

ANALISIS SENTIMEN PADA ULASAN LMS PEMBELAJARAN MENGGUNAKAN METODE NATURAL LANGUAGE PROCESSING

Januar Putra Hidayat*¹⁾, Ida Nurhaida²⁾

1. Informatika, Fakultas Teknik dan Desain, Universitas Pembangunan Jaya, Indonesia
2. Informatika, Fakultas Teknik dan Desain, Universitas Pembangunan Jaya, Indonesia

Article Info

Kata Kunci: Coursera, komentar, algoritma, ensemble

Keywords: Coursera, comment, algorithm, ensemble

Article history:

Received 3 October 2024
Revised 7 November 2024
Accepted 10 December 2024
Available online 1 March 2025

DOI :

<https://doi.org/10.29100/jupi.v10i1.5899>

* Corresponding author.

Corresponding Author

E-mail address:

januar.putrahidayat@student.upj.ac.id

ABSTRAK

Coursera merupakan sebuah LMS yang dapat di akses oleh siapa saja dan di mana saja. Bagi pengguna android dapat mengunduhnya melalui *Google Play Store*. Selain mengunduh pengguna juga dapat memberikan nilai dan komentar terhadap aplikasi yang di gunakan. Komentar masyarakat yang ada di *Google Play Store* cukup beragam dan memiliki sifat acak. Agar dapat mengetahui komentar yang di tulis oleh pengguna merupakan komentar positif atau negatif, maka dilakukan analisis sentimen terhadap komentar pengguna. Peneliti menggunakan data komentar dari 3 negara sebagai pembandingan tren komentar setiap negara. Selain itu juga peneliti menggunakan 3 algoritma berbeda untuk menghitung hasil dari komentar masyarakat. 3 algoritma yang digunakan ialah *Naive Bayes*, SVM, dan KNN. Untuk mendapatkan hasil yang maksimal dari analisis sentimen maka digunakan metode *ensemble* dengan menggabungkan 3 algoritma. Dengan menggunakan metode *ensemble* yang di dapatkan menjadi lebih optimal sehingga mengetahui tren komentar masyarakat dari 3 negara lebih mengarah ke positif. Dari ketiga negara, Amerika merupakan negara dengan presentase komentar positif tertinggi.

ABSTRACT

Coursera is an LMS that can be accessed by anyone and anywhere. Android users can download it via Google Play Store. Apart from downloading, users can also provide ratings and comments on the applications they use. Public comments on the Google Play Store are quite diverse and random in nature. In order to know whether comments written by users are positive or negative, sentiment analysis is carried out on user comments. Researchers used comment data from 3 countries as a comparison of comment trends for each country. Apart from that, researchers also used 3 different algorithms to calculate the results of public comments. The 3 algorithms used are Naive Bayes, SVM, and KNN. To get maximum results from sentiment analysis, an ensemble method is used by combining 3 algorithms. By using the ensemble method the results obtained are more optimal so that the trend in public comments from the 3 countries is more positive. Of the three countries, America is the country with the highest percentage of positive comments.

I. PENDAHULUAN

BERKEMBANGNYA teknologi dan informasi membuat metode dalam pembelajaran semakin berkembang. Metode pembelajaran yang semula hanya dapat dilakukan secara tatap muka kini dapat dilaksanakan secara daring. Pembelajaran secara daring ini dapat disebut *e-learning (electronic learning)*. *E-learning* merupakan proses pembelajaran menggunakan teknologi internet sebagai alat bantu dalam fasilitas, penyampaian, dan kemungkinan dalam pembelajaran jarak jauh [1].

Dalam melaksanakan kegiatan belajar secara daring perlu yang namanya LMS (*Learning Management System*). LMS merupakan sebuah perangkat lunak yang dapat digunakan oleh pengajar untuk manajemen, monitoring, administrasi dan distribusi materi pengajaran kepada murid. Dengan diberlakukannya sistem pelajaran secara *hybrid*, banyak sekolah dan perguruan tinggi mulai menggunakan LMS yang ada. Dengan adanya LMS membuat lembaga pendidikan mampu menyelenggarakan *e-learning* [2]. Apalagi menurut Pelaksana Tugas (Plt.) Direktur Jendral Pendidikan Tinggi Kementerian Pendidikan dan Kebudayaan (Kemendikbud) dalam acara temu media

Bincang Sore secara virtual (24/6/2020) dikatakan bahwa terdapat 8 juta mahasiswa dan 300 ribu dosen yang beralih ke dalam pembelajaran daring pada bulan Maret [3]. Seiring banyaknya sekolah maupun perguruan tinggi yang menggunakan LMS, mulai banyak layanan-layanan penyedia LMS yang beredar. Agar LMS banyak dipakai oleh lembaga pendidikan maka penyedia layanan LMS kini berlomba-lomba menyediakan fitur yang membantu dan memudahkan baik pengajar maupun murid. Salah satu contohnya ialah Coursera. Coursera sendiri merupakan aplikasi LMS yang memberikan kursus *online* dan sertifikasi kepada penggunanya. Coursera didirikan pada 2012 oleh Daphne Koller dan Andrew Ng, 2 orang profesor dalam ilmu komputer dari Universitas Stanford.

Terlepas dari fungsinya, setiap LMS pasti memiliki kelebihan dan kekurangan. Kelebihan yang ada memang dapat membuat LMS menjadi lebih unggul. Namun, kekurangan dari LMS biasanya menimbulkan berbagai komentar negatif dari masyarakat khususnya dalam media sosial. Media sosial sendiri merupakan tempat bagi masyarakat bebas berkespresi dan mengungkapkan berbagai hal baik itu positif maupun negatif. Begitu juga dengan komentar terhadap layanan LMS. Komentar yang diucapkan oleh pengguna media sosial terkadang dapat memberikan informasi kepada masyarakat lain baik itu positif atau negatif. Seperti contoh dari penelitian yang diadakan di UIN Malang terkait LMS *Moodle* dimana dalam penelitian tersebut terdapat 64 responden dengan rincian 48 komentar positif dan 16 komentar negatif [4]. Komentar-komentar inilah yang seringkali digunakan oleh penyedia layanan aplikasi untuk menjadi bahan pertimbangan mereka kedepannya.

