

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

**Analisis Sentimen Tanggapan Masyarakat Terhadap Kenaikan
Biaya Haji Tahun 2023 Menggunakan Metode
K- Nearest Neighbor (KNN)**

TUGAS AKHIR

Disusun Sebagai Salah Satu Syarat
Untuk Memperoleh Gelar Sarjana Teknik
Pada Jurusan Teknik Informatika

Oleh

HAFSYAH
NIM. 11950125075



UIN SUSKA RIAU

FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI SULTAN SYARIF KASIM RIAU
PEKANBARU
2024



LEMBAR PERSETUJUAN

Analisis Sentimen Tanggapan Masyarakat Terhadap Kenaikan Biaya Haji Tahun 2023 Menggunakan Metode K- Nearest Neighbor (KNN)

TUGAS AKHIR

Oleh

HAFSYAH
NIM. 11950125075

Telah diperiksa dan disetujui sebagai Laporan Tugas Akhir
di Pekanbaru, pada tanggal 04 Januari 2024

Pembimbing I,

DR. Elin Haerani, S.T., M.Kom
NIP. 19810523 200710 2 003

UIN SUSKA RIAU

- Hak Cipta Dilindungi Undang-Undang**
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
 2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

LEMBAR PENGESAHAN

Analisis Sentimen Tanggapan Masyarakat Terhadap Kenaikan Biaya Haji Tahun 2023 Menggunakan Metode K- Nearest Neighbor (KNN)

Oleh

HAFSYAH

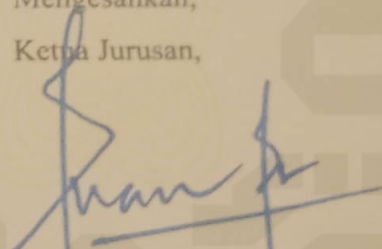
NIM. 11950125075

Telah dipertahankan di depan sidang dewan penguji
sebagai salah satu syarat untuk memperoleh gelar Sarjana Teknik
pada Universitas Islam Negeri Sultan Syarif Kasim Riau

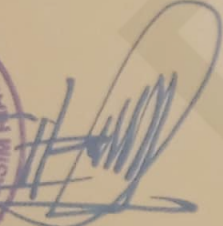
Pekanbaru, 04 Januari 2024

Mengesahkan,

Ketua Jurusan,


IWAN ISKANDAR, M.T.

NIP. 19821216 201503 1 003


DR. HARTONO, M.PD.

NIP. 19640301 199203 1 003

Dewan Penguji

Ketua : Iwan Iskandar, M.T.

Pembimbing I : DR. Elin Haerani, S.T., M. Kom.

Penguji I : Novriyanto, S.T., M. Sc

Penguji II : Fadhilah Syafria, S.T., M. Kom.

- Hak Cipta Dilindungi Undang-Undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
 2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



SURAT PERNYATAAN

© Hak cipta milik UIN Suska Riau
 State Islamic University of Sultan Syarif Kasim Riau

Hak Cipta dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Saya yang bertandatangan di bawah ini:

Nama : Hafsyah
 NIM : 11950125075
 Tempat, Tgl. Lahir : Tangun, 27 Desember 1999
 Fakultas : Sains dan Teknologi
 Prodi : Teknik Informatika
 Judul Jurnal :

Analisis Sentimen Tanggapan Masyarakat Terhadap Kenaikan Biaya Haji Tahun 2023 Menggunakan Metode K- Nearest Neighbor (KNN)

Menyatakan dengan sebenar-benarnya bahwa:

- 1. Penulisan jurnal dengan judul sebagaimana tersebut di atas adalah hasil pemikiran dan penelitian saya sendiri.
- 2. Semua kutipan pada karya tulis saya ini sudah disebutkan sumbernya.
- 3. Oleh karena itu jurnal saya ini, saya nyatakan bebas dari plagiat.
- 4. Apabila di kemudian hari terbukti terdapat plagiat dalam penulisan jurnal saya tersebut, maka saya bersedia menerima sanksi sesuai peraturan perundang-undangan.

Demikianlah Surat Pernyataan ini saya buat dengan penuh kesadaran dan tanpa paksaan dari pihak manapun juga.

Pekanbaru, 08 Januari 2024
 Yang membuat pernyataan



Hafsyah
 NIM. 11950125075

Analisis Sentimen Tanggapan Masyarakat Terhadap Kenaikan Biaya Haji Tahun 2023 Menggunakan Metode K- Nearest Neighbor (KNN)

Hafsyah, Elin Haerani, Novriyanto, Fadhilah Syafria*

Sains dan Teknologi, Teknik Informatika, Universitas Islam Negeri Sultan Syarif Kasim Riau, Pekanbaru, Indonesia

Email: ¹11950125075@students.uin-suska.ac.id, ²elin.haerani@uin-suska.ac.id, ^{3,*}novriyanto@uin-suska.ac.id,

⁴fadhilah.syafria@uin-suska.ac.id

Email Penulis Korespondensi: ¹11950125075@students.uin-suska.ac.id

Abstrak—Pemerintah Indonesia menerapkan kebijakan kenaikan biaya ibadah haji pada tahun 2023, namun kebijakan tersebut banyak menuai komentar positif dan negatif di kalangan masyarakat.. Komentar masyarakat diambil dari jejaring media sosial Twitter, karena memuat banyak informasi sehingga menarik minat sebagian besar masyarakat. Dengan adanya kenaikan biaya haji tahun 2023 maka perlu dilakukan analisis sentimen. Penelitian ini Menggunakan metode K-Nearest Neighbor karena mudah dalam penerapannya dan data yang digunakan dibagi menjadi dua kelas, positif dan negatif. Hasil penelitian penerapan metode K-Nearest Neighbor dalam analisis sentimen kenaikan biaya haji tahun 2023 dengan menggunakan 3.000 data yang diambil dari komentar Twitter. Data tweet yang digunakan, terdapat 1866 komentar positif dan 415 komentar negatif dan jumlah seluruh data bersih 2281, dilihat dari banyaknya data positif dibandingkan data negatif, mendapatkan nilai akurasi sebesar 81,17% pada pembagian data 70:30, 79,87% pada pembagian data 80:20, 77,73% pada pembagian data 90:10. Sementara, nilai akurasi tertinggi adalah 81,17% dengan precision 82,48%, recall 97,67%, F1- Score 89,43%. Pada penelitian ini terdapat lebih banyak respon positif, hal ini membuktikan bahwa kenaikan biaya haji tahun 2023 menggunakan metode K-Nearest Neighbor (KNN) dapat diterima oleh masyarakat.

