

**ANALISIS SENTIMEN TERHADAP QATAR SEBAGAI TUAN
RUMAH PIALA DUNIA FIFA 2022 PADA MEDIA SOSIAL
TWITTER MENGGUNAKAN ALGORITMA C4.5**

SKRIPSI



**Oleh:
DIMITRI EV HAMUS SIROJJ
NIM. 18680025**

**PROGRAM STUDI PERPUSTAKAAN DAN ILMU INFORMASI
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2023**

**ANALISIS SENTIMEN TERHADAP QATAR SEBAGAI TUAN RUMAH
PIALA DUNIA FIFA 2022 PADA MEDIA SOSIAL TWITTER
MENGUNAKAN ALGORITMA C4.5**

SKRIPSI

Oleh:

DIMITRI EV HAMUS SIROJJ

NIM. 18680025

Diajukan kepada:

Fakultas Sains dan Teknologi

Universitas Islam Negeri Maulana Malik Ibrahim Malang

Untuk memenuhi salah satu persyaratan dalam

Memperoleh Gelar Sarjana Sains dan Informasi (S.S.I.)

**PROGRAM STUDI PERPUSTAKAAN DAN ILMU INFORMASI
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2023**

LEMBAR PERSETUJUAN

**ANALISIS SENTIMEN TERHADAP QATAR SEBAGAI TUAN RUMAH
PIALA DUNIA FIFA 2022 PADA MEDIA SOSIAL TWITTER
MENGUNAKAN ALGORITMA C4.5**

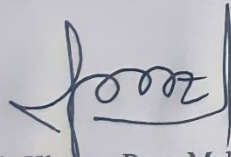
SKRIPSI

Oleh:
DIMITRI EV HAMUS SIROJJ
NIM: 18680025

Telah diperiksa dan disetujui untuk diuji:

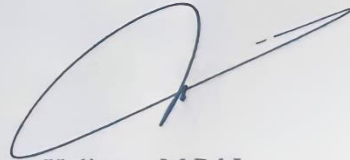
Tanggal: 11 Oktober 2023

Pembimbing I




Fakhri Khusnu Reza Mahfud, M.Kom.
NIP. 199005062019031007

Pembimbing II



Yulianto, M.Pd.I.
NIP. 198707122019031005

Mengetahui,
Ketua Program Studi Perpustakaan dan Ilmu Informasi
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang



Dr. Ir. Mokhammad Amin Hariyadi, MT
NIP. 196701182006011001

LEMBAR PENGESAHAN

ANALISIS SENTIMEN TERHADAP QATAR SEBAGAI TUAN RUMAH PIALA DUNIA FIFA 2022 PADA MEDIA SOSIAL TWITTER MENGUNAKAN ALGORITMA C4.5

SKRIPSI

Oleh:

DIMITRI EV HAMUS SIROJJ

NIM: 18680025

Telah dipertahankan di depan Dewan Penguji Skripsi
dan dinyatakan diterima sebagai salah satu persyaratan
untuk memperoleh Gelar Sarjana Sains Informasi (S.S.I.)
pada tanggal 11 Oktober 2023

Susunan Dewan Penguji

Ketua Penguji : Dr. Ir. Mokhammad Amin Hariyadi, MT
NIP. 196701182005011001

Anggota Penguji I : Wahyu Hariyanto, M.M
NIP. 198907212019031007

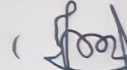
Anggota Penguji II : Fakhris Khusnu Reza Mahfud, M.Kom
NIP. 199005062019031007

Anggota Penguji III : Yulianto, M.Pd.I
NIP. 198707122019031005

Tanda Tangan

()

()

()

()

Mengetahui dan mengesahkan,
Ketua Program Studi Perpustakaan dan Ilmu Informasi
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang



Dr. Ir. Mokhammad Amin Hariyadi, MT
NIP. 196701182005011001

PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan di bawah ini:

Nama : Dimitri Ev Hamus Sirojj

NIM : 18680025

Prodi : Perpustakaan dan Ilmu Informasi

Fakultas : Sains dan Teknologi

Judul Skripsi : Analisis Sentimen Terhadap Qatar Sebagai Tuan Rumah Piala Dunia FIFA 2022 Pada Media Sosial Twitter Menggunakan Algoritma C4.5

Menyatakan bahwa skripsi ini ditulis dengan tulisan saya sendiri melalui penelitian dan observasi secara langsung, tidak ada campur tangan penulisan dari pihak lain, menyalin karya orang lain, pengambilan data orang lain, ataupun mencuri dan mengakui ide orang lain sebagai tulisan saya pribadi. Seluruh data rujukan dan sumber informasi telah dikutip dengan benar dan jelas yang tercantumkan pada daftar pustaka.

Saya memahami bahwa jika ada pelanggaran dalam aspek apa pun dari kajian skripsi ini dapat menyebabkan diskualifikasi saya sebagai sarjana. Oleh karena itu, apabila suatu saat jika terbukti bahwa tulisan skripsi ini merupakan plagiasi, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Malang, 23 November 2023
Pembuat Pernyataan,



Dimitri Ev Hamus Sirojj
NIM. 18680025

KATA PENGANTAR

Assalamualaikum Wr. Wb.

Puji syukur kehadirat Allah SWT yang telah melimpahkan Rahmat serta hidayah-Nya, sehingga peneliti dapat menyelesaikan studi dan skripsi ini dengan baik dan lancar. Shalawat serta salam tercurahkan kepada Rasulullah Muhammad SAW beserta sahabat-sahabatnya.

Peneliti menyadari bahwa banyak pihak yang telah mendukung dalam menyelesaikan penulisan skripsi ini. Oleh karena itu, peneliti mengucapkan terima kasih kepada:

1. Kedua orang tua serta keluarga besar yang selalu memberikan dukungan moral, materi maupun spiritual sehingga penulis dapat merasakan dan menyelesaikan pendidikan perkuliahan hingga akhir dengan baik.
2. Bapak Dr. Ir. M. Amin Hariyadi, M.T selaku Ketua Program Studi Perpustakaan dan Ilmu Informasi Fakultas Sains dan Teknologi UIN Maulana Malik Ibrahim Malang, serta menjadi Dosen Penguji I yang telah mendampingi dan mengawasi setiap tahap pengerjaan skripsi hingga selesai.
3. Bapak Fakhris Khusnu Reza Mahfud, M.Kom selaku Dosen Pembimbing I yang telah memberikan waktu, bimbingan, nasehat, dan seluruh dukungannya selama proses pengerjaan skripsi, sehingga tulisan ini dapat terselesaikan dengan baik.
4. Bapak Yulianto, M.Pd.I selaku Dosen Pembimbing II yang telah meluangkan waktunya untuk membimbing pengerjaan skripsi ini.
5. Bapak Wahyu Hariyanto, M.M selaku Dosen Penguji II yang telah mendampingi dan mengawasi setiap tahap pengerjaan skripsi hingga selesai.
6. Ibu Nita Sita Mudawamah, M.IP selaku dosen wali yang membimbing peneliti dari awal perkuliahan sampai akhir.
7. Seluruh dosen dan staff Program Studi Perpustakaan dan Ilmu Informasi yang telah membagikan ilmu dan pengalamannya.
8. Dua ahli dari ahli ilmu bahasa dan ilmu sosial yang mendukung berjalannya penelitian dalam pengerjaan skripsi.
9. Seluruh sahabat peneliti yang turut mendukung dalam pengerjaan skripsi.

10. Seluruh teman peneliti.

Peneliti menyadari bahwa dalam penyusunan skripsi ini masih terdapat kekurangan dan peneliti berharap semoga skripsi ini dapat memberikan manfaat kepada para pembaca khususnya bagi peneliti secara pribadi. Amin Yaa Rabbal Alamin.

Wassalamu'alaikum Wr. Wb.

Malang, 4 Oktober 2023

Peneliti

DAFTAR ISI

HALAMAN JUDUL.....	i
HALAMAN LEMBAR PERSETUJUAN	ii
HALAMAN LEMBAR PENGESAHAN	iii
HALAMAN PERNYATAAN KEASLIAN TULISAN	iv
KATA PENGANTAR	v
DAFTAR ISI.....	vii
DAFTAR TABEL.....	viii
DAFTAR GAMBAR	ix
ABSTRAK	x
ABSTRACT.....	xi
مستخلص البحث.....	xii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Identifikasi Masalah.....	5
1.3 Tujuan Penelitian	5
1.4 Manfaat Penelitian	5
1.5 Batasan Masalah	5
1.6 Sistematika Penulisan	6
BAB II TINJAUAN PUSTAKA.....	7
2.1 Tinjauan Pustaka	7
2.2 Landasan Teori.....	10
BAB III METODE PENELITIAN.....	26
3.1 Jenis Penelitian.....	26
3.2 Desain Penelitian	26
3.3 Sumber Data.....	53
3.4 Pengumpulan Data	53
BAB IV HASIL DAN PEMBAHASAN	56
4.1 Hasil Penelitian	56
4.2 Pembahasan.....	74
BAB V PENUTUP.....	79
5.1 Kesimpulan	79
5.2 Saran	79
DAFTAR PUSTAKA	81
LAMPIRAN.....	91

DAFTAR TABEL

Tabel 2.1: Contoh <i>Confusion Matrix</i>	16
Tabel 3.1: <i>Case Folding</i>	28
Tabel 3.2: <i>Punctuation</i>	28
Tabel 3.3: Emoticon	29
Tabel 3.4: URL.....	29
Tabel 3.5: <i>Hashtag</i>	29
Tabel 3.6: <i>Username</i>	30
Tabel 3.7: <i>Number</i>	30
Tabel 3.8: <i>Whitespace</i>	31
Tabel 3.9: <i>Filtering</i>	31
Tabel 3.10: <i>Tokenizing</i>	31
Tabel 3.11: <i>Stemming</i>	32
Tabel 3.12: Contoh Data Mentah	32
Tabel 3.13: Proses <i>Text Pre-processing</i>	32
Tabel 3.14: Jumlah Frekuensi Fitur	33
Tabel 3.15: Perhitungan <i>Term Frequency</i>	34
Tabel 3.16: Perhitungan <i>Inverse Document Frequency</i>	36
Tabel 3.17: Perhitungan TF-IDF.....	36
Tabel 3.18: Contoh Data Latih.....	40
Tabel 3.19: Nilai Data Latih Pertama	41
Tabel 3.20: Perhitungan <i>Gain</i> untuk Akar Pohon Keputusan.....	43
Tabel 3.21: Nilai Data Latih Kedua	43
Tabel 3.22: Perhitungan <i>Gain</i> untuk <i>Internal Node</i> Pertama.....	45
Tabel 3.23: Nilai Data Latih Ketiga	46
Tabel 3.24: Perhitungan <i>Gain</i> untuk <i>Internal Node</i> Kedua	48
Tabel 3.25: Nilai Data Latih Keempat	49
Tabel 3.26: Perhitungan <i>Gain</i> untuk <i>Internal Node</i> Ketiga	49
Tabel 3.27: Contoh Evaluasi Model dengan <i>Confusion Matrix</i>	50
Tabel 3.28: <i>Scraper Source Code</i>	54
Tabel 4.1: Sampel Hasil <i>Pre-Processing</i> Tahap <i>Cleaning</i>	57
Tabel 4.2: Sampel Hasil <i>Pre-Processing</i> Tahap <i>Stopwords</i>	58
Tabel 4.3: Sampel Hasil <i>Pre-Processing</i> Tahap <i>Stemming</i>	59
Tabel 4.4: Sampel Hasil <i>Pre-Processing</i> Tahap <i>Tokenizing</i>	60
Tabel 4.5: <i>Root</i> dan <i>Internal Node</i> Rasio 70:30.....	62
Tabel 4.6: Pola Susunan Kata sebagai Keputusan Rasio 70:30.....	64
Tabel 4.7: <i>Root</i> dan <i>Internal Node</i> Rasio 80:20.....	66
Tabel 4.8: Pola Susunan Kata sebagai Keputusan Rasio 80:20.....	69
Tabel 4.9: Hasil Evaluasi <i>Confusion Matrix</i>	72

DAFTAR GAMBAR

Gambar 3.1: Alur Penelitian	26
Gambar 3.2: Pola <i>Root</i>	43
Gambar 3.3: Pola <i>Internal Node</i> Pertama Sebagai Cabang dari <i>Root</i>	46
Gambar 3.4: Pola <i>Internal Node</i> Kedua Sebagai Cabang dari <i>Root</i>	48
Gambar 3.5: Pola Pembelajaran C4.5	50
Gambar 3.6: Dataset Hasil <i>Scraping</i>	55
Gambar 4.1: Hasil Pelabelan.....	56
Gambar 4.2: Tabel <i>Matrix</i> Rasio 70:30.....	73
Gambar 4.3: Tabel <i>Matrix</i> Rasio 80:20.....	73
Gambar 4.4: Diagram Perbandingan Persentase Rasio 70:30 dengan Rasio 80:20.....	75

ABSTRAK

Sirojj, Dimitri Ev Hamus. 2023. **Analisis Sentimen terhadap Qatar sebagai Tuan Rumah Piala Dunia FIFA 2022 pada Media Sosial Twitter Menggunakan Algoritma C4.5. Skripsi. Program Studi Perpustakaan dan Ilmu Informasi Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang.**
Pembimbing: (I) Fakhris Khusnu Reza Mahfud, M.Kom. (II) Yulianto, M.Pd.I.

Kata kunci: Analisis Sentimen, *Text Mining*, C4.5, *Decision Tree*, Twitter, Qatar, Piala Dunia 2022

Pada tahun 2022 terdapat pesta besar pertandingan sepak bola yaitu piala dunia FIFA 2022 yang diselenggarakan di Qatar. Berbagai dampak positif muncul karena Qatar menjadi tuan rumah piala dunia. Namun, terdapat beberapa masalah dan pro kontra dalam persiapan hingga menjelang pelaksanaan, seperti perlakuan pihak penyelenggara yang tidak manusiawi terhadap pekerja migran, bahkan terdapat survei yang menyatakan bahwa banyak responden yang ingin penyelenggaraan pertandingan piala dunia FIFA 2022 ini tidak dilaksanakan di Qatar. Oleh karena itu, perlu dilakukan analisis sentimen terhadap Qatar sebagai tuan rumah piala dunia FIFA 2022 supaya mengetahui seberapa banyak antara dukungan dengan persepsi buruk masyarakat Indonesia mengenai penyelenggaraan piala dunia ini. Metode yang digunakan pada penelitian ini adalah C4.5 yang termasuk dalam algoritma pohon keputusan, serta evaluasi kinerja modelnya menggunakan *Confusion Matrix*. Tujuan penelitian ini adalah untuk mengetahui nilai akurasi dari algoritma C4.5 dalam analisis sentimen terhadap Qatar sebagai tuan rumah piala dunia FIFA 2022 pada media sosial Twitter. Berdasarkan hasil dari evaluasi model, diketahui bahwa nilai tertinggi ada pada rasio 80:20 yang memiliki akurasi sebesar 59%, sedangkan pada rasio 70:30 memiliki akurasi sebesar 56%. Berdasarkan dari hasil pengujian data, ditemukan adanya permasalahan yang menyebabkan hasil akurasi terbilang kurang maksimal yaitu pada proses *pre-processing* pada tahap *stopwords* dan *stemming*.

ABSTRACT

Sirojj, Dimitri Ev Hamus. 2023. **Sentiment Analysis on Qatar as the FIFA World Cup 2022 Host on Twitter Social Media Using the C4.5 Algorithm. Thesis. Library and Information Science Department. Faculty of Science and Technology. Universitas Islam Negeri Maulana Malik Ibrahim Malang.**
Advisors: (I) Fakhris Khusnu Reza Mahfud, M.Kom. (II) Yulianto, M.Pd.I.

Keywords: *Sentiment Analysis, Text Mining, C4.5, Decision Tree, Twitter, Qatar, FIFA World Cup 2022.*

In 2022, a grand football tournament, the FIFA World Cup 2022, was held in Qatar. Various positive impacts emerged as Qatar hosted the World Cup. However, there were several issues and controversies in the preparation leading up to the event, such as the inhumane treatment of migrant workers by the organizers. A survey indicated that many respondents preferred the FIFA World Cup 2022 not to be held in Qatar. Therefore, it is necessary to conduct a sentiment analysis on Qatar as the FIFA World Cup 2022 host to understand the extent of support and negative perceptions among the Indonesian public regarding the event. The method used in this research is C4.5, which belongs to the decision tree algorithm and evaluates the performance of the model using Confusion matrix. This research aims to determine the accuracy of the C4.5 algorithm in sentiment analysis towards Qatar as the host of the FIFA World Cup 2022 on the Twitter social media platform. Based on the model evaluation result, it is found that the highest score was in the 80:20 ratio which has an accuracy at 59%, while the 70:30 ratio has an accuracy at 56%. Based on the testing results, there are identified issues causing the accuracy to be less than optimal, particularly in the pre-processing stage involving stopwords and stemming.

مستخلص البحث

سراج، ديميتري إيف هاموس. ٢٠٢٣. تحليل المشاعر تجاه قطر كمضيف لكأس العالم لكرة القدم ٢٠٢٢ على وسائل التواصل الاجتماعي "تويتز" باستخدام خوارزمية C4.5. البحث الجامعي. قسم علوم المكتبات والمعلومات، كلية العلوم والتكنولوجيا بجامعة مولانا مالك إبراهيم الإسلامية الحكومية مالانج. المشرف الأول: فخريس حسن رزا محفوظ، الماجستير. المشرف الثاني: يوليانتو، الماجستير.

الكلمات الرئيسية: تحليل المشاعر، التنقيب عن النص، C4.5، شجرة القرار، تويتز، كأس العالم ٢٠٢٢.

في عام ٢٠٢٢، هناك حفلة كبيرة من مباريات كرة القدم، وهي كأس العالم لكرة القدم ٢٠٢٢ التي تقام في قطر. ظهرت تأثيرات إيجابية مختلفة لأن قطر استضافت كأس العالم. ومع ذلك، هناك العديد من المشاكل والإيجابيات والسلبيات في التحضير حتى الفترة التي تسبق التنفيذ، مثل معاملة المنظمين غير الإنسانية للعمال المهاجرين، وهناك حتى دراسة استقصائية تفيد بأن العديد من المستجيبين يريدون عدم إقامة مباراة كأس العالم لكرة القدم ٢٠٢٢ في قطر. لذلك، من الضروري إجراء تحليل المشاعر تجاه قطر كمضيف لكأس العالم لكرة القدم ٢٠٢٢ من أجل معرفة مقدار الدعم الموجود بين التصور السيئ للشعب الإندونيسي فيما يتعلق بعقد كأس العالم هذه. الطريقة المستخدمة في هذه الدراسة هي C4.5 والتي تم تضمينها في خوارزمية شجرة القرار، وكذلك تقييم أداء النموذج باستخدام المصفوفة الإرباك. الهدف من هذا البحث هو تحديد الدقة لخوارزمية C4.5 في تحليل المشاعر تجاه قطر كمضيف لكأس العالم لكرة القدم ٢٠٢٢ على وسائل التواصل الاجتماعي "تويتز". بناء على نتائج تقييم النموذج، من المعروف أن أعلى قيمة بنسبة ٨٠:٢٠ لها دقة ٥٩٪، بينما نسبة الدقة ٧٠:٣٠ هي ٥٦٪. بناء على نتائج معالجة البيانات، وجد أن هناك مشاكل تسببت في أن تكون نتائج الدقة منخفضة إلى حد ما، وتحديدًا في عملية المعالجة المسبقة في مرحلتها الكلمات المتوقفة والمجدعية.

BAB I PENDAHULUAN

1.1 Latar Belakang

Saat ini dunia diramaikan dengan pertandingan olahraga sepak bola piala dunia FIFA 2022. Sepak bola menjadi olahraga yang paling digemari oleh masyarakat dunia dengan jumlah lebih dari 3.5 miliar orang di dunia (World Atlas, 2022). Pada piala dunia saat ini, Qatar dipilih menjadi tuan rumah penyelenggara piala dunia FIFA 2022. Secara umum, piala dunia memberi dampak positif terhadap kepentingan negara. Kepentingan negara tersebut meliputi pertahanan (*Defence Interests*), ekonomi (*Economic Interests*), tata dunia (*World Order Interests*), dan ideologi (*Ideological Interests*) (Nuechterlein, 1976).

Selain itu, terdapat dampak positif yang dialami masyarakat internasional secara umum seperti meningkatnya bisnis *merchandise* dan kedai yang memberi fasilitas nonton bareng piala dunia. Namun, terdapat beberapa masalah dalam persiapan hingga menjelang pelaksanaan, yaitu Qatar memperlakukan pekerja migran yang bekerja untuk membangun sarana dan prasarana piala dunia dibawah standar manusia, seperti minimnya fasilitas listrik, saluran pembuangan limbah, dan air mengalir (Ambarita et al., 2022). Adapun menurut survei internasional yang dilakukan oleh www.statista.com yang menunjukkan bahwa 69.1% menyetujui untuk tidak diadakan di Qatar, 17.47% menyetujui diadakan di Qatar, 13.43% memilih untuk netral (*2022 FIFA World Cup Global Opinion on Qatar as Location*, 2022).

Adanya kesan baik dan buruk ini menjadi hal yang penting dan menarik untuk dikaji, karena supaya memberi kejelasan terhadap kecenderungan emosi masyarakat yang sebenarnya. Pada penelitian ini menggunakan kesan atau opini masyarakat Indonesia secara khusus terhadap Qatar sebagai tuan rumah piala dunia FIFA 2022 melalui Twitter. Masyarakat Indonesia sendiri berada ditingkat kedua yang memiliki persentase 77% setelah Nigeria yang memiliki persentase 83% sebagai penduduk yang antusias dan menyukai olahraga sepak bola (Databoks, 2017). Oleh karena itu, opini yang berbahasa Indonesia di Twitter terkait negara Qatar sebagai tuan rumah piala dunia FIFA 2022 yang akan dijadikan data.

Pada penelitian ini, media sosial Twitter digunakan sebagai tempat penambangan teks atau pengumpulan opini publik terkait Qatar sebagai tuan rumah piala dunia FIFA 2022. Pada September 2021 hingga September 2022 penggunaan Twitter di Indonesia berada di urutan keempat setelah Instagram, YouTube, dan Facebook (StatCounter Global Stats, 2022). Dari survei tersebut dapat diketahui bahwa Twitter tergolong media sosial yang sering digunakan di Indonesia. Selain itu, masyarakat memilih untuk melakukan pengungkapan diri, emosi bahagia atau sedih, sekedar berkeluh kesah, membagi dan mencari informasi, pemikiran atau pendapat di Twitter karena dibandingkan dengan media sosial lain yang lebih mementingkan visual, sedangkan Twitter lebih cenderung pada unggahan berupa teks (Zaskya et al., 2021). Penyampaian pandangan melalui Twitter sangat mempengaruhi perilaku masyarakat dalam berbagai hal (Mahfud et al., 2020). Oleh karena itu media sosial Twitter dipilih oleh peneliti sebagai tempat penggalian data teks.

Analisis sentimen sendiri memiliki tujuan untuk mengetahui kecenderungan emosi terhadap opini atau persepsi suatu topik tertentu. Analisis sentimen dilakukan supaya mengetahui dengan jelas persentase kecenderungan positif, negatif, maupun netral. Oleh karena itu, perlu ditentukan kecenderungannya dengan cara atau ilmu tertentu seperti analisis sentimen yang bertujuan agar dapat mengelola perbedaan yang terjadi. Ada kaitannya hal tersebut dengan Qur'an Surat Hud ayat 118-119 tentang perbedaan, perselisihan, dan keharusan seseorang untuk mengelola hal tersebut:

وَلَوْ شَاءَ رَبُّكَ لَجَعَلَ النَّاسَ أُمَّةً وَاحِدَةً وَلَا يَزَالُونَ مُخْتَلِفِينَ ﴿١١٨﴾ إِلَّا مَنْ رَّحِمَ رَبُّكَ وَلِذَلِكَ خَلَقَهُمْ وَتَمَّتْ كَلِمَةُ رَبِّكَ لَأَمْلَأَنَّ جَهَنَّمَ مِنَ الْجِنَّةِ وَالنَّاسِ أَجْمَعِينَ ﴿١١٩﴾

Artinya:

“Jika Tuhanmu menghendaki, tentu Dia akan menjadikan manusia umat yang satu. Namun, mereka senantiasa berselisih, kecuali orang yang dirahmati oleh Tuhanmu. Menurut (kehendak-Nya) itulah Allah menciptakan mereka. Kalimat (keputusan) Tuhanmu telah tetap, Aku pasti akan memenuhi (neraka) Jahanam (dengan pendurhaka) dari kalangan jin dan manusia semuanya” (Q.S. Hud ayat 118-119).

Ayat tersebut menjelaskan bahwa perbedaan adalah suatu fitrah dari manusia yang telah diatur oleh Allah sehingga manusia dapat memilih sesuka hatinya. Hal tersebut wajar dan pasti ada, sehingga perlu adanya pengelolaan.

Seperti menurut Quraish Shihab dalam kitab tafsirnya al-Misbah menerangkan bahwa:

Allah tidak menghendaki semua manusia sejak dahulu sampai nanti hanya satu umat saja, dalam arti lain adalah yang memiliki satu agama, satu pendapat, dan satu kecenderungan. Mereka berselisih menurut cara berpikir dan bahkan dengan hawa nafsu kecuali orang yang mendapat petunjuk dari Allah. Sedangkan dalam bukunya yang berjudul *Yang Hilang Dari Kita : Akhlak* menegaskan bahwa yang dimaksud orang yang dirahmati dan diberi petunjuk adalah mampu mengelola perbedaan tersebut (Shihab, 2016). Pengelolaan dalam konteks ini adalah supaya mengetahui hal yang dominan sebagai solusi ketika ada suatu topik yang dibahas baik dalam musyawarah maupun yang lain. Selain itu, juga agar saling menghargai setiap perbedaan dan memunculkan toleransi ketika berselisih baik dalam masyarakat maupun forum (Shihab, 2001).

Tafsir ayat tersebut dapat dipahami bahwa penting adanya kemampuan mengelola jika terjadi adanya perbedaan pendapat diantara masyarakat. Dalam kajian ini, salah satu upaya yang dilakukan adalah melakukan analisis sentimen untuk menentukan kecenderungan atau hal dominan antara positif, negatif, dan netral, sehingga mengetahui secara jelas hal yang sudah maupun sedang terjadi.

Dalam penelitian ini, menggunakan algoritma C4.5 yang termasuk dari algoritma *decision tree*. Algoritma C4.5 dipilih karena dalam visualisasi hasil penggalian data berbentuk seperti pohon yang memiliki hirarki sehingga mudah untuk dipahami (Khan dkk., 2020). Algoritma C4.5 sudah bisa menangani data bertipe numerik yang sebelumnya tidak dapat ditangani oleh algoritma ID3 (Santoso dkk., 2020). Selain itu, berdasarkan *literatur review* tentang analisis sentimen bahwa algoritma *decision tree* termasuk C4.5 masih jarang digunakan untuk analisis sentimen, sekaligus memberikan saran agar mulai memperbanyak penelitian analisis sentimen dengan algoritma *decision tree* termasuk C4.5 (Singgalen, 2021).

Pada penelitian analisis sentimen terhadap *review* hotel menunjukkan bahwa akurasi algoritma C4.5 tertinggi diantara algoritma lain yaitu 96,94 %, Naive Bayes dengan optimasi sebesar 95,91%, Naive Bayes tanpa optimasi sebesar 89,98%, dan Support Vector Machine sebesar 89,86% (Taufik, 2018). Didukung

dengan penelitian tentang perbandingan metode Naive Bayes, KNN dan Decision Tree terhadap analisis sentimen transportasi KRL *commuter line*, algoritma Naive Bayes, dengan akurasi sebesar 80,00%, dan presisi 66,67 %, algoritma KNN memiliki akurasi sebesar 80%, dan presisi 100%, sedangkan tertinggi adalah pada algoritma Decision Tree yang memiliki akurasi sebesar 100%, dan precision 100% (Romadloni et al., 2019).