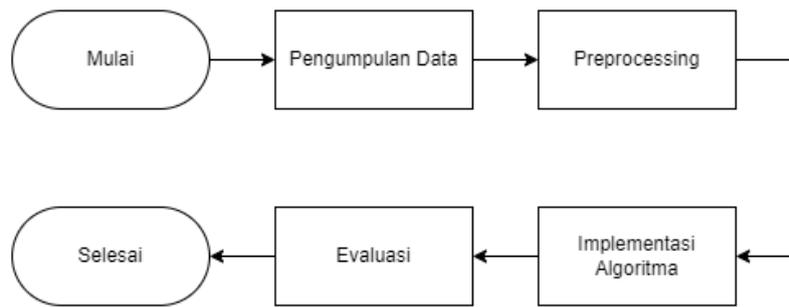
Untuk melihat dan mengklasifikasikan komentar pengguna terhadap aplikasi coursera, peneliti menggunakan platform *Google Play Store* sebagai data untuk analisis sentimen masyarakat terhadap coursera. Analisis Sentimen proses dalam memahami dan mengolah data yang berbentuk teks secara otomatis guna mendapatkan informasi yang dicari [5]. Dengan adanya analisis sentimen juga dapat melihat opini publik terhadap sebuah isu [6]. *Google Play Store* merupakan tempat bagi pengguna android untuk mengunduh aplikasi. Selain mengunduh aplikasi, pengguna juga dapat memberikan komentar dan penilaian terhadap aplikasi tersebut. Komentar dan penilaian inilah yang peneliti gunakan sebagai data dalam penelitian. Karena komentar yang terdapat dalam *Google Play Store* cenderung beragam dan tidak teratur, maka digunakan metode analisis sentimen agar dapat mengklasifikasikan komentar-komentar tersebut [7].

Peneliti menggunakan metode *Naive Bayes*, *Support Vector Machine* (SVM), dan *K-Nearest Neighbors* (KNN) dalam melakukan perhitungan analisis sentimen. Penggunaan ketiga metode ini dikarenakan tingginya tingkat akurasi dibanding penggunaan metode lainnya seperti *Decision Tree* dan *Neural Networks*. Dalam jurnal yang dilakukan oleh Xhemali, ia mengatakan bahwa metode *Naive Bayes* memiliki akurasi sebesar 95,18%. Nilai ini lebih tinggi dibanding akurasi yang didapat dari metode *Decision Tree* dan *Neural Network* [8]. Selain itu juga SVM memiliki keunggulan dimana data nonlinear dapat dipisahkan serta memiliki dimensi luas dengan memanfaatkan fungsi kernel agar bekerja lebih efektif [9]. Berdasarkan penelitian sebelumnya terkait analisis sentimen amazon pada kolom komentar *Google Play Store* menggunakan algoritma *Naive Bayes* mendapatkan akurasi sebesar 88,37% untuk data 80:20 [10]. Lalu jika melihat penelitian terkait analisis sentimen halodoc didapatkan hasil algoritma *Naive Bayes* memiliki akurasi 92,50%, SVM memiliki akurasi 93%, dan KNN memiliki akurasi 95% [11].

Berdasarkan latar belakang di atas peneliti menggunakan 3 algoritma karena bertujuan untuk menggabungkan ketiga algoritma tersebut dengan metode *ensemble*. Metode *ensemble* merupakan metode menggabungkan algoritma dan melakukan voting untuk mendapatkan hasil terbaik. Hasil inilah yang nantinya akan digunakan sebagai hasil dalam analisis sentimen. Adanya analisis sentimen terhadap coursera dengan data yang didapat dari *Google Play Store* dapat berguna untuk melihat pendapat masyarakat terhadap coursera. Pendapat ini nantinya akan di klasifikasikan apakah pendapat tersebut merupakan positif, netral atau negatif. Dengan adanya analisis sentimen ini nantinya calon pengguna atau bahkan lembaga pendidikan yang ingin menggunakan coursera sebagai LMS pengajaran dapat melihat hal yang positif dan negatif dari aplikasi coursera serta dapat melihat penilaian masyarakat baik dari Indonesia maupun luar negeri terkait aplikasi tersebut.

II. METODE PENELITIAN

Dalam melakukan penelitian, peneliti menggunakan sejumlah tahapan-tahapan. Berikut alur penelitian yang akan dilakukan seperti pada gambar 1.



Gambar 1 Diagram Alir Penelitian

A. Pengumpulan Data

Data ulasan pada LMS Coursera yang digunakan berasal dari *Google Play Store*. Data tersebut diambil dengan cara crawling menggunakan bahasa pemrograman python. Penggunaan *Google Play Store* sebagai sumber data dikarenakan penelitian terdahulu juga menggunakan aplikasi tersebut sebagai sumber data untuk melakukan analisis sentimen [12]. Untuk mendapatkan ulasan yang variatif, peneliti mengambil data dari 3 negara, yaitu Indonesia, Inggris, dan Amerika Serikat. Total data yang di dapat ialah dan data tersebut tersimpan dengan format CSV.

Peneliti menggunakan *library* dari *google-play-scraper* untuk dapat melakukan pengumpulan data. Penggunaan *library* ini dikarenakan mudahnya dalam melakukan pengumpulan data. Selain itu data yang dikumpulkan cukup lengkap dan pengguna dapat menentukan parameter apa saja yang akan dicari, seperti: negara, bahasa, aplikasi dalam *Google Play Store*, dan lain-lain.

B. Preprocessing

Pada tahap preprocessing, data yang telah dikumpulkan selanjutnya diolah ke dalam beberapa tahapan agar data dapat diberi label dan di implementasikan dengan algoritma. Preprocessing sendiri merupakan sebuah tahapan dalam penghapusan kesalahan dan ketidakberaturan data mentah agar lebih siap untuk dianalisis [13].