Kata Kunci: Kenaikan Biaya Haji Tahun 2023; Biaya Haji; Twitter; Analisis Sentimen; K-Nearest Neighbor;

Abstract—The Indonesian government implemented a policy of increasing the cost of Hajj in 2023, but the policy has attracted many positive and negative comments among the public. Public comments are taken from the social media network Twitter, because it contains a lot of information so that it attracts the interest of most people. With the increase in Hajj costs in 2023, it is necessary to conduct sentiment analysis. This study uses the K-Nearest Neighbor method because it is easy to apply and the data used are divided into two classes, positive and negative. The results of research on the application of the K-Nearest Neighbor method in sentiment analysis of the increase in Hajj costs in 2023 using 3,000 data taken from Twitter comments. The tweet data used, there were 1866 positive comments and 415 negative comments and the total net data of 2281, judging from the amount of positive data compared to negative data, obtained an accuracy value of 81.17% in 70:30 data sharing, 79.87% in 80:20 data sharing, 77.73% in 90:10 data sharing. Meanwhile, the highest accuracy value was 81.17% with 82.48% precision, 97.67% recall, F1- Score 89.43%. In this study, there were more positive responses, this proves that the increase in Hajj costs in 2023 using the K-Nearest Neighbor (KNN) method can be accepted by the community.

Keywords: Hajj Cost Increase in 2023; Hajj Expenses; Twitter; Sentiment Analysis; K-Nearest Neighbor;

1. PENDAHULUAN

Berdasarkan data indonesiabaik.id kenaikan biaya haji dari tahun 2022 ke tahun 2023 sangat meningkat, yakni dari 39,8 juta menjadi 49,8 juta dalam kurun waktu satu tahun. Sehingga dengan adanya kenaikan biaya haji tersebut menjadikan adanya komentar positif dan negatif dikalangan calon jemaah haji ataupun masyarakat yang ingin melaksanakan haji. Diketahui sejak tahun 2014 biaya penyelenggaraan ibadah haji (BPIH) terus mengalami kenaikan. Namun, setiap tahun tahun presentasi besaran nilai manfaat dan Bipih pada BPIH tidak selalu sama. BPIH ini adalah besaran dana yang dipakai pemerintah sebagai operasional jemaah ketika melakukan ibadah haji. Sedangkan Bipih adalah bagian dari BPIH yang berarti besaran dana yang harus dibayar jemaah sebelum menunaikan ibadah haji.[1]

Twitter adalah sebuah situs *website* yang dioperasikan oleh *Twitter Inc.* yang menawarkan jaringan sosial berupa mikroblog sehingga memungkinkan penggunaanya untuk mengirim dan membaca pesan tweet. *Tweet* menampilkan beberapa spesifikasi seperti sebuah teks tulisan hingga 140 kata yang ditampilkan pada halaman profil pemilik akun, serta dapat dilihat secara publik, namun pengirim dapat membatasi pengiriman pesan ke daftar teman-teman mereka saja. Media sosial twitter digunakan untuk melakukan analisis sentimen terhadap berita yang sedang dibahas berdasarkan opini warganet dari twitter [2]. Penelitian ini menggunakan media sosial twitter sebagai sumber penelitian. Karena informasi dan data kenaikan biaya haji lebih banyak didapatkan di *twitter* dibandingkan media sosial lain.

Analisis sentimen merupakan cabang ilmu dari penelitian text mining atau data mining. Analisis sentimen bertujuan untuk mengekstrak atribut dari sebuah komentar (opini, sentimen, dan emosi) yang diekspresikan secara tekstual[3]. Penelitian ini merupakan penelitian kepustakaan (*library research*) atau studi kepustakaan yang memusatkan pencarian data hanya pada dokumen, arsip, maupun suatu kebijakan dan ketetapan aturan yang berlaku pada tahun penelitian.[4]

Penelitian ini penulis memutuskan untuk menggunakan metode *K-Nearest Neighbor (KNN)*. Alasan penulis menggunakan metode K-Nearest Neighbor karena mudah diimplementasikan, data yang digunakan memiliki label sehingga memudahkan dalam proses pengelompokan kedalam kelas yang paling sesuai. K-Nearest Neighbor memiliki kelebihan dalam aspek kemudahan menerjemahkan hasil, waktu perhitungan dan akurasi dari prediksi dibandingkan metode lainnya[5]. Data ini diperoleh dari *Twitter*. Data *Twitter* yang dikumpulkan sebanyak 3000 tweets. Data *Twitter* yang telah diperoleh selanjutnya akan diolah dengan preprocessing text yaitu dengan proses cleaning,



casefolding, tokenizing, normalisasi, stopword removal, stemming. selanjutnya data akan di filter untuk diambil mana data yang memiliki sentimen positif dan mana yang memiliki sentimen negatif[6].

Metode *K-Nearest Neighbor* yang ditujukan untuk mengolah dan mengklasifikasikan opini sehingga akan diketahui klasifikasi dari opini tersebut (positif atau negatif)[7]. Data review diambil dalam bahasa Indonesia yang memiliki kecenderungan positif atau negatif untuk membantu usaha dalam menganalisis kenaikan biaya haji menggunakan metode *K-Nearest Neighbor*[8]. Sentimen akan diklasifikasikan menjadi dua yaitu positif dan negatif. Untuk mengklasifikasikannya tersebut bisa menggunakan metode *K-Nearest Neighbor* (KNN) yang merupakan sebuah metode machine learning yang dapat melakukan klasifikasi terhadap objek berdasarkan dataset atau data latih berdasarkan jarak yang paling dekat dengan objek tersebut.