Berdasarkan permasalahan yang telah diuraikan, untuk mengetahui kecenderungan opini masyarakat Indonesia terhadap Qatar sebagai tuan rumah piala dunia FIFA 2022, maka penelitian yang berjudul Analisis Sentimen Terhadap Qatar Sebagai Tuan Rumah Piala Dunia FIFA 2022 Pada Media Sosial Twitter Menggunakan Algoritma C4.5 ini menarik dan penting untuk dilakukan.

Dalam penelitian ini akan melakukan analisis sentimen positif, negatif, dan netral terhadap tweet dari hasil penambangan teks dengan topik Qatar sebagai tuan rumah piala dunia FIFA 2022. Menggunakan tiga kelas sentimen karena opini bukan hanya bersifat positif dan negatif saja, melainkan opini juga dapat berbentuk netral untuk diketahui mana yang paling bernilai besar atau dominan (Perdana et al., 2022). Adapun dalam proses setiap tahapnya menggunakan bahasa pemrograman *python*, *Jupyter Notebook* sebagai IDE dan proses manual. Dimulai dari *scraping* data menggunakan *library python* yaitu *Snsrape*. Selanjutnya melakukan penghapusan terhadap *tweet* yang terduplikasi dengan Microsoft Excel. Kemudian melakukan pelabelan data oleh ahli bahasa. Selanjutnya melakukan *pre-processing* (*cleaning, filtering, tokenizing, dan stemming*), pada tahap *cleaning* dan *tokenizing* menggunakan *library natural language toolkit* (NLTK), sedangkan pada tahap *filtering* dan *stemming* menggunakan *library* Sastrawi. Setelah itu, melakukan pembobotan istilah dengan *term frequency-inverse document frequency* (TF-IDF) sebagai pembobotan fitur (kata) dalam setiap dokumen, karena himpunan data teks memiliki ribuan kata unik yang dapat muncul dalam dua kelas atau lebih sehingga perlu adanya pembobotan (Mahfud & Tjahyanto, 2017). Kemudian melakukan pemrosesan data dengan *machine learning* yaitu algoritma C4.5 menggunakan *library scikit-learn*. Hasil evaluasi yaitu berupa persentase performa model (akurasi dan F1-Score). Penggunaan evaluasi akurasi dan F1-Score karena digunakan untuk menghitung *true* dan *false* pada data yang kelasnya tidak seimbang, supaya

mengetahui perbedaan persentase (Rais et al., 2019). Selain itu, data dari hasil klasifikasi akan divisualisasikan dalam bentuk diagram persentase kecenderungan positif, negatif dan netral, serta visualisasi bentuk pohon keputusan. Adanya penelitian ini diharapkan dapat mengetahui bagaimana kecenderungan persepsi masyarakat Indonesia secara umum melalui media sosial Twitter.

1.2 Identifikasi Masalah

Bedasarkan latar belakang yang sudah dijabarkan, adapun identifikasi masalah pada penelitian ini adalah “bagaimana akurasi, *recall*, presisi, dan *F1-Score* dari algoritma C4.5 dalam analisis sentimen terhadap Qatar sebagai tuan rumah piala dunia FIFA 2022 pada media sosial Twitter?”

1.3 Tujuan Penelitian

Adapun tujuan dilaksanakannya penelitian ini adalah untuk mengetahui nilai akurasi, *recall*, presisi, dan *F1-Score* dari algoritma C4.5 dalam analisis sentimen terhadap Qatar sebagai tuan rumah piala dunia FIFA 2022 pada media sosial Twitter.

1.4 Manfaat Penelitian

Manfaat penelitian terdapat dua hal yaitu secara teoritis dan praktis. Manfaat secara teoritis dari hasil penelitian ini adalah menjadi referensi penelitian dan pengembangan lebih lanjut terkait analisis sentimen dengan algoritma C4.5 baik menggunakan opini di Twitter maupun media lain. Adapun manfaat secara praktis yaitu diharapkan dapat memberikan manfaat untuk evaluasi terhadap negara yang menjadi tuan rumah piala dunia sehingga negara penyelenggara selanjutnya dapat semakin baik dalam mempersiapkan dan melaksanakan piala dunia sepak bola, serta memberikan gambaran sentimen masyarakat Indonesia sebagai pengguna Twitter terhadap negara Qatar yang menjadi tuan rumah piala dunia FIFA 2022.

1.5 Batasan Masalah

Adapun untuk menghindari meluasnya ruang lingkup penelitian sehingga tidak terfokus, perlu adanya batasan masalah pada penelitian. Batasan masalah yang digunakan pada penelitian ini sebagai berikut:

1. Data yang digunakan adalah *tweets* yang telah disaring dengan kata kunci “qatar sebagai tuan rumah”

2. *Tweets* berbahasa Indonesia yang diposting mulai dari 9 Desember 2010 sampai dengan 18 Desember 2022.

1.6 Sistematika Penulisan

Sistematika penulisan dibuat agar mempermudah pembaca untuk memahami alur penulisan pada penelitian ini. Sistematika yang dibuat terdiri dari lima bab antara lain:

BAB I Pendahuluan

Pada bab I memuat sub bab seperti latar belakang, identifikasi masalah, tujuan penelitian, manfaat penelitian, batasan masalah, dan sistematika penulisan.

BAB II Tinjauan Pustaka

Pada bab II terdapat beberapa penelitian terdahulu yang terkait serta beberapa teori, konsep dan definisi sebagai pendukung penelitian ini yang terkait analisis sentimen, klasifikasi, *text mining*, *supervised learning*, macam-macam metode *Stopwords* dan *Stemming*, pembobotan istilah dengan *term frequency-inverse document frequency* (TF-IDF), C4.5, *K-fold cross validation*, *confusion matrix*, Twitter, Snsrape, Jupyter Notebook, dan Python.

BAB III Metode Penelitian

Pada bab III berisi tentang jenis penelitian, desain penelitian, waktu dan tempat penelitian, sumber data, dan pengumpulan data.

BAB IV Hasil dan Pembahasan

Pada bab IV membahas tentang hasil penelitian yang dilakukan berdasarkan identifikasi masalah yang telah dirumuskan dan dibatasi dengan beberapa hal.

BAB V Kesimpulan dan Saran

Pada bab V menjadi penutup yang berisi tentang kesimpulan dari hasil analisis dan saran dari peneliti agar penelitian selanjutnya semakin dikembangkan.

BAB II

TINJAUAN PUSTAKA DAN LANDASAN TEORI

2.1 Tinjauan Pustaka

Penelitian mengenai analisis sentimen sudah banyak dilakukan terutama menggunakan media sosial Twitter, seperti penelitian yang berjudul “Analisis Sentimen Twitter Menggunakan *Text Mining* dengan Algoritma *Naïve Bayes Classifier*”. Penelitian ini memiliki tujuan mengklasifikasi data *tweet* menjadi dua kelas sentimen positif dan negatif dari opini masyarakat Indonesia terhadap pemilihan kepala daerah Jawa Barat. Metode yang digunakan adalah algoritma *Naïve Bayes Classifier* dengan menggunakan 300 *tweet* yang dibagi 70% (200 *tweet*) untuk data latih dan 30% (100 *tweet*) untuk data uji. Hasil dari penelitian ini menunjukkan bahwa data uji sebanyak 100 *tweet* tersebut tergolong sentimen negatif dengan rincian terdapat 32 data bersentimen positif dan 68 data bersentimen negatif, serta hasil evaluasi kinerja *machine learning* algoritma *Naïve Bayes Classifier* mendapatkan akurasi sebesar 84% (Sudiantoro & Zuliarso, 2018).

Adapun penelitian kedua mengenai analisis sentimen olahraga sepak bola menggunakan Twitter yang berjudul “Analisis Sentimen Tentang Opini Terhadap Performa Timnas Sepak Bola Indonesia Pada Twitter”. Penelitian ini dilakukan karena adanya pertandingan kualifikasi piala dunia 2022 Indonesia melawan Malaysia tahun 2019 dan final *Seagames* 2019 Indonesia melawan Vietnam. Tujuan dari penelitian ini adalah memberi kemudahan kepada masyarakat untuk menilai *tweet* bernilai positif atau negatif. Pada penelitian ini menggunakan metode *Naïve Bayes Classifier* dengan menggunakan tiga skenario pengujian akurasi pemrosesan data dengan *machine learning*, yaitu komposisi 70% data latih dan 30% data uji, komposisi 80% data latih dan 20% data uji, serta komposisi 90% data latih dan 10% data uji. Data diambil melalui Twitter API dengan *hashtag* #InfoTimnas, #TimnasDay, #TimnasIndonesia, #TimnasJuara, dan #KitaGaruda, sehingga menghasilkan 530 *tweet* dengan rincian 424 *tweet* untuk data Indonesia melawan Malaysia dan 106 *tweet* untuk data Indonesia melawan Vietnam. Kemudian hasil dari *crawling* data dilabeli secara manual oleh 3 jurnalis sepak bola dengan dua kelas, yaitu positif dan negatif. Hasil dari penelitian ini menunjukkan bahwa

terdapat kecenderungan sentimen negatif, yaitu sentimen pemain berjumlah 119 *tweet* negatif dan 50 *tweet* positif, sentimen pelatih berjumlah 25 *tweet* negatif dan 5 positif, dan sentimen federasi terdapat 3 *tweet* positif dan 2 *tweet* negatif. Sedangkan, hasil dari evaluasi kinerja *machine learning* dengan komposisi 70% data latih dan 30% data uji memiliki akurasi 78%, komposisi 80% data latih dan 20% data uji memiliki akurasi 84%, serta komposisi 90% data latih dan 10% data uji memiliki akurasi 87%. Dapat diketahui dari pengujian tiga skenario tersebut bahwa algoritma *Naïve Bayes Classifier* bergantung kepada banyaknya persentase data latih sehingga akurasi cenderung mengalami peningkatan (Astiningrum et al., 2020).

Selain itu, pada penelitian ketiga terdapat komparasi algoritma dalam analisis sentimen pertandingan piala dunia 2014 di Brazil yang berjudul “*Sentiment Analysis on Twitter Data of World Cup Soccer Tournament Using Machine Learning*”. Tujuan dari penelitian ini adalah untuk membandingkan kinerja beberapa algoritma *machine learning* untuk *natural language processing* (NLP). Metode yang digunakan sebagai komparasi adalah *Naïve Bayes Classifier* (NBC), *Support Vector Machine* (SVM), *Random Forest* (RF), dan K-Nearest Neighbors (KNN) dengan memakai berbagai rasio data latih dan data uji, seperti 60:40, 70:30, 80:20, dan 90:10. Pengambilan data dalam penelitian ini menggunakan *hashtag*, seperti #brazil2014, #worldcup, dan *hashtag* pertandingan seperti #ALGRUS Algeria vs Russia dan lainnya. Data awal *crawling* tersedia lebih dari 3.5 juta *tweet*, namun karena keterbatasan sumber daya komputasi, peneliti memutuskan untuk membatasi *crawling* data selama 2 hari yang menghasilkan 24.335 *tweet*. Kelas sentimen dalam penelitian ini dibagi menjadi tiga, yaitu positif, negatif, dan netral. Hasil dari penelitian ini, algoritma NBC memiliki akurasi paling tinggi dengan rasio 90:10 sebesar 88.17%, KNN dengan rasio 80:20 memiliki akurasi sebesar 87.49%, RF dengan rasio 60:40 memiliki akurasi sebesar 85.32%, dan SVM dengan rasio 90:10 memiliki akurasi sebesar 41.77% (Patel & Passi, 2020).

Kemudian, pada penelitian keempat terdapat analisis sentimen berbahasa Arab mengenai piala dunia 2018 yang berjudul “*Sentiment analysis of comments in social media*”. Tujuan dari penelitian ini adalah untuk mengetahui sentimen dari fitur tekstual dan fitur emoji dari *tweet* berbahasa Arab yang diposting selama

penyelenggaraan Piala Dunia 2018. Metode yang digunakan adalah *lexicon-based* serta teknik pengumpulan data menggunakan 10 *hashtag* yang pada setiap *hashtag* memiliki 100 *tweet* sehingga jumlah total data terdapat 1000 *tweet*. Hasil dari penelitian ini adalah *tweet* berbahasa Arab akan lebih akurat jika terdapat kombinasi teks dan emoji. Selain itu, ditemukan bahwa sentimen dianggap netral oleh mesin ketika isi *tweet* berbahasa Inggris tetapi menulis dengan alfabet Arab. Hal tersebut karena tidak teridentifikasi dalam *lexicon* sehingga tidak dikenali oleh mesin (Alrumaih et al., 2020).

Kelima terdapat penelitian yang berjudul “Analisis Sentimen Pada Tweet dengan Tagar #bpjsrasarentenir Menggunakan Metode Support Vectore Machine (SVM)”. Penelitian ini dilakukan karena ramainya isu BPJS di Indonesia pada tahun 2019 sehingga menjadi *trending topic* di Twitter. Penelitian ini memiliki tujuan menentukan kelas positif dan negatif dan mengklasifikasi sentimen yang berhubungan dengan topik kenaikan iuran BPJS. Metode yang digunakan dalam penelitian ini adalah algoritma *Support Vector Machine* (SVM). Adapun tahapan analisis sentimen mulai dari prapemrosesan, seperti *Case folding*, *Tokenizing*, *Filtering*, dan *Stemming*. Setelah itu melakukan pembobotan fitur dalam dokumen dengan TF-IDF. Data dikumpulkan dengan cara *crawling* menggunakan *tools* Rapid Miner Studio yang mendapatkan 300 *tweet* dengan *hashtag* #bpjsrasarentenir dan diberi rasio 70:30 untuk data latih dan data uji. Hasil dari penelitian ini adalah membuat sistem yang dapat memberi nilai sentimen positif atau negatif yang dibangun menggunakan bahasa Python 3 (Santoso, 2021).

Dari lima penelitian terdahulu yang telah dipaparkan, kesamaan penelitian tersebut dengan penelitian ini adalah analisis sentimen menggunakan media sosial Twitter sebagai tempat pengumpulan data. Namun, sisi perbedaan dapat diketahui mulai dari tidak adanya penggunaan *hashtag* sebagai cara mengumpulkan data pada penelitian ini, *trending topic* yang berbeda (topik dan waktu), terdapat sebagian topik yang sama membahas piala dunia namun berbeda bahasa, pada penelitian ini terfokus pada sentimen yang menjadi tuan rumah penyelenggara piala dunia FIFA 2022, dan metode *machine learning* yang digunakan pada penelitian ini adalah algoritma C4.5.

2.2 Landasan Teori

2.2.1 Analisis Sentimen

Analisis sentimen atau *natural language processing* (NLP) adalah kombinasi antara bidang ilmu bahasa dan ilmu komputer yang berupaya untuk menentukan polaritas positif atau negatif yang terdapat dalam teks opini secara otomatis (Taboada, 2016). Adapun analisis sentimen memiliki nama lain seperti *opinion mining* (penambangan opini), *opinion extraction* (ekstraksi opini), *sentiment mining* (penambangan sentimen), *subjectivity analysis* (analisis subjektivitas), *affect analysis* (analisis pengaruh), *emotion analysis* (analisis emosi), *review mining* (penambangan ulasan).

Analisis sentimen memiliki tujuan memberikan evaluasi sebuah opini, sikap, penilaian dan emosi terhadap produk, layanan, organisasi, individu, masalah, peristiwa, topik seperti (politik, sains dan olahraga) (Liu, 2012). Data sentimen tersebut bisa didapat melalui opini atau komentar dari blog, *microblog*, Twitter, Wordpress, Kaskus, Facebook, LinkedIn, dan Email (Ramdhani & Rahim, 2014).

2.2.2 Klasifikasi

Klasifikasi adalah proses menemukan sebuah model atau fungsi yang menjelaskan dan membedakan label data yang memiliki tujuan memprediksi kelas yang belum diketahui (G. T. Santoso, 2021). Dalam arti lain, teknik klasifikasi dalam analisis sentimen bertujuan untuk mengelompokkan dan memprediksi kelas dokumen berdasarkan fitur dari sampel yang ada. Pada proses klasifikasi, sampel yang digunakan sebagai pendefinisian kelas data sehingga menjadi model prediksi yang sesuai disebut *data training* (data latih), sedangkan *data testing* (data uji) digunakan untuk mengetahui kinerja model yang dijadikan prediksi. Ada beberapa metode yang biasanya dijadikan sebagai pengklasifikasi yaitu *Decision Tree*, K-Nearest Neighbor (K-NN), *Naïve Bayes*, *Random Forest* dan *Support Vector Machine* (SVM) (Maulidah et al., 2020).

2.2.3 Text Mining

Text mining adalah cabang studi dari *data mining* yang mana menganalisis dokumen tekstual selama periode waktu tertentu dengan menggunakan berbagai metode (Feldman & Sanger, 2007). *Text mining* berbeda dengan *data mining* dari jenis data yang digunakan untuk proses ekstraksi, *text mining* menggunakan data

yang tidak terstruktur sedangkan *data mining* menggunakan data terstruktur dalam bentuk database. Terdapat dua prosedur dalam proses *text mining*, yaitu *text pre-processing* untuk pembersihan sekaligus persiapan klasifikasi, dan *knowledge extraction* untuk tujuan pemrosesan (Yan et al., 2022).

Text pre-processing bertujuan untuk mengubah data teks menjadi bentuk yang diperlukan dalam proses klasifikasi. Adapun *text pre-processing* memiliki beberapa proses seperti *cleaning*, *filtering*, *tokenizing*, dan *stemming* (Mahawardana et al., 2022). Sedangkan *knowledge extraction* adalah mencakup beberapa tujuan, yaitu mulai dari klasifikasi teks, pengelompokan teks, ekstraksi informasi, dan pencarian informasi (Yan et al., 2022).

2.2.4 Supervised Learning

Supervised learning adalah metode pemrosesan data *machine learning* yang terawasi atau terdapat label yang telah ditentukan pada setiap baris data (Rosid et al., 2022). Mesin dilatih untuk mempelajari input *data training* yang telah ditetapkan labelnya sehingga dapat mengidentifikasi keterkaitan data dengan labelnya. Algoritma *machine learning* yang termasuk *supervised learning* adalah *Convolutional Neural Network-Support Vector Machine (CNN-SVM)*, *Deep Learning*, *Decision Stump*, *C4.5*, *J4.8*, *Bagging*, *AdaBoost*, *Logistic Regression*, *Linear Regression*, dan *Naive Bayes Classifier* (Roihan et al., 2020).

2.2.5 Macam-macam Metode Stopwords dan Stemming

Pada tahap *stopwords* atau *stoplist wordlist* memiliki tujuan pembersihan kata yang tidak relevan untuk dijadikan indeks klasifikasi. Ada beberapa daftar *stopwords* berbahasa Indonesia yang disusun oleh Fadillah Z. Tala, Damian Doyle, dan Wibisono (Rahutomo et al., 2019). Damian Doyle dan Wibisono mengikuti daftar dari Tala, namun sedikit dimodifikasi. Sedangkan metode yang digunakan pada proses *stopwords* ada beberapa macam, yaitu seperti *Term Based Random Sampling (TBRS)* dan *stopword Tala* (Rinandyaswara et al., 2022). Selain itu, Sastrawi juga dapat dijadikan *stopword removal* untuk dokumen berbahasa Indonesia (Rosid et al., 2020).

Pada tahap *stemming* terdapat beberapa macam metode yang dapat digunakan, seperti *Porter*, *Snowball*, dan Sastrawi. Metode *Porter* dan *Snowball* dapat menggunakan *library NLTK* untuk proses *stemming*, hal ini dapat dilihat di

situs www.nltk.org yang berisi dokumentasi kode di *library* NLTK. Sedangkan metode Sastrawi dapat dilihat di situs github.com/sastrawi/sastrawi. *Stemmer* Sastrawi memakai algoritma Nazief dan Adriani dalam proses pengembalian kepada kata dasar (Rahutomo et al., 2019).

Metode *Porter* serta Sastrawi dijelaskan dalam penelitian yang berjudul “Perbandingan Algoritma *Stemming Porter* dengan Algoritma Nazief & Adriani untuk *Stemming* Dokumen Teks Bahasa Indonesia” (Agusta, 2009) dan metode *Snowball* dijelaskan dalam penelitian yang berjudul “Implementasi *Text Mining* Terhadap Analisis Sentimen Masyarakat Dunia di Twitter Terhadap Kota Medan Menggunakan *K-Fold Cross Validation* dan *Naïve Bayes Classifier*” (Ridwansyah, 2022).

a. Metode *Porter*

Algoritma ini digunakan bekerja untuk menghapus partikel, kata ganti kepemilikan, menghapus awalan dan akhiran.

b. Metode Sastrawi

- 1) Cari kata yang akan distem dalam kamus. Jika ditemukan maka diasumsikan bahwa kata tersebut adalah *root word*. Maka algoritma berhenti.
- 2) *Inflection Suffixes* (“-lah”, “-kah”, “-ku”, “-mu”, atau “-nya”) dibuang. Jika berupa *particles* (“-lah”, “-kah”, “-tah” atau “pun”) maka langkah ini diulangi lagi untuk menghapus *Possesive Pronouns* (“-ku”, “-mu”, atau “-nya”), jika ada.
- 3) Hapus *Derivation Suffixes* (“-i”, “-an” atau “-kan”). Jika kata ditemukan di kamus, maka algoritma berhenti. Jika tidak maka ke langkah 3a
 - 3a, jika “-an” telah dihapus dan huruf terakhir dari kata tersebut adalah “-k”, maka “-k” juga ikut dihapus. Jika kata tersebut ditemukan dalam kamus maka algoritma berhenti. Jika tidak ditemukan maka lakukan langkah 3b.
 - 3b, akhiran yang dihapus (“-i”, “-an” atau “-kan”) dikembalikan, lanjut ke langkah 4.
- 4) Hapus *Derivation Prefix*. Jika pada langkah 3 ada sufiks yang dihapus maka pergi ke langkah 4a, jika tidak pergi ke langkah 4b.

- 4a, periksa tabel kombinasi awalan-akhiran yang tidak diijinkan. Jika ditemukan maka algoritma berhenti, jika tidak pergi ke langkah 4b.
- 4b, For $i = 1$ to 3, tentukan tipe awalan kemudian hapus awalan. Jika *root word* belum juga ditemukan lakukan langkah 5, jika sudah maka algoritma berhenti. Catatan: jika awalan kedua sama dengan awalan pertama algoritma berhenti.

5) Melakukan Recoding.

6) Jika semua langkah telah selesai tetapi tidak juga berhasil maka kata awal diasumsikan sebagai *root word*. Proses selesai.

c. Metode *Snowball*

Algoritma ini digunakan untuk *stemming* beberapa bahasa Internasional (*multi-language*) untuk memperoleh kata dasar dengan menghapus imbuhan awalan kata, kata sisipan kata, akhiran kata, serta kombinasi kata antara awalan kata dan akhiran. Selain ketiga metode yang telah disebut, ada Lovins dan Paice/Husk sebagai algoritma *stemming* juga (Asmara et al., 2016).

2.2.6 Pembobotan Istilah dengan *term frequency-inverse document frequency (TF-IDF)*

Pembobotan istilah dimaksudkan agar menjadikan data yang tidak terstruktur menjadi terstruktur. Ada beberapa metode pembobotan istilah atau fitur dalam setiap dokumen seperti term frequency (TF), term frequency-inverse document frequency (TF-IDF), Best Match 25 (BM25), term frequency-inverse gravity moment (TF-IGM), dan TF-IDF-inverse class frequency (TF-IDF-ICF) (Kim et al., 2022). Namun, metode TF-IDF adalah pembobotan yang sering digunakan dalam *information retrieval* dan *text mining* (Chamira, 2022). Metode pembobotan TF-IDF adalah skema yang populer digunakan (Carvalho & Guedes, 2020).

TF-IDF adalah gabungan dari aspek lokal dan global dari setiap istilah (Mahfud & Tjahyanto, 2017). TF adalah ukuran kemunculan berulangannya sebuah istilah dalam sebuah dokumen, serta dalam seluruh dokumen yang ada dalam korpus (Saadah et al., 2013). TF memiliki beberapa formulasi perhitungan seperti TF biner atau *boolean*, TF murni, TF normalisasi atau *max*, dan TF logaritmik

(Saputra et al., 2022). Sedangkan IDF adalah *inverse* frekuensi dokumen yang mengandung kata tersebut (Wahyuni et al., 2017).

Pada penentuan nilai TF-IDF adalah dengan rumus $w = tf \times idf$. Pada penelitian ini memakai perhitungan TF murni, untuk mendapatkan nilai TF murni adalah berdasarkan kemunculan istilah dalam dokumen, jika terdapat 5 maka bernilai 5, dan jika tidak muncul maka bernilai 0 (Stiawan & Hidayat, 2019). Sedangkan, rumus nilai IDF ditentukan melalui rumus $IDF = \log \left(\frac{n}{df} \right)$ (Fahrudin et al., 2022).

2.2.7 C4.5

Algoritma C4.5 ini dikenalkan oleh Quinlan pada tahun 1996 yang mana menjadi pengembangan dari algoritma ID3 sebagai *classifier* dan termasuk dari algoritma *Decision Tree*. *Decision Tree* memiliki berbagai model algoritma, seperti *Classification and Regression Tree* (CART), *Iterative Dichotomiser 3* atau *Induction Of Decision Tree* (ID3), dan C4.5 (Pengembangan dari ID3) (Rufiyanto dkk., 2021). Selain itu, ada dua algoritma lain yang termasuk *Decision Tree* yaitu *Decision Stump* dan *Random Forest* (Fiqri dkk., 2020).

Pada perhitungan algoritma C4.5, langkah pertama yang perlu dilakukan adalah menghitung *entropy* data latih yang telah disiapkan. *Entropy* digunakan untuk menghitung ketidakpastian dari sekumpulan atribut dari dataset yang mana semakin tinggi *entropy* suatu atribut maka semakin tinggi ketidakpastian (perbedaan keputusan) (Essra, 2016). Berikut rumus persamaan 2.1 yang digunakan sebagai pengukuran *entropy*:

$$Entropy(S) = \sum_{i=1}^n - p_i * \log_2 p_i \quad (2.1)$$

S : Himpunan kasus

n : Jumlah partisi S atau total data

p_i : Jumlah *value* yang terdapat *term* dibagi dengan total dokumen

Langkah selanjutnya untuk menentukan atribut sebagai akar harus berdasarkan nilai *gain* tertinggi dari berbagai atribut yang ada. Berikut rumus persamaan 2.2 untuk menghitung nilai *gain*:

$$Gain(S, A) = Entropy(s) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i) \quad (2.2)$$

S : Himpunan kasus

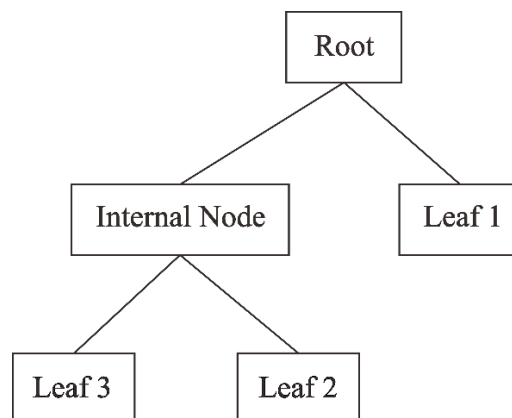
A : Atribut

n : Jumlah partisi atribut A

$|S_i|$: Jumlah kasus pada partisi ke i atau jumlah masing-masing kelas terhadap $value$ dalam sebuah atribut

$|S|$: Jumlah keseluruhan dataset

Alur algoritma pohon keputusan C4.5 jika divisualisasikan seperti pada gambar 2.1 di bawah.