- 1) *Cleaning Data* merupakan tahapan dalam menghapus atau menghilangkan data teks yang tidak penting. Proses ini ditujukan agar data yang digunakan memiliki kualitas yang tinggi dan terhindar dari teks-teks tidak penting [14]. Karakter yang tidak memiliki bobot seperti “@”, “%”, “\$”, “!”, “?” dan lainnya akan di hapuskan.
- 2) *Case Folding* adalah sebuah proses dalam mengubah huruf yang ada di dalam teks menjadi huruf kecil semua. Tahapan ini berguna untuk mempermudah dalam pengolahan teks karena semua huruf merupakan huruf kecil [15].
- 3) *Tokenization* adalah proses dalam pemisahan setiap kata dari sebuah kalimat utuh. Proses ini bertujuan untuk mempermudah klasifikasi nantinya [16].
- 4) *Stopword* merupakan tahapan penghapusan kata tidak bermakna atau memiliki bobot rendah sehingga dapat mempersingkat kalimat tanpa mengurangi maknanya [17]. Kata seperti “banyak”, “yang”, “untuk”, ini”, “itu” dan lainnya.
- 5) *Stemming* adalah proses penghapusan kata berimbuhan dan mengubah kata berimbuhan menjadi kata dasar [18]. Untuk data yang didapat dari Indonesia program menggunakan *library sastrawi* yang sudah sesuai dengan kaidah bahasa Indonesia. Sedangkan untuk data yang di dapat dari Amerika Serikat dan Inggris menggunakan library NLTK.

```
1 import pandas as pd
2 import re
3 import nltk
4
5 from nltk.tokenize import word_tokenize
6 from nltk.corpus import stopwords
7 from nltk.stem import PorterStemmer
8 from nltk.stem import WordNetLemmatizer
9
```

Gambar 2 Library untuk Preprocessing

Dalam python untuk melakukan tahap *preprocessing* dapat menggunakan berbagai library namun pada penelitian kali ini peneliti menggunakan *library* re dan nltk untuk melakukan tahap *preprocessing*. Gambar 2 merupakan *library* yang di gunakan oleh peneliti.

C. Implementasi Algoritma

Setelah data diberikan label tahap berikutnya ialah melakukan implementasi algoritma. Peneliti menggunakan 3 algoritma yaitu *Naive Bayes*, *Support Vector Machine*, dan *K-Nearest Neighbors*. Ketiga algoritma ini nantinya akan dilakukan proses *ensemble* atau pemilihan hasil terbaik dengan penggabungan dari beberapa model yang ada dengan cara voting hasil [19].

1) *Naive Bayes* merupakan sebuah algoritma dalam klasifikasi yang berasal dari teorema Bayes [20]. *Naive Bayes* menggunakan metode probabilitas dan statistik dalam melakukan pengklasifikasian. Metode ini memungkinkan untuk memprediksi peluang berdasarkan pengalaman. Dalam menentukan kemungkinan kejadian, *Naive Bayes* menggunakan probabilitas.

$$P(A|B) = \frac{(P(B|A) * P(A))}{P(B)}$$

Keterangan:

$P(A|B)$ = Peluang kejadian A yang diberikan oleh B

$P(A|B)$ = Peluang kejadian B yang diberikan oleh A

$P(A)$ = Peluang dari A

$P(B)$ = Peluang dari B

Algoritma ini digunakan karena memiliki keunggulan di mana *Naive Bayes* menggunakan jumlah data latih dari data terkecil untuk ditentukan parameter yang diperlukan proses klasifikasi [21].

2) *Support Vector Machine* merupakan salah satu metode yang biasa digunakan dalam klasifikasi dan regresi. Untuk mendapatkan *hyperlane* (fungsi pemisah) yang optimal dalam memisahkan observasi dengan nilai variable target berbeda digunakan teknik SVM [22]. Algoritma ini di gunakan karena memiliki ruang dimensi yang besar serta memiliki performa yang bagus apabila mendapati data beratribut banyak [23]. SVM memiliki kelebihan dalam hal akurasi jika dibanding dengan algoritma lainnya. Namun tingkat akurasi dari algoritma ini dipengaruhi juga oleh parameter C (*cost*) dan *gamma*. Untuk mendapatkan hasil yang optimal maka penentuan parameter harus sesuai seperti nilai C yang tidak boleh terlalu lebar atau kecil dan nilai *gamma* yang tidak boleh terlalu tinggi atau terlalu rendah. Apabila kedua parameter ini tidak sesuai maka akan terdapat banyak kesalahan dalam melakukan prediksi.

Dalam implementasi algoritma SVM pertama-tama perlu memanggil library LinearSVC yang berasal dari modul sklearn.svm. Library ini berguna untuk melakukan perhitungan SVM. Data yang telah diolah selanjutnya masuk ke dalam pembagian data latih dan data tes dengan skala 80:20. Ubah teks menjadi vektor untuk memudahkan algoritma melakukan perhitungan. Data yang sudah menjadi vektor selanjutnya akan dihitung sesuai dengan perhitungan SVM.

```
1 for c in [0.01, 0.05, 0.25, 0.5, 0.75, 1]:
2     svm = LinearSVC(C=c)
3     svm.fit(x_train, y_train)
4     print('Akurasi untuk c = %s: %s' %(c, accuracy_score(y_test, svm.predict(x_test))))

Akurasi untuk c = 0.01: 0.9371069182389937
Akurasi untuk c = 0.05: 0.9371069182389937
Akurasi untuk c = 0.25: 0.9245283018867925
Akurasi untuk c = 0.5: 0.9182389937106918
Akurasi untuk c = 0.75: 0.9119496855345912
Akurasi untuk c = 1: 0.9056603773584906
```

Gambar 3 Hasil parameter C

3) *K-Nearest Neighbors* merupakan sebuah metode klasifikasi yang didasarkan dengan data training yang memiliki jarak paling dekat [24]. KNN sendiri adalah algoritma klasifikasi yang biasa digunakan untuk memprediksi kejadian yang akan datang [25]. Untuk menjalankan algoritma KNN perlu menambahkan library *KNeighborsClassifier* dari modul sklearn.neighbors. Parameter K yang digunakan sebagai uji ialah K=3.

```
27 X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
28
29 knn = KNeighborsClassifier(n_neighbors=3)
30
31 knn.fit(X_train, y_train)
32
33 y_pred = knn.predict(X_test)
```