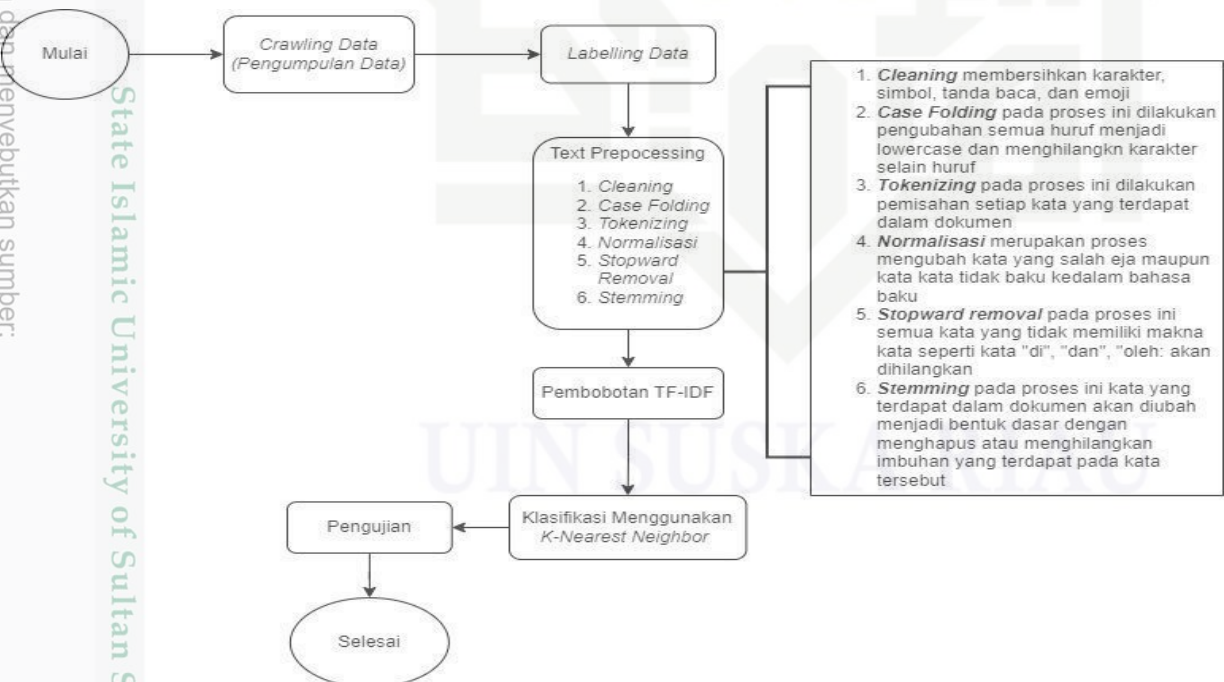
Berdasarkan penelitian (F. Istighfarizky, N. A. Sanjaya ER, I. M. Widiartha, L. G. Astuti, I. G. N. A. C. Putra, and I. K. G. Suhartana) "Klasifikasi Jurnal menggunakan Metode KNN dengan Mengimplementasikan Perbandingan Seleksi Fitur,". Kesimpulan dari penelitian bahwa implementasi seleksi fitur dapat meningkatkan performa precision, recall, F1-Score, dan akurasi dari metode *K-Nearest Neighbor* dalam mengklasifikasikan jurnal dan seleksi fitur Chi-Square lebih unggul daripada seleksi fitur Gini Index. Pada eksperimen seleksi fitur Chi-Square, kombinasi threshold 1% dengan parameter nilai $k=6$ adalah kombinasi yang dipilih menjadi model terbaik. Threshold ini menyeleksi sekitar 31 fitur dengan nilai Chi-Square tertinggi. Pada pengujian data baru, model *K-Nearest Neighbor* dengan seleksi fitur *Chi-Square* menghasilkan performa precision, recall, F1-Score, dan akurasi secara berturut-turut yaitu 85%, 83.3%, 88.2%, dan 92.3%. Pada eksperimen seleksi fitur Gini Index, kombinasi threshold 1% dengan parameter nilai $k=4$ adalah kombinasi yang dipilih menjadi model terbaik. Threshold ini menyeleksi sekitar 31 fitur dengan nilai Gini Index tertinggi. Pada pengujian data baru, model *K-Nearest Neighbor* dengan seleksi fitur Gini Index menghasilkan performa precision, recall, F1-Score, dan akurasi secara berturut-turut yaitu 81.2%, 80.3%, 81.6%, dan 86.6%[9].

Penelitian ini melakukan analisis sentimen dengan menggunakan metode *K-Nearest Neighbor* (KNN) dengan cara *crawling* data *twitter* dan melabelkan data secara manual. Kemudian mencari hasil akurasi dilakukan pengujian menggunakan *library python* dan *confusion matrix*[10]. Penelitian ini juga bermanfaat untuk mengumpulkan sentimen masyarakat bertanggapan positif dan negatif terhadap kenaikan biaya haji tahun 2023[11].

2. METODOLOGI PENELITIAN

2.1 Langkah-Langkah Penelitian

Langkah-langkah dalam riset ini melibatkan penerapan metode *K-Nearest Neighbor*, yang dijelaskan melalui alur proses dalam desain penelitian berikut ini:



Gambar 1. Desain Penelitian

2.2 Pengumpulan data

Pada pengumpulan data ini dilakukan labelling data, selanjutnya yaitu *text preprocessing* (*cleaning, case folding, tokenizing, normalisasi, stopword removal, stemming*).



