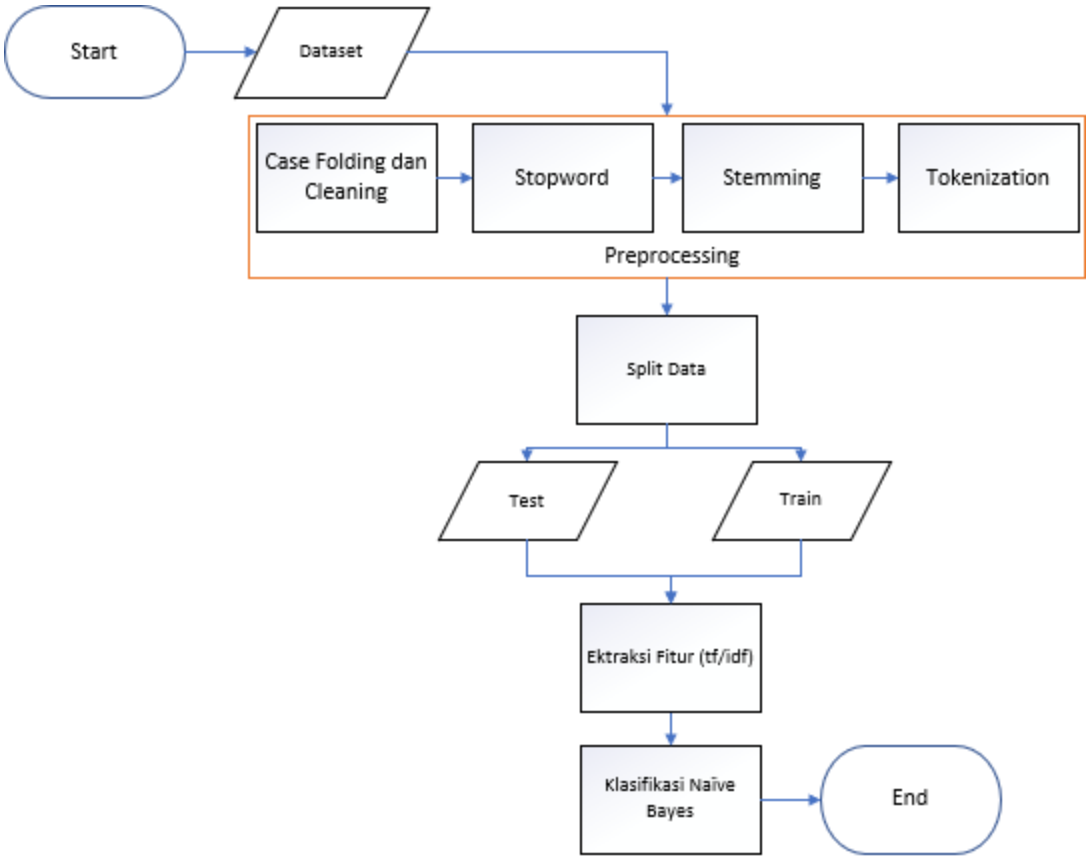
$$Recall = \frac{TP}{TP+FN}$$

3. Accuracy adalah perhitungan data yang bernilai true dibanding dengan keseluruhan jumlah data. Akurasi juga sering disebut sebagai tingkat kedekatan antara nilai prediksi dengan nilai aktual. Untuk menghitung nilai accuracy menggunakan persamaan

$$accuracy = \frac{TP+TN}{TP+TN+FP+FN}$$

III. METODE

Proses yang akan dilakukan pada penelitian ini digambarkan dengan diagram alir berikut:



GAMBAR 3.1
DIAGRAM ALUR PROSES PENELITIAN.

A. Dataset

Dataset yang digunakan adalah 50.000 data dari kumpulan review film didalam website IMDb yang telah dikumpulkan oleh Andrew Lee Maas yang merupakan ilmuan riset Machine Learning, Deep Learning, Natural Language Processing ternama yang merupakan bagian dari grup proyek Apple . Kemudian data diolah menjadi 2 bagian yaitu review postif dan review negatif. Dimana didapat 25.000 review postif dan 25.000 review negatif. Kemudian data dibagi menjadi 2 yaitu data train dan data test.