Gambar 4 Penentuan parameter K

D. Evaluasi

Pada tahap terakhir ini, peneliti akan melakukan evaluasi terhadap hasil dari model dan teknik yang telah diterapkan. Evaluasi yang dilakukan ialah menggunakan metode *Confusion Matrix*. *Confusion Matrix* adalah metode yang digunakan untuk mengukur kinerja dari sebuah metode yang digunakan untuk klasifikasi [26]. Untuk menghitung hasil evaluasi terdapat 4 atribut, yaitu:

1. TP (*True Positive*)
Merupakan jumlah data yang memiliki nilai positif dan diklasifikasikan sebagai data positif.
2. FP (*False Positive*)
Merupakan jumlah data yang memiliki nilai negatif namun diklasifikasikan sebagai data positif.
3. TN (*True Negative*)
Merupakan jumlah data yang memiliki nilai negatif dan diklasifikasikan sebagai data negatif.
4. FN (*False Negative*)
Merupakan jumlah data yang memiliki nilai positif namun diklasifikasikan sebagai data negatif.

A) Accuracy

Accuracy merupakan jumlah data yang diprediksi benar baik itu data positif atau negatif [27]. Berikut rumus perhitungan *accuracy*:

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

B) Precision

Precision merupakan ketepatan prediksi data bernilai positif [27]. Berikut rumus perhitungan *precision*:

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

C) Recall

Recall merupakan jumlah akurasi model dalam melakukan prediksi dengan nilai positif [27]. Berikut perhitungan rumus *recall*:

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

D) F1 – Score

F1 – Score merupakan perhitungan dengan mengkalikan *precision* dengan *recall* yang kemudian dibagi dengan nilai *precision* dijumlahkan dengan *recall* [27]. Hal ini bertujuan untuk mendapatkan keseimbangan antara *precision* dan *recall*. Berikut perhitungan rumus *F1 – score*:

$$F1 - score = \frac{2 \times presisi \times recall}{presisi + recall}$$

E. Ensemble

Ensemble merupakan sebuah metode dalam machine learning dalam penggabungan beberapa algoritma menjadi 1 untuk mendapatkan hasil prediksi berdasarkan mayoritas suara [28]. Dalam penelitian ini, peneliti menggunakan *ensemble* untuk menggabungkan algoritma *Naive Bayes*, SVM, dan KNN untuk melakukan analisis sentimen masyarakat yang berasal dari komentar di *Google Play Store* terkait *Coursera*. Peneliti menggunakan *ensemble* agar hasil yang di dapatkan lebih optimal dimana dari 3 algoritma yang digunakan akan ditarik 1 kesimpulan guna menentukan hasil akhir apakah komentar tersebut termasuk ke dalam komentar positif atau negatif.

III. HASIL DAN PEMBAHASAN

A. Pengumpulan Data

Data yang digunakan berasal dari komentar pengguna terhadap coursera melalui platform *Google Play Store* didapatkan melalui proses yang dinamakan crawling. Proses ini dilakukan melalui platform *Google Colab* yang memiliki basis bahasa *Python*. Dalam proses crawling peneliti hanya mengambil nama pengguna, ulasan pengguna, dan rating dari pengguna terhadap coursera. Hasil yang pengguna dapatkan ialah 992 data untuk Amerika Serikat,

258 data untuk Indonesia, dan 797 data untuk Inggris. Data yang didapatkan ini kemudian disimpan dalam format CSV untuk memudahkan peneliti dalam mengolah data tersebut.

TABEL I
 DATA YANG TELAH DIKUMPULKAN

Ulasan
Thank you so much for your commitment to making learning accessible to everyone. The application is very convenient and deserves more than 5 stars. And thank you very much again!
senang diberi kesempatan untuk belajar dan menimba ilmu baru di platform digital ini senang diberi kesempatan untuk belajar dan menimba ilmu baru di platform digital ini
happens 3 times and I need at least an 80 to pass, this is a problem. I don't even get the chance to pass the quiz. It wastes my time, and I'm hoping this review will bring attention to the matter and it will be resolved. 2 stars until this is resolved.

B. Preprocessing

Data ulasan yang telah dikumpulkan selanjutnya akan diolah dalam tahap preprocessing. Berikut beberapa tahapan dalam preprocessing

1. Cleaning

Cleaning adalah proses penghapusan data teks yang tidak penting. Pada tahap ini juga karakter yang tidak berhubungan dengan bobot sebuah teks akan dihapuskan juga seperti “@”, “%”, “(”, “)” dan lainnya, Tabel II berisikan contoh penerapan tahap cleaning.

TABEL II
 CLEANING

Sebelum	Sesudah
Coursera is the Best Free prime learning app in the market , [Big companies® also do accepts it's certifications].	Coursera is the Best Free prime learning app in the market Big companies also do accepts it s certifications
terima kasih banyak untuk coursera, memberikan kursus gratis untuk saya dapat mengembangkan skill saya, sedih rasanya ketika say berumur 22 baru tau ada platform seperti ini, sekali lagi saya ucapkan terima kasih banyak atas bantuan keuangannya, saya akan memanfaatkan ini sebaik baiknya ????	terima kasih banyak untuk coursera memberikan kursus gratis untuk saya dapat mengembangkan skill saya sedih rasanya ketika say berumur baru tau ada platform seperti ini sekali lagi saya ucapkan terima kasih banyak atas bantuan keuangannya saya akan memanfaatkan ini sebaik baiknya
I think coursera is good a good platform for learning and and it's very convenient for someone and easy easy learning and I'm glad to be part of coursera I really want to appreciate coursera for this opportunity I'm also giving me a chance to learn again it is a great honour and privilege to be introduced to coursera	I think coursera is good a good platform for learning and and it s very convenient for someone and easy easy learning and I'm glad to be part of coursera I really want to appreciate coursera for this opportunity I m also giving me a chance to learn again it is a great honour and privilege to be introduced to coursera

2. Case Folding

Case Folding merupakan tahapan dimana semua huruf akan diubah menjadi huruf kecil untuk mempermudah proses pengolahan data. Tabel III merupakan contoh teks yang telah diubah hurufnya.