Gambar 2.1: Pola Pohon Keputusan
(Sumber: Hasil olah data peneliti, 2023)

Keterangan:

1) *Root* (Akar)

Atribut yang memiliki nilai *gain* tertinggi pada perhitungan awal akan menjadi akar

2) *Internal Node* (Perantara)

Atribut cabang dari *root* yang menjadi keputusan selanjutnya

3) *Leaf* (Daun)

Keputusan akhir dari percabangan

Algoritma tersebut menghitung *entropy* total dan *gain*, atribut yang memiliki *gain* tertinggi akan menjadi akar. Selanjutnya melakukan perhitungan ulang dengan eliminasi dokumen dan atribut yang telah menjadi *gain*. Perhitungan akan dilakukan berulang hingga bertemu *entropy* sama dengan nol.

2.2.8 Confusion Matrix

Confusion matrix digunakan sebagai pengevaluasi kinerja dari *machine learning* (Lubis et al., 2022). Evaluasi ini meliputi akurasi, *recall*, presisi, dan F1 *score*. Gambaran skema evaluasi ini terdapat pada tabel 2.1 di bawah.

Tabel 2.1: Contoh Confusion Matrix

		Prediksi	
		Negatif	Positif
Aktual	Negatif	True Negatif (TN)	False Positif (FP)
	Positif	False Negatif (FN)	True Positif (TP)

(Sumber: Hasil olah data peneliti, 2023)

Keterangan:

- 1) TN : Total data prediksi negatif yang sesuai dengan data aktual (hasil prediksi negatif dari *machine learning* sama dengan data opini yang berlabel negatif)
- 2) TP : Total data prediksi positif yang sesuai dengan data aktual (hasil prediksi positif dari *machine learning* sama dengan data opini yang berlabel positif)
- 3) FN : Total data prediksi negatif tetapi tidak sesuai dengan data aktual (hasil prediksi negatif dari *machine learning*, tetapi data opini menunjukkan label positif)
- 4) FP : Total data prediksi positif tetapi tidak sesuai dengan data aktual (hasil prediksi positif dari *machine learning*, tetapi data opini menunjukkan label negatif)

Evaluasi kinerja *machine learning* memiliki beberapa poin seperti akurasi, *recall*, presisi, dan F1-Score sebagai berikut:

- 1) Akurasi adalah tingkat kemiripan data yang diprediksi dengan data aktual, tingkat akurasi dihitung dengan rumus persamaan 2.3.

$$\frac{TP+TN}{TP+TN+FP+FN} \quad (2.3)$$

- 2) *Recall* adalah tingkat keberhasilan sistem dalam temu kembali informasi, hasil *recall* dapat ditemukan dengan rumus persamaan 2.4.

$$\frac{TP}{TP+FN} \quad (2.4)$$

- 3) Presisi adalah relevansi dokumen yang diambil oleh sistem, untuk mencari presisi dapat dihitung dengan rumus persamaan 2.5.

$$\frac{TP}{TP+FP} \quad (2.5)$$

- 4) F1 *Score* adalah perbandingan antara rata-rata presisi dan recall yang dibobotkan, hal ini dapat ditemukan dengan rumus persamaan 2.6.

$$\frac{2(Presisi)(Recall)}{(Presisi+Recall)} \quad (2.6)$$

2.2.9 Twitter

Twitter adalah media sosial yang dibuat oleh Jack Dorsey pada tahun 2006 yang sekarang tahun 2022 dipegang kepemilikannya oleh Elon Musk. Twitter memberikan fitur terhadap pengguna untuk mengirim postingan berbentuk teks dengan batas maksimal 140 karakter, sedangkan pada tahun 2017 meningkat menjadi 280 karakter. keinginan penggunanya. Unggahan *tweet* terdiri dari beberapa bagian yaitu emoticon, URLs, re-tweet (RT) untuk mengunggah ulang *tweet* diberanda twitter, @ untuk menyebutkan pengguna lain, # untuk hashtag topik yang sama (Wang et al., 2012). Adapun selain itu ada komentar dan *quote tweet* (QT) dan *reply*. Menurut survei dari www.statista.com, jumlah pengguna twitter di dunia dari tahun 2019 terdapat 312.7 juta, 2020 terdapat 347.6 juta, 2021 terdapat 362.4 juta, dan 2022 terdapat 368.4 juta (*Twitter: Number of Users Worldwide from 2019 to 2022*, 2022).

2.2.10 Snsrape

Snsrape adalah *library* python yang digunakan untuk *scraping* data pada beberapa media sosial seperti Facebook, Instagram, Mastodon, Reddit, Telegram, Twitter, Weibo dan VKontakte tanpa menggunakan API untuk penambangannya. Snsrape dapat diunduh, dan mengetahui tutorial penggunaan serta informasi lebih lanjut di <https://github.com/JustAnotherArchivist/snsrape>. Adapun rincian yang dapat digali datanya sebagai berikut:

- 1) Facebook : profil pengguna, grup, dan komunitas (alias posting pengunjung)
- 2) Instagram : profil pengguna, tagar, dan lokasi
- 3) Mastodon : profil pengguna dan toots (tunggal atau utas)
- 4) Reddit : pengguna, subreddit, dan pencarian (via Pushshift)

- 5) Telegram : saluran
- 6) Twitter : pengguna, profil pengguna, tagar, pencarian, tweet (utas tunggal atau sekitarnya), daftar posting, dan tren
- 7) VKontakte : profil pengguna
- 8) Weibo : profil pengguna

2.2.11 Jupyter Notebook

Jupyter notebook adalah IDE aplikasi berbasis web yang dibangun menggunakan bahasa pemrograman python dengan fungsi untuk pengeditan, pengujian, pengemasan, dan pembangunan perangkat lunak. Aplikasi ini dijalankan melalui *localhost* komputer (Pratama & Mirza, 2021). File dari jupyter notebook berekstensi *ipynb* yang dapat berisi kode komputer dan *rich text element* seperti paragraf, persamaan matematik, gambar dan tautan (Setiabudidaya, 2018). Informasi dan cara memasang jupyter notebook di komputer ada di laman <https://jupyter.org/>. Jupyter notebook dapat dikonversi ke sejumlah format output standar terbuka, seperti HTML, PDF, LaTeX, slide presentasi, Markdown, ReStructuredText, dan Python (Firdaus, 2020).

2.2.12 Python

Python adalah bahas pemrograman yang dibuat oleh Guido van Rossum yang memiliki interpretatif multifungsi dengan filosofi perancangan yang berfokus pada tingkat keterbacaan kode. Python diklaim sebagai bahasa yang menggabungkan kapabilitas, kemampuan, dengan sintaksis kode yang sangat jelas (Saragih, 2016).

Selain itu, bahasa pemrograman python memiliki multifungsi dalam penggunaan seperti digunakan untuk *command line interface (CLI)*, *internet of things (IOT)*, aplikasi *mobile*, *website*, *graphical user interface (GUI)* desktop, *game*, robotika, *data mining*, *hacking*, hingga artifisial intelijen (AI) (Wardana, 2019). Ada beberapa sistem operasi yang mendukung kerja kode Python antara lain adalah Linux/Unix, Mac OSX, Java Virtual Machine, Windows, OS/2, Palm, Amiga dan Symbian (untuk produkproduk Nokia) (Habibi et al., 2017).

2.2.13 Fikih Sosial

Kehidupan bermasyarakat tidak bisa ditinggalkan oleh manusia. Manusia hidup berkelompok saling membutuhkan dan saling mengisi itu adalah yang disebut

mahluk sosial (Asir, 2014). Kehidupan bermasyarakat dalam agama Islam diatur agar tidak ada kerugian antar sesama dan mewujudkan kesejahteraan sosial. Hal tersebut terdapat pada pembahasan fikih sosial, yang mana didefinisikan sebagai sebuah proses dan hasil perumusan hukum syariat yang dieksplorasi secara mendetail untuk kemaslahatan atau kepentingan masyarakat (Yulianto, 2017). Pelaksanaan hidup bermasyarakat tentu ada hal yang perlu dihindari seperti sifat sombong dan buruk sangka. Kedua sifat tersebut dijelaskan sebagai berikut.

a. Sombong

1) Definisi dan dalil-dalil tentang sombong

Salah satu dari sifat tercela adalah sombong. Sifat tersebut sangat dibenci pada semua kalangan dalam hidup bersosial. Akibat dari melakukan sifat tersebut dapat merugikan berbagai pihak termasuk diri sendiri. Terdapat definisi dalam sebuah potongan Hadits yang menyatakan tentang bentuk dari kesombongan sebagai berikut.

الْكِبْرُ بَطْرُ الْحَقِّ وَغَمَطُ النَّاسِ

“Sombong adalah menolak kebenaran dan meremehkan orang lain.” (HR. Muslim no. 91).

Hadits tersebut menyebutkan dua bentuk hakikat kesombongan, yaitu menolak kebenaran dan merendahkan sesama manusia. Definisi dua bentuk kesombongan ini merujuk kepada tiga aspek, yaitu sombong kepada Allah, sombong kepada Rasulnya, dan sombong kepada manusia. Maksud dari menolak kebenaran adalah tidak menerima apa yang telah diperintah dan dilarang oleh Allah serta Rasulnya, sedangkan maksud dari merendahkan sesama manusia adalah melakukan penghinaan dan merasa paling tinggi diantara yang lain (Taufikurrahman, 2017). Allah membenci serta melarang sifat sombong ini, seperti yang terdapat dalam al-Qur’an surat Luqman ayat 18.

وَلَا تُصَعِّرْ خَدَّكَ لِلنَّاسِ وَلَا تَمْشِ فِي الْأَرْضِ مَرَحًا إِنَّ اللَّهَ لَا يُحِبُّ كُلَّ مُخْتَالٍ فَخُورٍ

“Dan janganlah kamu memalingkan mukamu dari manusia (karena sombong) dan janganlah kamu berjalan di muka bumi dengan angkuh. Sesungguhnya Allah tidak menyukai orang-orang yang sombong lagi membanggakan diri.” (QS. Luqman ayat 18).

Pada perspektif tafsir al-Munir dan al-Misbah menyatakan bahwa ayat tersebut memberikan arahan agar sesama manusia harus saling menunjukkan wajah yang ceria dan bahagia, menjauhi kesombongan, senantiasa rendah hati, dan memiliki perkataan yang baik (Hasanatudaroini, 2020). Oleh karena itu, dalam menjalani hidup bersosial perlu memperhatikan beberapa poin dari tafsir tersebut agar tidak mendatangkan kerugian bagi diri sendiri dan orang lain, sehingga akan menciptakan kelompok sosial yang sehat dan diberkahi Allah. Adapun dari al-Qur'an surat Luqman ayat 18, ditegaskan dengan Hadits berikut.

وَإِنَّ اللَّهَ أَوْحَىٰ إِلَيَّ أَنْ تَوَاضَعُوا حَتَّىٰ لَا يَفْخَرَ أَحَدٌ عَلَىٰ أَحَدٍ وَلَا يَبْتَغِيَ أَحَدٌ عَلَىٰ أَحَدٍ

“*Sesungguhnya Allah mewahyukan kepadaku agar kalian bersikap rendah hati hingga tidak seorang pun yang bangga atas yang lain dan tidak ada yang berbuat aniaya terhadap yang lain.*” (HR. Muslim no. 2865).

Hadits di atas menunjukkan bahwa diharuskan bagi manusia terutama sesama kaum muslim agar dapat saling merendahkan hati sehingga dapat menjaga diri dari perbuatan yang tercela. Hal ini mengarah pada perintah dan larangan, yaitu perintah untuk memiliki sifat rendah hati dan menjauhi sifat sombong.

2) Bentuk-bentuk alasan seseorang melakukan kesombongan

Sifat sombong tidak memandang usia maupun jenis kelamin pada manusia, tetapi hal ini berpotensi terjadi karena adanya sebuah kelebihan pada diri seseorang yang tidak dimiliki orang lain (Norazizah et al., 2021). Sifat ini muncul karena memiliki berbagai alasan, yaitu karena kapasitas ilmu pengetahuan yang dipunya, ibadah dan amal saleh yang dilaksanakan, derajat keturunan, keindahan paras yang dimiliki, kekuasaan atau wewenang, serta golongan yang dianut (Taufikurrahman, 2020). Selain itu, dapat juga dikarenakan harta yang melimpah (Haqie, 2013).

3) Hukum sombong

Dengan memahami dalil dari al-Qur'an surat Luqman ayat 18 dan Hadits riwayat Muslim nomor 2865 yang telah disebut, dapat disimpulkan dengan kaidah usul fikih tentang asal sifat kata perintah dan larangan. Kaidah tersebut menyatakan bahwa asal dari kata perintah adalah menunjukkan hukum wajib, sedangkan asal dari kata larangan adalah menunjukkan hukum haram (Djalil, 2014).

Dalam surat Luqman ayat 18 terdapat dua kata kerja larangan yang menjadi dasar, yaitu *وَلَا تُصَوِّرْ* yang berarti janganlah kamu memalingkan dan *وَلَا تَمْشِ* yang

berarti janganlah kamu berjalan. Larangan memalingkan wajah karena sombong adalah kaidah untuk dihukumi haram atas melakukan perbuatan tersebut, sedangkan larangan berjalan memiliki kata penjelas yang menunjukkan keadaan untuk larangan tersebut adalah مَرَحًا yang berarti keangkuhan, ini yang menjadi kaidah untuk dihukumi haram jika melakukan perbuatan tersebut.

Dalam Hadits riwayat Muslim nomor 2865 terdapat kata kerja perintah تَوَاضَعُوا yang berarti rendahkanlah hati kalian atau hendaklah bersikap rendah hati. Kata kerja tersebut menjadi dasar hukum bahwa wajib untuk memiliki sifat rendah hati yang lawan dari sifat sombong. Memahami dari kaidah tersebut memberikan kesimpulan bahwa sombong adalah sifat yang diharamkan, tetapi sebaliknya memiliki kerendahan hati adalah sebuah kewajiban.

b. Buruk sangka

1) Definisi dan dalil-dalil tentang buruk sangka

Buruk sangka termasuk dari perbuatan tercela dalam tatanan kehidupan sosial. Oleh karena itu, perbuatan ini harus dihindari supaya dapat menciptakan hubungan kemasyarakatan yang baik. Buruk sangka didefinisikan sebagai suatu anggapan, sikap, pendapat buruk tanpa adanya bukti konkrit, sehingga bertentangan atau menunjukkan hal yang sebaliknya terhadap hal yang disangkakan (Nanda, 2015). Adapun macam-macam dari buruksangka adalah buruk sangka kepada diri sendiri, buruk sangka kepada orang lain, buruk sangka kepada keadaan, buruk sangka kepada Allah (Putri, 2019). Hal ini tegas dilarang dalam al-Qur'an surat al-Hujurat ayat 12.

يَا أَيُّهَا الَّذِينَ آمَنُوا اجْتَنِبُوا كَثِيرًا مِّنَ الظَّنِّ إِنَّ بَعْضَ الظَّنِّ إِثْمٌ وَلَا تَجَسَّسُوا وَلَا يَغْتَبَ بَعْضُكُم بَعْضًا أَيُحِبُّ أَحَدُكُمْ أَنْ يَأْكُلَ لَحْمَ أَخِيهِ مَيْتًا فَكَرِهْتُمُوهُ وَاتَّقُوا اللَّهَ إِنَّ اللَّهَ تَوَّابٌ رَّحِيمٌ

“Wahai orang-orang yang beriman! Jauhilah banyak dari prasangka, sesungguhnya sebagian prasangka itu dosa dan janganlah kamu mencari-cari kesalahan orang lain dan janganlah ada di antara kamu yang menggunjing sebagian yang lain. Apakah ada di antara kamu yang suka memakan daging saudaranya yang sudah mati? Tentu kamu merasa jijik. Dan bertakwalah kepada Allah, sesungguhnya Allah Maha Penerima tobat, Maha Penyayang.” (QS. Al-Hujurat ayat 12).

Awal dari ayat tersebut menyatakan bahwa kebanyakan dari prasangka itu akan menimbulkan dosa, sehingga Allah memberi peringatan untuk menjauhinya.

Perilaku buruk sangka ini ditegaskan dalam sebuah Hadits kalau buruk sangka itu adalah perkataan yang paling dusta.

إِيَّاكُمْ وَالظَّنَّ فَإِنَّ الظَّنَّ أَكْذَبُ الْحَدِيثِ

“Jauhilah kamu sekalian pada sifat berburuk sangka, karena sesungguhnya berburuk sangka adalah perkataan yang paling dusta.” (HR. Bukhari dan Muslim).

Hasil dari memahami ayat al-Qur’an dan Hadits di atas tentang buruk sangka, dapat disimpulkan bahwa prasangka buruk adalah termasuk perbuatan tercela dan perbuatan yang harus ditinggalkan karena terdapat perintah untuk menjauhinya.

2) Bentuk-bentuk alasan seseorang melakukan buruk sangka

Sifat buruk sangka memiliki penyebab kemunculan (Hidayat, 2013), sebagai berikut.

- Tidak pada kategori yang sama (baik dari jalan pikiran, status sosial, maupun kelompok).
- Individu yang memiliki dominasi sosial tinggi (otoriter) cenderung bersikap negatif karena selalu ingin memiliki terlihat dominan.
- Fanatisme individu terhadap nilai-nilai dari kelompoknya.
- Ambisi untuk selalu berkompetisi, meliputi perjuangan untuk memperoleh pekerjaan, tempat hunian yang layak, pendidikan yang unggul, atau menyangkut kepemilikan yang lain. Pada akhirnya kompetisi yang berkepanjangan akan memicu pandangan negatif terhadap kelompok lain (Elfariani, 2021).

3) Hukum berburuk sangka

Kerjasama antar anggota masyarakat merupakan sebuah keharusan supaya terjaga dan tercapainya kepentingan bersama yang salah satunya dengan dilarangnya buruk sangka terhadap sesama (Lajnah Pentashihan Mushaf Al-Qur’an, 2010). Bahkan sifat buruk sangka ini menurut sebagian ulama lebih berbahaya daripada berdusta (Gymnastiar, 2003), hal ini ditinjau dari seperti Hadits tentang buruk sangka di atas. Selain itu, terdapat kata perintah dalam surat Hujurat ayat 12, yaitu اجْتَنِبُوا yang berarti jauhilah kamu sekalian. Bahkan setelah kata tersebut ditegaskan dengan kata كَثِيرًا yang berarti banyak, hal ini mewakili arti dari keharusan untuk menjauhi segala jenis prasangka karena sebagian besar dari

prasangka itu adalah dosa. Hukum atas sifat dan perilaku buruk sangka ini adalah haram untuk dilakukan.

c. Kesamaan agama membentuk dukungan

1) Kesamaan agama dan dalil-dalilnya

Identitas sosial dibentuk berdasarkan berbagai hal, salah satunya dengan adanya persamaan ideologi keagamaan (Ikhsan, 2017). Persamaan ini membentuk pola pandangan atau opini yang sama juga terhadap sebuah peristiwa yang sedang maupun telah terjadi, namun dalam kesamaan ideologi keagamaan juga tidak bisa menjamin memiliki pandangan yang sama karena masing-masing individu memiliki berbagai penafsiran dan pemahaman yang berbeda-beda (Anggundia, 2017). Al-Qur'an terdapat surat tentang sesama orang beriman adalah saudara sebagai berikut.

إِنَّمَا الْمُؤْمِنُونَ إِخْوَةٌ فَأَصْلِحُوا بَيْنَ أَخَوَيْكُمْ ۗ وَاتَّقُوا اللَّهَ لَعَلَّكُمْ تُرْحَمُونَ

“Orang-orang beriman itu sesungguhnya bersaudara. Sebab itu damaikanlah (perbaikilah hubungan) antara kedua saudaramu itu dan takutlah terhadap Allah, supaya kamu mendapat rahmat.” (QS. Al-Hujurat ayat 10).

Adapun Hadits yang memberitahu pentingnya persaudaraan dan keharusan untuk saling menuatkan sebagai berikut.

الْمُؤْمِنُ لِلْمُؤْمِنِ كَالْبُنْيَانِ يَتَدُّ بِغَضِّهِ بَعْضًا

“Seorang mukmin dengan mukmin lainnya seperti satu bangunan, sebagiannya menguatkan sebagian yang lain.” (HR. Bukhari dan Muslim).

Hasil dari memahami ayat al-Qur'an dan Hadits di atas tentang kesamaan agama dalam konteks agama Islam, dapat disimpulkan bahwa opini yang bersifat mendukung dapat terjadi ketika seorang yang memiliki pemahaman mengenai ukhuwah Islamiyah yang terdapat dalam al-Qur'an maupun Hadits.

2) Bentuk-bentuk ukhuwah Islamiyah

Ukhuwah Islamiyah terbentuk atas dasar keimanan kepada Allah, rasul, beserta kitabnya. Dalam menjalankannya perlu adanya beberapa hal yang harus diperhatikan untuk dihindari agar bisa terjaga ikatan persaudaraan (Anshori, 2016), beberapa hal tersebut disebutkan antara lain.

- Saling mengolok
- Menghina dan mencaci dengan kata-kata yang menyakitkan
- Memanggil dengan nama atau sebutan yang buruk

- Buruk sangka
- Mencari kesalahan
- Menggunjing

Enam hal yang telah tersebut perlu dihindari supaya menciptakan persatuan, kekuatan, dan kasih sayang terhadap sesama ikatan persaudaraan.

2.2.14 Fikih Teknologi

Twitter merupakan salah satu dari teknologi informasi yang berbasis media sosial. Sebagai teknologi, media sosial ini tentu terdapat hal baik maupun buruk tergantung dengan penggunaannya. Oleh karena itu, dalam hukum Islam teknologi bersifat sebagai perantara yang tidak memiliki hukum, kecuali sesuai dengan tujuan penggunaannya. Hal tersebut terdapat dalam dua Hadits yang menerangkan pisau sebagai alat yang mana ada dalam kitab Hadits shahih Muslim. Kedua Hadits tersebut dipaparkan sebagai berikut.

إِنَّ اللَّهَ كَتَبَ الْإِحْسَانَ عَلَى كُلِّ شَيْءٍ فَإِذَا قَتَلْتُمْ فَأَحْسِنُوا الْقِتْلَةَ وَإِذَا ذَبَحْتُمْ فَأَحْسِنُوا الذَّبْحَ وَلِيُحَدِّثَ أَحَدَكُمْ شَفْرَتَهُ فَلْيُرِّخْ ذَبِيحَتَهُ

“*Sesungguhnya Allah memerintahkan agar berbuat baik terhadap segala sesuatu. Jika kalian hendak membunuh, maka bunuhlah dengan cara yang baik. Jika kalian hendak menyembelih, maka sembelihlah dengan cara yang baik. Hendaklah kalian menajamkan pisaunya dan senangkanlah hewan yang akan disembelih.*” (HR. Muslim no. 1955).

مَنْ أَشَارَ إِلَى أَخِيهِ بِحَدِيدَةٍ فَإِنَّ الْمَلَائِكَةَ تَلْعَنُهُ حَتَّىٰ وَإِنْ كَانَ أَخَاهُ لِأَبِيهِ وَأُمِّهِ

“*Barangsiapa menodongkan senjata tajam kepada saudaranya, maka para malaikat akan melaknatnya sampai dia meninggalkan perbuatan tersebut, walaupun saudara tersebut adalah saudara kandung seapak dan seibu.*” (HR. Muslim, no. 2616).

Terdapat dua kasus yang berbeda dari penggunaan pisau atau senjata tajam dari kedua hadis di atas. Hadits yang pertama menyatakan bahwa pisau digunakan untuk menyembelih binatang harus dengan kondisi tajam terlebih dahulu, sedangkan pada Hadits kedua menyebutkan bahwa akan mendapatkan laknat dari malaikat jika senjata tajam tersebut digunakan hanya sekedar mengacungkan atau mengarahkannya kepada manusia terlebih yang seiman dan masih tergolong keluarga kandung.

a. Teknologi sebagai media

Dapat dipahami dari dua Hadits tentang pisau tersebut bahwa pisau atau senjata tajam disebut sebagai wasilah yang dihukumi sesuai tujuan dari

penggunaanya. Jika digunakan untuk kebaikan maka akan dihukumi mubah bahkan wajib, sedangkan jika digunakan untuk keburukan maka akan dihukumi makruh bahkan haram. Terdapat kaidah fikih mengenai hukum dari benda sebagai media (Yulianto, 2021).

لِلْوَسَائِلِ حُكْمُ الْمَقَاصِدِ

“*Hukum perantara itu adalah hukum tujuan (penggunaan) dari perantara itu.*”

Kaidah tersebut memberikan pengertian bahwa alat tidak memiliki hukum kecuali pada tujuan pemakaiannya. Dapat disimpulkan bahwa platform Twitter sebagai teknologi media sosial tidak memiliki hukum dalam pemakaian, kecuali jika digunakan untuk hal buruk dapat dihukumi makruh bahkan haram atau jika digunakan untuk kebaikan, maka dapat dihukumi sunnah bahkan wajib.

b. Tujuan pemakaian teknologi dan hukumnya

Hal tersebut jika dari sudut pandang Islam terhadap pemakaian teknologi sebagai media sosial dijelaskan bahwa pemanfaatan teknologi dengan tujuan masalah primer hukumnya wajib, masalah sekunder hukumnya sunah, mafsadah primer hukumnya haram, dan mafsadah sekunder hukumnya makruh (Suprpto & Yulianto, 2023). Adapun demikian itu ditegaskan jika pemanfaatan dan pengembangan teknologi itu untuk menjaga agama, jiwa, akal, keturunan, serta harta, maka hukumnya *fardu ain* (wajib).

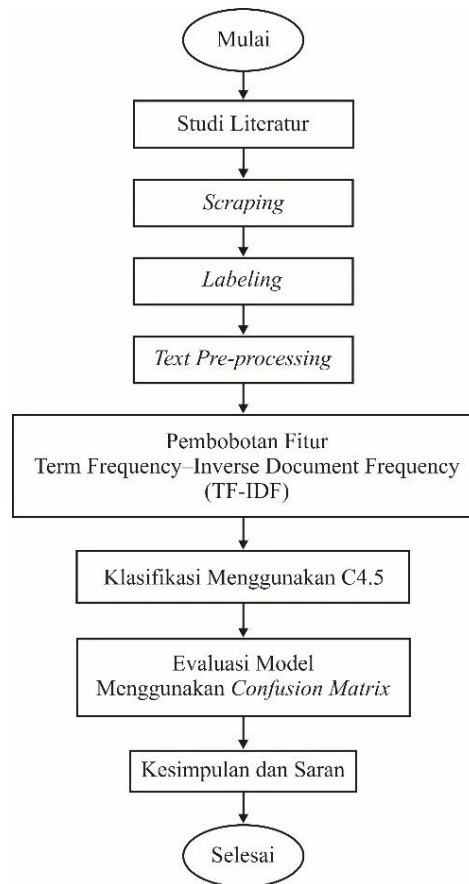
BAB III METODE PENELITIAN

3.1 Jenis Penelitian

Penelitian ini memakai pendekatan *text mining*. *Text mining* adalah proses teknik ekstraksi informasi penting dan bermakna dari kumpulan data yang besar (Tavana et al., 2022). Pada pendekatan ini, data teks yang berjumlah banyak dan tidak terstruktur diekstrak menjadi data terstruktur dan dianalisis sesuai dengan tujuannya. Adapun dalam penelitian ini menggunakan *text mining* sebagai klasifikasi data yang memakai algoritma C4.5.

3.2 Desain Penelitian

Penelitian ini terdiri atas tujuh tahapan, yaitu studi literatur, *scraping* data, *labeling*, *text pre-processing*, pembobotan fitur, klasifikasi, dan evaluasi model. Adapun desain atau alur penelitian serta penjelasannya terdapat di gambar 3.1.



Gambar 3.1: Alur Penelitian
(Sumber: Hasil olah data peneliti, 2023)

3.2.1 Studi Literatur

Tahap pertama adalah mencari segala kepustakaan yang berkaitan dengan analisis sentimen menggunakan berbagai algoritma *machine learning* terutama C4.5. Kemudian, mengamati fenomena-fenomena yang sedang terjadi di dunia melalui berita *online*, *trending topic* di Twitter dan website statistik sehingga akan dijadikan topik pada penelitian terkait penambahan data teks yang ada di Twitter. Topik yang ditentukan adalah terkait Qatar menjadi tuan rumah piala dunia FIFA 2022.