2.3 Pembobotan Tf-Idf

Sebuah metode yang merupakan *integrasi antar term frequency* (TF), dan *invers document frequency* (IDF). Untuk mencari representasi nilai dari tiap tiap dokumen dari suatu kumpulan data training (*training set*) dimana nantinya dibentuk suatu vektor antara dokumen dengan kata (*document with terms*)[15].

$$IDF = \log \log \frac{D}{DF} \tag{1}$$

$$TF - IDF = TF * IDF \tag{2}$$

Keterangan:

- D : jumlah dokumen pada data latih
- DF : jumlah dokumen yang memiliki term
- TF : jumlah kemunculan term pada suatu dokumen
- IDF : jumlah dokumen yang mengandung term

2.4 Klasifikasi Menggunakan K-Nearest Neighbor

Metode klasifikasi objek untuk data latih dengan memilih kelas fitur berdasarkan jarak terdekat dengan data latih, algoritma ini dapat mengklasifikasikan objek baru. Kn merupakan metode klasifikasi dengan menghitung nilai *k* (*neighbor*)[16]. K adalah titik kepala objek, nilai k akan dibandingkan untuk menentukan klasifikasi objek. Model *k-nearest neighbor* pada penelitian ini diimplementasikan dalam bahasa *python* menggunakan fungsi *K-Nearest Neighbor Classifier* dari *library sklearn*[17].

$$dis(x_1, x_2) = \sqrt{\sum_{i=0}^n (x_{1i} - x_{2i})^2} \tag{3}$$

Keterangan:

- x_1 : sample data
- x_2 : data uji
- i : variabel data

2.5 Pengujian

Pada tahap pengujian ini menggunakan metode tingkat akurasi dengan tabel *Confussion Matrix*. Dari pengumpulan data awal hingga penilaian akhir. Dengan mengukur kinerja menggunakan matriks konfusi, akurasi dicapai sehingga dapat melakukan pengujian. Berbagai data *tweet* dari pengguna *Twitter* terkait permasalahan yang diangkat peneliti digunakan untuk mengetahui keakuratan KNN melalui matriks konfusi. Dengan memeriksa pustaka matriks konfusi *Python*, hal ini dilakukan dengan menghitung nilai presisi, presisi, dan perolehan untuk TP, TN, FP, dan FN[18].

Tabel 1. Confussion Matrix

	Actual positive	Actual Negative
Predicted Positive	TP	FP
Predicted Negative	FN	TN

Keterangan:

1. TP: jumlah total yang memiliki nilai prediksi dan nilai aktual adalah sama
2. FP: jumlah total yang nilai prediksinya positif namun nilai sebenarnya negatif
3. FN: jumlah total nilainya prediksinya negatif namun nilai sebenarnya positif
4. TN: jumlah total yang memiliki nilai prediksi dan nilai aktual adalah negatif

Dibawah ini adalah rumus tahapan perhitungan *accuracy*, *precision*, *recall*, dan *F1-score*.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} * 100 \tag{4}$$

$$Precision = \frac{TP}{TP+FP} * 100 \tag{5}$$

$$Recall = \frac{TP}{TP+FN} * 100 \tag{6}$$

$$F1 - Score = 2 * \frac{Precision*Recall}{Precision+Recall} \tag{7}$$

3. HASIL DAN PEMBAHASAN

3.1 Mengumpulkan Data

Proses pengambilan data atau informasi dari *platform* media sosial *twitter* proses ini dilakukan dengan menggunakan *API* (*aplication programming interface*) dengan kata kunci “kenaikan biaya haji dan biaya haji 2023” menggunakan bahasa Indonesia. Pada proses pengumpulan data ini digunakan bahasa *python*.



2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

2023-02-16 03:34:35+00:00,1.62606E+18,kristariyanto68,,in,,,"b'DOK! Biaya Penyelenggaraan Ibadah Haji 2023 Final, Segini Besar dan Rinciannya https://t.co/BWB4LtgFUW"	
2023-02-16 03:32:11+00:00,1.62606E+18,abu_abros,,in,,,"b'Biaya Penyelenggaraan Ibadah Haji (BPIH) 1444 H/2023 M. https://t.co/XfcRkoGCE"	
2023-02-16 03:30:19+00:00,1.62606E+18,MNCTrijayaBdg,"Bandung, Indonesia",in,,,"b'Dukung Biaya Haji 2023 kenaikan, Calon Jemaah Haji: Tak Adil Jika Ambil Subsidi Hak Jemaah Tunggu https://t.co/...	
2023-02-16 03:30:19+00:00,1.62606E+18,MNCTrijayaBdg,"Bandung, Indonesia",in,,,"b'Dukung Biaya Haji 2023 kenaikan, Calon Jemaah Haji: Tak Adil Jika Ambil Subsidi Hak Jemaah Tunggu https://t.co/...	
2023-02-16 03:30:00+00:00,1.62606E+18,sandra_hut,,in,,,"b'Pemerintah dan Komisi VIII DPR RI menyetujui biaya haji 2023 yang perlu ditanggung jemaah sebesar Rp 49.812.700,26. Ini berlaku untuk https://t.co/...	
2023-02-16 03:21:02+00:00,1.62606E+18,womanindonesia,"Jakarta Selatan, DKI Jakarta",in,,,"b'Biaya Haji Tahun 2023 Resmi Jadi Rp49,8 Juta per Jemaah https://t.co/my5h3uIiCY"	
2023-02-16 03:19:56+00:00,1.62606E+18,Media_Asuransi,Jakarta,in,,,"b'Rapat Panitia Kerja (Panja) Komisi VIII DPR RI bersama Kementerian Agama (Kemenag) sepakati Biaya Perjalanan Ibadah Haji https://t.co/...	
2023-02-16 03:19:49+00:00,1.62606E+18,jawapos,Jakarta,in,,,"b'Sah! Biaya Haji 2023 Yang Ditanggung Jemaah Rp 49,8 Juta, Begini Rinciannya https://t.co/o1B18q2ARK"	
2023-02-16 03:18:49+00:00,1.62606E+18,897GlobalRadio,,in,["GlobalNews"],,"b'Dukung Biaya Haji 2023 kenaikan, Calon Jemaah Haji: Tak Adil Jika Ambil Subsidi Hak Jemaah Tunggu https://t.co/kzQ: https://t.co/kzQ:	
2023-02-16 03:18:49+00:00,1.62606E+18,897GlobalRadio,,in,["GlobalNews"],,"b'Dukung Biaya Haji 2023 kenaikan, Calon Jemaah Haji: Tak Adil Jika Ambil Subsidi Hak Jemaah Tunggu https://t.co/kzQ: https://t.co/kzQ:	
2023-02-16 03:15:03+00:00,1.62606E+18,tvOneNews,Pulo Gadung,,in,,,"b'Tok, Biaya Haji 2023 kenaikan Jadi Rp 49,8 Juta https://t.co/2r31v24vW0"	
2023-02-16 03:15:03+00:00,1.62606E+18,tvOneNews,Pulo Gadung,,in,,,"b'Tok, Biaya Haji 2023 kenaikan Jadi Rp 49,8 Juta https://t.co/2r31v24vW0"	
2023-02-16 03:14:40+00:00,1.62606E+18,RILISonline,Jakarta,in,,,"b'Tok! DPR Tetapkan Biaya Haji 2023 Sebesar Rp49,8 Juta https://t.co/M1tj5Q9M7L https://t.co/PXdv1V3EGc"	
2023-02-16 03:12:03+00:00,1.62605E+18,DaulainAhmad,,in,,,"b'Biaya Haji Reguler 2023 Rp90 Juta, Menag Ingatkan Keberlangsungan Nilai Manfaat https://t.co/OOmTim139n"	
2023-02-16 03:01:30+00:00,1.62605E+18,Beritasatu,Jakarta,in,,,"b'Ini Perincian Biaya Haji 2023 Sebesar Rp 90,05 Juta https://t.co/Vznpnytp24"	
2023-02-16 03:00:37+00:00,1.62605E+18,id_resonansi,Yogyakarta,in,,,"b'Pemerintah dan DPR Sepakati Biaya Perjalanan Haji Tahun 2023 sebesar 49,8 Juta https://t.co/9DnMdMR17L"	
2023-02-16 03:00:01+00:00,1.62605E+18,voaindonesia,"Washington, DC",in,,,"b'Tweeps, pemerintah & https://t.co/...	

