

PERBANDINGAN METODE KLASIFIKASI NAÏVE BAYES, DECISION TREE, RANDOM FOREST TERHADAP ANALISIS SENTIMEN KENAIKAN BIAYA HAJI 2023 PADA MEDIA SOSIAL YOUTUBE

Muhammad Yasir¹, Robertus Suraji²

Fakultas Ilmu Komputer, Universitas Bhayangkara Jakarta Raya
Email: muhammad.yasir@dsn.ubharajaya.ac.id, robertus.suraji@dsn.ubharajaya.ac.id

ABSTRAK

Kata kunci:

Analisis Sentimen, biaya haji, youtube, naïve bayes, decision tree, random forest

Euforia Penyelenggaraan ibadah haji tahun 2023 menjadi trending perbincangan yang ramai di media sosial, salah satunya pada media youtube. Pemerintah kerajaan arab saudi menerapkan kebijakan untuk menurunkan biaya paket haji sebesar 30% dari tahun sebelumnya dan tidak ada pembatasan usia jemaah haji bahkan mendapatkan penambahan kuota sebanyak 221.000 jemaah haji untuk indonesia. Bertolak belakang dengan kebijakan pemerintah arab saudi, kementerian agama membuat suatu wacana yang kontradiktif dengan mengusulkan kenaikan biaya haji dengan komposisi pembiayaan menjadi 70% biaya perjalanan haji (BIPIH) dan 30% nilai manfaat pada usulan biaya haji tahun 2023. Usulan biaya haji tersebut di sahkan menjadi suatu kebijakan melalui Kepres Nomor 7 Tahun 2023 tentang biaya penyelenggaraan ibadah haji tahun 1444 hijriah/2023. Hal ini menimbulkan berbagai respon masyarakat terhadap kebijakan tersebut. Tujuan penelitian ini melakukan analisis sentimen masyarakat dengan cara melakukan klasifikasi komentar positif dan negatif pada media sosial youtube terhadap kebijakan kenaikan biaya perjalan haji 2023. Penelitian ini melakukan perbandingan hasil akurasi terhadap beberapa metode klasifikasi seperti *naïve bayes*, *decision tree* dan *random forest*. Dari penelitian ini didapatkan hasil akurasi naïve bayes sebesar 90%, *Decision Tree* sebesar 83%, *Random Forest* sebesar 87%. Diharapkan dengan adanya analisis sentimen klasifikasi komentar youtube ini dapat memberikan gambaran bagi pemerintah melihat respon penilaian dari masyarakat dalam membuat suatu kebijakan.

ABSTRACT

Keywords:

Sentiment analysis, pilgrimage costs, youtube, naïve bayes, decision tree, random forest

Euphoria The implementation of the 2023 Hajj pilgrimage has become a trending topic of conversation on social media, one of which is on YouTube. The government of the Kingdom of Saudi Arabia implemented a policy to reduce the cost of Hajj packages by 30% from the previous year and there was no age limit for pilgrims and even received an additional quota of 221,000 pilgrims for Indonesia. Contrary to the policy of the Saudi Arabian government, the Ministry of Religion created a contradictory discourse by proposing an increase in the costs of the Hajj with the composition of the financing being 70% of the costs of the Hajj trip (BIPIH) and 30% of the value of the benefits in the proposed Hajj costs in 2023. The proposed Hajj costs were ratified become a policy through Presidential Decree Number 7 of 2023 regarding the costs of organizing the pilgrimage in 1444 hijriah/2023. This gave rise to various public responses to the policy. The purpose of this study was to analyze public sentiment by classifying

positive and negative comments on YouTube social media regarding the policy of increasing the cost of the 2023 Hajj pilgrimage. This research compared the results of the accuracy of several classification methods such as naïve bayes, decision trees and random forest. From this research, the accuracy of naïve Bayes was obtained by 90%, Decision Tree by 83%, Random Forest by 87%. It is hoped that with the sentiment analysis of the classification of YouTube comments, it can provide an overview for the government to see the response of the public's assessment in making a policy.

PENDAHULUAN

Euforia Penyelenggaraan ibadah haji tahun 2023 menjadi trending perbincangan yang ramai di media sosial, salah satunya pada media youtube. Dimana pemerintah yang diwakili oleh kementerian agama telah menandatangani kesepakatan dengan kerajaan arab saudi terkait penyelenggaraan ibadah haji 1444 H / 2023 Masehi, kesepakatan tersebut ditanda tangani oleh Menteri Agama Yaquut Cholil Qoumas bersama Menteri Haji Dan Umrah Arab Saudi Tawfiq F Al Rabiah di Jeddah (Kemenag, 2023b).