B. Preprocessing

Tahapan preprocessing adalah tahapan untuk membersihkan, menata dan menstruktur data mentah yang tidak terstruktur dan memiliki banyak noise yang berupa tanda baca atau kalimat yang tidak berarti. Preprocessing data memiliki 4 tahap:

1. Case Folding dan Cleaning

Merupakan proses untuk membuat semua text menjadi lowercase dan menghilangkan tanda baca {full stops (.), commas (,), question marks (?), exclamation marks (!), colons (:), semi-colons (;), apostrophes (') and speech marks (","")}, tag- tag HTML, angka dan bahkan emoji dikarenakan dataset ini merupakan hasil dari website IMDb. Proses cleaning dilakukan dengan menggunakan regex atau bisa juga disebut regular expression. Regex itus sendiri adalah konstruksi dalam suatu bahasa untuk mencocokkan teks berdasarkan pola tertentu, terutama untuk kasus-kasus kompleks

TABEL 3.1
CONTOH HASIL CLEANING DAN CASE FOLDING

Tweet Sebelum *Cleaning*	*Tweet* Sesudah *Cleaning*
One of the other reviewers has mentioned that after watching just 1 Oz episode you'll be hooked. They are right, as this is exactly what happened with me. The first thing that struck me about Oz was its brutality and unflinching scenes of violence, which set in right from the word GO. Trust me, this is not a show for the faint hearted or timid. This show pulls no punches with regards to drugs, sex or violence. Its is hardcore, in the classic use of the word.....	one of the other reviewers has mentioned that after watching just oz episode youll be hooked they are right as this is exactly what happened with methe first thing that struck me about oz was its brutality and unflinching scenes of violence which set in right from the word go trust me this is not a show for the faint hearted or timid this show pulls no punches with regards to drugs sex or violence its is hardcore in the classic use of the word..

2. Stopword removal

Stop words removal merupakan proses untuk membuang kata-kata yang umum atau dianggap tidak penting dalam pemrosesan data. Misalnya, kata-kata yang terdapat pada library NLTK(Natural Language Toolkit) di python dan kata yang frekuensi kemunculan sangat banyak.

List NLTK:

{'ourselves', 'hers', 'between', 'yourself', 'but', 'again', 'there', 'about', 'once', 'during', 'out', 'very', 'having', 'with', 'they', 'own', 'an', 'be', 'some', 'for', 'do', 'its', 'yours', 'such', 'into', 'of', 'most', 'itself', 'other', 'off', 'is', 's', 'am', 'or', 'who', 'as', 'from', 'him', 'each', 'the', 'themselves', 'until', 'below', 'are', 'we', 'these', 'your', 'his', 'through', 'don', 'nor', 'me', 'were', 'her', 'more', 'himself', 'this', 'down', 'should', 'our', 'their', 'while', 'above', 'both', 'up', 'to', 'ours', 'had', 'she', 'all', 'no', 'when', 'at', 'any', 'before', 'them', 'same', 'and', 'been', 'have', 'in', 'will', 'on', 'does', 'yourselves', 'then', 'that', 'because', 'what', 'over', 'why', 'so', 'can', 'did', 'not', 'now', 'under', 'he', 'you', 'herself', 'has', 'just', 'where', 'too', 'only', 'myself', 'which', 'those', 'i', 'after', 'few', 'whom', 't', 'being', 'if', 'theirs', 'my', 'against', 'a', 'by', 'doing', 'it', 'how', 'further', 'was', 'here', 'than'}

TABEL 3.2
CONTOH HASIL REMOVE STOPWORD

Tweet Sebelum *Stopword*	*Tweet* Sesudah *Stopword*
one of the other reviewers has mentioned that after watching just oz episode youll be hooked they are right as this is exactly what happened with methe first thing that struck me about oz was its brutality and unflinching scenes of violence which set in right from the word go trust me this is not a show for the faint hearted or timid this show pulls no punches with regards to drugs sex or violence its is hardcore in the classic use of the word....	reviewers mentioned watching oz episode youll hooked right exactly happened methe first thing struck oz brutality unflinching scenes violence set right word go trust show faint hearted timid show pulls punches regards drugs sex violence hardcore classic use word…

3. Stemming
Stemming yaitu mencari kata dasar dengan mengilangkan kata imbuhan seperti '-ers', '-ed', 'ing' dan mengelompokannya dari text data yang telah digunakan di penelitian ini. Dalam proses ini dilakukan juga menggunakan bantuan library nltk PorterStemmer pada bahasa pemrograman.