Gambar 2. Hasil Data Tweet

3.1.1 Labelling Data

Pada tahap pelabelan data ini dilakukan secara manual dan menggunakan validator (Guru Bahasa Indonesia) data yang digunakan adalah 3000 data tweet. Dan untuk pelabelan ini dibagi menjadi dua kelas yaitu positif dan negatif, setelah dilakukan pelabelan pada data terdapat banyak data yang ganda, jadi jumlah data yang didapatkan yaitu 2281 data bersih, yang jumlah data positif adalah 1866 dan jumlah data negatif 415. Pelabelan ini dilakukan dengan memberi label pada setiap komentar dari pengguna jejaring sosial media twitter.

Username	Komentar	Label
KompasData	b'Biaya haji tahun 2023 kenaikan tinggi dibandingkan sebelumnya, yaitu total Rp 84 juta - Rp 96 juta. Dari jumlah itu, Rp 44,3 juta - Rp 55,9 juta ditanggung langsung jemaah. Biaya mahal ini mesti diikuti peningkatan layanan. \nhttps://t.co/ODKmJhmQk\n\n#Opini #AdadiKompas'	Negatif
hariankompas	b'Akun TikTok @hariankompas dalam #AADK edisi Selasa, 11 April 2023 mengulas tentang pemerintah telah menetapkan nominal biaya penyelenggaraan ibadah haji tahun ini untuk setiap embarkasi.\n\nhttps://t.co/QHmu8eDWY4\n\n#Haji2023 #AdadiKompas https://t.co/y7phmc5MOj'	Negatif
yahya_id	b'Biaya Haji Tahun 2023 Tiap Daerah Berbeda, Embarkasi Surabaya Paling Mahal, Aceh Paling Rendah https://t.co/2xrlALx0CW'	Negatif
BPKHRI	b'Keppres (Keppres) Nomor 7 Tahun 2023 tentang Biaya Penyelenggaraan Ibadah Haji Tahun 1444 Hijriah/2023 diatur bahwa BPIH bersumber dari Bipih dan nilai manfaat setoran Bipih jemaah haji reguler. \n\nBipih sendiri diperoleh dari tiga sumber yakni jemaah haji, petugas haji daerah'	Positif
BPKHRI	b'Assalamualaikum #HajiMuda,\n\nBPIH 1444 H/2023 M rata-rata Rp90.050.637,26 per jemaah haji reguler. BPIH merupakan biaya riil yang dibutuhkan setiap jemaah untuk dapat menjalankan ibadah haji, sedangkan Bipih adalah biaya perjalanan ibadah haji yang dibayarkan calon jemaah. \n\nDalam https://t.co/NAqczV39Zk'	Negatif
radiopatria	b'Kantor Kemenag Kabupaten Blitar menyebut BPIH untuk Embarkasi Surabaya paling mahal dibanding daerah lain, karena faktor pesawat terbang yang akan digunakan.\n\nJoko Widodo Presiden telah mengeluarkan Keppres Nomor 7 Tahun 2023 untuk Biaya Penyelenggaraan Ibadah Haji (BPIH) tidak https://t.co/1HvX4uNvri'	Negatif
hariankompas	b'Pemerintah telah menetapkan nominal biaya penyelenggaraan ibadah haji tahun 1444 Hijriah/2023 Masehi untuk setiap embarkasi atau tempat pemberangkatan. \n\nhttps://t.co/u3ephBiYc9\n\n#Haji2023 #AdadiKompas @Kemenag_RI'	Positif

Gambar 3. Hasil Pelabelan Data

3.2 Text preprocessing

Pada tahap ini peneliti melakukan tahapan preprocessing yang bertujuan untuk membersihkan data.