TABEL III
 CASE FOLDING

Sebelum	Sesudah
Coursera is the Best Free prime learning app in the market , [Big companiesÂ® also do accepts it's certifications]. (US)	coursera is the best free prime learning app in the market big companies also do accepts it s certifications
terima kasih banyak untuk coursera, memberikan kursus gratis untuk saya dapat mengembangkan skill saya, sedih rasanya ketika say berumur 22 baru tau ada platform seperti ini, sekali lagi saya ucapkan terima kasih banyak atas bantuan keuangannya, saya akan memanfaatkan ini sebaik baiknya ????	terima kasih banyak untuk coursera memberikan kursus gratis untuk saya dapat mengembangkan skill saya sedih rasanya ketika say berumur baru tau ada platform seperti ini sekali lagi saya ucapkan terima kasih banyak atas bantuan keuangannya saya akan memanfaatkan ini sebaik baiknya
I think coursera is good a good platform for learning and and it s very convenient for someone and easy easy learning and I'm glad to be part of coursera I really want to appreciate coursera for this opportunity I m also giving me a chance to learn again it is a great honour and privilege to be introduced to coursera	i think coursera is good a good platform for learning and and it s very convenient for someone and easy easy learning and I'm glad to be part of coursera i really want to appreciate coursera for this opportunity i m also giving me a chance to learn again it is a great honour and privilege to be introduced to coursera

3. Tokenization

Tokenization adalah tahapan pemecahan setiap kata dalam sebuah kalimat. Tabel IV merupakan contoh kalimat yang telah dipecah.

TABEL IV
 TOKENIZATION

Sebelum	Sesudah
coursera is the best free prime learning app in the market big companies also do accepts it s certifications	['coursera', 'is', 'the', 'best', 'free', 'prime', 'learning', 'app', 'in', 'the', 'market', 'big', 'companies', 'also', 'do', 'accepts', 'it', 's', 'certifications']
terima kasih banyak untuk coursera memberikan kursus gratis untuk saya dapat mengembangkan skill saya sedih rasanya ketika say berumur baru tau ada platform seperti ini sekali lagi saya ucapkan terima kasih banyak atas bantuan keuangannya saya akan memanfaatkan ini sebaik baiknya	['terima', 'kasih', 'banyak', 'coursera', 'memberikan', 'kursus', 'gratis', 'untuk', 'mengembangkan', 'skill', 'saya', 'sedih', 'rasanya', 'ketika', 'say', 'berumur', 'tau', 'ada', 'platform', 'seperti', 'ini', 'sekali', 'lagi', 'saya', 'ucapkan', 'terima', 'kasih', 'banyak', 'atas', 'bantuan', 'keuangannya', 'saya', 'manfaatkan', 'ini', 'sebaik', 'baiknya']
i think coursera is good a good platform for learning and and it s very convenient for someone and easy easy learning and i m glad to be part of coursera i really want to appreciate coursera for this opportunity i m also giving me a chance to learn again it is a great honour and privilege to be introduced to coursera	['i', 'think', 'coursera', 'is', 'good', 'a', 'good', 'platform', 'from', 'learning', 'and', 'it', 's', 'very', 'convenient', 'for', 'someone', 'and', 'easy', 'easy', 'learning', 'and', 'i', 'm', 'glad', 'to', 'be', 'part', 'of', 'coursera', 'i', 'really', 'appreciate', 'coursera', 'for', 'opportunity', 'i', 'm', 'also', 'giving', 'me', 'a', 'chance', 'to', 'learning', 'again', 'it', 'is', 'great', 'honour', 'and', 'privilege', 'to', 'introduced', 'to', 'coursera']

4. Stopword

Stopword adalah proses penghapusan kata-kata yang tidak memiliki makna atau berbobot rendah. Tabel V adalah contoh proses *stopword*.

TABEL V
STOPWORD

Sebelum	Sesudah
'coursera', 'is', 'the', 'best', 'free', 'prime', 'learning', 'app', 'in', 'the', 'market', 'big', 'companies', 'also', 'do', 'accepts', 'it', 's', 'certifications']	coursera free prime learning app market big companies accepts certifications
['terima', 'kasih', 'banyak', 'coursera', 'memberikan', 'kursus', 'gratis', 'untuk', 'mengembangkan', 'skill', 'saya', 'sedih', 'rasanya', 'ketika', 'say', 'berumur', 'tau', 'ada', 'platform', 'seperti', 'ini', 'sekali', 'lagi', 'saya', 'ucapkan', 'terima', 'kasih', 'banyak', 'atas', 'bantuan', 'keuangannya', 'saya', 'manfaatkan', 'ini', 'sebaik', 'baiknya']	terima kasih coursera kursus gratis mengembangkan skill sedih berumur tau platform ucapkan terima kasih bantuan keuangannya manfaatkan baiknya
['i', 'think', 'coursera', 'is', 'good', 'a', 'good', 'platform', 'from', 'learning', 'and', 'it', 's', 'very', 'convenient', 'for', 'someone', 'and', 'easy', 'easy', 'learning', 'and', 'i', 'm', 'glad', 'to', 'be', 'part', 'of', 'coursera', 'i', 'really', 'appreciate', 'coursera', 'for', 'opportunity', 'i', 'm', 'also', 'giving', 'me', 'a', 'chance', 'to', 'learning', 'again', 'it', 'is', 'great', 'honour', 'and', 'privilege', 'to', 'introduced', 'to', 'coursera']	oursera platform learn convenient easy easy learning glad part coursera appreciate coursera opportunity giving chance learning great honour privilege introduced coursera

5. Stemming

Stemming adalah proses pengolahan data dimana setiap kata berimbuhan akan dihapuskan kata imbuhan dan diubah menjadi kata dasar. Tabel VI merupakan contoh proses stemming.