TABEL 3.3
DATA HASIL STEMMING

Tweet Sebelum *Stemming*	*Tweet* Sesudah *Stemming*
reviewers mentioned watching oz episode youll hooked right exactly happened methe first thing struck oz brutality unflinching scenes violence set right word go trust show faint hearted timid show pulls punches regards drugs sex violence hardcore classic use word…	review mention watch oz episode youll hook right exactli happen methe first thing struck oz brutal unflinch scene violence set right word go trust show faint heart timid show pull punch regard drug sex violence hardcore classic use word….

4. Tokenization
Merupakan proses untuk memecahkan kalimat untuk menjadi beberapa bagian yang dinamakan token. Sebuah token dapat dianggap menjadi satu bentuk sebuah kata, frasa, atau suatu elemen yang berarti. Dalam proses ini dilakukan juga menggunakan bantuan library nltk pada bahasa pemrograman.

Tweet Sebelum *Tokenizaton*	*Tweet* Sesudah *Tokenization*
review mention watch oz episode youll hook right exactli happen methe first thing struck oz brutal unflinch scene violence set right word gotrust show faint heart timid show pull punch regard drug sex violence hardcore classic use word….	[review, mention, watch, oz, episode, youll, hook, right, exactli, happen, methe, first, thing, struck, oz, brutal, unflinch, scene, violence, set, right, word, go, trust, show, faint, heart, timid, show, pull, punch, regard, drug, sex, violence, hardcore, classic, use, word,…]

E. Ekstraksi Fitur

Dalam kasus ini setelah semua data dikumpulkan dan hasil data tersebut telah melewati proses preprocessing, proses selanjutnya adalah membuat fitur yang berguna untuk mempermudah proses pengklasifikasian data tweet tersebut biasa proses ini dibuat dengan proses ekstraksi fitur. Pada proses ekstraksi fitur ini terdapat dua proses yang dilakukan yaitu proses pembuatan word vector dalam proses ini sistem akan mengubah suatu teks menjadi representasi vector dan proses pembobotan kata menggunakan Tf-Idf (term frequency inverse document frequency). TF atau Term Frequency itu sendiri adalah banyaknya frekuensi kemunculan kata dari suatu term dalam dokumen bersangkutan, sedangkan IDF atau Inverse Document Frequency adalah perhitungan dari bagaimana term disebarkan atau didistribusikan secara luas dalam koleksi dokumen yang bersangkutan. Setalah semua kata diproses dan berubah menjadi vektor kata, selanjutnya adalah proses pemberian bobot dari setiap kata pada setiap kalimat atau dokumen menggunakan Tf-Idf (term frequencyinverse document frequency).

Pada proses ekstraksi fitur, proses pertama yang dilakukan oleh sistem yaitu mengubah dataset menjadi suatu representasi vector dengan menggunakan library yang sudah disediakan oleh Phyton yang bernama library CountVectorizer.

Sebagai contoh penelitian menggunakan komentar positif dan negatif, diantarnya:

Positive Class :
(Doc1) “I thought this was a wonderful way to spend time on a too hot summer weekend, sitting in the air conditioned theater and watching a light-hearted comedy.”
(Doc2) “One of the other reviewers has mentioned that after watching just 1 Oz episode you'll be hooked.”
(Doc3) “I though this film is wonderful,it the first film i had watched at the cinema the picture was dark in places i very nervous it was back in 74/75”

Negative Class :
(Doc1) “I am a big fan. I like movies of all types. I think this is arguably the worst I've ever seen. I get that it follows the book closely, which raises the point that not everything”
(Doc2) “I couldn't believe this worst movie was actually made at all. I think with the worst actors you could find”
(Doc3) “I think Micheal Ironsides acting career must be over, if he has to star in this sort of low budge crap.”

Setelah sistem melakukan preprocessing dapat diambil kata dari kalimat di atas

Setelah tahapan di atas dari setiap dokumen ditampilkan mejadi sebuah vector dengan elemen, ketika kata tersebut terdapat di dalam dokumen maka diberikan nilai 1, jika tidak ada maka diberikan nilai 0. Sebagai contoh terdapat pada Tabel 3.4 di bawah.