3.2.2 Scraping

Tahap kedua adalah melakukan *scraping* data teks atau opini yang ada di Twitter. Dalam tahap ini menggunakan Jupyter Notebook dan *library* Snsrape sebagai pengambilan data sekaligus pembuatan dataset. Pada proses *information retrieval* dikendalikan dengan beberapa filter, seperti teks berbahasa Indonesia, kata kunci “qatar sebagai tuan rumah”.

3.2.3 Labeling

Tahap ketiga adalah pelabelan atau pemberian kelas pada data yang telah terbentuk dataset. Pemberian kelas pada penelitian ini dilakukan secara manual yang memiliki tiga kelas yaitu positif, negatif, dan netral. Pelabelan dilakukan oleh ahli yang berprofesi sebagai guru bahasa Indonesia di Madrasah Aliyah al-Irtiqo’ Malang dan guru ilmu sosial di SMA MBS Ki Bagus Hadikusumo Bogor.

3.2.4 Text Pre-processing

Tahap keempat adalah melakukan pembersihan terhadap data, berbagai atribut yang kurang berdampak pada proses klasifikasi akan dihilangkan supaya mesin memahami teks dengan akurasi tinggi. Data mentah yang diproses pada tahap ini diharapkan menjadi dokumen yang memiliki kualitas tinggi dalam proses klasifikasi.

Pada tahap *text pre-processing* memiliki beberapa urutan dalam pembersihan data, yaitu mulai dari *cleaning*, *filtering*, *tokenizing*, dan *stemming*.

- a) *Cleaning* memiliki peran pengurangan *noise* dengan eliminasi beberapa hal, seperti angka, URL, email, simbol, dan karakter khusus (Muttaqin & Bachtiar, 2016). Adapun proses detailnya seperti *case folding*, *punctuation*, *emoticon*, URL, *hashtag*, *username*, *number*, dan *whitespace*.

- *Case folding* berperan mengubah ukuran teks menjadi serupa dalam huruf kecil atau *lowercase* (Azzahra & Wibowo, 2020), karena dalam data yang tidak terstruktur banyak terjadi inkonsistensi pola dari data. Contoh dari proses *case folding* pada tabel 3.1.

Tabel 3.1: Case Folding

Data mentah	Belanda Resmi Jadikan Qatar Sebagai Tuan Rumah Terburuk Dalam Sejarah Piala Dunia! https://t.co/tqYVLMlDPi
Hasil <i>case folding</i>	belanda resmi jadikan qatar sebagai tuan rumah terburuk dalam sejarah piala dunia! https://t.co/tqyvlmldpi

(Sumber: Hasil olah data peneliti, 2023)

- Eliminasi *punctuation* merupakan penghapusan tanda baca seperti tanda tanya (?), tanda seru (!), koma (,) dan sebagainya (Filcha & Hayaty, 2019). Proses ini juga berperan penting dalam proses *cleaning* data. Contoh dari proses eliminasi tanda baca pada tabel 3.2.

Tabel 3.2: Punctuation

Data mentah	Itu juga alasan yang bikin Qatar (sebagai tuan rumah) mendukung banget Maroko. Setiap mereka main—apalagi menang—ada bendera Maroko di mana-mana. Para pendukung Maroko juga berasa ada di rumah/negara sendiri.
Hasil eliminasi tanda baca	Itu juga alasan yang bikin Qatar sebagai tuan rumah mendukung banget Maroko Setiap mereka main apalagi menang ada bendera Maroko di mana mana Para pendukung Maroko juga berasa ada di rumah negara sendiri

(Sumber: Hasil olah data peneliti, 2023)

- Eliminasi *emoticon*, karena terkadang *emoticon* yang dipakai tidak mengekspresikan sebuah emosi yang sesungguhnya. Kemungkinan pemahaman dan interpretasi pengguna dalam arti dan penggunaan *emoticon* sangat beragam. Hal tersebut dapat berupa ikon yang sesuai dengan hal yang disampaikan atau ikon yang tidak melambangkan apa yang disampaikan (Rico & Fuady, 2014). Contoh dari proses eliminasi *emoticon* pada tabel 3.3.

Tabel 3.3: Emoticon

Data mentah	Pildun Qatar emang penuh dengan kejutan 😊😊 ((Even Qatar sebagai tuan rumah is also kinda...A suprise))
Hasil eliminasi <i>emoticon</i>	Pildun Qatar emang penuh dengan kejutan ((Even Qatar sebagai tuan rumah is also kinda...A suprise))

(Sumber: Hasil olah data peneliti, 2023)

- Eliminasi URL sangat diperlukan, karena URL tidak memiliki arti untuk kalimat yang akan diproses dalam klasifikasi (Khairunnisa et al., 2021). Contoh dari proses eliminasi URL pada tabel 3.4.

Tabel 3.4: URL

Data mentah	Inilah klarifikasi ILO terkait berita kematian pekerja migran di sektor konstruksi Qatar, sejak FIFA menetapkan negara itu sebagai tuan rumah Piala Dunia 2022. https://t.co/yP9G8NEiO0
Hasil eliminasi URL	Inilah klarifikasi ILO terkait berita kematian pekerja migran di sektor konstruksi Qatar, sejak FIFA menetapkan negara itu sebagai tuan rumah Piala Dunia 2022.

(Sumber: Hasil olah data peneliti, 2023)

- Eliminasi *hashtag*, karena tidak memiliki arti dalam konteks topik dan hanya sebagai salah satu cara untuk temu kembali informasi (Hidayatullah et al., 2021). Contoh dari proses eliminasi *hashtag* pada tabel 3.5.

Tabel 3.5: Hashtag

Data mentah	Polisi Belgia menyelidiki dugaan kasus penyuapan politikus Uni Eropa terkait kemenangan Qatar sebagai tuan rumah Piala Dunia 2022. #Internasional #AdadiKompas https://t.co/l2f3bEdJy2
Hasil eliminasi <i>hashtag</i>	Polisi Belgia menyelidiki dugaan kasus penyuapan politikus Uni Eropa terkait kemenangan Qatar sebagai tuan rumah Piala Dunia 2022. https://t.co/l2f3bEdJy2

(Sumber: Hasil olah data peneliti, 2023)

- Eliminasi *username* atau *mention*, karena tidak memiliki arti dalam konteks topik sehingga termasuk sebagai komponen yang tidak berguna

dalam klasifikasi (Salim & Mayary, 2020). Contoh dari proses eliminasi *username* pada tabel 3.6.

Tabel 3.6: Username

Data mentah	@anonapostate @MichaelRainhard @dina_sulaeman Qatar sudah jelas menghargai event sepak bola terbesar di dunia justru kaum" lu yang gak menghargai qatar sebagai tuan rumah.
Hasil eliminasi <i>username</i>	Qatar sudah jelas menghargai event sepak bola terbesar di dunia justru kaum" lu yang gak menghargai qatar sebagai tuan rumah.

(Sumber: Hasil olah data peneliti, 2023)

- Eliminasi *number*, karena nomor dianggap kurang bernilai dalam analisis teks (Pradana, 2020). Contoh dari proses eliminasi *number* pada tabel 3.7.

Tabel 3.7: Number

Data mentah	Sebagai tuan rumah Piala Dunia 2022, Qatar punya cara khusus dalam memastikan keamanan. Turki menjadi salah satu negara yang membantu pengamanan dengan mengirim 100 petugas polisi operasi khusus, 50 spesialis bom, dan 80 anjing pelacak. #WorldcupQatar2022 #WorldCup #Golkar2024 https://t.co/HYap4luNbp
Hasil eliminasi <i>number</i>	Sebagai tuan rumah Piala Dunia, Qatar punya cara khusus dalam memastikan keamanan. Turki menjadi salah satu negara yang membantu pengamanan dengan mengirim petugas polisi operasi khusus, spesialis bom, dan anjing pelacak. #WorldcupQatar #WorldCup #Golkar https://t.co/HYap4luNbp

(Sumber: Hasil olah data peneliti, 2023)

- Eliminasi *whitespace* ada pada akhir tahap cleaning karena ada kemungkinan setelah penghapusan beberapa hal yang tidak penting dapat menyisakan spasi kosong (Zuhad & Wilantika, 2022). Contoh dari proses eliminasi *whitespace* pada tabel 3.8.

Tabel 3.8: Whitespace

Data mentah	Menghormati qatar sebagai tuan rumah #WorldcupQatar2022
Hasil eliminasi <i>whitespace</i>	Menghormati qatar sebagai tuan rumah #WorldcupQatar2022

(Sumber: Hasil olah data peneliti, 2023)

- b) *Filtering* atau dapat disebut *stopwords* menggunakan *library* sastrawi yang terdapat 804 kata, *filtering* berperan membuang semua kata tidak penting yang ada dalam sebuah kalimat (Indraini & Ernawati, 2022). Selain itu, ada beberapa tambahan daftar *stopwords* yang ada pada lampiran penelitian karena semakin banyak *stopwords* yang dieliminasi, maka semakin memiliki potensi peningkatan akurasi (Baskoro et al., 2021). Contoh pada tahap *filtering* pada tabel 3.9.

Tabel 3.9: Filtering

Data mentah	Kekalahan Qatar atas Ekuador semalam menjadikan Qatar sebagai tuan rumah Piala Dunia pertama yang kalah di pertandingan pembuka. Permainan khas Qatar, tiki-taka, tidak berjalan dengan baik dan tidak mencerminkan sebagai juara Asia.
Hasil <i>filtering</i>	Kekalahan Qatar Ekuador semalam menjadikan Qatar tuan rumah Piala Dunia pertama kalah pertandingan pembuka. Permainan khas Qatar, tiki-taka, berjalan baik mencerminkan juara Asia.

(Sumber: Hasil olah data peneliti, 2023)

- c) *Tokenizing* berperan sebagai pemisah kalimat menjadi pecahan kata atau token (Jamaluddin et al., 2020). Contoh pada tahap *tokenizing* pada tabel 3.10.

Tabel 3.10: Tokenizing

Data mentah	Hanya Qatar yang gagal cetak gol pertama sebagai tuan rumah https://t.co/P9eQGs04FA
Hasil <i>tokenizing</i>	Hanya Qatar yang gagal cetak gol pertama sebagai tuan rumah

(Sumber: Hasil olah data peneliti, 2023)

- d) *Stemming* menggunakan *library* Sastrawi yang mengambil berdasarkan kamus kata dasar dari <http://kateglo.lostfocus.org>. *Stemmer* Sastrawi dipilih karena lebih baik daripada *Porter* dan *Snowball* untuk *stemmer* dokumen berbahasa Indonesia (Hendrawan et al., 2021). *Stemming* memiliki peran sebagai pengembalian kata yang terdapat imbuhan menjadi kata dasar (Widhiasih et al.,

2023). Dalam sebuah kalimat tentu terdapat *suffix* maupun *prefix* pada setiap kata. Oleh karena itu, harus dikembalikan kepada kata dasarnya agar mesin mudah mengenali sehingga mempermudah proses klasifikasi. Contoh pada tahap *stemming* pada tabel 3.11.

Tabel 3.11: Stemming

Data mentah	Pembangunan Infrastruktur Qatar sebagai Tuan Rumah Piala Dunia Banyak Menarik Perhatian Berbagai Negara di Dunia
Hasil <i>stemming</i>	Bangun, Infrastruktur, Qatar, Tuan, Rumah, Piala, Dunia, Banyak, Tarik, Hati, Berbagai, Negara, Dunia

(Sumber: Hasil olah data peneliti, 2023)

3.2.5 Pembobotan Fitur Menggunakan *Term Frequency–Inverse Document Frequency* (TF-IDF)

Setelah melakukan *text pre-processing* adalah melakukan pembobotan istilah atau ekstraksi fitur dengan TF-IDF. Pembobotan istilah pada penelitian ini memakai memakai metode TF murni dalam formulasi menentukan nilai TF. Adapun dari segi pola perhitungan deret kata menggunakan karakter unigram atau token yang terdiri dari satu kata, karena memberikan hasil lebih baik daripada bigram dan trigram (Tiffani, 2020).

a) Data mentah

Tabel 3.12: Contoh Data Mentah

A	Piala dunia Qatar emang penuh dengan kejutan
B	Terkejut!! Qatar negara paling sukses sebagai tuan rumah PD sepanjang sejarah
C	Qatar Gagal Totalnya Sebagai Tuan Rumah Piala Dunia
D	Sudah Resmi! Qatar Tercatat Sebagai Tuan Rumah Piala Dunia Paling Buruk
E	Gagal Bedakan prestasi Qatar sebagai tuan rumah Pildun dan Qatar sebagai peserta Pildun

(Sumber: Hasil olah data peneliti, 2023)

b) Data setelah proses *text pre-processing*

Tabel 3.13: Proses *Text Pre-processing*

A	piala dunia qatar memang penuh kejut
B	Kejut qatar negara paling sukses tuan rumah panjang sejarah
C	qatar gagal total tuan rumah piala dunia

D	sudah resmi qatar catat tuan rumah piala dunia paling buruk
E	gagal beda prestasi qatar tuan rumah qatar peserta

(Sumber: Hasil olah data peneliti, 2023)

c) Menghitung TF, IDF, dan TF-IDF

- Jumlah fitur dalam setiap dokumen

Tabel 3.14: Jumlah Frekuensi Fitur

Dokumen 1		Dokumen 2		Dokumen 3		Dokumen 4		Dokumen 5	
Fitur	Jumlah	Fitur	Jumlah	Fitur	Jumlah	Fitur	Jumlah	Fitur	Jumlah
piala	1	piala	0	piala	1	piala	1	piala	0
dunia	1	dunia	0	dunia	1	dunia	1	dunia	0
qatar	1	qatar	1	qatar	1	qatar	1	qatar	1
memang	1	memang	0	memang	0	memang	0	memang	0
penuh	1	penuh	0	penuh	0	penuh	0	penuh	0
kejut	1	kejut	0	kejut	0	kejut	0	kejut	0
negara	0	negara	1	negara	0	negara	0	negara	0
paling	0	paling	1	paling	0	paling	0	paling	0
sukses	0	sukses	1	sukses	0	sukses	0	sukses	0
tuan	0	tuan	1	tuan	1	tuan	1	tuan	1
rumah	0	rumah	1	rumah	1	rumah	1	rumah	1
panjang	0	panjang	1	panjang	0	panjang	0	panjang	0
sejarah	0	sejarah	1	sejarah	0	sejarah	0	sejarah	0
gagal	0	gagal	0	gagal	1	gagal	0	gagal	0
total	0	total	0	total	1	total	0	total	0
sudah	0	sudah	0	sudah	0	sudah	1	sudah	0
resmi	0	resmi	0	resmi	0	resmi	1	resmi	0
catat	0	catat	0	catat	0	catat	1	catat	0
buruk	0	buruk	0	buruk	0	buruk	1	buruk	0
beda	0	beda	0	beda	0	beda	0	beda	1
prestasi	0	prestasi	0	prestasi	0	prestasi	0	prestasi	1
peserta	0	peserta	0	peserta	0	peserta	0	peserta	1

(Sumber: Hasil olah data peneliti, 2023)

- Menghitung TF

Tabel 3.15:Perhitungan Term Frequency

No.	Dokumen 1	Dokumen 2	Dokumen 3	Dokumen 4	Dokumen 5
1	tf (“piala”, D1) = $1/6 = 0.166$	tf (“piala”, D2) = $0/8 = 0$	tf (“piala”, D3) = $1/7 = 0.142$	tf (“piala”, D4) = $1/9 = 0.111$	tf (“piala”, D5) = $0/6 = 0$
2	tf (“dunia”, D1) = $1/6 = 0.166$	tf (“dunia”, D2) = $0/8 = 0$	tf (“dunia”, D3) = $1/7 = 0.142$	tf (“dunia”, D4) = $1/9 = 0.111$	tf (“dunia”, D5) = $0/6 = 0$
3	tf (“qatar”, D1) = $1/6 = 0.166$	tf (“qatar”, D2) = $1/8 = 0.125$	tf (“qatar”, D3) = $1/7 = 0.142$	tf (“qatar”, D4) = $1/9 = 0.111$	tf (“qatar”, D5) = $1/6 = 0.166$
4	tf (“memang”, D1) = $1/6 = 0.166$	tf (“memang”, D2) = $0/8 = 0$	tf (“memang”, D3) = $0/7 = 0$	tf (“memang”, D4) = $0/9 = 0$	tf (“memang”, D5) = $0/6 = 0$
5	tf (“penuh”, D1) = $1/6 = 0.166$	tf (“penuh”, D2) = $0/8 = 0$	tf (“penuh”, D3) = $0/7 = 0$	tf (“penuh”, D4) = $0/9 = 0$	tf (“penuh”, D5) = $0/6 = 0$
6	tf (“kejut”, D1) = $1/6 = 0.166$	tf (“kejut”, D2) = $0/8 = 0$	tf (“kejut”, D3) = $0/7 = 0$	tf (“kejut”, D4) = $0/9 = 0$	tf (“kejut”, D5) = $0/6 = 0$
7	tf (“negara”, D1) = $0/6 = 0$	tf (“negara”, D2) = $1/8 = 0.125$	tf (“negara”, D3) = $0/7 = 0$	tf (“negara”, D4) = $0/9 = 0$	tf (“negara”, D5) = $0/6 = 0$
8	tf (“paling”, D1) = $0/6 = 0$	tf (“paling”, D2) = $1/8 = 0.125$	tf (“paling”, D3) = $0/7 = 0$	tf (“paling”, D4) = $1/9 = 0.111$	tf (“paling”, D5) = $0/6 = 0$
9	tf (“sukses”, D1) = $0/6 = 0$	tf (“sukses”, D2) = $1/8 = 0.125$	tf (“sukses”, D3) = $0/7 = 0$	tf (“sukses”, D4) = $0/9 = 0$	tf (“sukses”, D5) = $0/6 = 0$
10	tf (“tuan”, D1) = $0/6 = 0$	tf (“tuan”, D2) = $1/8 = 0.125$	tf (“tuan”, D3) = $1/7 = 0.142$	tf (“tuan”, D4) = $1/9 = 0.111$	tf (“tuan”, D5) = $1/6 = 0.166$
11	tf (“rumah”, D1) = $0/6 = 0$	tf (“rumah”, D2) = $1/8 = 0.125$	tf (“rumah”, D3) = $1/7 = 0.142$	tf (“rumah”, D4) = $1/9 = 0.111$	tf (“rumah”, D5) = $1/6 = 0.166$
12	tf (“panjang”, D1) = $0/6 = 0$	tf (“panjang”, D2) = $1/8 = 0.125$	tf (“panjang”, D3) = $0/7 = 0$	tf (“panjang”, D4) = $0/9 = 0$	tf (“panjang”, D5) = $0/6 = 0$

No.	Dokumen 1	Dokumen 2	Dokumen 3	Dokumen 4	Dokumen 5
13	tf (“sejarah”, D1) = 0/6 = 0	tf (“sejarah”, D2) = 1/8 = 0.125	tf (“sejarah”, D3) = 0/7 = 0	tf (“sejarah”, D4) = 0/9 = 0	tf (“sejarah”, D5) = 0/6 = 0
14	tf (“gagal”, D1) = 0/6 = 0	tf (“gagal”, D2) = 0/8 = 0	tf (“gagal”, D3) = 1/7 = 0.142	tf (“gagal”, D4) = 0/9 = 0	tf (“gagal”, D5) = 0/6 = 0
15	tf (“total”, D1) = 0/6 = 0	tf (“total”, D2) = 0/8 = 0	tf (“total”, D3) = 1/7 = 0.142	tf (“total”, D4) = 0/9 = 0	tf (“total”, D5) = 0/6 = 0
16	tf (“sudah”, D1) = 0/6 = 0	tf (“sudah”, D2) = 0/8 = 0	tf (“sudah”, D3) = 0/7 = 0	tf (“sudah”, D4) = 1/9 = 0.111	tf (“sudah”, D5) = 0/6 = 0
17	tf (“resmi”, D1) = 0/6 = 0	tf (“resmi”, D2) = 0/8 = 0	tf (“resmi”, D3) = 0/7 = 0	tf (“resmi”, D4) = 1/9 = 0.111	tf (“resmi”, D5) = 0/6 = 0
18	tf (“catat”, D1) = 0/6 = 0	tf (“catat”, D2) = 0/8 = 0	tf (“catat”, D3) = 0/7 = 0	tf (“catat”, D4) = 1/9 = 0.111	tf (“catat”, D5) = 0/6 = 0
19	tf (“buruk”, D1) = 0/6 = 0	tf (“buruk”, D2) = 0/8 = 0	tf (“buruk”, D3) = 0/7 = 0	tf (“buruk”, D4) = 1/9 = 0.111	tf (“buruk”, D5) = 0/6 = 0
20	tf (“beda”, D1) = 0/6 = 0	tf (“beda”, D2) = 0/8 = 0	tf (“beda”, D3) = 0/7 = 0	tf (“beda”, D4) = 0/9 = 0.111	tf (“beda”, D5) = 1/6 = 0.166
21	tf (“prestasi”, D1) = 0/6 = 0	tf (“prestasi”, D2) = 0/8 = 0	tf (“prestasi”, D3) = 0/7 = 0	tf (“prestasi”, D4) = 0/9 = 0	tf (“prestasi”, D5) = 1/6 = 0.166
22	tf (“peserta”, D1) = 0/6 = 0	tf (“peserta”, D2) = 0/8 = 0	tf (“peserta”, D3) = 0/7 = 0	tf (“peserta”, D4) = 0/9 = 0	tf (“peserta”, D5) = 1/6 = 0.166

(Sumber: Hasil olah data peneliti, 2023)

- Menghitung IDF

Tabel 3.16: Perhitungan *Inverse Document Frequency*

No.	Fitur	IDF
1	piala	$\text{idf}(\text{"piala"}, D) = \log(3/5) = 0.221$
2	dunia	$\text{idf}(\text{"dunia"}, D) = \log(3/5) = 0.221$
3	qatar	$\text{idf}(\text{"qatar"}, D) = \log(5/5) = 0$
4	memang	$\text{idf}(\text{"memang"}, D) = \log(1/5) = 0.698$
5	penuh	$\text{idf}(\text{"penuh"}, D) = \log(1/5) = 0.698$
6	kejut	$\text{idf}(\text{"kejut"}, D) = \log(1/5) = 0.698$
7	negara	$\text{idf}(\text{"negara"}, D) = \log(1/5) = 0.698$
8	paling	$\text{idf}(\text{"paling"}, D) = \log(2/5) = 0.397$
9	sukses	$\text{idf}(\text{"sukses"}, D) = \log(1/5) = 0.698$
10	tuan	$\text{idf}(\text{"tuan"}, D) = \log(4/5) = 0.096$
11	rumah	$\text{idf}(\text{"rumah"}, D) = \log(4/5) = 0.096$
12	panjang	$\text{idf}(\text{"panjang"}, D) = \log(1/5) = 0.698$
13	sejarah	$\text{idf}(\text{"sejarah"}, D) = \log(1/5) = 0.698$
14	gagal	$\text{idf}(\text{"gagal"}, D) = \log(1/5) = 0.698$
15	total	$\text{idf}(\text{"total"}, D) = \log(1/5) = 0.698$
16	sudah	$\text{idf}(\text{"sudah"}, D) = \log(1/5) = 0.698$
17	resmi	$\text{idf}(\text{"resmi"}, D) = \log(1/5) = 0.698$
18	catat	$\text{idf}(\text{"catat"}, D) = \log(1/5) = 0.698$
19	buruk	$\text{idf}(\text{"buruk"}, D) = \log(1/5) = 0.698$
20	beda	$\text{idf}(\text{"beda"}, D) = \log(1/5) = 0.698$
21	prestasi	$\text{idf}(\text{"prestasi"}, D) = \log(1/5) = 0.698$
22	peserta	$\text{idf}(\text{"peserta"}, D) = \log(1/5) = 0.698$

(Sumber: Hasil olah data peneliti, 2023)

- Menghitung TF-IDF

Tabel 3.17: Perhitungan TF-IDF

No.	Fitur	TF-IDF
1	piala	$\text{tf-idf}(\text{"piala"}, D1, D) = 0.166 \times 0.221 = 0.036686$
		$\text{tf-idf}(\text{"piala"}, D2, D) = 0 \times 0.221 = 0$
		$\text{tf-idf}(\text{"piala"}, D3, D) = 0.142 \times 0.221 = 0.031382$
		$\text{tf-idf}(\text{"piala"}, D4, D) = 0.111 \times 0.221 = 0.024531$
		$\text{tf-idf}(\text{"piala"}, D5, D) = 0 \times 0.221 = 0$

No.	Fitur	TF-IDF
2	dunia	tf-idf ("dunia", D1, D) = $0.166 \times 0.221 = 0.036686$
		tf-idf ("dunia", D2, D) = $0 \times 0.221 = 0$
		tf-idf ("dunia", D3, D) = $0.142 \times 0.221 = 0.031382$
		tf-idf ("dunia", D4, D) = $0.111 \times 0.221 = 0.024531$
		tf-idf ("dunia", D5, D) = $0.166 \times 0.221 = 0$
3	qatar	tf-idf ("qatar", D1, D) = $0.166 \times 0 = 0$
		tf-idf ("qatar", D2, D) = $0.125 \times 0 = 0$
		tf-idf ("qatar", D3, D) = $0.142 \times 0 = 0$
		tf-idf ("qatar", D4, D) = $0.111 \times 0 = 0$
		tf-idf ("qatar", D5, D) = $0.166 \times 0 = 0$
4	memang	tf-idf ("memang", D1, D) = $0.166 \times 0.698 = 0.115868$
		tf-idf ("memang", D2, D) = $0 \times 0.698 = 0$
		tf-idf ("memang", D3, D) = $0 \times 0.698 = 0$
		tf-idf ("memang", D4, D) = $0 \times 0.698 = 0$
		tf-idf ("memang", D5, D) = $0 \times 0.698 = 0$
5	penuh	tf-idf ("penuh", D1, D) = $0.166 \times 0.698 = 0.115868$
		tf-idf ("penuh", D2, D) = $0 \times 0.698 = 0$
		tf-idf ("penuh", D3, D) = $0 \times 0.698 = 0$
		tf-idf ("penuh", D4, D) = $0 \times 0.698 = 0$
		tf-idf ("penuh", D5, D) = $0 \times 0.698 = 0$
6	kejut	tf-idf ("kejut", D1, D) = $0.166 \times 0.698 = 0.115868$
		tf-idf ("kejut", D2, D) = $0 \times 0.698 = 0$
		tf-idf ("kejut", D3, D) = $0 \times 0.698 = 0$
		tf-idf ("kejut", D4, D) = $0 \times 0.698 = 0$
		tf-idf ("kejut", D5, D) = $0 \times 0.698 = 0$
7	negara	tf-idf ("negara", D1, D) = $0 \times 0.698 = 0$
		tf-idf ("negara", D2, D) = $0.125 \times 0.698 = 0.08725$
		tf-idf ("negara", D3, D) = $0 \times 0.698 = 0$
		tf-idf ("negara", D4, D) = $0 \times 0.698 = 0$
		tf-idf ("negara", D5, D) = $0 \times 0.698 = 0$
8	paling	tf-idf ("paling", D1, D) = $0 \times 0.397 = 0$
		tf-idf ("paling", D2, D) = $0.125 \times 0.397 = 0.049625$
		tf-idf ("paling", D3, D) = $0 \times 0.397 = 0$