1	Komentar	cleaning	casefolding	tokenisasi	normalisasi	stopwordremoval	stemming
1	b'Siapkan Diri Anda untuk Merespon Panggilan Haji dengan Biaya Mulai dari 200 Jutaan. Berangkat Tahun Ini bersama para Habib, dan kyai yang sudah berpengalaman! \n-----\n-----\n\nhttps://t.co/6nGbwZVSWww https://t.co/8wCJFiiWqA'	b Siapkan Diri Anda untuk Merespon Panggilan Haji dengan Biaya Mulai dari 200 Jutaan Berangkat Tahun Ini bersama para Habib dan kyai yang sudah berpengalaman n n	b siapkan diri anda untuk merespon panggilan haji dengan biaya mulai dari jutaan berangkat tahun ini bersama para habib dan kyai yang sudah berpengalaman n n	b, siapkan, diri, anda, untuk, merespon, panggilan, haji, dengan, biaya, mulai, dari, jutaan, berangkat, tahun, ini, bersama, para, habib, dan, kyai, yang, sudah, berpengalaman, dan, dan,	b, siapkan, diri, anda, untuk, merespon, panggilan, haji, dengan, biaya, mulai, dari, jutaan, berangkat, tahun, ini, bersama, para, habib, dan, kyai, yang, sudah, berpengalaman, n, dan, dan,	siapkan, merespon, panggilan, haji, biaya, jutaan, berangkat, tahun, habib, kyai, berpengalaman,	merespon, panggil, haji, biaya, juta, berangkat, tahun, habib, kyai, alam,
2							

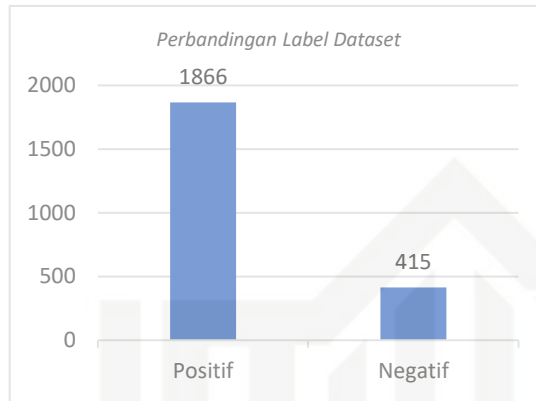
Gambar 4. Proses Text Preprocessing



Setelah text preprocessing selanjutnya dilakukan TF-IDF. TF-IDF ini merupakan proses pemberian bobot term pada dokumen. Pembobotan ini digunakan nantinya oleh algoritma machine learning untuk klasifikasi dokumen.

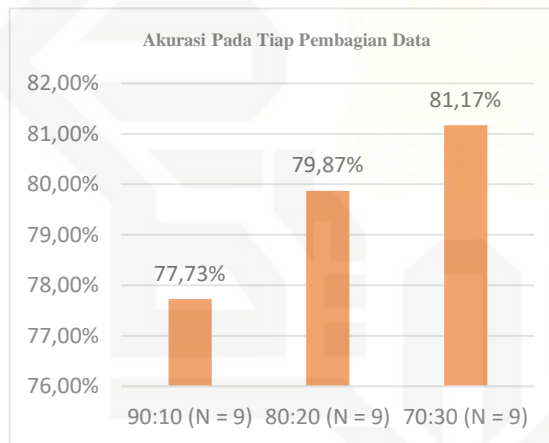
3.3 Hasil Pengujian

Hasil pengujian pada penelitian ini, peneliti menggunakan 3000 data yang berfungsi data uji dan data latih. Data *tweet* yang sudah diseleksi kemudian diberi label secara manual dengan menggunakan validator dan memperoleh dua label yaitu positif dan negatif, dengan kelas positif berjumlah 1866 dan jumlah kelas negatif 415.



Gambar 5. Perbandingan Label Dataset

Pada hasil pengujian ini dilakukan dengan tiga percobaan, percobaan yang pertama dilakukan menggunakan perbandingan data latih yaitu 90% dan data uji 10% terdapat akurasi 77,73%, percobaan yang kedua yaitu data latih 80% dan data uji 20% terdapat akurasi 79,87%, percobaan yang ketiga dengan data latih 70% dan data uji 30% terdapat akurasi 81,17%. Akurasi yang didapatkan pada ketiga perbandingan tersebut ditampilkan pada gambar berikut.



Gambar 6. Grafik Perbandingan Akurasi

Tabel 2. Perbandingan Akurasi Dengan Nilai N Yang Berbeda

	Pengujian 70:30		Pengujian 80:20		Pengujian 70:30	
	N-Neighbor	Akurasi	N-Neighbor	Akurasi	N-Neighbor	Akurasi
3		77,81%	3	75,11%	3	75,11%
5		79%	5	76,81%	5	76,42%
7		79,71%	7	77,24%	7	74,67%
9		81,17%	9	79,87%	9	77,73%