TABEL 3.4
CONTOH PEMBOBOTAN POSITIF

	Thought	Was	Review	Mention	Film	Wonderful	Watch
Doc1	1	1	0	0	0	1	1
Doc2	0	0	1	1	0	0	1
Doc3	1	2	0	0	2	1	1

TABEL 3.5
CONTOH PEMBOBOTAN NEGATIF

	Movie	Was	Carrier	Worst	Over	Think	Believe
Doc1	1	0	0	1	0	1	0
Doc2	1	1	0	1	0	1	1
Doc3	0	0	1	0	1	1	0

Proses perhitungan bobot kata dilakukan dengan proses awal menghitung TF atau Term Frequencyterlebih dahulu.

TABEL 3.6
PROSES PERHITUNGAN TF POSITIF (TERM FREQUENCY)

T (Term)	Doc1	Doc2	Doc3
Thought	1	0	1
Was	1	0	2
Review	0	1	0
Mention	0	1	0
Film	0	0	2
Wonderful	1	0	1
Watch	1	1	1

T (Term)	Doc1	Doc2	Doc3
Movie	1	1	0
Was	0	1	0
Carrier	0	0	1
Worst	1	1	0
Over	0	0	1
Think	1	1	1
Believe	0	1	0

Setelah proses perhitungan bobot TF selesai selanjutnya dilakukan proses menentukan DF atau Document Frequency yaitu dengan banyaknya term (t) muncul dalam semua dokumen. Maka akan memperoleh hasil seperti Tabel 3.7 di bawah

TABEL 3.7
PROSES PERHITUNGAN DF (DOCUMENT FREQUENCY)

T (Term)	DF
Thought	2
Was	3
Review	1
Mention	1
Film	2
Wonderful	2
Watch	3

T (Term)	DF
Movie	2
Was	1
Carrier	1
Worst	2
Over	1
Think	3
Believe	1

Kemudian proses menghitung nilai IDF (Inverse Document Frequency) dengan cara menghitung nilai dari log hasil D atau jumlah dokumen dalam contoh kasus ini ada 3 dokumen, dari 3 dokumen tersebut dibagi dengan nilai DF (Document Frequency).

TABEL 4.1
PROSES IDF (INVERSE DOCUMENT FREQUENCY)

T (Term)	DF	D/DF	IDF
Thought	2	1,5	0,18
Was	3	1	0
Review	1	3	0,48
Mention	1	3	0,48
Film	2	1,5	0,18
Wonderful	2	1,5	0,18
Watch	3	1	0

T (Term)	DF	D/DF	IDF
Movie	2	1,5	0,18
Was	1	3	0,48
Carrier	1	3	0,48
Worst	2	1,5	0,18
Over	1	3	0,48
Think	3	1	0
Believe	1	3	0,48

Setelah mendapatkan nilai IDF (Inverse Document Frequency), selanjutnya dilanjutkan dengan menghitung TF-IDF.

TABEL 4.2
CONTOH PROSES PERHITUNGAN TF-IDF

T	TF			DF	IDF	W = TF*IDF		
	Doc 1	Doc 2	Doc 3			Doc 1	Doc 2	Doc 3
Thought	1	0	1	2	0,18	0,18	0	0,18
Was	1	0	2	3	0	0	0	0
Review	0	1	0	1	0,48	0	0,48	0
Mention	0	1	0	1	0,48	0	0,48	0
Film	0	0	2	2	0,18	0	0	0,35
Wonderful	1	0	1	2	0,18	0,18	0	0,18
Watch	1	1	1	3	0	0	0	0

T	TF			DF	IDF	W = TF*IDF		
	Doc 1	Doc 2	Doc 3			Doc 1	Doc 2	Doc 3
Movie	1	1	0	3	0,18	0,18	0,18	0
Was	0	1	0	1	0,48	0	0,48	0
Carrier	0	0	1	1	0,48	0	0	0,48
Worst	1	1	0	2	0,18	0,18	0,18	0
Over	0	0	1	1	0,48	0	0	0,48
Think	1	1	1	3	0	0	0	0
Believe	0	1	0	1	0,48	0	0,48	0

Hasil dari word vector yang sudah mendapatkan bobot dapat dilihat pada Tabel 4.3 di bawah.