No.	Fitur	TF-IDF
		tf-idf ("paling", D4, D) = $0.111 \times 0.397 = 0.044067$
		tf-idf ("paling", D5, D) = $0 \times 0.397 = 0$
9	sukses	tf-idf ("sukses", D1, D) = $0 \times 0.698 = 0$
		tf-idf ("sukses", D2, D) = $0.125 \times 0.698 = 0.08725$
		tf-idf ("sukses", D3, D) = $0 \times 0.698 = 0$
		tf-idf ("sukses", D4, D) = $0 \times 0.698 = 0$
		tf-idf ("sukses", D5, D) = $0 \times 0.698 = 0$
10	tuan	tf-idf ("tuan", D1, D) = $0 \times 0.096 = 0$
		tf-idf ("tuan", D2, D) = $0.125 \times 0.096 = 0.012$
		tf-idf ("tuan", D3, D) = $0.142 \times 0.096 = 0.013632$
		tf-idf ("tuan", D4, D) = $0.111 \times 0.096 = 0.010656$
		tf-idf ("tuan", D5, D) = $0.166 \times 0.096 = 0.015936$
11	rumah	tf-idf ("rumah", D1, D) = $0 \times 0.096 = 0$
		tf-idf ("rumah", D2, D) = $0.125 \times 0.096 = 0.012$
		tf-idf ("rumah", D3, D) = $0.142 \times 0.096 = 0.013632$
		tf-idf ("rumah", D4, D) = $0.111 \times 0.096 = 0.010656$
		tf-idf ("rumah", D5, D) = $0.166 \times 0.096 = 0.015936$
12	panjang	tf-idf ("panjang", D1, D) = $0 \times 0.698 = 0$
		tf-idf ("panjang", D2, D) = $0.125 \times 0.698 = 0.08725$
		tf-idf ("panjang", D3, D) = $0 \times 0.698 = 0$
		tf-idf ("panjang", D4, D) = $0 \times 0.698 = 0$
		tf-idf ("panjang", D5, D) = $0 \times 0.698 = 0$
13	sejarah	tf-idf ("sejarah", D1, D) = $0 \times 0.698 = 0$
		tf-idf ("sejarah", D2, D) = $0.125 \times 0.698 = 0.08725$
		tf-idf ("sejarah", D3, D) = $0 \times 0.698 = 0$
		tf-idf ("sejarah", D4, D) = $0 \times 0.698 = 0$
		tf-idf ("sejarah", D5, D) = $0 \times 0.698 = 0$
14	gagal	tf-idf ("gagal", D1, D) = $0 \times 0.698 = 0$
		tf-idf ("gagal", D2, D) = $0 \times 0.698 = 0$
		tf-idf ("gagal", D3, D) = $0.142 \times 0.698 = 0.013632$
		tf-idf ("gagal", D4, D) = $0 \times 0.698 = 0$
		tf-idf ("gagal", D5, D) = $0 \times 0.698 = 0$
15	total	tf-idf ("total", D1, D) = $0 \times 0.698 = 0$

No.	Fitur	TF-IDF
		tf-idf ("total", D2, D) = $0 \times 0.698 = 0$
		tf-idf ("total", D3, D) = $0142 \times 0.698 = 0.013632$
		tf-idf ("total", D4, D) = $0 \times 0.698 = 0$
		tf-idf ("total", D5, D) = $0 \times 0.698 = 0$
16	sudah	tf-idf ("sudah", D1, D) = $0 \times 0.698 = 0$
		tf-idf ("sudah", D2, D) = $0 \times 0.698 = 0$
		tf-idf ("sudah", D3, D) = $0 \times 0.698 = 0$
		tf-idf ("sudah", D4, D) = $0.111 \times 0.698 = 0.077478$
		tf-idf ("sudah", D5, D) = $0 \times 0.698 = 0$
17	resmi	tf-idf ("resmi", D1, D) = $0 \times 0.698 = 0$
		tf-idf ("resmi", D2, D) = $0 \times 0.698 = 0$
		tf-idf ("resmi", D3, D) = $0 \times 0.698 = 0$
		tf-idf ("resmi", D4, D) = $0.111 \times 0.698 = 0.077478$
		tf-idf ("resmi", D5, D) = $0 \times 0.698 = 0$
18	catat	tf-idf ("catat", D1, D) = $0 \times 0.698 = 0$
		tf-idf ("catat", D2, D) = $0 \times 0.698 = 0$
		tf-idf ("catat", D3, D) = $0 \times 0.698 = 0$
		tf-idf ("catat", D4, D) = $0.111 \times 0.698 = 0.077478$
		tf-idf ("catat", D5, D) = $0 \times 0.698 = 0$
19	buruk	tf-idf ("buruk", D1, D) = $0 \times 0.698 = 0$
		tf-idf ("buruk", D2, D) = $0 \times 0.698 = 0$
		tf-idf ("buruk", D3, D) = $0 \times 0.698 = 0$
		tf-idf ("buruk", D4, D) = $0.111 \times 0.698 = 0.077478$
		tf-idf ("buruk", D5, D) = $0 \times 0.698 = 0$
20	beda	tf-idf ("beda", D1, D) = $0 \times 0.698 = 0$
		tf-idf ("beda", D2, D) = $0 \times 0.698 = 0$
		tf-idf ("beda", D3, D) = $0 \times 0.698 = 0$
		tf-idf ("beda", D4, D) = $0.111 \times 0.698 = 0.077478$
		tf-idf ("beda", D5, D) = $0 \times 0.698 = 0$
21	prestasi	tf-idf ("prestasi", D1, D) = $0 \times 0.698 = 0$
		tf-idf ("prestasi", D2, D) = $0 \times 0.698 = 0$
		tf-idf ("prestasi", D3, D) = $0 \times 0.698 = 0$
		tf-idf ("prestasi", D4, D) = $0 \times 0.698 = 0$

No.	Fitur	TF-IDF
		tf-idf (“prestasi”, D5, D) = 0.166 x 0.698 = 0.115868
22	peserta	tf-idf (“peserta”, D1, D) = 0 x 0.698 = 0
		tf-idf (“peserta”, D2, D) = 0 x 0.698 = 0
		tf-idf (“peserta”, D3, D) = 0 x 0.698 = 0
		tf-idf (“peserta”, D4, D) = 0 x 0.698 = 0
		tf-idf (“peserta”, D5, D) = 0.166 x 0.698 = 0.115868

(Sumber: Hasil olah data peneliti, 2023)

3.2.6 Klasifikasi Menggunakan C4.5

Setelah data melewati tahap *pre-processing* dan pembobotan istilah, selanjutnya membagi data menjadi 2 yaitu data latih dan data uji dengan rasio 70:30 dan rasio 80:20, karena kedua rasio tersebut dapat dikatakan memiliki tingkat kesamaan akurasi yang sama tingginya pada penggunaan algoritma C4.5 (Asshiddiqi & Lhaksana, 2020). Sebelum mengklasifikasi data uji, algoritma C4.5 perlu mempelajari model dari data latih yang telah diberi label atau kelas. Pada tabel 3.15 adalah contoh data latih yang akan dipelajari oleh mesin dengan algoritma C4.5.

Tabel 3.18: Contoh Data Latih

Kode Dokumen	Dokumen Data Latih	Kelas/Label
D1	piala dunia qatar memang penuh kejut	positif
D2	kejut qatar negara paling sukses tuan rumah panjang sejarah	positif
D3	qatar gagal total tuan rumah piala dunia	negatif
D4	sudah resmi qatar catat tuan rumah piala dunia paling buruk	negatif
D5	gagal beda prestasi qatar tuan rumah qatar peserta	netral

(Sumber: Hasil olah data peneliti, 2023)

Selanjutnya atribut (kata) yang terpilih akan diberi nilai 0 jika tidak ada dalam dokumen (*tweet*), sedangkan atribut yang ada dalam dokumen akan diberi nilai 1. Berikut pada tabel 3.16 adalah nilai dari atribut tiap dokumen.

Tabel 3.19: Nilai Data Latih Pertama

KD	Atribut							Kelas
	kejut	sukses	gagal	resmi	buruk	prestasi	peserta	
D1	1	0	0	0	0	0	0	pos
D2	1	1	0	0	0	0	0	pos
D3	0	0	1	0	0	0	0	neg
D4	0	0	1	1	1	0	0	neg
D5	0	0	1	0	0	1	1	net

(Sumber: Hasil olah data peneliti, 2023)

Perhitungan *entropy* untuk keseluruhan data latih:

Entropy total =

$$\left(-\frac{1}{5}\right) \times \log_2 \left(\frac{1}{5}\right) + \left(-\frac{2}{5}\right) \times \log_2 \left(\frac{2}{5}\right) + \left(-\frac{1}{5}\right) \times \log_2 \left(\frac{1}{5}\right) + \left(-\frac{3}{5}\right) \times \log_2 \left(\frac{3}{5}\right) + \left(-\frac{3}{5}\right) \times \log_2 \left(\frac{3}{5}\right) = 2.341$$

Perhitungan *entropy* dan gain pada setiap atribut untuk menentukan simpul akar:

a) Kejut

$$\text{Entropy atribut "kejut" yang bernilai (0)} = \left(-\frac{1}{3}\right) \times \log_2 \left(\frac{1}{3}\right) + \left(-\frac{2}{3}\right) \times \log_2 \left(\frac{2}{3}\right) = 0.918$$

$$\text{Entropy atribut "kejut" yang bernilai (1)} = \left(-\frac{2}{2}\right) \times \log_2 \left(\frac{2}{2}\right) = 0$$

$$\text{Gain (total, kejut)} = (2.341) - \left(\frac{3}{5}\right) \times (0.918) + \left(\frac{2}{5}\right) \times (0) = 1.790$$

b) Sukses

$$\text{Entropy atribut "sukses" yang bernilai (0)} = \left(-\frac{1}{4}\right) \times \log_2 \left(\frac{1}{4}\right) + \left(-\frac{2}{4}\right) \times \log_2 \left(\frac{2}{4}\right) + \left(-\frac{1}{4}\right) \times \log_2 \left(\frac{1}{4}\right) = 1.5$$

$$\text{Entropy atribut "sukses" yang bernilai (1)} = \left(-\frac{1}{1}\right) \times \log_2 \left(\frac{1}{1}\right) = 0$$

$$\text{Gain (total, sukses)} = (2.341) - \left(\frac{4}{5}\right) \times (1.5) + \left(\frac{1}{5}\right) \times (0) = 1.141$$

c) Gagal

$$\text{Entropy atribut "gagal" yang bernilai (0)} = \left(-\frac{2}{2}\right) \times \log_2 \left(\frac{2}{2}\right) = 0$$

$$\text{Entropy atribut "gagal" yang bernilai (1)} = \left(-\frac{2}{3}\right) \times \log_2 \left(\frac{2}{3}\right) + \left(-\frac{1}{3}\right) \times \log_2 \left(\frac{1}{3}\right) = 0.918$$

$$\text{Gain (total, gagal)} = (2.341) - \left(\frac{2}{5}\right) \times (0) + \left(\frac{3}{5}\right) \times (0.918) = 1.790$$

d) Resmi

$$\text{Entropy atribut "resmi" yang bernilai (0)} = \left(-\frac{2}{4}\right) \times \log_2 \left(\frac{2}{4}\right) + \left(-\frac{1}{4}\right) \times \log_2 \left(\frac{1}{4}\right) + \left(-\frac{1}{4}\right) \times \log_2 \left(\frac{1}{4}\right) = 1.5$$

$$\text{Entropy atribut "resmi" yang bernilai (1)} = \left(-\frac{1}{1}\right) \times \log_2 \left(\frac{1}{1}\right) = 0$$

$$\text{Gain (total, resmi)} = (2.341) - \left(\frac{4}{5}\right) \times (1.5) + \left(\frac{1}{5}\right) \times (0) = 1.141$$

e) Buruk

$$\text{Entropy atribut "buruk" yang bernilai (0)} = \left(-\frac{2}{4}\right) \times \log_2 \left(\frac{2}{4}\right) + \left(-\frac{1}{4}\right) \times \log_2 \left(\frac{1}{4}\right) + \left(-\frac{1}{4}\right) \times \log_2 \left(\frac{1}{4}\right) = 1.5$$

$$\text{Entropy atribut "buruk" yang bernilai (1)} = \left(-\frac{1}{1}\right) \times \log_2 \left(\frac{1}{1}\right) = 0$$

$$\text{Gain (total, buruk)} = (2.341) - \left(\frac{4}{5}\right) \times (1.5) + \left(\frac{1}{5}\right) \times (0) = 1.141$$

f) Prestasi

$$\text{Entropy atribut "prestasi" yang bernilai (0)} = \left(-\frac{2}{4}\right) \times \log_2 \left(\frac{2}{4}\right) + \left(-\frac{2}{4}\right) \times \log_2 \left(\frac{2}{4}\right) = 1$$

$$\text{Entropy atribut "prestasi" yang bernilai (1)} = \left(-\frac{1}{1}\right) \times \log_2 \left(\frac{1}{1}\right) = 0$$

$$\text{Gain (total, gagal)} = (2.341) - \left(\frac{4}{5}\right) \times (1) + \left(\frac{1}{5}\right) \times (0) = 1.541$$

g) Peserta

$$\text{Entropy atribut "peserta" yang bernilai (0)} = \left(-\frac{2}{4}\right) \times \log_2 \left(\frac{2}{4}\right) + \left(-\frac{2}{4}\right) \times \log_2 \left(\frac{2}{4}\right) = 1$$

$$\text{Entropy atribut "peserta" yang bernilai (1)} = \left(-\frac{1}{1}\right) \times \log_2 \left(\frac{1}{1}\right) = 0$$

$$\text{Gain (total, gagal)} = (2.341) - \left(\frac{4}{5}\right) \times (1) + \left(\frac{1}{5}\right) \times (0) = 1.541$$

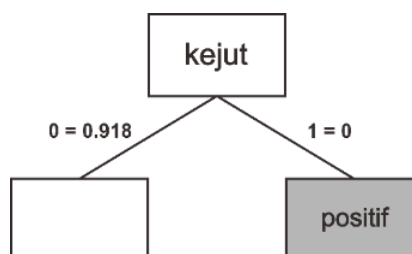
Perhitungan *gain* untuk membuat akar pohon keputusan ada pada tabel 3.17. Gain pertama ini adalah yang menjadi *root* pada algoritma pohon keputusan C4.5. Kata yang sudah ditetapkan menjadi *root* adalah sebagai kata yang paling dominan dalam seluruh dokumen.

Tabel 3.20: Perhitungan *Gain* untuk Akar Pohon Keputusan

Atribut	Neg(0)	Pos(0)	Net(0)	<i>Entropy</i>	Neg(1)	Pos(1)	Net(1)	<i>Entropy</i>	<i>Gain</i>
kejut	2	1	0	0.918	0	1	1	0	1.790
sukses	2	1	1	1.5	0	1	0	0	0.892
gagal	0	2	0	0	2	0	1	0.918	1.790
resmi	1	2	1	1.5	1	0	0	0	0.892
buruk	1	2	1	1.5	1	0	0	0	0.892
prestasi	2	2	0	1	0	0	1	0	1.292
peserta	2	2	0	1	0	0	1	0	1.292
Gain tertinggi : 1.790									
Root : kejut = positif									

(Sumber: Hasil olah data peneliti, 2023)

Hasil dari perhitungan nilai *gain* memberikan pola pohon seperti pada gambar 3.2 yang terdapat kata beserta bobot dan kelas yang diputuskan oleh sistem.

**Gambar 3.2: Pola Root**

(Sumber: Hasil olah data peneliti, 2023)

Perhitungan akan berulang untuk menentukan cabang selanjutnya, karena nilai *entropy* dari atribut “kejut” sebagai *root* dengan nilai 0 belum sama dengan 0 atau masih bernilai 0.918. Pada perhitungan selanjutnya, dokumen yang hanya terdapat satu atribut klasifikasi dan atribut itu adalah “kejut”, maka dokumen tersebut tereliminasi dari perhitungan karena telah terprediksi hasilnya. Namun, atribut “kejut” tidak tereliminasi karena masih ada dokumen yang mengandung atribut tersebut. Seperti pada tabel 3.18 yang telah meniadakan dokumen yang berkode D1.

Tabel 3.21: Nilai Data Latih Kedua

KD	Atribut							Kelas
	kejut	sukses	gagal	resmi	buruk	prestasi	peserta	
D2	1	1	0	0	0	0	0	pos
D3	0	0	1	0	0	0	0	neg
D4	0	0	1	1	1	0	0	neg
D5	0	0	1	0	0	1	1	net

(Sumber: Hasil olah data peneliti, 2023)

Perhitungan *entropy* untuk keseluruhan data latih:

Entropy total =

$$\left(-\frac{2}{4}\right) \times \log_2 \left(\frac{2}{4}\right) + \left(-\frac{1}{4}\right) \times \log_2 \left(\frac{1}{4}\right) + \left(-\frac{3}{4}\right) \times \log_2 \left(\frac{3}{4}\right) + \left(-\frac{3}{4}\right) \times \log_2 \left(\frac{3}{4}\right) = 1.622$$

Perhitungan *entropy* dan gain pada setiap atribut untuk menentukan cabang dari akar:

a) Kejut

$$\text{Entropy atribut "kejut" yang bernilai (0)} = \left(-\frac{1}{3}\right) \times \log_2 \left(\frac{1}{3}\right) + \left(-\frac{2}{3}\right) \times \log_2 \left(\frac{2}{3}\right) = 0.918$$

$$\text{Entropy atribut "kejut" yang bernilai (1)} = \left(-\frac{1}{1}\right) \times \log_2 \left(\frac{1}{1}\right) = 0$$

$$\text{Gain (total, kejut)} = (1.622) - \left(\frac{3}{4}\right) \times (0.918) + \left(\frac{1}{4}\right) \times (0) = 0.933$$

b) Sukses

$$\text{Entropy atribut "sukses" yang bernilai (0)} = \left(-\frac{2}{3}\right) \times \log_2 \left(\frac{2}{3}\right) + \left(-\frac{1}{3}\right) \times \log_2 \left(\frac{1}{3}\right) = 0.918$$

$$\text{Entropy atribut "sukses" yang bernilai (1)} = \left(-\frac{1}{1}\right) \times \log_2 \left(\frac{1}{1}\right) = 0$$

$$\text{Gain (total, sukses)} = (1.622) - \left(\frac{3}{4}\right) \times (0.918) + \left(\frac{1}{4}\right) \times (0) = 0.933$$

c) Gagal

$$\text{Entropy atribut "gagal" yang bernilai (0)} = \left(-\frac{1}{1}\right) \times \log_2 \left(\frac{1}{1}\right) = 0$$

$$\text{Entropy atribut "gagal" yang bernilai (1)} = \left(-\frac{2}{3}\right) \times \log_2 \left(\frac{2}{3}\right) + \left(-\frac{1}{3}\right) \times \log_2 \left(\frac{1}{3}\right) = 0.918$$

$$\text{Gain (total, sukses)} = (1.622) - \left(\frac{3}{4}\right) \times (0.918) + \left(\frac{1}{4}\right) \times (0) = 0.933$$

d) Resmi

$$\text{Entropy atribut "resmi" yang bernilai (0)} = \left(-\frac{1}{3}\right) \times \log_2 \left(\frac{1}{3}\right) + \left(-\frac{1}{3}\right) \times \log_2 \left(\frac{1}{3}\right) + \left(-\frac{1}{3}\right) \times \log_2 \left(\frac{1}{3}\right) = 1.584$$

$$\text{Entropy atribut "resmi" yang bernilai (1)} = \left(-\frac{1}{1}\right) \times \log_2 \left(\frac{1}{1}\right) = 0$$

$$\text{Gain (total, resmi)} = (1.622) - \left(\frac{3}{4}\right) \times (1.584) + \left(\frac{1}{4}\right) \times (0) = 0.434$$

e) Buruk

$$\text{Entropy atribut "resmi" yang bernilai (0)} = \left(-\frac{1}{3}\right) \times \log_2 \left(\frac{1}{3}\right) + \left(-\frac{1}{3}\right) \times \log_2 \left(\frac{1}{3}\right) + \left(-\frac{1}{3}\right) \times \log_2 \left(\frac{1}{3}\right) = 1.584$$

$$\text{Entropy atribut "resmi" yang bernilai (1)} = \left(-\frac{1}{1}\right) \times \log_2 \left(\frac{1}{1}\right) = 0$$

$$\text{Gain (total, resmi)} = (1.622) - \left(\frac{3}{4}\right) \times (1.584) + \left(\frac{1}{4}\right) \times (0) = 0.434$$

f) Prestasi

$$\text{Entropy atribut "prestasi" yang bernilai (0)} = \left(-\frac{1}{3}\right) \times \log_2 \left(\frac{1}{3}\right) + \left(-\frac{2}{3}\right) \times \log_2 \left(\frac{2}{3}\right) = 0.918$$

$$\text{Entropy atribut "prestasi" yang bernilai (1)} = \left(-\frac{1}{1}\right) \times \log_2 \left(\frac{1}{1}\right) = 0$$

$$\text{Gain (total, gagal)} = (1.622) - \left(\frac{3}{4}\right) \times (0.918) + \left(\frac{1}{4}\right) \times (0) = 0.933$$

g) Peserta

$$\text{Entropy atribut "prestasi" yang bernilai (0)} = \left(-\frac{1}{3}\right) \times \log_2 \left(\frac{1}{3}\right) + \left(-\frac{2}{3}\right) \times \log_2 \left(\frac{2}{3}\right) = 0.918$$

$$\text{Entropy atribut "prestasi" yang bernilai (1)} = \left(-\frac{1}{1}\right) \times \log_2 \left(\frac{1}{1}\right) = 0$$

$$\text{Gain (total, gagal)} = (1.622) - \left(\frac{3}{4}\right) \times (0.918) + \left(\frac{1}{4}\right) \times (0) = 0.933$$

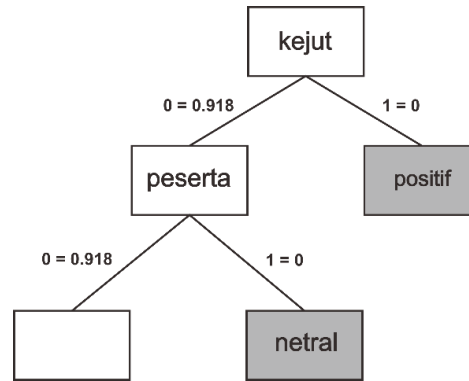
Perhitungan *gain* untuk membuat cabang pertama pohon keputusan ada pada tabel 3.19.

Tabel 3.22: Perhitungan Gain untuk Internal Node Pertama

Atribut	Neg(0)	Pos(0)	Net(0)	Entropy	Neg(1)	Pos(1)	Net(1)	Entropy	Gain
kejut	2	0	1	0.918	0	1	0	0	0.933
sukses	2	0	1	0.918	0	1	0	0	0.933
gagal	0	1	0	0	2	0	1	0.918	0.933
resmi	1	1	1	1.584	1	0	0	0	0.434
buruk	1	1	1	1.584	1	0	0	0	0.434
prestasi	2	1	0	0.918	0	0	1	0	0.933
peserta	2	1	0	0.918	0	0	1	0	0.933
Gain tertinggi : 0.933									
Internal node pertama : peserta = netral									

(Sumber: Hasil olah data peneliti, 2023)

Hasil dari perhitungan nilai *gain* memberikan pola cabang pohon seperti pada gambar 3.3.



Gambar 3.3: Pola Internal Node Pertama Sebagai Cabang dari Root
(Sumber: Hasil olah data peneliti, 2023)

Perhitungan akan berulang untuk menentukan cabang selanjutnya, karena nilai *entropy* dari atribut “peserta” sebagai *internal node* dengan nilai 0 belum sama dengan 0 atau masih bernilai 0.918. Pada perhitungan selanjutnya, atribut “peserta” tereliminasi karena sudah tidak ada dalam dokumen manapun. Atribut “peserta” terprediksi netral karena memiliki *entropy* 0 terhadap nilai 1. Seperti pada tabel 3.20 yang meniadakan atribut “peserta” dan tidak meniadakan satu dokumen sama sekali.

Tabel 3.23: Nilai Data Latih Ketiga

KD	Atribut						Kelas
	kejut	sukses	gagal	resmi	buruk	prestasi	
D2	1	1	0	0	0	0	pos
D3	0	0	1	0	0	0	neg
D4	0	0	1	1	1	0	neg
D5	0	0	1	0	0	1	net

(Sumber: Hasil olah data peneliti, 2023)

Perhitungan *entropy* untuk keseluruhan data latih:

Entropy total =

$$\left(-\frac{2}{4}\right) \times \log_2 \left(\frac{2}{4}\right) + \left(-\frac{1}{4}\right) \times \log_2 \left(\frac{1}{4}\right) + \left(-\frac{3}{4}\right) \times \log_2 \left(\frac{3}{4}\right) + \left(-\frac{2}{4}\right) \times \log_2 \left(\frac{2}{4}\right) = 1.811$$

Perhitungan *entropy* dan gain pada setiap atribut untuk menentukan cabang dari cabang akar:

a) Kejut

$$\begin{aligned} \text{Entropy atribut "kejut" yang bernilai (0)} &= \left(-\frac{2}{3}\right) \times \log_2 \left(\frac{2}{3}\right) + \\ &\left(-\frac{1}{3}\right) \times \log_2 \left(\frac{1}{3}\right) = 0.918 \end{aligned}$$

$$\text{Entropy atribut "kejut" yang bernilai (1)} = \left(-\frac{1}{1}\right) \times \log_2 \left(\frac{1}{1}\right) = 0$$

$$\text{Gain (total, sukses)} = (1.811) - \left(\frac{3}{4}\right) \times (0.918) + (0) = 1.122$$

b) Sukses

$$\text{Entropy atribut "sukses" yang bernilai (0)} = \left(-\frac{2}{3}\right) \times \log_2 \left(\frac{2}{3}\right) + \left(-\frac{1}{3}\right) \times \log_2 \left(\frac{1}{3}\right) = 0.918$$

$$\text{Entropy atribut "sukses" yang bernilai (1)} = \left(-\frac{1}{1}\right) \times \log_2 \left(\frac{1}{1}\right) = 0$$

$$\text{Gain (total, sukses)} = (1.811) - \left(\frac{3}{4}\right) \times (0.918) + (0) = 1.122$$

c) Gagal

$$\text{Entropy atribut "gagal" yang bernilai (0)} = \left(-\frac{1}{1}\right) \times \log_2 \left(\frac{1}{1}\right) = 0$$

$$\text{Entropy atribut "gagal" yang bernilai (1)} = \left(-\frac{2}{3}\right) \times \log_2 \left(\frac{2}{3}\right) + \left(-\frac{1}{3}\right) \times \log_2 \left(\frac{1}{3}\right) = 0.918$$

$$\text{Gain (total, sukses)} = (1.811) - \left(\frac{3}{4}\right) \times (0.918) + (0) = 1.122$$

d) Resmi

$$\text{Entropy atribut "resmi" yang bernilai (0)} = \left(-\frac{1}{3}\right) \times \log_2 \left(\frac{1}{3}\right) + \left(-\frac{1}{3}\right) \times \log_2 \left(\frac{1}{3}\right) + \left(-\frac{1}{3}\right) \times \log_2 \left(\frac{1}{3}\right) = 1.584$$

$$\text{Entropy atribut "resmi" yang bernilai (1)} = \left(-\frac{1}{1}\right) \times \log_2 \left(\frac{1}{1}\right) = 0$$

$$\text{Gain (total, resmi)} = (1.811) - \left(\frac{3}{4}\right) \times (1.584) + \left(\frac{1}{4}\right) \times (0) = 0.623$$

e) Buruk

$$\text{Entropy atribut "resmi" yang bernilai (0)} = \left(-\frac{1}{3}\right) \times \log_2 \left(\frac{1}{3}\right) + \left(-\frac{1}{3}\right) \times \log_2 \left(\frac{1}{3}\right) + \left(-\frac{1}{3}\right) \times \log_2 \left(\frac{1}{3}\right) = 1.584$$

$$\text{Entropy atribut "resmi" yang bernilai (1)} = \left(-\frac{1}{1}\right) \times \log_2 \left(\frac{1}{1}\right) = 0$$

$$\text{Gain (total, resmi)} = (1.811) - \left(\frac{3}{4}\right) \times (1.584) + \left(\frac{1}{4}\right) \times (0) = 0.623$$

f) Prestasi

$$\text{Entropy atribut "prestasi" yang bernilai (0)} = \left(-\frac{2}{3}\right) \times \log_2 \left(\frac{2}{3}\right) + \left(-\frac{1}{3}\right) \times \log_2 \left(\frac{1}{3}\right) = 0.918$$

$$\text{Entropy atribut "prestasi" yang bernilai } (1) = \left(-\frac{1}{1}\right) \times \log_2 \left(\frac{1}{1}\right) = 0$$

$$\text{Gain (total, resmi)} = (1.811) - \left(\frac{3}{4}\right) \times (0.918) + (0) = 1.122$$

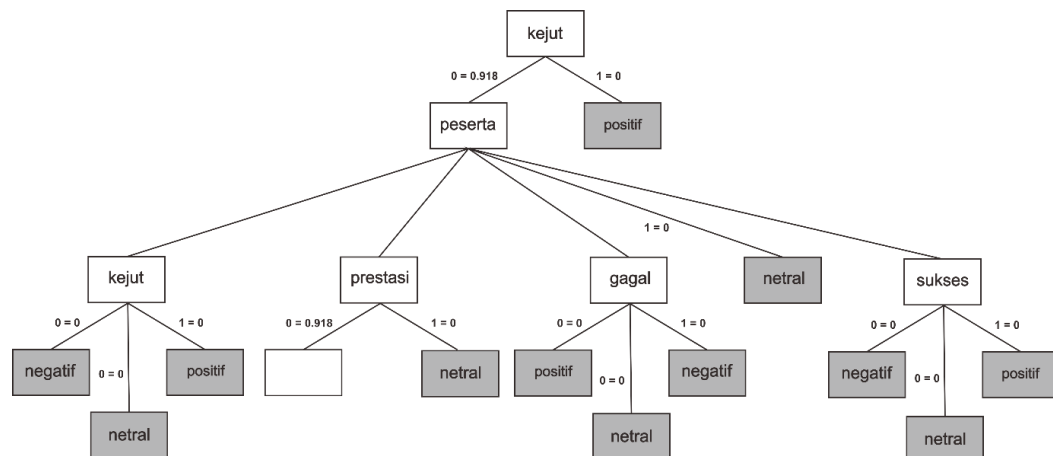
Perhitungan *gain* untuk membuat cabang kedua pohon keputusan ada pada tabel 3.21.