Berikut dilampirkan gambar kodingan pada pengujian 70%:30%, penjelasan dari gambar kodingan dibawah ini yaitu pada tanda merah adalah data uji yang digunakan *test_size = 0.3* berarti data pelatihan yang digunakan yaitu 70% dan data uji yang digunakan yaitu 30%. Kemudian pada tanda hijau adalah model klasifikasi K-Nearest Neighbor (*KNeighborsClassifier*) menggunakan data pelatihan *KNeighborsClassifier (n_neighbor = 9)*. Dan pada tanda biru adalah hasil dari pengujian akurasi, akurasi yang didapatkan pada pengujian ini yaitu 81,17%.



```
#data = ubahdata(fitur)
from sklearn.model_selection import train_test_split
x_train, x_test, y_train, y_test = train_test_split(X, kelas, test_size=0.3, random_state=0)
x_test

<685x3319 sparse matrix of type '<class 'numpy.float64''>'
with 8628 stored elements in Compressed Sparse Row format>

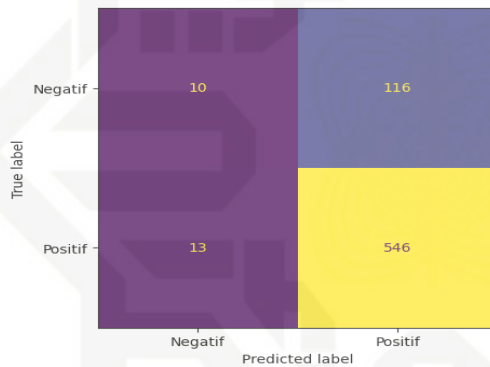
from sklearn.neighbors import KNeighborsClassifier
knn = KNeighborsClassifier(n_neighbors=9)
knn_model = knn.fit(x_train, y_train)
knn_predict = knn_model.predict(x_test)
#print(knn_predict)

from sklearn.metrics import accuracy_score
score_knn = accuracy_score(knn_predict, y_test)
print("Accuracy with KNN: ",score_knn)
print(accuracy_score(y_test, knn_predict)*100)
#print(knn_predict)

Accuracy with KNN: 0.8116788321167884
81.16788321167884
```

Gambar 8. Kodingan Pada Pengujian 70:30

Pada gambar dibawah ini, ditampilkan *confussion matrix* hasil pengujian 70%:30% dengan nilai True Positif (TP) 546, True Negatif (TN) 10, False Positif (FP) 116, False Negatif (FN) 13, maka di dapatkan akurasi yaitu 81,17%.



Gambar 9: Confussion Matrix

Pengujian akurasi

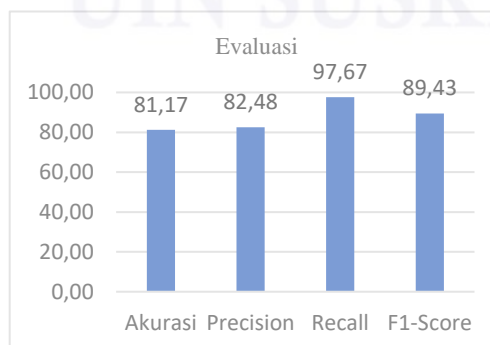
$$= \frac{546+10}{546+10+116+13} \times 100\%$$

$$= \frac{556}{685} \times 100\%$$

$$= 81,17\%$$

Tabel 3. Perbandingan Nilai K

Nilai K	Confussion Matrix	Akurasi
1	70%:30%	81,17%
3	80%:20%	79,87%
3	90%:10%	77,73%



Gambar 9. Hasil Evaluasi

1. Hak Cipta Dilindungi Undang-Undang
 2. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
 2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



Selanjutnya dilakukan tahapan perhitungan nilai yang telah dibangun menggunakan metode *K-Nearest Neighbor*, dan dilakukan evaluasi dengan melihat *confussion matrix* dan menghitung nilai *accuracy*, *precision*, *recall* dan *F1-Score*.

Hasil evaluasi pada akurasi tertinggi (81,17%).

$$\begin{aligned} \text{Accuracy} &= \frac{546+10}{546+10+116+13} \times 100\% \\ &= \frac{556}{685} \times 100\% \\ &= 81,17\% \end{aligned}$$

$$\begin{aligned} \text{Precision} &= \frac{546}{546+116} \times 100\% \\ &= \frac{546}{662} \times 100\% \\ &= 82,48\% \end{aligned}$$

$$\begin{aligned} \text{Recall} &= \frac{546}{546+13} \times 100\% \\ &= \frac{546}{559} \times 100\% \\ &= 97,67\% \end{aligned}$$

$$\begin{aligned} \text{F1 - Score} &= 2 * \frac{82,48 * 97,67}{82,48 + 97,67} \\ &= 2 * \frac{8.005.8216}{180.15} \\ &= 89,43\% \end{aligned}$$

4. KESIMPULAN

Berdasarkan data *tweet* yang digunakan, terdapat 1866 komentar positif dan 415 komentar negatif, dilihat dari banyaknya data positif dibandingkan data negatif, hasil penelitian mengenai penerapan metode *K-Nearest Neighbor* (KNN) dalam analisis sentimen terhadap kenaikan biaya haji tahun 2023 dengan menggunakan 3000 data yang diambil dari komentar *twitter* mendapatkan nilai akurasi sebesar 81,17% pada pembagian data 70:30, 79,87% pada pembagian data 80:20, 77,73% pada pembagian data 90:10. Sementara, nilai akurasi tertinggi adalah 81,17% karena data latih yang digunakan adalah data yang mudah dikenali oleh mesin, dengan *precision* 82,48%, *recall* 97,67%, *F1 - Score* 89,43%. Pada penelitian ini terdapat lebih banyak respon positif, hal ini membuktikan bahwa kenaikan biaya haji tahun 2023 menggunakan metode *K-Nearest Neighbor* (KNN) dapat diterima oleh masyarakat.