TABEL 4.3
CONTOH WORD VECTOR YANG SUDAH DIBOBOTKAN

	Thought	Was	Review	Mention	Film	Wonderful	Watch
Doc1	0,18	0	0	0	0	0,18	0
Doc2	0	0	4,8	4,8	0	0	0
Doc3	0,18	0	0	0	0,35	0,18	0

	Movie	Was	Carrier	Worst	Over	Think	Believe
Doc1	0,18	0	0	0,18	0	0	0
Doc2	0,18	0,48	0	0,18	0	0	0,48
Doc3	0	0	4,8	0	4,8	0	0

Dokumen yang telah diubah menjadi word vector selanjutnya akan dihitung menggunakan rumus TF-IDF, dengan menggunakan rumus ini maka akan menghasilkan word vector yang memiliki nilai yang sudah terbobot. Kemudian untuk melakukan proses pengklasifikasian diperlukan hasil dari data yang sudah diolah dari proses sebelumnya yaitu hasil dari proses preprocessing dan hasil dari pembobotan kata dengan Tf-Idf. Jika proses pembobotan selesai maka dataset dapat digunakan dalam perhitungan klasifikasi Naïve Bayes Classifier.

Proses Klasifikasi Naive Bayes

1. P(Positif | Thought, Was, Review, Mention, Film, Wonderful, Watch) =

P(Positif) x P(Thought|Positif) x P(Was|Positif) x P(Review|Positif) x P(Mention|Positif) x P(Film|Positif) x P(Wonderful|Positif) x P(Wacth|Positif)

2. P(Negatif | Movie, Was, Carrier, Worst, Over, Think, Believe) =

P(Negatif) x P(Movie|Negatif) x P(Was|Negatif) x P(Carrier|Negatif) x P(Worst|Negatif) x P(Over|Negatif) x P(Think|Negatif) x P(Believe|Negatif)

IV. HASIL DAN PEMBAHASAN

A. Hasil Pengujian

Setelah proses Tf-Idf didapat kelas Positif dan Negatif yang memiliki term tinggi. Lalu akan dicari probabilitas kemunculan kata dengan term tinggi tadi pada seluruh dokumen.

TABEL 4.1
TERM KATA DALAM REVIEW

	Wonderful	Thought	Movie	Worst
Positif	5	3	1	0
Negatif	1	1	3	3

Menghitung probabilitas:

Pr(Negatif) = $\frac{Kelas\ Negatif}{Data\ Train} = \frac{2}{4} = 0{,}5$

Pr(Positif) = $\frac{Kelas\ Positif}{Data\ Train} = \frac{2}{4} = 0{,}5$

Menghitung probabilitas setiap kata:

Pr(Thought|Positif) = $\frac{4}{9} = 0{,}45$

Pr(Wonderful|Positif) = $\frac{6}{9} = 0{,}67$

Pr(Movie|Positif) = $\frac{2}{9} = 0{,}22$

Pr(Worst|Positif) = $\frac{1}{9} = 0{,}11$

Pr(Thought|Negatif) = $\frac{2}{8} = 0{,}25$

Pr(Wonderful|Negatif) = $\frac{2}{8} = 0{,}25$

Pr(Movie|Negatif) = $\frac{4}{8} = 0{,}5$

Pr(Worst|Negatif) = $\frac{4}{8} = 0{,}5$

Selanjutnya adalah melakukan perkalian nilai pada probabilitas setiap kelas uji pada contoh review:

Doc1 : “I thought that Mukhsin has wonderfull written. Its not just about entertainment.”

P(X|Positif) =Pr(Posistif) x Pr(Thought|Positif) x Pr(Wonderful|Positif)
= 0,5 x 0,45 x 0,67
= 0,15

P(X|Negatif) = Pr(Positif) x Pr(Thought|Negatif) x Pr(Wonderful|Negatif)
= 0,5 x 0,25 x 0,25
= 0,03

Berarti review Doc1 termasuk kelas **Positif**

Tahapan akhir setelah melakukan semua proses pengklasifikasian, maka barulah bisa menghitung dari performa dari algoritme yang dipergunakan. Untuk mengetahui tingkatan dari performa Algoritme Naive Bayes, maka dilakukan pengujian terhadap model. Hasil dari klasifikasi nantinya akan ditampilkan dalam bentuk confusion matrix. Tabel yang ditampilkan di dalam confusion matrix ini terdiri dari kelas predicted dan juga kelas actual. Model dari confusion matrix ini dapat dilihat pada Tabel dibawah