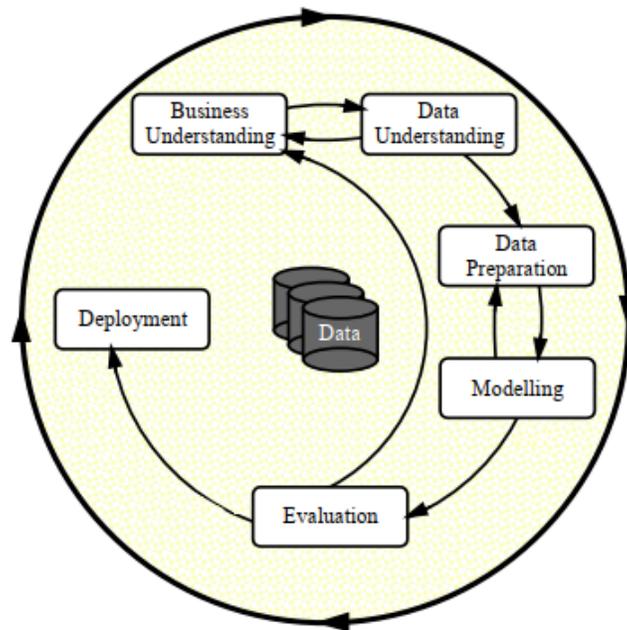
Hasil kesepakatan kerjasama tersebut, pemerintah kerajaan arab saudi memberikan relaksasi usia berhaji pada Indonesia yaitu tidak ada pembatasan usia jemaah haji seperti tahun-tahun sebelumnya. Kemudian Indonesia juga mendapatkan kuota haji sebanyak 221.000 jemaah haji, kuota itu terdiri atas 203.320 jemaah haji reguler, dan 17.680 jemaah haji khusus, artinya kuota keberangkatan jemaah haji indonesia naik 100% dibanding dengan tahun 2022. Akan tetapi kementerian agama mengusulkan kenaikan biaya perjalanan haji (BIPIH) dari Rp39.886.009 tahun 2022 menjadi Rp69.193.734 sebagai usulan tahun 2023 dengan dalih pengalokasian komposisi pembiayaan menjadi 70% Bipih dan 30% nilai manfaat pada usulan biaya haji tahun 2023, dan di sahkan menjadi suatu kebijakan melalui Kepmenag Nomor 352 Tahun 2023 tentang biaya perjalanan haji reguler tahun 1444 H/ 2023 M dan penggunaan nilai manfaat, serta Keppres Nomor 7 Tahun 2023 tentang biaya penyelenggaraan ibadah haji tahun 1444 hijriah/2023 (Kemenag, 2023a).

Dampak dari wacana usulan kenaikan biaya haji tersebut menjadi trending topik perbincangan di dunia maya, salah satunya pada media sosial youtube. Dimana diketahui indonesia merupakan pengguna media sosial youtube terbesar ke 3 di dunia(Mahmudan, 2011). Oleh karena itu, diperlukanlah suatu analisa data dalam menilai dan mengklasifikasi berbagai respon komentar masyarakat dengan melakukan analisis sentimen masyarakat terhadap usulan kenaikan biaya haji tahun 2023 di media sosial youtube (Hidayat, 2015). Tujuan penelitian ini melakukan analisis sentimen dengan cara melakukan klasifikasi komentar positif dan negatif pada media sosial youtube terhadap kebijakan kenaikan biaya perjalan haji 1444 H/ 2023 Masehi. Penelitian ini menggunakan data komentar youtube yang di *crawling* dari salah satu channel berita nasional, hasil

crawling tersebut berhasil mendapatkan 1014 komentar. Penelitian ini melakukan *experiment research* menggunakan beberapa metode seperti *naïve bayes*, *decision tree* dan *random forest* dalam melakukan klasifikasi komentar *youtube* kemudian melakukan perbandingan hasil akurasi terbaik dari proses klasifikasi tersebut. Diharapkan dengan adanya penelitian analisis sentimen ini ini dapat memberikan gambaran bagi pemerintah melihat respon penilaian dari masyarakat dalam membuat suatu kebijakan (Fiarni et al., 2016).

METODE

Dalam penelitian ini penulis menggunakan metodologi CRISP-DM, dalam perkembangannya, CRISP-DM diakui sebagai metodologi data mining yang komprehensif untuk kebutuhan suatu proyek bisnis, dan telah menjadi salah satu metode yang paling banyak digunakan dalam bidang ilmu data. CRISP-DM telah menjadi standar *de facto* untuk proyek-proyek analisis, penambangan data, dan ilmu data (Schröer et al., 2021).