REFERENCES

- Nurhanisah, Yuli. "Biaya Haji Indonesia Tahun 2023." Indonesia Baik.Id Diakses Tanggal 24 Oktober 2023. <https://indonesiabaik.id/infografis/biaya-haji-indonesia-tahun-2023>
- M. Riefky and W. Pramesti, "Sentiment Analysis of Southeast Asian Games (SEA Games) in Philippines 2019 Based on Opinion of Internet User of Social Media Twitter with K-Nearest Neighbor and Support Vector Machine," *J. Mat. Stat. dan Komputasi*, vol. 17, no. 1, pp. 26–41, 2020, doi: 10.20956/jmsk.v17i1.9947.
- F. Astuti, R. M. Candra, S. Agustian, and S. Ramadhani, "Klasifikasi Sentimen Masyarakat Terhadap Pemerintah Terkait Penerapan Kebijakan New Normal Menggunakan Metode K-Nearest Neighbor," *J. Nas. Komputasi dan Teknol. Inf.*, vol. 5, no. 3, pp. 531–538, 2022, doi: 10.32672/jnkti.v5i3.4455.
- A. Y. Permana and M. Makmun, "Analisis Sentimen pada Teks Opini Penilaian Kinerja Dosen dengan Pendekatan Algoritma KNN," *J. Ilm. Komputasi*, vol. 19, no. 1, pp. 39–50, 2020, doi: 10.32409/jikstik.19.1.154.
- V. Alvian, A. Nilogiri, and H. Azizah, "Klasifikasi Siswa Berprestasi Menggunakan Metode K-Nearest Neighbor (KNN) Pada SMA Negeri 2 Situbondo," *J. Smart Teknol.*, vol. 3, no. 6, pp. 602–611, 2022, [Online]. Available: <http://jurnal.unmuhjember.ac.id/index.php/JSTH>.
- S. S. Salim and J. Mayary, "Analisis Sentimen Pengguna Twitter Terhadap Dompot Elektronik Dengan Metode Lexicon Based Dan K – Nearest Neighbor," *J. Ilm. Inform. Komput.*, vol. 25, no. 1, pp. 1–17, 2020, doi: 10.35760/ik.2020.v25i1.2411.
- Normah, B. Rifai, S. Vambudi, and R. Maulana, "Analisa Sentimen Perkembangan Vtuber Dengan Metode Support Vector Machine Berbasis SMOTE," *J. Tek. Komput. AMIK BSI*, vol. 8, no. 2, pp. 174–180, 2022, doi: 10.31294/jtk.v4i2.
- R. Putri Fitrianti, A. Kurniawati, D. Agustien, J. Sistem Informasi, and F. Ilmu Komputer dan Teknologi Informasi, "Implementasi Algoritma K-Nearest Neighbor Terhadap Analisis Sentimen Review Restoran Dengan Teks Bahasa Indonesia," *Semin. Nas. Apl. Teknol. Inf.*, pp. 1907–5022, 2019.
- J. Homepage, S. R. Cholil, T. Handayani, R. Prathivi, and T. Ardianita, "IJCIT (Indonesian Journal on Computer and Information Technology) Implementasi Algoritma Klasifikasi K-Nearest Neighbor (KNN) Untuk Klasifikasi Seleksi Penerima Beasiswa," *IJCIT (Indonesian J. Comput. Inf. Technol.)*, vol. 6, no. 2, pp. 118–127, 2021.
- T. Wiratama Putra, A. Triayudi, and A. Andrianingsih, "Analisis Sentimen Pembelajaran Daring Menggunakan Metode Naïve Bayes, KNN, dan Decision Tree," *J. JTIK (Jurnal Teknol. Inf. dan Komunikasi)*, vol. 6, no. 1, pp. 20–26, 2022, doi: 10.35870/jtik.v6i1.368.
- S. N. Arafah and F. Fathoni, "Sentiment Analysis Pada Masyarakat Terhadap LRT Kota Palembang Menggunakan Metode Improved K-Nearest Neighbor," *J. Media Inform. Budidarma*, vol. 6, no. 3, p. 1554, 2022, doi: 10.30865/mib.v6i3.4434.
- N. S. P. Juana, E. Haerani, F. Syafrina, and E. Budianita, "Analisis Sentimen Tanggapan Masyarakat Terhadap Calon Presiden 2024 Ridwan Kamil Menggunakan Metode Naive Bayes Classifier," *J. Sist. Komput. dan Inform.*, vol. 4, no. 4, p. 570, 2023, doi: 10.30865/json.v4i4.6168.
- S. Rahayu, Y. MZ, J. E. Bororing, and R. Hadiyat, "Implementasi Metode K-Nearest Neighbor (K-NN) untuk Analisis Sentimen Kepuasan Pengguna Aplikasi Teknologi Finansial FLIP," *Edumatic J. Pendidik. Inform.*, vol. 6, no. 1, pp. 98–106, 2022, doi: 10.29408/edumatic.v6i1.5433.
- A. Firdaus, "Aplikasi Algoritma K-Nearest Neighbor pada Analisis Sentimen Omicron Covid-19," *J. Ris. Stat.*, pp. 85–92, 2022, doi: 10.29313/jrs.v2i2.1148.
- N. K. Widayanti, I. K. G. Darma Putra, and N. K. Dwi Rusjyanthi, "Seleksi Fitur Bobot Kata dengan Metode TFIDF untuk Ringkasan Bahasa Indonesia," *J. Ilm. Merpati (Menara Penelit. Akad. Teknol. Informasi)*, vol. 6, no. 2, p. 119, 2018, doi: 10.24843/jim.2018.v06.i02.p06.
- A. R. Isnain, J. Supriyanto, and M. P. Kharisma, "Implementation of K-Nearest Neighbor (K-NN) Algorithm For Public Sentiment Analysis of Online Learning," *IJCCS (Indonesian J. Comput. Cybern. Syst.)*, vol. 15, no. 2, p. 121, 2021, doi:



10.22146/ijccs.65176.

[17] Febryananda, Ade Rizky. "Mengenal Algoritma Klasifikasi K-Nearest Neighbor dalam Analisis Sentimen pada Data Ulasan suatu Produk." SRK. diakses tanggal 07 maret 2023. https://lab_adrk.ub.ac.id/id/mengenal-algoritma-klasifikasi-k-nearest-neighbor-dalam-analisis-sentimen-pada-data-ulasan-suatu-produk/

[18] Fatma, Hayati, E. Haerani, F. Syafria, and E. Budianita, "Analisis Sentimen Tanggapan Masyarakat Terhadap Calon Presiden 2024 Ridwan Kamil Menggunakan Metode K-Nearest Neighbor" J. Sist. Komput. dan Inform., vol. 8, no. 2, p. 139, 2023, doi: 10.30865/json.v4i4.6168.

Disetujui dan Diprofilkan oleh
Dipindangi Undang-Undang

Revisi 1.0
Copyright milik UIN Suska Riau

State Islamic University of Sultan Syarif Kasim Riau



2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.