Tabel 3.24: Perhitungan Gain untuk Internal Node Kedua

Atribut	Neg(0)	Pos(0)	Net(0)	Entropy	Neg(1)	Pos(1)	Net(1)	Entropy	Gain
kejut	2	0	1	0	0	1	0	0	1.122
sukses	2	0	1	0	0	1	0	0	1.122
gagal	0	1	0	0	2	0	1	0	1.122
resmi	1	1	1	1.584	1	0	0	0	0.623
buruk	1	1	1	1.584	1	0	0	0	0.623
prestasi	2	1	0	0.918	0	0	1	0	1.122
Gain tertinggi : 1.122									
Internal node ke 2 : "prestasi = netral", "kejut = positif", "sukses = positif", "gagal negatif/netral"									

(Sumber: Hasil olah data peneliti, 2023)

Hasil dari perhitungan nilai *gain* memberikan pola cabang pohon seperti pada gambar 3.4.



Gambar 3.4: Pola Internal Node Kedua Sebagai Cabang dari Root

(Sumber: Hasil olah data peneliti, 2023)

Perhitungan akan dilanjutkan, karena nilai *entropy* dari atribut "prestasi" sebagai *internal node* kedua dengan nilai 0 belum sama dengan 0. Kata "kejut, sukses, dan prestasi" dieliminasi karena masing-masing telah menjadi internal node yang kedua dan memiliki nilai *entropy* 0. Oleh karena itu, dokumen yang mengandung kata tersebut juga dieliminasi karena sudah tidak ada kata terkait yang akan dihitung.

Tabel 3.25: Nilai Data Latih Keempat

KD	Atribut			Kelas
	gagal	resmi	buruk	
D4	0	1	1	neg
D5	1	0	0	net

(Sumber: Hasil olah data peneliti, 2023)

Perhitungan *entropy* untuk keseluruhan data latih:

Entropy total =

$$\left(-\frac{2}{2}\right) \times \log_2 \left(\frac{2}{2}\right) = 0$$

Perhitungan *entropy* dan gain pada setiap atribut untuk menentukan cabang dari cabang akar:

a) Gagal

$$\text{Entropy atribut "resmi" yang bernilai (0)} = \left(-\frac{1}{1}\right) \times \log_2 \left(\frac{1}{1}\right) = 0$$

$$\text{Entropy atribut "resmi" yang bernilai (1)} = \left(-\frac{1}{1}\right) \times \log_2 \left(\frac{1}{1}\right) = 0$$

$$\text{Gain (total, resmi)} = (0) - (0) = 0$$

b) Resmi

$$\text{Entropy atribut "resmi" yang bernilai (0)} = \left(-\frac{1}{1}\right) \times \log_2 \left(\frac{1}{1}\right) = 0$$

$$\text{Entropy atribut "resmi" yang bernilai (1)} = \left(-\frac{1}{1}\right) \times \log_2 \left(\frac{1}{1}\right) = 0$$

$$\text{Gain (total, resmi)} = (0) - (0) = 0$$

c) Buruk

$$\text{Entropy atribut "buruk" yang bernilai (0)} = \left(-\frac{1}{1}\right) \times \log_2 \left(\frac{1}{1}\right) = 0$$

$$\text{Entropy atribut "buruk" yang bernilai (1)} = \left(-\frac{1}{1}\right) \times \log_2 \left(\frac{1}{1}\right) = 0$$

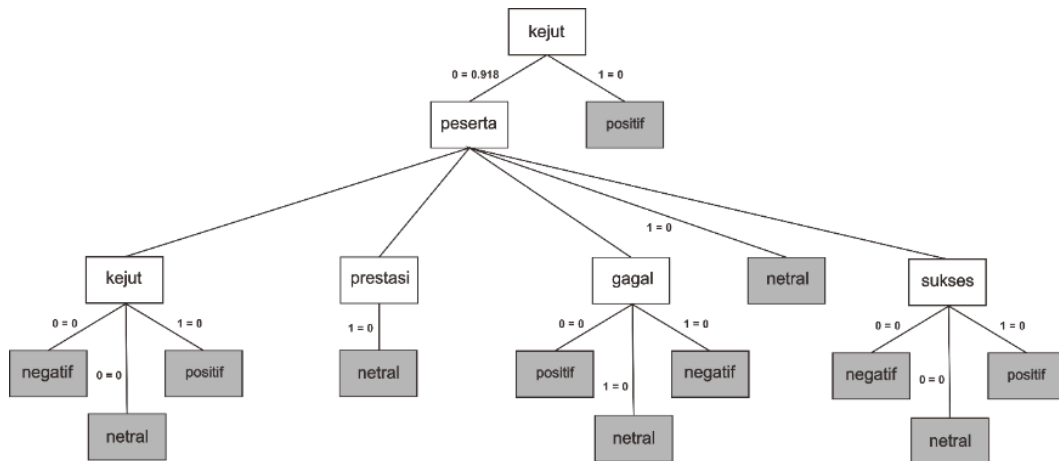
$$\text{Gain (total, resmi)} = (0) - (0) = 0$$

Tabel 3.26: Perhitungan Gain untuk Internal Node Ketiga

Atribut	Neg(0)	Pos(0)	Net(0)	Entropy	Neg(1)	Pos(1)	Net(1)	Entropy	Gain
gagal	1	0	0	0	0	0	1	0	0
resmi	0	0	1	0	1	0	0	0	0
buruk	0	0	1	0	1	0	0	0	0
Gain tertinggi : 0									
Internal node ke 3 : tidak ada									

(Sumber: Hasil olah data peneliti, 2023)

Hasil dari perhitungan nilai *gain* memberikan pola seperti pada gambar 3.5.



Gambar 3.5: Pola Pembelajaran C4.5
(Sumber: Hasil olah data peneliti, 2023)

Perhitungan telah terhenti pada *internal node* kedua, karena semua nilai *entropy* dari atribut telah sama dengan 0. Gambar 3.5 akan menjadi model pembelajaran bagi mesin untuk menentukan prediksi dari proses pengklasifikasian data uji.

3.2.7 Evaluasi Model Menggunakan *Confusion Matrix*

Evaluasi model dilakukan agar mengetahui performa baik dari model baik dari akurasi, *recall*, maupun presisi. Dalam penelitian ini, terdapat tiga baris dan tiga kolom yang memuat *true* dan *false* dari negatif, *true* dan *false* dari positif, dan *true* dan *false* dari netral. Hal tersebut seperti pada tabel 3.22.

Tabel 3.27: Contoh Evaluasi Model dengan *Confusion Matrix*

		Prediksi		
		Negatif	Positif	Netral
Aktual	Negatif	TNeg (200)	FPos (10)	FNet (10)
	Positif	FNeg (5)	TPos (350)	FNet (30)
	Netral	FNeg (10)	FPos (5)	TNet (100)

(Sumber: Hasil olah data peneliti, 2023)

Keterangan:

- TNeg : *True* Negatif
- FNeg : *False* Negatif
- TPos : *True* Positif

- FPos : *False* Positif
- TNet : *True* Netral
- FNet : *False* Netral

3.2.7.1 Akurasi

Akurasi yaitu menghitung kemiripan antara data aktual dengan hasil prediksi. Hal tersebut dihitung dengan rumus sebagai berikut:

$$\frac{TPos+TNeg+TNet}{TPos+TNeg+TNet+FPos+FNeg+FNet} \quad (3.1)$$

$$\frac{350 + 200 + 100}{350 + 200 + 100 + 15 + 15 + 40}$$

$$\frac{650}{720} = 0.90$$

3.2.7.2 Recall

Recall menghitung tingkat keberhasilan temu kembali informasi yang dilakukan oleh sistem. Pada penelitian ini terdapat tiga kelas yaitu positif, negatif, dan netral. Oleh karena itu, masing-masing kelas perlu dihitung tingkat keberhasilan temu kembali informasinya. Adapun perhitungan *recall* sebagai berikut:

a) *Recall* positif

$$\frac{TPos}{TPos+FNeg+FNet} \quad (3.2)$$

$$\frac{350}{350 + 15 + 40} = \frac{350}{405} = 0.86$$

b) *Recall* negatif

$$\frac{TNeg}{TNeg+FPos+FNet} \quad (3.3)$$

$$\frac{200}{200 + 15 + 40} = \frac{200}{255} = 0.78$$

c) *Recall* netral

$$\frac{TNet}{TNet+FPos+FNeg} \quad (3.4)$$

$$\frac{100}{100 + 15 + 15} = \frac{100}{130} = 0.76$$

3.2.7.3 Presisi

Presisi yaitu menghitung tingkat kemiripan data yang diambil oleh sistem. Perhitungan presisi dilakukan terhadap masing-masing kelas. Adapun perhitungan presisi sebagai berikut:

- a) Presisi positif

$$\frac{TPos}{TPos + FPos} \quad (3.5)$$

$$\frac{350}{350 + 15} = 0.95$$

- b) Presisi negatif

$$\frac{TNeg}{TNeg + FNeg} \quad (3.6)$$

$$\frac{200}{200 + 15} = 0.93$$

- c) Presisi netral

$$\frac{TNet}{TNet + FNet} \quad (3.7)$$

$$\frac{100}{100 + 40} = 0.71$$

3.2.7.4 F1-Score

F1-Score adalah mencari nilai gabungan atau rata-rata dari recall dan presisi. Adapun perhitungan F1-Score sebagai berikut:

$$\frac{2(Presisi)(Recall)}{(Presisi + Recall)} \quad (3.8)$$

- a) F1-Score positif

$$\frac{2(0.95)(0.86)}{(1,81)}$$

$$\frac{2(0,81)}{(1,81)} = 0.89$$

b) F1-Score negatif

$$\frac{2(0.93)(0.78)}{(1,71)}$$

$$\frac{2(0,72)}{(1,71)} = 0.84$$

c) F1-Score netral

$$\frac{2(0.71)(0.76)}{(1.47)}$$

$$\frac{2(0.53)}{(1.47)} = 0.72$$

3.3 Sumber Data

Data yang digunakan dalam penelitian ini adalah kumpulan *tweet*, *quote tweet*, dan *reply* yang terkait topik Qatar sebagai tuan rumah piala dunia FIFA 2022 di media sosial Twitter. Himpunan data yang diambil adalah *tweet* mulai dari 2010 hingga 2022, karena penunjukkan Qatar sebagai tuan rumah piala dunia adalah pada tanggal 9 Desember 2010 dan pelaksanaan pada 21 November 2022 sampai 18 Desember 2022 (*Profile of the FIFA World Cup Qatar 2022, 2022*). *Tweet* yang diperoleh adalah sebanyak 2925.

3.4 Pengumpulan Data

Penelitian ini menggunakan metode *scraping* untuk temu kembali informasi sekaligus pengumpulan data. Untuk memperoleh data tersebut, terdapat beberapa alat seperti Jupyter Notebook dan Snsrape. Adapun tampilan *source code* untuk pengambilan data pada tabel 3.24.

Tabel 3.28: Scraper Source Code

Line	Source Code
1	<code>import snsrape.modules.twitter as sntwitter</code>
2	<code>import pandas as pd</code>
3	<code>pd.options.display.max_colwidth = 500</code>
4	<code>query = 'qatar sebagai tuan rumah lang:id until:2022-12-18 since:2010-12-9'</code>
5	<code>tweets = []</code>
6	<code>limit = 10000</code>
7	<code>try:</code>
8	<code>for tweet in sntwitter.twittersearchscraper(query=query).get_items():</code>
9	<code>if len(tweets) == limit:</code>
10	<code>break</code>
11	<code>else:</code>
12	<code>tweets.append([tweet.date, tweet.content])</code>
13	<code>df = pd.dataframe(tweets, columns=['datetime', 'content'])</code>
14	<code>except exception as e:</code>
15	<code>print(e)</code>

(Sumber: Hasil olah data peneliti, 2023)

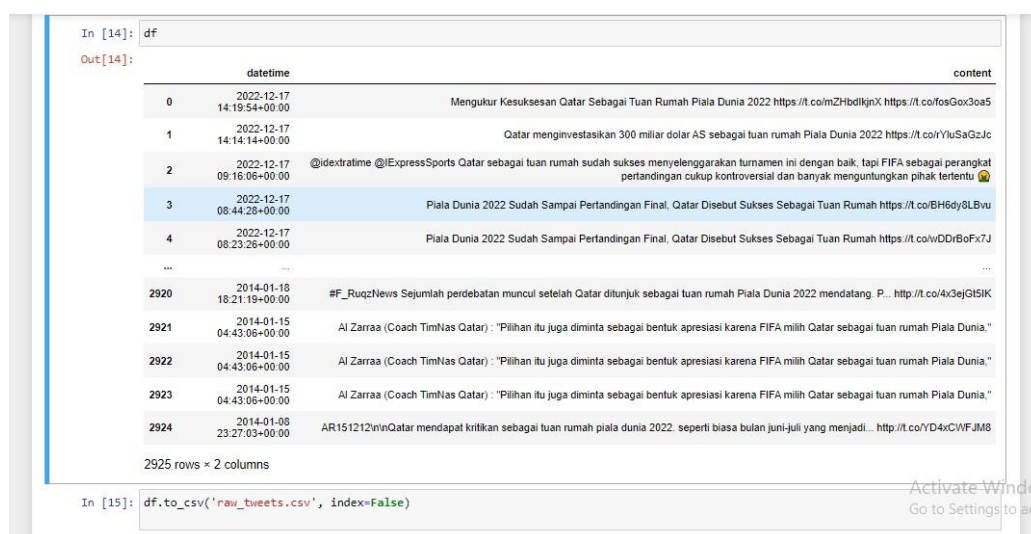
Source code di atas dapat dijelaskan, yaitu masukkan modul *scraper* milik Snsrape yaitu `snsrape.modules.twitter` dan inisialisasi sebagai `sntwitter`. Setelah itu, masukkan modul `pandas` untuk visualisasi dan inisialisasi sebagai `pd`. Selanjutnya, mengatur ukuran kolom yang akan ditampilkan `pandas` dengan `pd.options.display.max_colwidth = 500`. Penampilan baris data tidak dimaksimalkan sejumlah data yang diperoleh, karena sudah terwakili dengan *default* sepuluh baris data yang tampil. Kemudian, buat variabel `query` untuk mempersingkat kode pemanggilan saat digunakan untuk *information retrieval*, seperti `query = 'qatar sebagai tuan rumah lang:id until:2022-12-18 since:2010-12-9'`. Setelah itu, membuat variabel `tweets` dengan *list* kosong. Selanjutnya, membuat variabel `limit` untuk membatasi jumlah data yang akan diambil yaitu 10000.

Kemudian membuat perulangan, dengan `tweet` bernilai `true` yang ada di `twittersearchscraper(query=query).get_items()`: dalam modul `sntwitter`, jika jumlah item sudah terpenuhi sama dengan `limit` maka *break*. Jika tidak terdapat kesalahan dalam kode, maka *insert* `tweet.date` dan `tweet.content` dalam *array*. Selanjutnya buat variabel `df` untuk menampilkan data yang telah dipanggil dengan dua kolom yang berisi `datetime` dan `content`. Selanjutnya membuat `except` untuk memberikan

validasi jika ada kesalahan. Jika semua kondisi terpenuhi, maka data akan ditampilkan dengan kode print.

Library Snscape ini mengambil data berupa kalimat opini yang sesuai dengan *filter* tanpa menggunakan API, namun pada akun twitter yang *private* tidak akan ikut teridentifikasi oleh *library* Snscape. Terdapat banyak jenis akun pengguna yang berkomentar mengenai Qatar sebagai tuan rumah piala dunia FIFA 2022, seperti akun *platform* berita, akun perorangan, dan akun *menfess* atau *base*. Berbagai opini pengguna Twitter masing-masing dapat terlihat keberagaman ekspresi melalui sebuah teks atau kalimat. Susunan variasi kata yang beragam ini nanti yang akan dipilah mana yang termasuk opini positif, opini negatif, maupun opini netral menggunakan algoritma C4.5.

Hasil dari *scraping* dari media sosial Twitter menunjukkan jumlah *tweet* yang didapat, tanggal waktu *upload tweet*, dan konten atau *tweet* yang terkait ada pada gambar 3.5.



```
In [14]: df
Out[14]:
```

	datetime	content
0	2022-12-17 14:19:54+00:00	Mengukur Kesuksesan Qatar Sebagai Tuan Rumah Piala Dunia 2022 https://t.co/mZHbdIknX https://t.co/fosGox3oa5
1	2022-12-17 14:14:14+00:00	Qatar menginvestasikan 300 miliar dolar AS sebagai tuan rumah Piala Dunia 2022 https://t.co/YyU5aGzJc
2	2022-12-17 09:16:06+00:00	@idextralme @iExpressSports Qatar sebagai tuan rumah sudah sukses menyelenggarakan turnamen ini dengan baik, tapi FIFA sebagai perangkat pertandingan cukup kontroversial dan banyak menguntungkan pihak tertentu 🤔
3	2022-12-17 08:44:28+00:00	Piala Dunia 2022 Sudah Sampai Pertandingan Final. Qatar Disebut Sukses Sebagai Tuan Rumah https://t.co/BH6dy9LBvu
4	2022-12-17 08:23:26+00:00	Piala Dunia 2022 Sudah Sampai Pertandingan Final. Qatar Disebut Sukses Sebagai Tuan Rumah https://t.co/vDDrBoF7J
...
2920	2014-01-18 18:21:19+00:00	#F_RuqzNews Sejumlah perdebatan muncul setelah Qatar ditunjuk sebagai tuan rumah Piala Dunia 2022 mendatang P.. http://t.co/4x3eJGISIK
2921	2014-01-15 04:43:06+00:00	Al Zarraa (Coach TimNas Qatar) : "Pilihan itu juga diminta sebagai bentuk apresiasi karena FIFA pilih Qatar sebagai tuan rumah Piala Dunia."
2922	2014-01-15 04:43:06+00:00	Al Zarraa (Coach TimNas Qatar) : "Pilihan itu juga diminta sebagai bentuk apresiasi karena FIFA pilih Qatar sebagai tuan rumah Piala Dunia."
2923	2014-01-15 04:43:06+00:00	Al Zarraa (Coach TimNas Qatar) : "Pilihan itu juga diminta sebagai bentuk apresiasi karena FIFA pilih Qatar sebagai tuan rumah Piala Dunia."
2924	2014-01-08 23:27:03+00:00	AR151212w/nQatar mendapat kritikan sebagai tuan rumah piala dunia 2022. seperti biasa bulan juni-juli yang menjadi... http://t.co/YD4xQWfJM8

2925 rows x 2 columns

```
In [15]: df.to_csv('raw_tweets.csv', index=False)
```

Gambar 3.6: Dataset Hasil Scraping
(Sumber: Hasil olah data peneliti, 2023)

Gambar 3.5 menampilkan tiga kolom yang berisi nomor data, waktu dan tanggal *tweet* terpublikasi, serta teks atau kalimat dari *tweet*. Data dengan jumlah 2925 *tweet* ditampilkan dengan sepuluh baris yang mewakili, namun data tersebut telah dikonversi menjadi dokumen berformat .CSV, sehingga dapat dilihat keseluruhan data yang diperoleh. Setelah itu, data yang sudah terkumpul ini dikonversikan menjadi format .xlsx supaya memudahkan proses pelabelan data oleh 2 orang ahli.

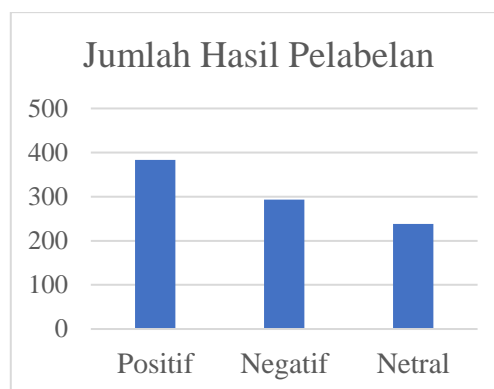
BAB IV HASIL DAN PEMBAHASAN

4.1 Hasil Penelitian

Pada bagian ini, memaparkan hasil dari pengujian *machine learning* sekaligus pengulasan terkait implementasi metode C4.5 untuk mengetahui sentimen masyarakat Indonesia mengenai Qatar sebagai tuan rumah piala dunia 2022. Sebelum melakukan pengujian, terdapat tahap *pre-processing* dan pembagian data menjadi data latih dan data uji yang memakai dua rasio, yaitu 70:30 dan 80:20. Pengujian pada penelitian ini dilakukan untuk menghitung nilai akurasi, *recall*, presisi, dan *F1-Score*. Berikut adalah hasil dari penelitian yang dilakukan.

4.1.1 Hasil Pelabelan Data

Pada penelitian ini, menggunakan dataset yang jumlah awal 2925 *tweet* menjadi 914 *tweet* karena terdapat banyak duplikasi data sehingga perlu adanya penghapusan data yang terduplikasi. Data dari hasil pembersihan tersebut diberi label oleh ahli bahasa Indonesia dan ahli ilmu sosial dengan memutuskan jenis kelas sentimen yaitu positif, negatif, atau netral. Hasil dari pelabelan dapat dikelompokkan dan dihitung jumlah masing-masing kelas seperti diagram pada gambar 4.1.



Gambar 4.1: Hasil Pelabelan
(Sumber: Hasil olah data peneliti, 2023)

Gambar 4.1 memberikan visualisasi berupa diagram batang yang mana garis horizontal menunjukkan jenis kelas sentimen, sedangkan garis vertikal menunjukkan jumlah pada masing-masing kelas sentimen. Kelas sentimen positif memiliki jumlah 383, kelas sentimen negatif memiliki jumlah 293, dan kelas sentimen netral berjumlah 238.

4.1.2 Hasil *Pre-processing*

Data yang telah dilabeli oleh ahli akan dibersihkan untuk memaksimalkan proses klasifikasi. Pada proses ini menggunakan *regular expression* (regex) untuk mencari dan mengganti partikel yang tidak diperlukan, seperti URL, *emoticon*, *hashtag*, dan *username/mention* akan diganti dengan *string* kosong. Selain itu, juga memanfaatkan *library* Sastrawi untuk proses *stopwords* dan *stemming*. Adapun pada proses tokenizing memanfaatkan *library* NLTK dan pembuatan dataframe menggunakan *pandas*. Hasil *cleaning*, *stopwords*, *stemming*, dan *tokenizing* sebagai berikut.

Tabel 4.1: Sampel Hasil *Pre-Processing* Tahap *Cleaning*

No	Cleaning
1	qatar menginvestasikan miliar dolar as sebagai tuan rumah piala dunia
2	qatar sebagai tuan rumah sudah sukses menyelenggarakan turnamen ini dengan baik tapi fifa sebagai perangkat pertandingan cukup kontroversial dan banyak menguntungkan pihak tertentu
3	piala dunia sudah sampai pertandingan final qatar disebut sukses sebagai tuan rumah
4	qatar memanfaatkan posisi sebagai tuan rumah piala dunia dan kehebatan ekonominya untuk memaksimalkan pengaruhnya baik dalam politik regional maupun di panggung global baca trias kuncahyono di opini Kompas
5	prancis pasti ada di final pemilihan qatar sebagai tuan rumah pildun gk lepas dari kontribusi prancis hubungan ekonomi bilateral kedua negara juga sangat kuat
6	sebagai tuan rumah dimana ini merupakan kali pertama piala dunia digelar di tanah arab pemerintah qatar pun habis habisan dalam mempersiapkannya mewujudkan kemegahan acara dana sebesar usd miliar pun digelontorkan
7	bertemu teman lama chief investment officer qatar investment authority untuk asia sheikh faisal thani al thani kerja sama di bidang olahraga dengan qatar sebagai tuan rumah piala dunia akan bermanfaat bagi kita dalam menggelar piala dunia u tahun depan
8	seri paduka menzahirkan ucapan tahniah kepada sheikh tamim dan kerajaan qatar sebagai tuan rumah atas kejayaan menganjur kejohanan bola sepak piala dunia fifa qatar
9	tahun ini gue dukung tim dari negara maroco gue optimis tim jagoan gue akan jadi juara karena semangat qatar sebagai tuan rumah merasuk ke jiwa pemain qatar menang kalah pilih main bersih
10	qatar dapat pujian yang meluas sebagai tuan rumah karena banyak hal yang baik dan berbeda dari piala dunia sebelumnya wajar jika banyak yg mencari kesalahannya apalagi salah satu yg menonjol amp keras melarang bendera pelangi

(Sumber: Hasil olah data peneliti, 2023)

Pada tabel 4.1 menampilkan hasil penghapusan URL, *emoticon*, *hashtag*, *username/mention*, *whitespace*, serta menyalurkan semua huruf menjadi *lowercase*. Setelah itu, melakukan *stopwords* dengan hasil yang ada pada tabel 4.2.

Tabel 4.2: Sampel Hasil Pre-Processing Tahap Stopwords

No	Stopwords
1	qatar menginvestasikan miliar dolar as tuan rumah piala dunia
2	qatar tuan rumah sukses menyelenggarakan turnamen fifa perangkat pertandingan kontroversial menguntungkan
3	piala dunia pertandingan final qatar sukses tuan rumah
4	qatar memanfaatkan posisi tuan rumah piala dunia kehebatan ekonominya memaksimalkan pengaruhnya politik regional panggung global baca trias kuncahyono opini kompas
5	prancis final pemilihan qatar tuan rumah pildun gk lepas kontribusi prancis hubungan ekonomi bilateral negara kuat
6	tuan rumah dimana piala dunia digelar tanah arab pemerintah qatar habis habisan mempersiapkannya mewujudkan kemegahan acara dana usd miliar digelontorkan
7	bertemu teman chief investment officer qatar investment authority asia sheikh faisal thani al thani bidang olahraga qatar tuan rumah piala dunia bermanfaat menggelar piala dunia tahun
8	seri paduka menzahirkan ucapan tahniah sheikh tamim kerajaan qatar tuan rumah kejayaan menganjur kejohanan bola sepak piala dunia fifa qatar
9	tahun gue dukung tim negara maroco gue optimis tim jagoan gue juara semangat qatar tuan rumah merasuk jiwa pemain qatar menang kalah pilih main bersih
10	qatar pujian meluas tuan rumah berbeda piala dunia wajar yg mencari kesalahannya salah yg menonjol amp keras melarang bendera pelangi

(Sumber: Hasil olah data peneliti, 2023)

Pada tabel 4.2 menampilkan hasil penghapusan kata yang tidak diperlukan. Pada proses ini, terdapat beberapa stopwords yang tidak ikut terhapus karena berbentuk singkatan dan kata slang. Setelah itu, melakukan stemming dengan hasil yang ada pada tabel 4.3.