Gambar 1. Tahapan CRISP-DM (Schröer et al., 2021)

Tahapan penelitian menggunakan CRISP-DM sebagai berikut:

A. Business Understanding

Tahapan ini dimulai dengan memahami tujuan dan persyaratan proyek dari berbagai sudut pandang dan mengubah pengetahuan yang di peroleh menjadi definisi masalah dan rencana awal yang dirancang untuk mencapai tujuan (Tyagi & Tripathi, 2019).

B. Data Understanding

Pada tahapan ini, Proses pemahaman data dimulai dengan pengumpulan data dan dilanjutkan dengan mengidentifikasi data, mengidentifikasi masalah kualitas data, atau menemukan subset menarik dari data tersembunyi untuk membentuk hipotesis.

C. Data Preparation

Tahap persiapan data mencakup semua kegiatan untuk membangun kumpulan data akhir dari data mentah asli. seperti memilih tabel, record dan atribut, preprocessing data, membuat atribut baru dan mengubah data untuk pemodelan.

D. Modeling

Pada tahapan ini, berbagai teknik pemodelan dipilih dan diterapkan, dan parameter dikalibrasi ke nilai optimal, dalam penelitian ini, pemodelan menggunakan metode klasifikasi naïve bayes, decision tree dan random forest

E. Evaluation

Pada tahapan ini, sudah membangun suatu model analisis data, sebelum lanjut ke implementasi akhir, model dilakukan evaluasi untuk memastikan bahwa model yang dibangun mencapai tujuan bisnis dengan benar. Dalam penelitian ini, evaluasi model menggunakan confusion matriks

F. Deployment

Pada tahapan ini, penulis melakukan pengujian model dan membuat visualisasi dari kumpulan data

HASIL DAN PEMBAHASAN

Pada tahap ini menjelaskan mengenai pembahasan dan langkah-langkah peneliti dalam melakukan analisis sentimen terhadap kebijakan kenaikan biaya perjalanan haji 1444 H/ 2023 Masehi sebagai berikut:

1. Business Understanding

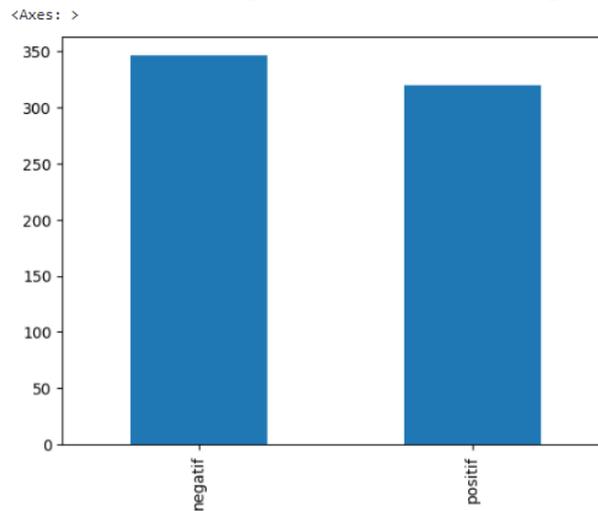
Tahap ini berfokus pada pemahaman tujuan dan persyaratan dan kemudian mengubah pengetahuan ini menjadi definisi masalah. Dalam hal ini, peneliti melakukan riset dan observasi melakukan pencarian trending topik di media sosial youtube, kemudian melakukan eksperimen membangun model analisis sentimen dengan menggunakan berbagai metode klasifikasi, kemudian melakukan komparasi dengan penelitian sebelumnya dengan mempelajari studi pustaka, studi literatur sejenis.

2. Data Understanding

Tahap ini berfokus pada pengumpulan data awal. Dalam hal ini peneliti melakukan pengumpulan data komentar youtube terkait dengan isu wacana kenaikan ibadah haji 2023, peneliti melakukan *crawling* data sebanyak 1.014 komentar yang didapatkan dari channel berita nasional Metro TV (https://www.youtube.com/watch?v=wKQR7q_INDU).

Proses *crawling* menggunakan *google colab python* mendapatkan data sebanyak 1.014 komentar, peneliti melakukan pembersihan dari data yang *missing value*, *redudance* dan *duplicate*.