TABEL 4.2
HASIL CONFUSION MATRIX

		Predict Class	
		Positive	***Negative***
Actual class	***Positive***	4339	638
	Negative	714	4309

Untuk mengetahui nilai dari akurasi model diperoleh dari banyak jumlah data yang tepat hasil klarifikasi dibagi dengan total dari data, seperti pada Gambar 4.1 di bawah

$$\text{Akurasi} = \frac{AA+BB}{AA+AB+BA+BB}$$

GAMBAR 4.1
RUMUS AKURASI

Dalam proses evaluasi model ini dilakukan setelah uji model telah selesai dilakukan. Evaluasi model berguna sebagai menghitung performa dari metode yang dipilih. Pada proses uji model ini akan menghasilkan confusion matrix dengan ukuran 2x2.

Nilai akurasi yang didapatkan dari pengujian model sebesar 86.48% yang proses perhitungannya berdasarkan jumlah nilai dari diagonal confusion matrix dibagi dengan seluruh jumlah data seperti berikut:

TABEL 4.3
MODEL CONFUSION MATRIX

		Predict Class	
		Class A	***Class B***
Actual class	***Class A***	AA	AB
	Class B	BA	BB

$$\text{Akurasi} = \frac{AA+BB}{AA+AB+BA+BB}$$

$$\text{Akurasi} = \frac{4339+4309}{4339+638+714+430} = 86.48\ \%$$

GAMBAR 4.2
HASIL PERHITUNGAN

Dengan diketahuinya nilai dari precision, recall, dan f-1 Score dalam kinerja di keseluruhan sistem, maka dapat mengetahui kemampuan dari sistem untuk mencari ketepatan atau kebenaran dari informasi yang diminta oleh pengguna dengan hasil jawaban yang dikeluarkan oleh sistem dan memberitahu tingkat keberhasilan dari suatu sistem dalam menentukan kembali suatu informasi atau nilai accuracy.

Hasil dari precision, recall, dan f-1 Score memiliki ukuran penilaian sebesar 0-1. Semakin tinggi nilai maka semakin baik, dalam artian semakin mendekati angka 1 nilai dari 0 maka sistem semakin baik. Hasil dari nilai precision, recall, dan f-1 Score di setiap kelas terdapat pada Tabel 4.9 di bawah.

TABEL 4.4
HASIL DARI NILAI PRECISION, RECALL, DAN F-1 SCORE

Jenis Klasifikasi	***Precision***	***Recall***	***F-1 Score***
Positif	0,86	0,87	0,87
Negatif	0,87	0,86	0,86

Adapun Hasil Akurasi, Nilai Precision, Recall, dan F-1 score dari data test dan data train :

1. 40000 Data Train 10000 Data Test = 86.48%

Jenis Klasifikasi	*Precision*	*Recall*	*F-1 Score*
Positif	0,86	0,87	0,87
Negatif	0,87	0,86	0,86

2. 35000 Data Train 15000 Data Test = 85.93%

Jenis Klasifikasi	*Precision*	*Recall*	*F-1 Score*
Positif	0,85	0,87	0,86
Negatif	0,87	0,85	0,86

3. 30000 Data Train 20000 Data Test = 85.82%

Jenis Klasifikasi	*Precision*	*Recall*	*F-1 Score*
Positif	0,85	0,87	0,86
Negatif	0,87	0,85	0,86

4. 25000 Data Train 25000 Data Test = 85.85%

Jenis Klasifikasi	*Precision*	*Recall*	*F-1 Score*
Positif	0,86	0,86	0,86
Negatif	0,86	0,85	0,86

B. Analisi Hasil Pengujian

Dapat dilihat dari hasil uji model dapat dilihat nilai precision, dan recall dari setiap kelas dapat dilihat tingkat kemampuan pemrosesan sistem dalam mencari tingkat ketepatan antara informasi yang diinginkan oleh pengguna sebagai kelas positif adalah “86%”, dan untuk kelas negatif adalah “87%”. Tingkat keberhasilan dari pemrosesan sistem dalam memperoleh kembali informasi kelas positif adalah “87%”, untuk kelas negatif adalah “86%” dan Accuration sebagai pembanding antara informasi yang dijawab oleh sistem dengan benar yaitu sebesar 86.48%.