TABEL VI
STEMMING

Sebelum	Sesudah
coursera free prime learning app market big companies accepts certifications	coursera free prime learn app market big company accepts certifications
terima kasih coursera kursus gratis mengembangkan skill sedih berumur tau platform ucapkan terima kasih bantuan keuangannya manfaatkan baiknya	terima kasih coursera kursus gratis kembang skill sedih umur tau platform ucap terima kasih bantu keangan manfaat baik
oursera platform learn convenient easy easy learning glad part coursera appreciate coursera opportunity giving chance learning great honour privilege introduced coursera	coursera platform learn convenient easy learn glad part coursera appreciate coursera opportunity give chance learn great honour privilege introduce coursera

C. Implementasi Algoritma

Setelah data selesai di olah langkah yang selanjutnya ialah membagi data menjadi 2 yaitu data latih dan data uji. Pembagian data latih dan data uji akan dibagi menjadi 80:20 dimana 80% data akan dijadikan data latih dan 20% data akan dijadikan data uji. Pembagian data latih dan data uji menjadi 80:20 dikarenakan tingkat akurasi daripembagian ini cukup tinggi jika melihat dari penelitian sebelumnya. Seperti contoh pada penelitian analisis sentimen aplikasi amazon melalui *Google Play Store* didapatkan hasil 88,37% untuk pembagian data 80:20 sedangkan untuk pembagian data seperti 70:30, 60:40 atau 90:10 memiliki akurasi dibawah 80:20 [10]. Dengan melihat tingginya akurasi dari pembagian data 80:20 maka peneliti menggunakan pembagian data tersebut. Setelah data dibagi menjadi 80:20 selanjutnya ialah penerapan algoritma *Naive Bayes*, SVM, dan KNN yang selanjutnya akan digabung dengan metode *ensemble*.

1. Naive Bayes

Pada algoritma *Naive Bayes*, dilakukan pengujian terhadap 797 data yang berasal dari Inggris, 992 data dari Amerika, dan 258 data dari Indonesia. Untuk melakukan uji test data dibagi menjadi 80:20. 80% data merupakan data training dan 20% data merupakan data test. Akurasi dari pengujian ini sebesar 0,8665 dimana untuk negara Inggris terdapat 717 komentar positif dan 70 komentar negatif, untuk Indonesia terdapat 177 komentar positif dan 76 komentar negatif, dan untuk Amerika terdapat 916 komentar positif dan 74 komentar negatif. Nilai akurasi 0,8665 menunjukkan bahwa data yang di prediksi oleh algoritma ini cukup baik karena dari seluruh data, 87% di antaranya berhasil di prediksi sesuai dengan klasifikasi komentar tersebut.

TABEL VII
 HASIL ANALISIS

Positif	Negatif
717	70
177	76
916	74

2. Support Vector Machine

Selanjutnya ialah penerapan algoritma menggunakan SVM. Sama seperti *Naive Bayes*, data yang digunakan untuk melakukan pengujian ialah data dari Inggris sebesar 797 data, Indonesia sebesar 258 data, dan Amerika sebesar 992 data. Semua data tersebut akan dibagi menjadi *data training* dan *data test* dengan perbandingan 80:20. 80% data digunakan sebagai *data training* dan 20% data digunakan sebagai *data test*. Setelah membagi data tahap selanjutnya yaitu mengimplementasikan SVM kepada dataset. Terhitung dengan perbandingan 80:20, akurasi yang didapat ialah 0,867 dengan hasil untuk negara Inggris ialah 728 komentar positif dan 63 komentar negatif, untuk negara Indonesia terdapat 233 komentar positif dan 25 komentar negatif, dan untuk negara Amerika terdapat 908 komentar positif dan 84 komentar negatif. Nilai akurasi 0,867 menunjukkan bahwa terdapat 87% data yang di prediksi secara benar oleh algoritma SVM.

TABEL VIII
 HASIL ANALISIS

Positif	Negatif
728	63
233	25
908	84

3. KNN

Terakhir yaitu melakukan pengujian dengan algoritma KNN. Seperti 2 algoritma sebelumnya dimana data yang digunakan berasal dari 3 negara yaitu Inggris sebanyak 797 data, Indonesia sebanyak 258 data, dan Amerika sebanyak 992 data. Setelah data tersebut diolah dan dilakukan perhitungan menggunakan algoritma KNN maka di dapatkan hasil untuk negara Inggris terdapat 788 komentar positif dan 4 komentar negatif, untuk negara Indonesia terdapat 254 komentar positif dan 4 komentar negatif, dan untuk negara Amerika terdapat 981 komentar positif dan 11 komentar negatif. Prediksi yang dilakukan oleh KNN memiliki nilai akurasi yang lebih rendah dibanding dengan 2 algoritma sebelumnya. Hal ini bisa dikarenakan penggunaan parameter K yang sedikit dan pembagian data latih dan data uji yang hanya terfokus pada 80:20 sehingga hasil dari penggunaan algoritma KNN kurang maksimal.

TABEL IX
 HASIL ANALISIS

Positif	Negatif
788	4
254	4
981	11

D. Ensemble

Tahap selanjutnya ialah tahap ensemble atau penggabungan algoritma untuk mendapatkan hasil terbaik. Untuk mendapatkan hasil terbaik dari ketiga algoritma ialah dengan menggabungkan algoritma *Naive Bayes*, SVM, dan KNN. Cara menggabungkan ketiga algoritma ini ialah dengan melihat hasil dari ketiga algoritma kemudian melakukan voting untuk menentukan hasil yang akan digunakan. dengan menggunakan *ensemble* didapatkan hasil sebagai berikut: untuk negara Inggris terdapat 780 komentar positif dan 17 komentar negatif, untuk negara

Indonesia terdapat 237 komentar positif dan 21 komentar negatif, dan untuk negara Amerika terdapat 967 komentar positif dan 25 komentar negatif.