Pada tahap ini juga peneliti berkoordinasi dengan ahli bahasa (*anator*) untuk membantu melabeli *dataset* (positif/negatif) yang telah dikumpulkan agar siap di olah ke proses selanjutnya Tujuannya adalah ahli atau pakar bahasa tersebut lebih memahami pemaknaan dari setiap teks yang tertulis (Hadna et al., 2016). Pembersihan dan pelabelan data tersebut didapatkan hasil sebanyak 666 komentar yang terdiri dari 346 komentar negatif dan 320 komentar positif.



3. Data Preparation

Tahap persiapan data mencakup semua kegiatan untuk membangun *dataset* akhir. Setelah pengumpulan dan pelabelan dataset, Pada tahap ini dilakukan *text preprocessing* data adalah sebagai berikut :

a. Tokenize

Tahap *Tokenize* merupakan proses memisahkan teks menjadi potongan-potongan kata yang menyusunnya. Dengan menggunakan `nltk.word_tokenize` pada python.

```
df['tokenize'] = df.apply(lambda row: nltk.word_tokenize(row['Text']), axis=1)
```

```
[6] # Tokenize
df['tokenize'] = df.apply(lambda row: nltk.word_tokenize(row['Text']), axis=1)
df.head(10)
```

	Text	label	tokenize
0	Mestinya Negara mendukung,menfasilitasi dan me...	negatif	[Mestinya, Negara, mendukung, ,, menfasilitasi...
1	pemerintah selalu membuat kegaduhan, keresahan...	negatif	[pemerintah, selalu, membuat, kegaduhan, ,, ke...
2	Harusnya kalau ongkos haji mau dinaikan, diber...	negatif	[Harusnya, kalau, ongkos, haji, mau, dinaikan,...
3	semoga ada penurunan biar jamaah yang berangka...	positif	[semoga, ada, penurunan, biar, jamaah, yang, b...
4	sadarlah pemerintah jangan buat kebijakan sepe...	negatif	[sadarlah, pemerintah, jangan, buat, kebijakan...
5	APABILA PEKERJAAN/JABATAN TIDAK SESUAI KEMAMPU...	negatif	[APABILA, PEKERJAAN/JABATAN, TIDAK, SESUAI, KE...
6	pemerintah sepertinya sengaja menghalangi jama...	negatif	[pemerintah, sepertinya, sengaja, menghalangi,...
7	semoga kita semuanya lancar serta dipermudahka...	positif	[semoga, kita, semuanya, lancar, serta, diperm...
8	Tolong diperjuangkan Pak, kasihan Yang Mau Ber...	negatif	[Tolong, diperjuangkan, Pak, ,, kasihan, Yang,...
9	Pihak saudi arab sangat bijak mempermudah sara...	negatif	[Pihak, saudi, arab, sangat, bijak, mempermudah...

b. Remove Punctuation

Pada tahap ini, penulis mengubah semua kata dalam kalimat menjadi huruf kecil menggunakan fungsi "lower case" python. Sedangkan Tahap *Remove Punctuation* merupakan Proses menghilangkan tanda baca atau simbol yang ada dalam data

```
#Case folding & remove punctuation
def clean_text(text):
    text = text.lower() #lowercase atau case folding
    text = re.sub('@^[^s]+', '', text) #remove username
    text = re.sub('\[.*?\]', '', text) # remove square brackets
    text = re.sub('((www\.[^s]+)|(https?:/[^s]+))', '', text) # remove URLs
    text = re.sub('[%s]' % re.escape(string.punctuation), '', text) # remove punctuation
    text = re.sub('\w*\d\w*', '', text)
    text = re.sub('[‘’“”...]', '', text)
    text = re.sub('\n', '', text)
    return text

remove_karakter = lambda x: clean_text(x)
```

```
df['remove_karakter'] = pd.DataFrame(df['Text'].apply(remove_karakter))
df.head(10)
```

	Text	label	tokenize	remove_karakter
0	Mestinya Negara mendukung,memfasilitasi dan me...	negatif	[Mestinya, Negara, mendukung, ,, memfasilitasi...	mestinya negara mendukungmemfasilitasi dan mer...
1	pemerintah selalu membuat kegaduhan, keresahan...	negatif	[pemerintah, selalu, membuat, kegaduhan, ,, ke...	pemerintah selalu membuat kegaduhan keresahan ...
2	Harusnya kalau ongkos haji mau dinaikan, diber...	negatif	[Harusnya, kalau, ongkos, haji, mau, dinaikan,...	harusnya kalau ongkos haji mau dinaikan diberl...
3	semoga ada penurunan biar jamaah yang berangkat...	positif	[semoga, ada, penurunan, biar, jamaah, yang, b...	semoga ada penurunan biar jamaah yang berangkat...
4	sadarlah pemerintah jangan buat kebijakan sepe...	negatif	[sadarlah, pemerintah, jangan, buat, kebijakan...	sadarlah pemerintah jangan buat kebijakan sepe...
5	APABILA PEKERJAAN/JABATAN TIDAK SESUAI KEMAMPU...	negatif	[APABILA, PEKERJAAN/JABATAN, TIDAK, SESUAI, KE...	apabila pekerjaanjabatan tidak sesuai kemampuan...
6	pemerintah sepertinya sengaja menghalangi jama...	negatif	[pemerintah, sepertinya, sengaja, menghalangi,...	pemerintah sepertinya sengaja menghalangi jama...
7	semoga kita semuanya lancar serta dipermudahka...	positif	[semoga, kita, semuanya, lancar, serta, diperm...	semoga kita semuanya lancar serta dipermudahka...