TABEL 4.5
PERBANDINGAN DENGAN PENELITIAN TERDAHULU

Penelitian	Dataset	Proses	Hasil
Palak Baid dkk (Naive Bayes, KNN, Random Forest)	2000 dataset 1000 data positive 1000 data negative	WEKA tool	Naive Bayes = 81.4% KNN = 55.30% Random Forest = 78.65%
Gurshobit Singh dkk	Data TMDB 10 review pertama	Based Opinion Mining	81.22%
Riko dkk (Naive Bayes)	4000 dataset	Gini Index	59.54%
Ahmad Zuili dkk (Naive Bayes)	1400 dataset 700 data positive 700 data negative	Chi Square	64.40%
Risky Novendri dkk (Naive Bayes)	1000 dataset	Tf-Idf	81%

Pada Tabel 4.5, menunjukkan hasil akurasi dan F1 yang berbeda-beda berdasarkan kondisi dataset, dan proses ekstraksi fitur yang berbeda. Penelitian Palak Baid dkk [5] untuk deteksi sentiment analisis dengan metode Naive Bayes, KNN, Random Forest, berhasil mendapatkan akurasi dan F1 sebesar 81.4% untuk Naïve Bayes dengan distribusi jumlah kelas yang cukup seimbang. Penelitian [6], jika dilihat dari jumlah dataset, dapat dibilang sebagai dataset yang digunakan tidak seimbang atau imbalanced tetapi menghasilkan akurasi 81.22% karena data yang dipakai sedikit. Lalu pada penelitian Riko dkk [8] dengan 4000 dataset dan penggunaan fitur Gini Index menghasilkan akurasi yang sangat rendah yaitu 59.54% jika dibandingkan dengan kondisi data pada penelitian Ahmad Zuili dkk [9] dengan data yang seimbang menghasilkan akurasi yang lebih tinggi , yaitu 64.40% Dibandingkan dengan penelitian yang dilakukan oleh Risky Novendri dkk. [10], penelitian ini menggunakan 1000 dataset yang didapat dari crawling data di

You Tube dengan fitur Tf-Idf, dimana data yang diproses tidak seimbang sehingga akurasi hanya 81%.

Perbandingan banyaknya data yang tidak seimbang dan penggunaan fitur pada penelitian yang diusulkan dengan penelitian-penelitian yang telah disebutkan diatas, dapat menjadi salah satu penyebab nilai akurasi dan F1 yang tidak dapat mencapai diatas 81%. Pada penelitian ini digunakan 50000 dataset dimana data seimbang 25000 positive dan 25000 negative. Dengan ukuran dataset yang cukup besar, digunakan lah fitur pemangkasan guna menghindari banyaknya dimensi agar ruang pencarian tidak tinggi dan dan tidak menyulitkan kinerja system serta membuat data konsisten. Fitur pemangkasan ini yaitu proses Tf-Idf sehingga pada penelitian ini menghasilkan akurasi yang akurat sebesar 86.48%. Terbukti pada penelitian Nurulhuda dkk [17] pada proses metode machine learning, suatu fitur direpresentasikan kedalam bentuk vektor. Nilai vektor didapatkan dari proses pembobotan term, pemilihan metode pembobotan term yang tepat dapat mempengaruhi performansi system. Terdapat beberapa metode pembobotan kata, diatara metode pembobotan tersebut, TF-IDF memiliki performansi terbaik.

V. KESIMPULAN

Pada penelitian ini dilakukan klasifikasi sentimen analisis review film dengan klasifikasi Naïve Bayes. Karena Naïve Bayes dapat bekerja dengan efisien dan bekerja cepat sebagai metode pengklasifikasian teks. Studi ini menggunakan dataset ulasan film, dataset 50.000 data. Dari pengolahan data yang telah dilakukan, penggunaan metode seleksi fitur yaitu Tf-Idf dapat meningkatkan akurasi pengklasifikasi Naïve Bayes. Data ulasan film dapat diklasifikasikan dengan baik menjadi ulasan positif dan ulasan negatif. Akurasi Naïve Bayes pada pengolahan data IMDb menggunakan Tf-Idf sebagai seleksi fitur mencapai 86.48%. Sehingga Naïve Bayes dengan fitur Tf-Idf pada masalah analisis klasifikasi sentimen review film terbukti memberikan akurasi yang akurat.