Jika melihat hasil akhir dari metode *ensemble* bisa dikatakan hasil yang di dapatkan oleh metode ini lebih baik

TABEL IX
HASIL ANALISIS

Positif	Negatif
780	17
237	21
967	25

daripada hanya menggunakan salah satu algoritma. Banyak prediksi salah pada 3 algoritma yang digunakan secara terpisah menjadi benar dengan menggabungkan ketiga algoritma tersebut. Meskipun tetap ada beberapa data yang terprediksi salah walaupun sudah menggunakan metode *ensemble*.

E. Evaluasi

Setelah melakukan implementasi algoritma *Naive Bayes* SVM, dan KNN. Langkah selanjutnya ialah tahap pengujian evaluasi. Untuk penggunaan algoritma *Naive Bayes* mendapatkan nilai *accuracy* sebesar 89%, *precision* sebesar 90%, *recall* sebesar 99%, dan *f1 - score* sebesar 94%. Untuk algoritma SVM mendapatkan nilai *accuracy* sebesar 90%, *precision* sebesar 91%, *recall* sebesar 99%, dan *f1 - score* sebesar 95%. Dan Untuk algoritma KNN mendapatkan nilai *accuracy* sebesar 88%, *precision* sebesar 88%, *recall* 99%, dan *f1 - score* sebesar 94%.

```
Naive Bayes
accuracy - 0.8928322784810125
precision - 0.8945204362592769
recall - 0.9957142857142858
f1 - 0.9423223938984588

SVM
accuracy - 0.9004113924050632
precision - 0.9068022982233508
recall - 0.9885300207039338
f1 - 0.9458266946682157

KNN
accuracy - 0.8789556962025318
precision - 0.8799294060370009
recall - 0.9985507246376812
f1 - 0.9354772825146747
```

Gambar 5 Hasil Evaluasi

Hasil dari keseluruhan tahap menunjukkan bahwa aplikasi *coursera* memiliki sentimen positif. Adanya sentimen positif ini akan menaikkan minat masyarakat untuk menggunakan aplikasi ini. Akurasi yang tinggi dari klasifikasi menunjukkan keberhasilan algoritma dalam memprediksi sentimen. Nilai presisi yang tinggi juga menunjukkan seberapa tepat algoritma memprediksi dengan benar nilai positif. Nilai recall yang tinggi menunjukkan bahwa algoritma tersebut memiliki kecenderungan memprediksi data sebagai positif. Untuk nilai F1 sendiri menunjukkan keseimbangan antara presisi dan recall.

Klasifikasi akan semakin baik apabila memiliki nilai akurasi yang tinggi, namun untuk mendapatkan akurasi yang tinggi terdapat sejumlah faktor yang mempengaruhi seperti pembagian data latih dan data tes, penentuan parameter, data yang digunakan, dan ketepatan penggunaan algoritma.

IV. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan dari 3 negara terkait komentar terhadap *coursera* melalui platform *Google Play Store*, komentar yang terlontar cenderung positif. Dari 797 data di Inggris terdapat 780 komentar positif dan 17 komentar negatif, dari 258 data di Indonesia terdapat 237 komentar positif dan 21 komentar negatif, dan dari 992 data dari Amerika terdapat 967 komentar positif dan 25 komentar negatif. Jika di hitung secara persentase, maka dapat disimpulkan bahwa negara Amerika cenderung lebih banyak komentar positif daripada Indonesia dan Inggris. Sebaliknya Indonesia memiliki persentase yang cukup rendah terhadap komentar positif terkait aplikasi *coursera*. Hasil ini merupakan hasil yang telah digabungkan menggunakan *ensemble*. Metode *ensemble* sangat cocok digunakan untuk mendapatkan hasil yang lebih baik daripada hanya menggunakan 1 algoritma. Untuk algoritma sayangnya algoritma KNN memiliki hasil yang kurang bagus dimana masih ada banyak komentar yang memiliki nilai FP dan FN. Hal ini tentu saja dapat menurunkan akurasi hasil dari *ensemble* karena hasil dari salah satu algoritma kurang optimal.