c. Stemming

Pada Tahap *Stemming*, Proses mengubah kata berimbuhan menjadi kata dasar. Pada tahap ini menggunakan *library* Python Sastrawi untuk menghilangkan kata-kata imbuhan dalam bahasa Indonesia ke bentuk dasarnya.

Perbandingan Metode Klasifikasi Naïve Bayes, Decision Tree, Random Forest Terhadap Analisis Sentimen Kenaikan Biaya Haji 2023 Pada Media Sosial Youtube

```
from Sastrawi.Stemmer.StemmerFactory import StemmerFactory
```

```
factory = StemmerFactory()
stemming = factory.create_stemmer()
```

```
output = [(stemming.stem(token)) for token in text]
```

```
df['stemming'] = output
df.head(10)
```

	Text	label	tokenize	remove_karakter	stemming
0	Mestinya Negara mendukung,menfasilitasi dan me...	negatif	[Mestinya, Negara, mendukung, .. menfasilitasi...	mestinya negara mendukungmenfasilitasi dan mer...	mesti negara mendukungmenfasilitasi dan rigan...
1	pemerintah selalu membuat kegaduhan, keresahan...	negatif	[pemerintah, selalu, membuat, kegaduhan, .. ke...	pemerintah selalu membuat kegaduhan keresahan ...	perintah selalu buat gaduh resah dan berat cal...
2	Harusnya kalau ongkos haji mau dinaikan, diber...	negatif	[Harusnya, kalau, ongkos, haji, mau, dinaikan, ..	harusnya kalau ongkos haji mau dinaikan diberl...	harus kalau ongkos haji mau naik laku bagi yan...
3	semoga ada penurunan biar jamaah yang berangka...	positif	[semoga, ada, penurunan, biar, jamaah, yang, b...	semoga ada penurunan biar jamaah yang berangka...	moga ada turun biar jamaah yang berangkat tahu...
4	sadarlah pemerintah jangan buat kebijakan sepe...	negatif	[sadarlah, pemerintah, jangan, buat, kebijakan...	sadarlah pemerintah jangan buat kebijakan sepe...	sadar perintah jangan buat bijak seperti itu b...
5	APABILA PEKERJAAN/JABATAN TIDAK SESUAI KEMAMPU...	negatif	[APABILA, PEKERJAAN/JABATAN, TIDAK, SESUAI, KE...	apabila pekerjaanjabatan tidak sesuai kemampua...	apabila pekerjaanjabatan tidak sesuai mampu tu...
6	pemerintah sepertinya sengaja menghalangi jama...	negatif	[pemerintah, sepertinya, sengaja, menghalangi,...	pemerintah sepertinya sengaja menghalangi jama...	perintah seperti sengaja halang jamaah tunai h...
7	semoga kita semuanya lancar serta dipermudahka...	positif	[semoga, kita, semuanya, lancar, serta, diperm...	semoga kita semuanya lancar serta dipermudahka...	moga kita semua lancar serta mudah segala urus...
8	Tolong diperjuangkan Pak, kasihan Yang Mau Ber...	negatif	[Tolong, diperjuangkan, Pak, ,, Kasihani, Yang,...	tolong diperjuangkan pak kasihan yang mau bera...	tolong juang pak kasihan yang mau berangkat on...
9	Pihak saudi arab sangat bijak mempermudah sara...	negatif	[Pihak, saudi, arab, sangat, bijak, memper muda...	pihak saudi arab sangat bijak memper mudah sara...	pihak saudi arab sangat bijak mudah sarana iba...