A. Saran

Dari hasil yang dikerjakan dalam penelitian ini masih mempunyai kekurangan dalam metode Naive Bayes Classifier untuk menentukan kemungkinan data. Diharapkan dalam penelitian berikutnya proses penelitian yang dilakukan dapat menggunakan metode atau algoritme klasifikasi yang lain yang berguna sebagai pembanding hasil untuk mencari nilai klasifikasi terbaik.

REFERENSI

[1] Hussein, D.M., 2016. A Survey on Sentiment Analysis Challenges, Cairo: Journal of King Saud University

[2] Khan, M.T., Durrani, M., Ali, A., Inayat, I., Khalid, S., Khan, H., 2016. Sentiment analysis and the complex natural language, Pakista: Complex adaptive system modeling.

[3] Dehkharghani R Mercan H Javeed A and Saygin Y, 2014 Expert Systems with Applications Sentimental causal rule discovery from Twitter Expert Syst. Appl. 41, 10 p. 4950–4958.

[4] Ahmad E Sazzad M A U Islam M T Azad M Islam S and Ali M H, 2017 Challenges, Comparative Analysis and a Proposed Methodology to Predict Sentiment from Movie Reviews using Machine Learning Big Data Analytics and Computational Intelligence (ICBDAC), 86-91.

[5] Baid, Palak & Gupta, Apoorva & Chaplot, Neelam. (2017). Sentiment Analysis of Movie Reviews Using Machine Learning Techniques. International Journal of Computer Applications. 179. 45-49. 10.5120/ijca2017916005.

[6] Brar, G.S., Sharma, A., 2018 Sentiment Analysis of Movie Review Using Supervised Machine Learning Techniques

[7] Potti, M.P., Kmar, M.D., Ram, N.S., Sandeep, P.V.R., Prasad, P.R.K., 2018. Sentiment Analysis On Movie Reviews Using NAÏVE BAYES Classifier

[8] Purnomoputra, R., Adiwijaya, A., & Novia Wisesty, U. (2019, November 15). Sentiment Analysis of Movie Review using Naïve Bayes Method with Gini Index Feature Selection.

[9] Amrullah, Ahmad & Anas, Andi & Adrian, Muh & Hidayat, Muh. Adrian Juniarta. (2020). Analisis Sentimen Movie Review Menggunakan Naive Bayes Classifier Dengan Seleksi Fitur Chi Square. 2. 40-44.

[10] Novendri, R., Callista, A. S. ., Pratama, D. N., & Puspita, C. E. . (2020). Sentiment Analysis of YouTube Movie Trailer Comments Using Naïve Bayes. Bulletin of Computer Science and Electrical Engineering, 1(1),26–32.

[11] Zulfa, I., & Winarko, E. (2017, July). Sentimen Analisis Tweet Berbahasa Indonesia dengan. IJCCS, 11, 2.

[12] Rozi, I. F., Pramono, S. H., & Dahlan, E. A. (2013). Implementasi opinion mining (analisis sentimen) untukekstraksi data opini publik pada perguruan tinggi. Jurnal EECCIS, 6(1), 37-43.

[13] [6] R. S. P. M. A. F. A. R. T. Lestari, "Analisis Sentimen Tentang Opini Pilkada Dki 2017 Pada Dokumen Twitter Berbahasa Indonesia Menggunakan Näive Bayes dan Pembobotan Emoji," Jurnal Pengembangan

Teknologi Informasi dan Ilmu Komputer, vol. 1, no. 12, pp. 1718-1724, 2017

[14] McCallum, A., & Nigam, K. 1998. A Comparison of Event Models for Naïve Bayes Text Classification.
Pittsburgh: Carnegie Mellon University

[15] Rahman, A., Wiranto, & Doewes, A. 2017. "Online News Classification Using Multinomial Naive Bayes".
ITSMART Vol. 6 No.1. Universitas Sebelas Maret

[16] Pendit, Putu Laxman., 2008. Perpustakaan Digital Dari A Sampai Z. Jakarta: Cita Karya Karsa Mandiri.

[17] N. Zainuddin dan A. Selamat, "Sentiment Analysis Using Support Vector Machine," International Conference on Computer, Communication, and Control Technology , pp. 333-337, 2014