DAFTAR PUSTAKA

- [1] Bullen and Janes, Making the Transition to E-Learning : Strategies, Information Science Pub, 2007.
- [2] E. Andari, "Implementasi Kurikulum Merdeka Belajar Menggunakan Learning Management System (LMS)," *Allimna: Jurnal Pendidikan Profesi Guru*, vol. 1.2, pp. 65-79, 2022.
- [3] I. Kamil and Y. E. Harususilo, "Dirjen Dikti: 70 Persen Mahasiswa dan Dosen Nilai Pembelajaran Daring Lebih Baik," *Kompas.com*, 2020.
- [4] M. N. Afif, H. M. Az-Zahra and D. Priharsari, "Evaluasi Pengalaman Pengguna pada LMS E-Learning UIN Malang menggunakan Metode UX Curve dari Sudut Pandang Pengajar (Studi Kasus: Dosen Jurusan Psikologi Universitas Islam Negeri Maulana Malik Ibrahim Malang)," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 7(4), pp. 1835-1845, 2023.
- [5] C. F. Hasri and D. Alita, "Penerapan Metode Naïve Bayes Classifier Dan Support Vector Machine Pada Analisis Sentimen Terhadap Dampak Virus Corona Di Twitter," *Jurnal Informatika dan Rekayasa Perangkat Lunak*, vol. 3(2), pp. 145-160, 2022.
- [6] R. Kurniawan and A. Apriliani, "Analisis Sentimen Masyarakat terhadap Virus Corona Berdasarkan Opini dari Twitter Berbasis Web Scraper," *Jurnal INSTEK (Informatika Sains dan Teknologi)*, vol. 5, no. 1, pp. 67-75, 2020.
- [7] F. Nufairi, N. Pratiwi and F. Herlando, "Analisis Sentimen pada Ulasan Aplikasi Threads di Google Play Store Menggunakan Algoritma Support Vector Machine," *JUPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, vol. 9(1), pp. 339-348, 2024.
- [8] D. Xhemalii, C. J. Hinde and R. G. Stone, "Naive Bayes vs Decision Trees vs. Neural Network in the Classification of Training Web Pages," *IJCSI International Journal of Computer Science Issues*, 4(1), 2009.
- [9] R. A. Saputra, D. Puspitasari and T. Baidawi, "Deteksi Kematangan Buah Melon Dengan Algoritma Support Vector Machine Berbasis Ekstraksi Fitur GLCM," *Jurnal Infotech*, vol. 4(2), pp. 200-206, 2022.
- [10] E. Hasibuan and E. A. Heriyanto, "Analisis Sentimen Pada Ulasan Aplikasi Amazon Shopping Di Google Play Store Menggunakan Naive Bayes Classifier," *Jurnal Teknik dan Science*, vol. 1, no. 3, pp. 13-24, 2022.
- [11] E. Indrayuni, A. Nurhadi and D. A. Krisyanti, "Implementasi Algoritma Naive Bayes, Support Vector Machine, dan K-Nearest Neighbors untuk Analisa Sentimen Aplikasi Halodoc," *Faktor Exacta*, vol. 14, no. 2, pp. 64-71, 2021.
- [12] A. Rahman, E. Utami and S. Sudarmawan, "Sentimen Analisis Terhadap Aplikasi pada Google Playstore Menggunakan Algoritma Naïve Bayes dan Algoritma Genetika," *Jurnal Komitika (Komputasi dan Informatika)*, vol. 5 (1), pp. 60-71, 2021.
- [13] N. Herlinawati, Y. Yuliani, S. Faizah, W. Gata and S. Samudi, "Analisis Sentimen Zoom Cloud Meetings di Play Store Menggunakan Naïve Bayes dan Support Vector Machine," *CESS (Journal of Computer Engineering, System and Science)*, Vol 5, No 2, 2020.
- [14] D. S. Putri and T. Ridwan, "Analisis Sentimen Ulasan Aplikasi Pospay Dengan Algoritma Support Vector Machine," *Jurnal Ilmiah Informatika*, pp. 32-40, 2023.
- [15] F. A. Indriyani and S. Faisal, "Analisis sentimen aplikasi tiktok menggunakan algoritma naïve bayes dan support vector machine," *TEKNOSAINS J. Sains, Teknol. dan Inform.*, pp. 176-184, 2023.
- [16] F. Herlando, A. R. Dzikrillan, F. Nufairi, E. Sinduningrum and M. Sholeh, "Analisis Perbandingan Sentimen dan Perbincangan Netizen Terhadap Twitter Pasca Pergantian Nama," *JUPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, vol. 9 (1), pp. 360-367, 2024.
- [17] D. Ardiansyah, A. Saepudin, R. Aryanti and E. Fitriani, "Analisis Sentimen Review Pada Aplikasi Media Sosial Tiktok Menggunakan Algoritma K-Nn Dan Svm Berbasis Pso," *Jurnal Informatika Kaputama (JIK)*, vol. 7(2), pp. 233-241, 2023.
- [18] S. S. Norman, D. R. Kusuma and L. N. Afifa, "Analisis Sentimen Tanggapan Pelanggan Indihome di Platform Sosial Media Facebook dan Twitter Menggunakan Support Vector Mesin dan Pendekatan Klasifikasi Naïve Bayes," *Jurnal Sains & Teknologi Fakultas Teknik Universitas Darma Persada*, vol. 13(1), pp. 124-133, 2023.
- [19] C. Lim and W. Goh, "The Application of an Ensemble Of Boosted Elman Networks to Time Series Prediction: A Benchmark Study," *J Comput Intell*, 2005.
- [20] M. F. Asnawi and A. P. Aji, "Sistem Pengelolaan Pembiayaan Kegiatan OSIS SMP Negeri 2 Batur Menggunakan Algoritma Naive Bayes Classifier dalam Pengklasifikasian Data Laporan," *Jurnal Ilmiah Teknik Dan Ilmu Komputer*, vol. 1(3), pp. 26-33, 2022.
- [21] F. Novianti and K. R. N. Wardani, "Analisis Sentimen Masyarakat Terhadap Data Tweet Traveloka Selama Rapid Test Antigen Menggunakan Algoritma Naïve Bayes," *JUPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, vol. 8(3), pp. 922-933, 2023.
- [22] G. Williams, *Data Mining with R: The Art of Excavating Data for Knowledge Discovery*, New York: Springer, 2011.
- [23] O. H. Rahman, G. Abdillah and A. Komarudin, "Klasifikasi Ujaran Kebencian pada Media Sosial Twitter Menggunakan Support Vector Machine," *jurnal resti (rekayasa sistem dan teknologi informasi)*, vol. 5(1), pp. 17-23, 2021.
- [24] A. D. A. Putra and S. Juanita, "Analisis Sentimen pada Ulasan pengguna Aplikasi Bibit Dan Bareksa dengan Algoritma KNN," *JATISI (Jurnal Teknologi Informatika dan Sistem Informasi)*, vol. 8(2), pp. 636-646, 2021.
- [25] M. A. Lusiano, S. M. Nasution and C. Setianingsih, "Implementation of the advanced traffic management system using k-nearest neighbor algorithm," *International Conference on Information Technology Systems and Innovation (ICITSI)*, pp. 149-154, 2020.
- [26] A. Budiman, J. C. Young and A. Suryadibrata, "Implementasi Algoritma Naive Bayes untuk Mengklasifikasikan Konten Twitter dengan Indikasi Depresi," *Jurnal Informatika: Jurnal Pengembangan IT*, vol. 6(2), pp. 133-138, 2021.
- [27] T. N. Wijaya, R. Indriati and M. N. Muzaki, "Analisis Sentimen Opini Publik Tentang Undang-Undang Cipta Kerja Pada Twitter," *Jambura Journal of Electrical and Electronics Engineering*, vol. 3(2), pp. 78-83, 2021.
- [28] M. Ardiansyah, "Model Ensemble Algoritma Naive Bayes Dan Random Forest Dalam Klasifikasi Penyakit Paru-paru Untuk Meningkatkan Akurasi," *SMARTLOCK: Jurnal Sains dan Teknologi*, vol. 2(2), pp. 32-38, 2023.