d. Modeling

Setelah dilakukan preprocessing data, pada tahap ini peneliti membangun suatu model klasifikasi analisis sentimen dengan memperbandingkan 3 model metode klasifikasi yaitu naïve bayes, decision tree, random forest.

```
[15] from sklearn.naive_bayes import MultinomialNB
      from sklearn.tree import DecisionTreeClassifier
      from sklearn.ensemble import RandomForestClassifier

      model1 = MultinomialNB().fit(X_train, y_train)
      model2 = DecisionTreeClassifier().fit(X_train, y_train)
      model3 = RandomForestClassifier().fit(X_train, y_train)

      prediksi_model1 = model1.predict(X_test)
      prediksi_model2 = model2.predict(X_test)
      prediksi_model3 = model3.predict(X_test)

      print('Akurasi model 1 adalah ', model1.score(X_test,y_test)*100,'%')
      print('Akurasi model 2 adalah ', model2.score(X_test,y_test)*100,'%')
      print('Akurasi model 3 adalah ', model3.score(X_test,y_test)*100,'%')

      Akurasi model 1 adalah 90.29850746268657 %
      Akurasi model 2 adalah 82.83582089552239 %
      Akurasi model 3 adalah 86.56716417910447 %
```

Perbandingan Metode Klasifikasi Naïve Bayes, Decision Tree, Random Forest Terhadap Analisis Sentimen Kenaikan Biaya Haji 2023 Pada Media Sosial Youtube

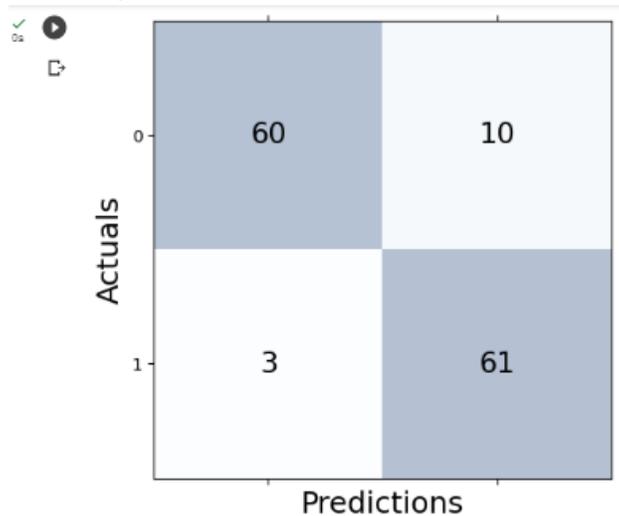
Dari klasifikasi tersebut menghasilkan akurasi sebagai berikut :

Metode	Hasil Akurasi
Naïve Bayes	90%
Decision Tree	83%
Random Forest	87%

e. Evaluation

Tahapan *evaluation* ini merupakan tahap laporan evaluasi model dengan menggunakan *confusion matrix* untuk mengukur performa klasifikasi dari ketiga metode sebagai berikut:

1) Confusion Matrix Naïve Bayes



```

Classification report:
              precision    recall  f1-score   support

   negatif    0.95     0.86     0.90         70
   positif    0.86     0.95     0.90         64

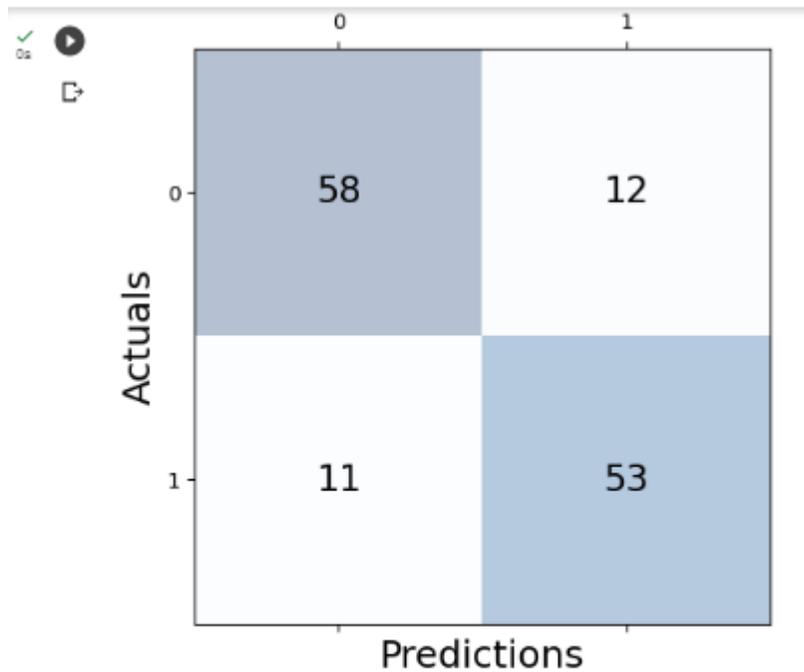
 accuracy          0.90         134
 macro avg         0.91     0.91     0.90         134
 weighted avg      0.91     0.90     0.90         134
    
```