Tabel 4.3: Sampel Hasil Pre-Processing Tahap Stemming

No	Stemmed
1	qatar investasi miliar dolar as tuan rumah piala dunia
2	qatar tuan rumah sukses selenggara turnamen fifa perangkat tanding kontroversial untung
3	piala dunia tanding final qatar sukses tuan rumah
4	qatar manfaat posisi tuan rumah piala dunia hebat ekonomi maksimal pengaruh politik regional panggung global baca trias kuncahyono opini Kompas
5	prancis final pilih qatar tuan rumah pildun gk lepas kontribusi prancis hubung ekonomi bilateral negara kuat
6	tuan rumah mana piala dunia gelar tanah arab perintah qatar habis habis siap wujud megah acara dana usd miliar gelontor
7	temu teman chief investment officer qatar investment authority asia sheikh faisal thani al thani bidang olahraga qatar tuan rumah piala dunia manfaat gelar piala dunia tahun
8	seri paduka menzahirkan ucap tahniah sheikh tamim raja qatar tuan rumah jaya anjur johan bola sepak piala dunia fifa qatar
9	tahun gue dukung tim negara maroco gue optimis tim jago gue juara semangat qatar tuan rumah rasuk jiwa main qatar menang kalah pilih main bersih
10	qatar puji luas tuan rumah beda piala dunia wajar yg cari salah salah yg tonjol amp keras larang bendera pelangi

(Sumber: Hasil olah data peneliti, 2023)

Pada tabel 4.3 menampilkan hasil pengembalian kepada kata dasar. Pada proses ini, semua berhasil dikembalikan kepada kata dasar. Kemudian, melakukan *Tokenizing* dengan hasil yang ada pada tabel 4.4.

Tabel 4.4: Sampel Hasil Pre-Processing Tahap Tokenizing

No	Tokenized
1	['qatar', 'investasi', 'miliar', 'dolar', 'as', 'tuan', 'rumah', 'piala', 'dunia']
2	['qatar', 'tuan', 'rumah', 'sukses', 'selenggara', 'turnamen', 'fifa', 'perangkat', 'tanding', 'kontroversial', 'untung']
3	['piala', 'dunia', 'tanding', 'final', 'qatar', 'sukses', 'tuan', 'rumah']
4	['qatar', 'manfaat', 'posisi', 'tuan', 'rumah', 'piala', 'dunia', 'hebat', 'ekonomi', 'maksimal', 'pengaruh', 'politik', 'regional', 'panggung', 'global', 'baca', 'trias', 'kuncayono', 'opini', 'kompas']
5	['prancis', 'final', 'pilih', 'qatar', 'tuan', 'rumah', 'pildun', 'gk', 'lepas', 'kontribusi', 'prancis', 'hubung', 'ekonomi', 'bilateral', 'negara', 'kuat']
6	['tuan', 'rumah', 'mana', 'piala', 'dunia', 'gelar', 'tanah', 'arab', 'perintah', 'qatar', 'habis', 'habis', 'siap', 'wujud', 'megah', 'acara', 'dana', 'usd', 'miliar', 'gelontor']
7	['temu', 'teman', 'chief', 'investment', 'officer', 'qatar', 'investment', 'authority', 'asia', 'sheikh', 'faisal', 'thani', 'al', 'thani', 'bidang', 'olahraga', 'qatar', 'tuan', 'rumah', 'piala', 'dunia', 'manfaat', 'gelar', 'piala', 'dunia', 'tahun']
8	['seri', 'paduka', 'menzahirkan', 'ucap', 'tahniah', 'sheikh', 'tamim', 'raja', 'qatar', 'tuan', 'rumah', 'jaya', 'anjur', 'johan', 'bola', 'sepak', 'piala', 'dunia', 'fifa', 'qatar']
9	['tahun', 'gue', 'dukung', 'tim', 'negara', 'maroco', 'gue', 'optimis', 'tim', 'jago', 'gue', 'juara', 'semangat', 'qatar', 'tuan', 'rumah', 'rasuk', 'jiwa', 'main', 'qatar', 'menang', 'kalah', 'pilih', 'main', 'bersih']
10	['qatar', 'puji', 'luas', 'tuan', 'rumah', 'beda', 'piala', 'dunia', 'wajar', 'yg', 'cari', 'salah', 'salah', 'yg', 'tonjol', 'amp', 'keras', 'larang', 'bendera', 'pelangi']

(Sumber: Hasil olah data peneliti, 2023)

Tabel 4.4 ini menampilkan hasil pemisahan setiap kata sehingga data yang tertokenisasi ini akan diberi bobot menggunakan TF-IDF. Tabel 4.1, tabel 4.2, tabel 4.3, dan tabel 4.4 ini menampilkan *dataframe* dengan mengambil lima contoh data teratas dengan total 1492 kata dari hasil pembersihan. Data tersebut disimpan dalam format .CSV yang siap untuk diberi bobot TF-IDF.

4.1.3 Hasil Pembobotan Kata

Pada proses ini, menggunakan fungsi *TfidfVectorizer* dari *library* sklearn untuk pembobotan kata. Hasil dari vektorisasi memunculkan nilai bobot berupa angka desimal. Pada dokumen sejumlah 914, terdapat 1492 kata yang terbobot. Hasil dari pembobotan ini yang digunakan sebagai klasifikasi pada pelatihan data.

4.1.4 Hasil Pelatihan Data

Pada tahap ini, sebelum melakukan pelatihan dan pengujian terhadap data yang telah terbobot dengan TF-IDF harus dibagi sesuai skenario rasio, yaitu 70:30 dan 80:20. Pada tahap ini melakukan *split* data dengan fungsi `train_test_split` yang ada di *library* sklearn. Fungsi tersebut digunakan untuk membagi data sesuai dengan rasio yang diinginkan, serta dapat berupa acak maupun tidak.

Variabel X digunakan sebagai seleksi baris dokumen dari 0 sampai 914 dan kolom fitur abadi sampai zona, sedangkan variabel Y digunakan untuk seleksi kolom label sentimen. *Split* data dengan rasio 70:30 menggunakan variabel `test_size` bernilai 0.3, sedangkan rasio 80:20 menggunakan variabel `test_size` bernilai 0.2. Variabel `test_size` tersebut berarti jumlah persentase data yang akan menjadi data uji, nilai 0.3 yang berarti 30% dari total dokumen dan nilai 0.2 yang berarti 20% dari total dokumen. Data dari masing-masing rasio tersebut tidak dibagi secara acak dengan `random_state` bernilai *None* dan `shuffle` bernilai *False*, yang berarti tidak ada pengacakan data sehingga data latih dan data uji tidak berubah, serta nilai evaluasi tidak akan berubah saat dieksekusi beberapa kali.

Pada rasio 70:30, data yang berjumlah total 914 ini terbagi menjadi dua, yaitu data latih sejumlah 639 dan data uji sejumlah 275, sedangkan pada *split* rasio 80:20 memperoleh data latih sejumlah 731 dan data uji sejumlah 183. Detail daftar data latih dan data uji pada kedua rasio ini terdapat pada lampiran.

Setelah melakukan proses *splitting* data, akan dilakukan proses pelatihan data dengan menggunakan fungsi `DecisionTreeClassifier` dari *library* sklearn. Sampel yang digunakan adalah hasil dari pembagian data sebelumnya yang tersimpan di variabel `x_train` sebagai fitur dan `y_train` sebagai labelnya. Pada model ini terdapat perbedaan hirarki cabang pohon keputusan antara rasio 70:30 dengan rasio 80:20. Penentuan jumlah hirarki atau kedalaman cabang dari pohon menggunakan `max_depth`, sedangkan untuk mencari `max_depth` terbaik adalah dengan `GridSearchCV` berdasarkan nilai akurasi, sehingga menghasilkan nilai *depth* terbaik pada rasio 70:30 adalah `max_depth` 8 dan pada rasio 80:20 adalah `max_depth` 9.

Setelah itu, jika sudah menampilkan keberhasilan klasifikasi, maka sudah dapat menampilkan visualisasi pohon keputusan data latih untuk melihat pola data

latih dengan matplotlib dan sklearn. Untuk membuat pohon keputusan menggunakan fungsi `tree.plot_tree` dengan memasukkan variabel yang menyimpan model klasifikasi. Pola data latih rasio 70:30 yang akan digunakan klasifikasi pada data uji. Fitur yang memiliki nilai gain tertinggi adalah fitur pada kolom 275 yaitu kata festival sebagai *root* dengan memiliki nilai *gain* sebesar 0.658. detail atribut beserta nilai dari *root* dan *internal node* terdapat pada tabel 4.5.

Tabel 4.5: *Root dan Internal Node Rasio 70:30*

<i>Depth</i>	Cabang Pohon	Atribut	<i>Value</i>	<i>Gain</i>	Kelas
0	Root	X 275	≤ 0.108	0.658	-
1	Internal node 1.1	X 1265	≤ 0.085	0.647	-
	Internal node 1.2	X 951	≤ 0.049	0.574	-
2	Internal node 1.1.1	X 565	≤ 0.08	0.0	Negatif
	Internal node 1.1.1	X 565	≤ 0.08	0.642	-
	Internal node 1.1.2	X 1488	≤ 0.069	0.0	Negatif
	Internal node 1.2.1	X 275	≤ 0.188	0.638	-
	Internal node 1.2.2	X 1016	≤ 0.119	0.34	-
3	Internal node 1.1.1.1	X 1403	≤ 0.099	0.638	-
	Internal node 1.2.1.1	X 1016	≤ 0.052	0.66	-
	Internal node 1.2.1.2	X 24	≤ 0.243	0.506	-
	Internal node 1.2.1.2	X 24	≤ 0.243	0.0	Netral
	Internal node 1.2.2.1	X 1139	≤ 0.113	0.185	-
	Internal node 1.2.2.1	X 1139	≤ 0.113	0.0	Netral
	Internal node 1.2.2.2	X 1106	≤ 0.106	0.611	-
	Internal node 1.2.2.2	X 1106	≤ 0.106	0.0	Negatif
4	Internal node 1.1.1.1.1	X 165	≤ 0.12	0.62	-
	Internal node 1.1.1.1.2	X 1403	≤ 0.115	0.635	-
	Internal node 1.2.1.1.1	X 1403	≤ 0.041	0.408	-
	Internal node 1.2.1.1.1	X 1403	≤ 0.041	0.0	Negatif
	Internal node 1.2.1.1.2	X 1357	≤ 0.129	0.661	-
	Internal node 1.2.1.1.2	X 1357	≤ 0.129	0.0	Negatif

<i>Depth</i>	Cabang Pohon	Atribut	<i>Value</i>	<i>Gain</i>	Kelas
	Internal node 1.2.1.2.1	X 123	≤ 0.241	0.464	-
	Internal node 1.2.1.2.1	X 123	≤ 0.241	0.0	Positif
	Internal node 1.2.2.1.1	X 526	≤ 0.179	0.133	-
	Internal node 1.2.2.1.1	X 526	≤ 0.179	0.0	Netral
	Internal node 1.2.2.2.1	X 240	≤ 0.208	0.0	Positif
5	Internal node 1.1.1.1.1.1	X 337	≤ 0.268	0.614	-
	Internal node 1.1.1.1.1.1	X 337	≤ 0.268	0.0	Negatif
	Internal node 1.1.1.1.1.2	X 687	≤ 0.317	0.198	-
	Internal node 1.1.1.1.1.2	X 687	≤ 0.317	0.198	Negatif
	Internal node 1.1.1.1.2.1	X 1271	≤ 0.142	0.5	-
	Internal node 1.1.1.1.2.2	X 240	≤ 0.19	0.654	-
	Internal node 1.2.1.1.1.1	X 1106	≤ 0.081	0.0	Netral
	Internal node 1.2.1.1.2.1	X 1265	≤ 0.207	0.648	-
	Internal node 1.2.1.1.2.1	X 1265	≤ 0.207	0.0	Positif
	Internal node 1.2.1.2.1.1	X 239	≤ 0.49	0.435	-
	Internal node 1.2.1.2.1.1	X 239	≤ 0.49	0.0	Positif
	Internal node 1.2.2.1.1.1	X 772	≤ 0.2	0.071	Negatif
6	Internal node 1.1.1.1.1.1.1	X 158	≤ 0.272	0.609	-
	Internal node 1.1.1.1.1.1.1	X 158	≤ 0.272	0.0	Netral
	Internal node 1.1.1.1.2.1.1	X 646	≤ 0.178	0.404	-
	Internal node 1.1.1.1.2.1.1	X 646	≤ 0.178	0.0	Positif
	Internal node 1.1.1.1.2.1.2	X 1427	≤ 0.385	0.0	Positif
	Internal node 1.1.1.1.2.2.1	X 944	≤ 0.177	0.557	-
	Internal node 1.1.1.1.2.2.1	X 944	≤ 0.177	0.0	Positif
	Internal node 1.1.1.1.2.2.2	X 1356	≤ 0.152	0.64	-
	Internal node 1.1.1.1.2.2.2	X 1356	≤ 0.152	0.0	Positif
	Internal node 1.2.1.1.2.1.1	X 944	≤ 0.08	0.658	-
	Internal node 1.2.1.2.1.1.1	X 204	≤ 0.309	0.401	-
	Internal node 1.2.1.2.1.1.1	X 204	≤ 0.309	0.0	Positif

<i>Depth</i>	Cabang Pohon	Atribut	<i>Value</i>	<i>Gain</i>	Kelas
7	Internal node 1.1.1.1.1.1.1.1	X 285	≤ 0.274	0.606	-
	Internal node 1.1.1.1.1.1.1.1	X 285	≤ 0.274	0.0	Negatif
	Internal node 1.1.1.1.2.1.1.1	X 848	≤ 0.156	0.37	-
	Internal node 1.1.1.1.2.1.1.1	X 848	≤ 0.156	0.0	Positif
	Internal node 1.1.1.1.2.2.1.1	X 1087	≤ 0.153	0.658	-
	Internal node 1.1.1.1.2.2.2.1	X 1016	≤ 0.14	0.593	-
	Internal node 1.2.1.1.2.1.1.1	X 1016	≤ 0.085	0.0	Positif
	Internal node 1.2.1.1.2.1.1.2	X 275	≤ 0.132	0.0	Netral
	Internal node 1.2.1.2.1.1.1.1	X 80	≤ 0.205	0.359	-
	Internal node 1.2.1.2.1.1.1.1	X 80	≤ 0.205	0.0	Positif
8	Internal node 1.1.1.1.1.1.1.1.1	X 285	≤ 0.274	0.0	Positif
	Internal node 1.1.1.1.2.1.1.1.1	X 848	≤ 0.156	0.0	Negatif
	Internal node 1.1.1.1.2.2.1.1.1	X 1087	≤ 0.153	0.0	Negatif
	Internal node 1.1.1.1.2.2.2.1.1	X 1016	≤ 0.14	0.0	Netral
	Internal node 1.2.1.2.1.1.1.1.1	X 80	≤ 0.205	0.0	Negatif

(Sumber: Hasil olah data peneliti, 2023)

Berdasarkan tabel 4.5 dapat diketahui bahwa model pada rasio 70:30 memiliki 32 pola yang menentukan sentimen dokumen hasil pelatihan data. Terdapat berbagai macam pola susunan kata untuk membuat keputusan mana yang tergolong dokumen positif, negatif, atau netral. Pada tabel 4.6 adalah kumpulan pola susunan kata dari hasil data latih.

Tabel 4.6: Pola Susunan Kata sebagai Keputusan Rasio 70:30

No	Pola Nomor Atribut	Konversi Menjadi Kata	Kelas
1	X 275, X 1265, X 565	festival, stuck, korup	Negatif
2	X 275, X 1265, X 1488	festival, stuck, workers	Negatif
3	X 275, X 951, X 275, X 24	festival, pildun, festival, alangkah	Netral
4	X 275, X 951, X 1016, X 1139	festival, pildun, putus, selamat	Netral
5	X 275, X 951, X 1016, X 1106	festival, pildun, putus, saksi	Negatif
6	X 275, X 951, X 275, X 1016, X 1403	festival, pildun, festival, putus, tuai	Negatif

No	Pola Nomor Atribut	Konversi Menjadi Kata	Kelas
7	X 275, X 951, X 275, X 1016, X 1357	festival, pildun, festival, putus, tengah	Negatif
8	X 275, X 951, X 275, X 24, X 123	festival, pildun, festival, alangkah, berontak	Positif
9	X 275, X 951, X 1016, X 1139, X 526	festival, pildun, putus, selamat, klub	Netral
10	X 275, X 951, X1016, 1106, X 240	festival, pildun, putus, saksi, dukung	Positif
11	X 275, X 1265, X 565, X 1403, X 165, X 337	festival, stuck, korup, tuai, buruk, halus	Negatif
12	X 275, X 1265, X 565, X 1403, X 165, X 687	festival, stuck, korup, tuai, buruk, menang	Negatif
13	X 275, X 951, X 275, X 1016, X 1403, X1106	festival, pildun, festival, putus, tuai, saksi	Netral
14	X 275, X 951, X 275, X 1016, X 1357, X 1265	festival, pildun, festival, putus, tengah, stuck	Positif
15	X 275, X 951, X 275, X 24, X 123, X 239	festival, pildun, festival, alangkah, berontak, duit	Positif
16	X 275, X 951, X 1016, X 1139, X 526, X 772	festival, pildun, Xputus, Xselamat, klub, netral	Negatif
17	X 275, X 1265, X 565, X 1403, X 165, X 337, X 158	festival, stuck, korup, tuai, buruk, halus, budaya	Netral
18	X 275, X 1265, X 565, X 1403, X 1403, X 1271, X 646	festival, stuck, korup, tuai, tuai, suka, lumayan	Positif
19	X 275, X 1265, X 565, X 1403, X 1403, X 1271, X 1427	festival, stuck, korup, tuai, tuai, suka, ujung	Positif
20	X 275, X 1265, X 565, X 1403, X 1403, X 240, X 944	festival, stuck, korup, tuai, tuai, dukung, petinggi	Positif
21	X 275, X 1265, X 565, X 1403, X 1403, X 240, X 1356	festival, stuck, korup, tuai, tuai, dukung, tenang	Positif
22	X 275, X 951, X 275, X 24, X 123, X 239, X 204	festival, pildun, festival, alangkah, berontak, duit, cuci	Positif

No	Pola Nomor Atribut	Konversi Menjadi Kata	Kelas
23	X 275, X 1265, X 565, X 1403, X 165, X 337, X 158, X 285	festival, stuck, korup, tuai, buruk, halus, budaya, gaduh	Negatif
24	X 275, X 1265, X 565, X 1403, X 1403, X 1271, X 646, X 848	festival, stuck, korup, tuai, tuai, suka, lumayan, operasional	Positif
25	X 275, X 951, X 275, X 1016, X 1357, X 1265, X 944, X 1016	festival, pildun, festival, putus, tengah, stuck, petinggi, putus	Positif
26	X 275, X 951, X 275, X 1016, X 1357, X 1265, X 944, X 275	festival, pildun, festival, putus, tengah, stuck, petinggi, festival	Netral
27	X 275, X 951, X 275, X 24, X 123, X 239, X 204, X 80	festival, pildun, festival, alangkah, berontak, duit, cuci, bahaya	Positif
28	X 275, X 1265, X 565, X 1403, X 165, X 337, X 158, X 285	festival, stuck, korup, tuai, buruk, halus, budaya, gaduh	Positif
29	X 275, X 1265, X 565, X 1403, X 1403, X 1271, X 646, X 848	festival, stuck, korup, tuai, tuai, suka, lumayan, operasional	Negatif
30	X 275, X 1265, X 565, X 1403, X 1403, X 240, X 944, X 1087	festival, stuck, korup, tuai, tuai, dukung, petinggi, rules	Negatif
31	X 275, X 1265, X 565, X 1403, X 1403, X 240, X 1356, X 1016	festival, stuck, korup, tuai, tuai, dukung, tenang, putus	Netral
32	X 275, X 951, X 275, X 24, X 123, X 239, X 204, X 80	festival, pildun, festival, alangkah, berontak, duit, cuci, bahaya	Negatif

(Sumber: Hasil olah data peneliti, 2023)

Setelah mendapatkan hasil data latih pada rasio 70:30, dilanjutkan pada proses mencari model data latih pada rasio 80:20 pada halaman *script* yang berbeda. Pola data latih rasio 80:20 yang akan digunakan klasifikasi pada data uji memiliki nilai *gain* tertinggi pada fitur nomor kolom 275 juga yaitu kata festival sebagai *root* dengan memiliki nilai *gain* sebesar 0.656. detail atribut beserta nilai dari *root* dan *internal node* terdapat pada tabel 4.7.

Tabel 4.7: *Root* dan *Internal Node* Rasio 80:20

<i>Depth</i>	Cabang pohon	Atribut	<i>Value</i>	<i>Gain</i>	Kelas
0	Root	X 275	≤ 0.108	0.656	-

<i>Depth</i>	Cabang pohon	Atribut	<i>Value</i>	<i>Gain</i>	Kelas
1	Internal node 1.1	X 1265	≤ 0.074	0.644	-
	Internal node 1.2	X 951	≤ 0.107	0.574	-
2	Internal node 1.1.1	X 565	≤ 0.077	0.639	-
	Internal node 1.1.2	X 1087	≤ 0.105	0.0	Negatif
	Internal node 1.2.1	X 1265	≤ 0.639	0.639	-
	Internal node 1.2.2	X 1016	≤ 0.119	0.324	-
3	Internal node 1.1.1.1	X 165	≤ 0.12	0.634	-
	Internal node 1.1.1.2	X 560	≤ 0.103	0.0	Negatif
	Internal node 1.2.1.1	X 275	≤ 0.188	0.622	-
	Internal node 1.2.1.2	X 212	≤ 0.203	0.0	Positif
	Internal node 1.2.2.1	X 427	≤ 0.187	0.175	-
	Internal node 1.2.2.1	X 427	≤ 0.187	0.0	Netral
	Internal node 1.2.2.2	X 1106	≤ 0.106	0.611	-
	Internal node 1.2.2.2	X 1106	≤ 0.106	0.0	Negatif
4	Internal node 1.1.1.1.1	X 1403	≤ 0.095	0.629	-
	Internal node 1.1.1.1.2	X 1406	≤ 0.263	0.0	Negatif
	Internal node 1.2.1.1.1	X 144	≤ 0.077	0.648	-
	Internal node 1.2.1.1.1	X 144	≤ 0.077	0.0	Positif
	Internal node 1.2.1.1.2	X 24	≤ 0.243	0.526	-
	Internal node 1.2.1.1.2	X 24	≤ 0.243	0.0	Netral
	Internal node 1.2.2.1.1	X 1406	≤ 0.134	0.124	-
	Internal node 1.2.2.1.1	X 1406	≤ 0.134	0.0	Netral
	Internal node 1.2.2.2.1	X 275	≤ 0.316	0.0	Positif
5	Internal node 1.1.1.1.1.1	X 337	≤ 0.269	0.607	-
	Internal node 1.1.1.1.1.1	X 337	≤ 0.269	0.0	Negatif
	Internal node 1.1.1.1.1.2	X 1087	≤ 0.151	0.651	-
	Internal node 1.2.1.1.1.1	X 275	≤ 0.184	0.631	-
	Internal node 1.2.1.1.1.1	X 275	≤ 0.184	0.0	Positif

<i>Depth</i>	Cabang pohon	Atribut	<i>Value</i>	<i>Gain</i>	Kelas
	Internal node 1.2.1.1.2.1	X 95	≤ 0.401	0.496	-
	Internal node 1.2.2.1.1.1	X 772	≤ 0.2	0.0	Negatif
6	Internal node 1.1.1.1.1.1.1	X 158	≤ 0.273	0.603	-
	Internal node 1.1.1.1.1.1.1	X 158	≤ 0.273	0.0	Netral
	Internal node 1.1.1.1.1.2.1	X 1356	≤ 0.096	0.619	-
	Internal node 1.1.1.1.1.2.1	X 1356	≤ 0.096	0.0	Positif
	Internal node 1.1.1.1.1.2.2	X 240	≤ 0.146	0.585	-
	Internal node 1.2.1.1.1.1.1	X 702	≤ 0.216	0.602	-
	Internal node 1.2.1.1.1.1.1	X 702	≤ 0.216	0.0	Positif
	Internal node 1.2.1.1.2.1.1	X 80	≤ 0.205	0.451	-
	Internal node 1.2.1.1.2.1.1	X 80	≤ 0.205	0.0	Positif
	Internal node 1.2.1.1.2.1.2	X 944	≤ 0.064	0.0	Positif
7	Internal node 1.1.1.1.1.1.1.1	X 240	≤ 0.063	0.6	-
	Internal node 1.1.1.1.1.2.1.1	X 239	≤ 0.26	0.612	-
	Internal node 1.1.1.1.1.2.1.1	X 239	≤ 0.26	0.0	Positif
	Internal node 1.1.1.1.1.2.2.1	X 38	≤ 0.206	0.32	-
	Internal node 1.1.1.1.1.2.2.1	X 38	≤ 0.206	0.0	Netral
	Internal node 1.1.1.1.1.2.2.2	X 1444	≤ 0.446	0.46	-
	Internal node 1.1.1.1.1.2.2.2	X 1444	≤ 0.446	0.0	Positif
	Internal node 1.2.1.1.1.1.1.1	X 237	≤ 0.096	0.556	-
	Internal node 1.2.1.1.1.1.1.1	X 237	≤ 0.096	0.0	Negatif
	Internal node 1.2.1.1.2.1.1.1	X 239	≤ 0.49	0.416	-
	Internal node 1.2.1.1.2.1.1.1	X 239	≤ 0.49	0.0	Positif
8	Internal node 1.1.1.1.1.1.1.1.1	X 1087	≤ 0.071	0.0	Positif
	Internal node 1.1.1.1.1.1.1.1.2	X 951	≤ 0.13	0.619	-
	Internal node 1.1.1.1.1.1.1.1.2	X 951	≤ 0.13	0.0	Negatif
	Internal node 1.1.1.1.1.2.1.1.1	X 1016	≤ 0.071	0.602	-
	Internal node 1.1.1.1.1.2.1.1.1	X 1016	≤ 0.071	0.0	Negatif
	Internal node 1.1.1.1.1.2.2.1.1	X 1403	≤ 0.398	0.0	Positif

<i>Depth</i>	Cabang pohon	Atribut	<i>Value</i>	<i>Gain</i>	Kelas
	Internal node 1.1.1.1.1.2.2.2.1	X 1257	≤ 0.196	0.346	-
	Internal node 1.1.1.1.1.2.2.2.1	X 1257	≤ 0.196	0.0	Negatif
	Internal node 1.2.1.1.1.1.1.1.1	X 1087	≤ 0.064	0.0	Netral
	Internal node 1.2.1.1.2.1.1.1.1	X 238	≤ 0.109	0.375	-
	Internal node 1.2.1.1.2.1.1.1.1	X 238	≤ 0.109	0.0	Netral
9	Internal node 1.1.1.1.1.1.1.1.2.1	X 951	≤ 0.13	0.0	Positif
	Internal node 1.1.1.1.1.2.1.1.1.1	X 1016	≤ 0.071	0.0	Positif
	Internal node 1.1.1.1.1.2.2.2.1.1	X 1257	≤ 0.196	0.0	Netral
	Internal node 1.2.1.1.2.1.1.1.1.1	X 238	≤ 0.109	0.0	Negatif

(Sumber: Hasil olah data peneliti, 2023)

Berdasarkan tabel 4.7 dapat diketahui bahwa model pada rasio 80:20 memiliki 34 pola yang menentukan sentimen dokumen hasil pelatihan data. Terdapat berbagai macam pola susunan kata untuk membuat keputusan mana yang tergolong dokumen positif, negatif, atau netral. Pada tabel 4.8 adalah kumpulan pola susunan kata dari hasil data latih.