Tabel Confusion Matrix Metode Naïve Bayes

Confusion Matrix Observed Class	Predicted Class		
		Positif	Negatif
	Positif	TP = 60	FN = 10
Negatif	FP = 3	TN = 61	

$$\text{Akurasi : } \frac{TP+TN}{TP+FP+TN+FN} = \frac{60+61}{60+10+3+61} = \frac{121}{134} = 0,9029$$

2) Confusion Matrix Decision Tree



```

Classification report:
              precision    recall  f1-score   support

   negatif     0.84      0.83      0.83        70
   positif     0.82      0.83      0.82        64

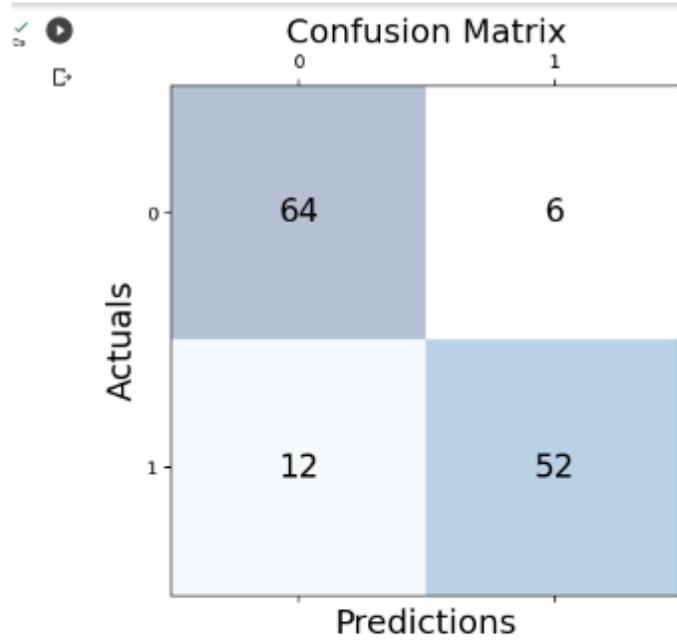
 accuracy              0.83        134
 macro avg              0.83      0.83      0.83       134
 weighted avg          0.83      0.83      0.83       134
    
```

Tabel Confusion Matrix Metode Decision Tree

Confusion Matrix Observed Class	Predicted Class		
		Positif	Negatif
Positif		TP = 58	FN = 12
Negatif		FP = 11	TN = 53

Akurasi : $\frac{TP+TN}{TP+FP+TN+FN} = \frac{58+53}{58+12+11+53} = \frac{121}{134} = 0,8283$

3) *Confusion Matrix Random Forest*



```

Classification report:
              precision    recall  f1-score   support

   negatif    0.84    0.91    0.88     70
   positif    0.98    0.81    0.85     64

 accuracy    0.87    0.87    0.87    134
 macro avg   0.87    0.86    0.86    134
 weighted avg 0.87    0.87    0.87    134
    
```

Confusion Matrix Observed Class	Predicted Class		
		Positif	Negatif
Positif		TP = 64	FN = 6
Negatif		FP = 12	TN = 52

$$\text{Akurasi} : \frac{TP+TN}{TP+FP+TN+FN} = \frac{64+52}{64+6+12+52} = \frac{116}{134} = 0,86567$$

f. Deployment

Setelah peneliti membangun dan mengevaluasi model analisis sentimen yang dibangun, pada tahap ini dilakukan pengujian atau uji prediksi suatu data teks. Dan visualiasasi teks menggunakan wordcloud.

1) Pengujian Data Teks

Pada model berikut, dilakukan pengujian data teks dengan menginputkan teks yang bernada sentimen negatif terhadap kebijakan kenaikan biaya haji 1444 H/ 2023 Masehi. Dan ketiga metode klasifikasi naïve bayes, decision tree, random forest memprediksi dengan benar data teks bersentimen negatif

```
29s ✓ 
input_text = input('Masukkan Opini anda mengenai kebijakan kenaikan ONH 2023: ')
vec_input_text = vec.transform([input_text])

print("Hasil Prediksi Model :")
print("Naive Bayes :", model1.predict(vec_input_text.toarray()))
print("Decision Tree :",model2.predict(vec_input_text.toarray()))
print("Random F0rest :",model3.predict(vec_input_text.toarray()))
```

↳ Masukkan Opini anda mengenai kebijakan kenaikan ONH 2023: sangat zolim ongkos haji naik bagi kami rakyat sangat berat pak
Hasil Prediksi Model :
Naive Bayes : ['negatif']
Decision Tree : ['negatif']
Random F0rest : ['negatif']

Pada model berikut, dilakukan pengujian data teks dengan menginputkan teks yang bernada sentimen positif terhadap kebijakan kenaikan biaya haji 1444 H/ 2023 Masehi. Dan ketiga metode klasifikasi naïve bayes, decision tree, random forest memprediksi dengan benar data teks bersentimen positif.

```
26s ✓ [21] input_text = input('Masukkan Opini anda mengenai kebijakan kenaikan ONH 2023: ')
vec_input_text = vec.transform([input_text])

print("Hasil Prediksi Model :")
print("Naive Bayes :", model1.predict(vec_input_text.toarray()))
print("Decision Tree :",model2.predict(vec_input_text.toarray()))
print("Random F0rest :",model3.predict(vec_input_text.toarray()))
```

Masukkan Opini anda mengenai kebijakan kenaikan ONH 2023: semoga kita semua diberi kemudahan untuk naik haji
Hasil Prediksi Model :
Naive Bayes : ['positif']
Decision Tree : ['positif']
Random F0rest : ['positif']

2) Visualiasi Wordcloud

Word cloud adalah visualisasi data teks yang menampilkan kata-kata dari kumpulan data teks di suatu model yang sudah dibangun. Visualisasi tersebut menjelaskan bahwa semakin besar ukuran suatu kata maka kata-kata tersebut yang frekuensi kemunculannya lebih banyak atau sering muncul dari suatu data teks.

Perbandingan Metode Klasifikasi Naïve Bayes, Decision Tree, Random Forest Terhadap Analisis Sentimen Kenaikan Biaya Haji 2023 Pada Media Sosial Youtube



Word cloud Positif



Word cloud Negatif



KESIMPULAN

Penelitian ini melakukan analisis sentimen dengan cara melakukan klasifikasi komentar positif dan negatif pada komentar di media sosial youtube mengenai kebijakan kenaikan ongkos biaya haji 1444 H / 2023 Masehi. Penelitian ini melakukan perbandingan hasil akurasi terhadap beberapa metode klasifikasi seperti *naïve bayes*, *decision tree* dan *random forest*. Dari penelitian ini didapatkan hasil akurasi *naïve bayes* sebesar 90%, *Decision Tree* sebesar 83%, *Random Forest* sebesar 87%. Diharapkan dengan adanya analisis sentimen klasifikasi komentar youtube ini dapat memberikan gambaran bagi pemerintah melihat respon penilaian dari masyarakat dalam membuat suatu kebijakan.

DAFTAR PUSTAKA

- Fiarni, C., Maharani, H., & Pratama, R. (2016). Sentiment analysis system for Indonesia online retail shop review using hierarchy Naive Bayes technique. *2016 4th International Conference on Information and Communication Technology (ICoICT)*, 1–6.
- Hadna, M. S., Santosa, P. I., & Winarno, W. W. (2016). Studi Literatur Tentang Perbandingan Metode Untuk Proses Analisis Sentimen Di Twitter. *Seminar Nasional Teknologi Informasi Dan Komunikasi, 2016*(Sentika), 57–64.
- Hidayat, A. N. (2015). Analisis Sentimen Terhadap Wacana Politik Pada Media Masa Online Menggunakan Algoritma Support Vector Machine Dan Naive Bayes. *Jurnal Elektronik Sistem Informasi Dan Komputer, 1*(1), 12–18.
- Kemenag. (2023a). *Keputusan Menteri no 352 th 2023*. 4.
- Kemenag. (2023b). *Kuota Haji 2023 Sebanyak 221 Ribu, Tidak Ada Pembatasan Usia*.
- Mahmudan, A. (2011). *Pengguna Youtube Indonesia Terbesar Ketiga di Dunia pada 2022*.
- Schröer, C., Kruse, F., & Gómez, J. M. (2021). A systematic literature review on applying CRISP-DM process model. *Procedia Computer Science, 181*(2019), 526–534. <https://doi.org/10.1016/j.procs.2021.01.199>
- Tyagi, P., & Tripathi, R. C. (2019). A review towards the sentiment analysis techniques for the analysis of twitter data. *Proceedings of 2nd International Conference on Advanced Computing and Software Engineering (ICACSE)*.