Tabel 4.8: Pola Susunan Kata sebagai Keputusan Rasio 80:20

No	Pola Nomor Atribut	Konversi Menjadi Kata	Kelas
1	X 275, X 1265, X 1087	festival, stuck, rules	Negatif
2	X 275, X 1265, X 565, X 560	festival, stuck, korup, kontribusi	Negatif
3	X 275, X 951, X 1265, X 212	festival, pildun, stuck, dana	Positif
4	X 275, X 951, X 1016, X 427	festival, pildun, putus, jawara	Netral
5	X 275, X 951, X 1016, X 1106	festival, pildun, putus, saksi	Negatif
6	X 275, X 1265, X 565, X 165, X 1406	festival, stuck, korup, buruk, tuding	Negatif
7	X 275, X 951, X 1265, X 275, X 144	festival, pildun, stuck, festival, boikot	Positif
8	X 275, X 951, X 1265, X 275, X 24	festival, pildun, stuck, festival, alangkah	Netral
9	X 275, X 951, X 1016, X 427, X 1406	festival, pildun, putus, jawara, tuding	Netral

No	Pola Nomor Atribut	Konversi Menjadi Kata	Kelas
10	X 275, X 1265, X 565, X 165, X 1403, X 337	festival, stuck, korup, buruk, tuai, halus	Negatif
11	X 275, X 951, X 1016, X 1106, X 275	festival, pildun, putus, saksi, festival	Positif
12	X 275, X 951, X 1265, X 275, X 144, X 275	festival, pildun, stuck, festival, boikot, festival	Positif
13	X 275, X 951, X 1016, X 427, X 1406, X 772	festival, pildun, putus, jawara, tuding, netral	Negatif
14	X 275, X 1265, X 565, X 165, X 1403, X 337, X 158	festival, stuck, korup, buruk, tuai, halus, budaya	Netral
15	X 275, X 1265, X 565, X 165, X 1403, X 1087, X 1356	festival, stuck, korup, buruk, tuai, rules, tenang	Positif
16	X 275, X 951, X 1265, X 275, X 144, X 275, X 702	festival, pildun, stuck, festival, boikot, festival, milih	Positif
17	X 275, X 951, X 1265, X 275, X 24, X 95, X 80	festival, pildun, stuck, festival, alangkah, baru, bahaya	Positif
18	X 275, X 951, X 1265, X 275, X 24, X 95, X 944	festival, pildun, stuck, festival, alangkah, baru, petinggi	Positif
19	X 275, X 1265, X 565, X 165, X 1403, X 1087, X 1356, X 239	festival, stuck, korup, buruk, tuai, rules, tenang, duit	Positif
20	X 275, X 1265, X 565, X 165, X 1403, X 1087, X 240, X 38	festival, stuck, korup, buruk, tuai, rules, dukung, analisis	Netral
21	X 275, X 1265, X 565, X 165, X 1403, X 1087, X 240, X 1444	festival, stuck, korup, buruk, tuai, rules, dukung, unsur	Positif
22	X 275, X 951, X 1265, X 275, X 144, X 275, X 702, X 237	festival, pildun, stuck, festival, boikot, festival, milih, duduk	Negatif
23	X 275, X 951, X 1265, X 275, X 24, X 95, X 80, X 239	festival, pildun, stuck, festival, alangkah, baru, bahaya, duit	Positif
24	X 275, X 1265, X 565, X 165, X 1403, X 337, X 158, X 240, X 1087	festival, stuck, korup, buruk, tuai, halus, budaya, dukung, rules	Positif

No	Pola Nomor Atribut	Konversi Menjadi Kata	Kelas
25	X 275, X 1265, X 565, X 165, X 1403, X 337, X 158, X 240, X 951	festival, stuck, korup, buruk, tuai, halus, budaya, dukung, pildun	Negatif
26	X 275, X 1265, X 565, X 165, X 1403, X 1087, X 1356, X 239, X 1016	festival, stuck, korup, buruk, tuai, rules, tenang, duit, putus	Negatif
27	X 275, X 1265, X 565, X 165, X 1403, X 1087, X 240, X 38, X 1403	festival, stuck, korup, buruk, tuai, rules, dukung, analisis, tuai	Positif
28	X 275, X 1265, X 565, X 165, X 1403, X 1087, X 240, X 1444, X 1257	festival, stuck, korup, buruk, tuai, rules, dukung, unsur, statement	Negatif
29	X 275, X 951, X 1265, X 275, X 144, X 275, X 702, X 237, X 1087	festival, pildun, stuck, festival, boikot, festival, milih, duduk, rules	Netral
30	X 275, X 951, X 1265, X 275, X 24, X 95, X 80, X 239, X 238	festival, pildun, stuck, festival, alangkah, baru, bahaya, duit, duga	Netral
31	X 275, X 1265, X 565, X 165, X 1403, X 337, X 158, X 240, 951	festival, stuck, korup, buruk, tuai, halus, budaya, dukung, pildun	Positif
32	X 275, X 1265, X 565, X 165, X 1403, X 1087, X 1356, X 239, X 1016	festival, stuck, korup, buruk, tuai, rules, tenang, duit, putus	Positif
33	X 275, X 1265, X 565, X 165, X 1403, X 1087, X 240, X 1444, X 1257	festival, stuck, korup, buruk, tuai, rules, dukung, unsur, statement	Netral
34	X 275, X 951, X 1265, X 275, X 24, X 95, X 80, X 239, X 238	festival, pildun, stuck, festival, alangkah, baru, bahaya, duit, duga	Negatif

(Sumber: Hasil olah data peneliti, 2023)

Setelah memperoleh hasil pelatihan, berarti model *machine learning* sudah siap digunakan sebagai alat untuk klasifikasi yang nanti dilakukan pada saat tahap pengujian data sekaligus evaluasi kinerja dari model.

4.1.5 Hasil Pengujian Data

Pada proses ini adalah tahap pengujian dan evaluasi data uji menggunakan *confusion matrix*. Evaluasi ini dilakukan agar mengetahui nilai kebenaran dari proses klasifikasi yang telah dilakukan oleh sistem. Untuk mengetahui nilai *true* dan *false* dari evaluasi model perlu menerapkan fungsi *confusion_matrix* pada library *sklearn*. Setelah itu, menerapkan fungsi *classification_report* sebagai penilai kinerja dan menampilkan persentase dari akurasi, *recall*, presisi, dan *F1-score* dari rasio 70:30 dan rasio 80:20. Hasil tersebut ada pada tabel 4.9.

Tabel 4.9: Hasil Evaluasi *Confusion Matrix*

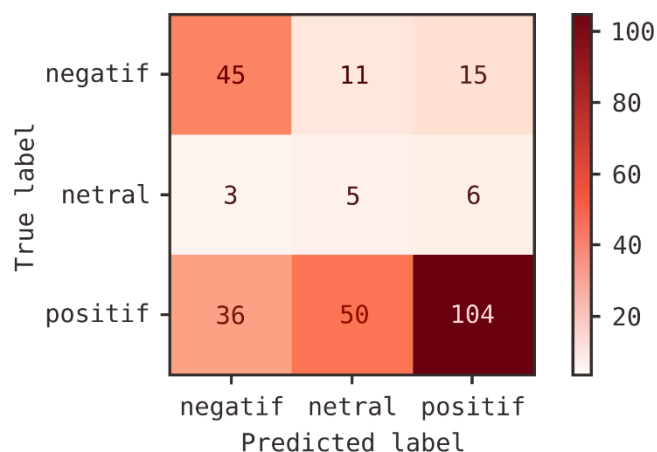
Rasio	Akurasi	Recall	Presisi	F1-Score
70:30	0.56	0.51	0.48	0.46
80:20	0.59	0.57	0.52	0.50

(Sumber: Hasil olah data peneliti, 2023)

Pada tabel 4.9 menunjukkan hasil *confusion matrix* bahwa skor tertinggi terdapat pada rasio 80:20 dengan menerapkan *max_depth* senilai 9 yang berdasarkan skor akurasi. Nilai *depth* 9 tersebut didapat dari pencarian *depth* terbaik antara *depth* 1 sampai *depth* 30 pada model *machine learning* rasio 80:20 ini dengan menerapkan fungsi *GridSearchCV* pada library *sklearn*. *Max_depth* menunjukkan batas maksimal kedalaman dari cabang pohon keputusan untuk melakukan klasifikasi. Penerapan jumlah cabang maksimal ini mempengaruhi persentase dari evaluasi sebuah model *machine learning* (Jondri, 2022). Pengujian pada rasio 80:20 ini memiliki akurasi sejumlah 59%, *recall* sejumlah 57%, presisi sejumlah 52%, dan *f1-score* sejumlah 50%. Berbeda pada rasio 70:30 yang memiliki persentase lebih rendah tetapi hanya memiliki selisih sedikit. Pada rasio 70:30 ini memiliki skor terbaik menggunakan *max_depth* senilai 8 yang sama perlakuan dengan rasio 80:20 dengan mencari *depth* terbaik antara *depth* 1 sampai *depth* 30. Pada rasio 70:30 memiliki akurasi sejumlah 56%, *recall* sejumlah 51%, presisi sejumlah 48%, dan *f1-score* sejumlah 46%.

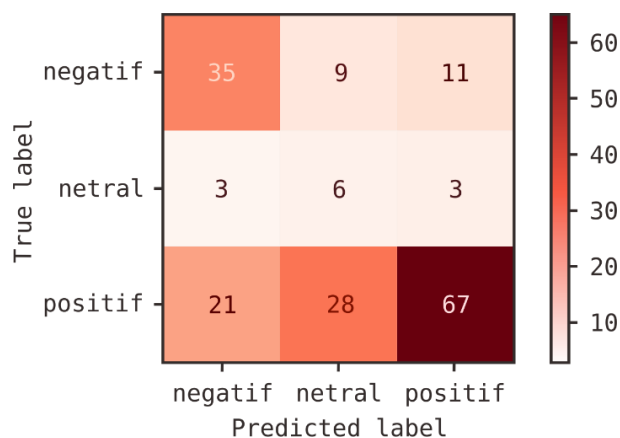
Semua persentase yang telah disebutkan berdasarkan perhitungan jumlah kecocokan prediksi nilai *true* dan *false* dari data aktual. Nilai *true* dan *false* tersebut divisualisasikan oleh *sklearn* berbentuk tabel *matrix* yang pada kotak kiri atas,

kotak tengah, dan kotak kanan bawah adalah berisi nilai *true* karena *true label* dan *predicted label* bertemu pada garis vertikal dan horizontal sebuah tabel *matrix*, sedangkan selain pada kotak yang telah disebut adalah bagian yang berisi nilai *false*. Pada gambar 4.2 merupakan visualisasi dari hasil evaluasi model dengan rasio 70:30.



Gambar 4.2: Tabel *Matrix* Rasio 70:30
(Sumber: Hasil olah data peneliti, 2023)

Pada gambar 4.2 dapat diketahui bahwa pada rasio 70:30 terdapat 154 nilai *true* yang terdiri dari 104 *positif*, 5 *netral*, dan 45 *negatif*. Adapun total nilai *false* ada 121 yang terdiri dari 86 *false positif*, 9 *false netral*, dan 26 *false negatif*. Pada tabel *matrix* ini, deret vertikal adalah *true label* atau data aktual, sedangkan deret horizontal adalah *predicted label* atau hasil dari prediksi model yang telah didapat pada saat pelatihan data. Setelah mendapat tabel *matrix* pada rasio 70:30, selanjutnya adalah visualisasi tabel *matrix* pada rasio 80:20 yang ada pada gambar 4.3.



Gambar 4.3: Tabel *Matrix* Rasio 80:20
(Sumber: Hasil olah data peneliti, 2023)

Pada gambar 4.3 dapat diketahui bahwa pada rasio 80:20 terdapat 108 nilai true yang terdiri dari 67 positif, 6 netral, dan 35 negatif. Adapun total nilai false ada 75 yang terdiri dari 49 false positif, 6 false netral, dan 20 false negatif.

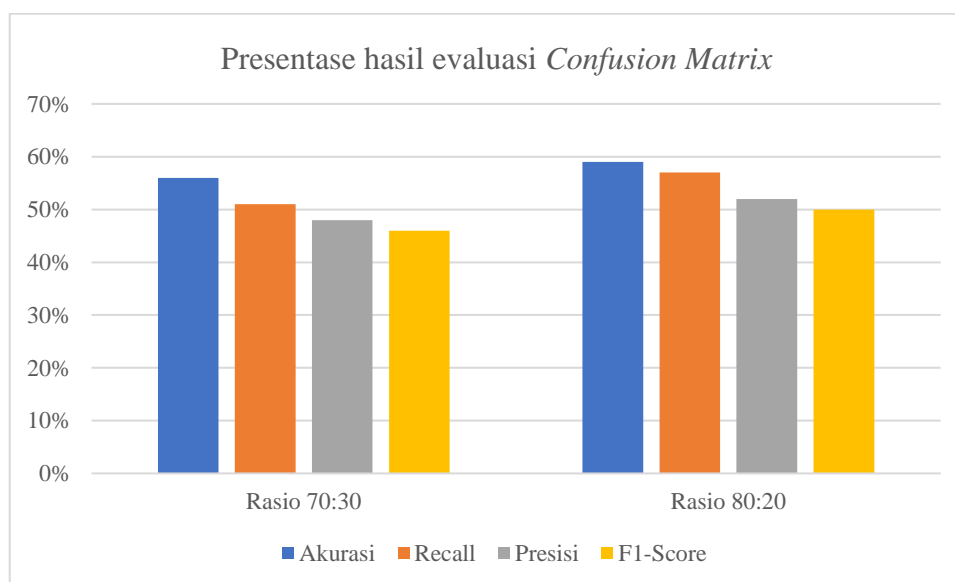
4.2 Pembahasan

Pada tahap awal, perihal penghapusan data yang terduplikasi menggunakan fitur penandaan dari Microsoft Excel mendapatkan 43 data yang terdeteksi duplikat, sedangkan masih banyak data yang sebenarnya juga terindikasi duplikat, namun belum terhapus sehingga dilakukan penandaan data secara manual yang mendapatkan 1968 data yang harus dihapus. Setelah itu, diperoleh data bersih dari duplikasi sejumlah 914 data. Kemudian, data tersebut diberi label oleh 2 ahli dari ilmu bahasa Indonesia dan ilmu sosial. Label pada sentimen positif memiliki jumlah paling banyak daripada label negatif yang berada pada urutan kedua dan label netral memiliki jumlah yang paling sedikit.

Dari hasil pelabelan tersebut, kemudian dilakukan pembersihan partikel pada teks yang tidak berguna untuk klasifikasi. Penghapusan terhadap *punctuation*, URL, *emoticon*, *hashtag*, *username/mention*, dan *whitespace* berhasil, serta dilakukan pengecilan huruf agar menjadi setara sehingga memudahkan proses klasifikasi. Pada tahap *cleaning* ini juga terdapat *newline* pada teks sehingga harus dihapus juga karena pada python terdeteksi sebagai tulisan (`\n`). Setelah data bersih dari beberapa partikel, kemudian melakukan penghapusan *stopwords* dengan memakai *library* sastrawi yang dalamnya memiliki modul penyimpan kata yang perlu dihapus. Setelah itu, dilakukan *stemming* dengan *library* sastrawi yang membutuhkan waktu beberapa menit dalam prosesnya, adapun kata yang tidak terdapat dalam daftar *stopwords* dan *stemming* tetap akan digunakan pada proses klasifikasi. Pada akhir *pre-processing* ini dilakukan tokenisasi yang nanti akan digunakan pada proses pembobotan kata.

Pembobotan dilakukan untuk memberi nilai terhadap kata yang berpengaruh. Dari hasil pembobotan terdapat 1492 kata yang telah diberi bobot oleh sistem. Nilai bobot mulai dari 0.0 hingga yang terbesar adalah 0.930. Kemudian data yang telah terbobot ini digunakan untuk proses klasifikasi pelatihan data. Klasifikasi dilakukan dengan memperhatikan banyaknya lapis pohon keputusan yang diterapkan. Pada rasio 70:30 memiliki kedalaman percabangan terbaik

berdasarkan nilai akurasi pada nilai 8 *depth*, sedangkan pada rasio 80:20 memiliki kedalaman percabangan terbaik pada nilai 9 *depth*. Pencarian skenario *depth* terbaik bisa dilakukan secara otomatis dengan menggunakan fitur dari sklearn, sehingga tidak perlu menguji satu persatu dengan skenario *depth* yang berbeda. Dari pelatihan, didapat 32 pola cabang pohon keputusan pada rasio 70:30 dengan susunan kata yang berbeda-beda dan 34 pola cabang pohon keputusan pada rasio 80:20 dengan susunan kata yang berbeda-beda pula. Setelah melakukan klasifikasi pada proses pelatihan, selanjutnya melakukan evaluasi model menggunakan *confusion matrix*. Pada evaluasi ini, menunjukkan adanya sedikit perbedaan nilai akurasi dan nilai F1-Score. Perbedaan tersebut divisualisasikan dengan diagram batang yang terdapat pada gambar 4.4.



Gambar 4.4: Diagram Perbandingan Persentase Rasio 70:30 dengan Rasio 80:20

(Sumber: Hasil olah data peneliti dari evaluasi *Confusion Matrix*, 2023)

Hasil dari pengujian dan evaluasi menunjukkan selisih antara dua rasio hanya sedikit. Pada penelitian ini, rasio 80:20 memiliki persentase lebih tinggi dibanding rasio 70:30, nilai tersebut dapat terlihat karena hasil dari evaluasi *confusion matrix*. Model *machine learning* yang digunakan pada penelitian ini berdasarkan grafik tertinggi dari rasio 80:20 dapat mengklasifikasi dengan akurasi 59% dari keseluruhan data yang diuji, dengan cara sistem menghitung total jumlah prediksi yang sesuai dengan data aktual dibagi dengan keseluruhan jumlah prediksi yang sesuai maupun yang tidak sesuai dengan data aktual. Adapun *recall* dari model

ini memperoleh 57% dari keseluruhan data yang diuji. Dari masing-masing kelas, sistem akan melakukan perhitungan nilai yang terprediksi *true* pada kelasnya dibagi dengan jumlah total dari nilai terprediksi *true* pada kelas itu sendiri dengan jumlah nilai kelas tersebut yang terprediksi menjadi kelas lain, atau dapat disebut *false prediction*. Persentase *recall* ini, didapat dari perhitungan nilai rata-rata dari keseluruhan kelas yang memiliki persentase temu kembali masing-masing yaitu kelas positif ditemukan sejumlah 58%, kelas negatif sejumlah 64%, dan kelas netral sejumlah 50%. Kemudian, hasil presisi yang diambil oleh sistem juga memperhatikan pada masing-masing kelas. Dari masing-masing kelas, sistem akan melakukan perhitungan nilai yang terprediksi *true* pada kelasnya dibagi dengan jumlah total dari nilai terprediksi *true* pada kelas itu sendiri dengan nilai yang terprediksi *false* pada kelas tersebut. Diperoleh nilai rata-rata sejumlah 52% dengan jumlah kelas positif 83%, kelas negatif sejumlah 59%, dan netral sejumlah 14%. Selain mengevaluasi performa model berdasarkan akurasi keseluruhan nilai *true* dan *false*, pada penelitian ini juga menggunakan alternatif evaluasi berdasarkan nilai akurasi pada masing-masing kelas yaitu *f1-score*, nilai ini menggabungkan antara *recall* dengan presisi. Evaluasi *f1-score* pada rasio 80:20 ini mendapat nilai rata-rata sejumlah 50% yang terdiri dari positif 68%, negatif 61%, dan netral 22%.

Adapun pada rasio 70:30 memiliki akurasi 56%, *recall* 51% yang didapat dari perhitungan rata-rata antara nilai positif 55%, nilai negatif 63%, dan netral 36%. Presisi sejumlah 48% yang didapat dari perhitungan rata-rata antara nilai positif 83%, negatif 54%, dan netral 8%. Rasio 70:30 ini memiliki nilai *f1-score* 46% dari perhitungan rata-rata antara nilai positif 66%, nilai negatif 58%, dan netral 12%.

Penelitian ini hanya murni menerapkan algoritma C4.5 tanpa adanya optimasi maupun normalisasi data pada proses *pre-processing*, sehingga data yang digunakan masih banyak yang belum bersih sepenuhnya. Terutama pada tahap *stopwords* dan *stemming*, pada saat percobaan ditemukan banyak kata yang seharusnya tereliminasi dan dikembalikan menjadi kata dasar, tetapi hal tersebut tidak bekerja dengan baik. Beberapa kata tersebut, seperti salah ketik, kata gaul, bukan bahasa Indonesia, kata berlebih, singkatan, dan kata lain yang tidak termuat

dalam kamus yang dimiliki *library* Sastrawi. Ada juga beberapa kata yang termuat dalam *library* Sastrawi, tetapi hal itu tidak bekerja.

Pada rasio 70:30, hasil dari 275 data yang diuji pada terdapat 154 nilai *true* sebagai representasi sentimen positif, negatif, dan netral. Pada nilai 154 tersebut, sentimen positif memiliki jumlah yang lebih banyak dari lainnya yaitu 104 opini, sentimen negatif memiliki 45 opini, dan sentimen netral memiliki 5 opini. Pada rasio 80:20, hasil dari 183 data yang diuji pada terdapat 108 nilai *true* sebagai representasi sentimen positif, negatif, dan netral. Pada nilai 108 tersebut, sentimen positif memiliki jumlah yang lebih banyak dari lainnya yaitu 67 opini, sentimen negatif memiliki 35 opini, dan sentimen netral memiliki 6 opini. Dapat dipahami baik dari rasio 70:30 maupun rasio 80:20 posisi sentimen positif lebih banyak daripada negatif, dan negatif lebih banyak daripada netral adalah sama dengan hasil pelabelan dari ahli yang mana posisi sentimen positif lebih unggul dari yang lain, negatif berada pada urutan kedua, dan sentimen netral yang paling sedikit jumlahnya. Kesimpulannya adalah kebanyakan masyarakat Indonesia adalah beropini positif terhadap Qatar sebagai tuan rumah piala dunia FIFA 2022.

Analisis sentimen memiliki tujuan untuk mengetahui bagaimana ekspresi orang dalam bentuk opini. Opini bukan hanya dalam bentuk verbal, melainkan dapat berbentuk nonverbal juga. Salah satu bentuk opini nonverbal tersebut adalah teks atau tulisan. Opini pada penelitian ini diperoleh dari Twitter dengan memilih topik piala dunia 2022. Pada persiapan dan penyelenggaraan piala dunia kali ini terdapat beberapa kontroversi sehingga masyarakat memiliki berbagai pandangan atas kontroversi tersebut. Tujuan dari analisis sentimen tersebut supaya dapat memilah dan membedakan mana opini yang baik dan yang buruk, sehingga hal tersebut dapat dijadikan evaluasi untuk masa depan yang lebih baik dengan harapan menjadikan opini baik untuk dioptimalkan dan menjadikan opini buruk untuk perbaikan penyelenggara piala dunia berikutnya. Adapun dalam pandangan islam, tujuan dari analisis sentimen ini seperti yang ada pada al-Qur'an Surat al-Baqarah ayat 42.

وَلَا تَلْبِسُوا الْحَقَّ بِالْبَاطِلِ وَتَكْتُمُوا الْحَقَّ وَأَنْتُمْ تَعْلَمُونَ

“Janganlah kamu campuradukkan kebenaran dengan kebatilan dan jangan pula kamu sembunyikan kebenaran, sedangkan kamu mengetahuinya” (Q.S. al-Baqarah ayat 42)

Ayat tersebut dijelaskan dalam tafsir al-Misbah bahwa pemuka agama dari bangsa Israel melakukan propaganda dengan pendapat-pendapat yang tidak sesuai dengan ajaran di kitab Taurat, mereka mengubah isi kitab Taurat dengan menambah pendapat-pendapat tersebut dan menyembunyikan perihal kenabian Muhammad SAW (Shihab, 2001). Ditinjau dari tafsir tersebut memang dalil ini ditujukan untuk bangsa Israel, namun hal tersebut juga bersifat cerita terdahulu untuk memberi pelajaran kepada umat sekarang yang mana secara makna hal tersebut juga ditujukan kepada umat Islam dari semua kalangan dan dalam bentuk kasus apapun (Lusiana, 2021). Dalam konteks tujuan dari analisis sentimen ini, berarti konsep klasifikasi untuk mengetahui dan memilah mana yang positif dan yang negatif agar tidak tercampur adalah selaras dengan ayat tersebut.

Demikian itu, supaya menjadi evaluasi agar pada masa yang akan datang menjadi lebih baik. Evaluasi juga sama seperti muhasabah yang berarti suatu pertimbangan terhadap nilai antara tindakan negatif dan positif (Nurbakhsh, dikutip dalam Netton, 2001). Adapun hadits yang menyatakan tentang pentingnya muhasabah atau evaluasi sebagai berikut.

مَنْ كَانَ يَوْمُهُ خَيْرًا مِنْ أَمْسِيهِ فَهُوَ رَاحٍ. وَمَنْ كَانَ يَوْمُهُ مِثْلَ أَمْسِيهِ فَهُوَ مَغْبُونٌ. وَمَنْ
كَانَ شَرًّا مِنْ أَمْسِيهِ فَهُوَ مَلْعُونٌ

“Barang siapa hari ini lebih baik dari hari kemarin, maka dia termasuk orang yang beruntung. Barang siapa yang hari ini sama dengan hari kemarin, maka dia termasuk orang yang merugi, dan barang siapa yang hari ini lebih buruk dari hari kemarin, maka dialah termasuk orang yang celaka” (HR. al-Hakim).

Dari hadits tersebut dapat diketahui bahwa keberuntungan atau kesuksesan didapat dari orang yang bisa memperbaiki keadaan yang sebelumnya terdapat kekurangan, sedangkan jika keadaan tersebut masih dengan perlakuan yang sama atau bahkan lebih buruk, maka akan mendapatkan kerugian dan kehancuran. Pada dasarnya, semua hal tersebut membutuhkan evaluasi agar dapat mengetahui mana yang harus dioptimalkan dan mana yang harus dihindari.

BAB V

PENUTUP

5.1 Kesimpulan

Berdasarkan pengujian dan evaluasi yang telah dilakukan, dapat disimpulkan bahwa hasil tertinggi antara dua skenario pengujian dan evaluasi performa sistem dalam memperoleh kinerja terbaik ada pada rasio 80:20 yang akurasi sebesar 59%, *recall* 57%, presisi 52%, dan *f1-score* 50%. Adapun pada rasio 70:30 memperoleh akurasi sejumlah 56%, *recall* 51%, presisi 48%, dan *f1-score* 46%. Dari persentase evaluasi model yang telah disebut, dapat diketahui bahwa algoritma C4.5 kurang maksimal dalam menangani data opini tentang Qatar sebagai tuan rumah piala dunia FIFA 2022 karena tidak bersihnya data pada tahap *pre-processing* khususnya pada *stopwords* dan *stemming* yang disebabkan adanya kekurangan *library* Sastrawi dalam melakukan *stopwords* dan *stemming*.

Pada penelitian ini, data hasil dari *scraping* mendapatkan tweet opini sejumlah 2925, namun setelah masuk proses penghapusan duplikasi menghasilkan 914 tweet opini yang akan dijadikan klasifikasi. Dari 914 data tersebut, terdapat 3 kelas sentimen yang masing-masing jumlah tweetnya adalah opini positif berjumlah 383, opini negatif berjumlah 293, dan opini netral berjumlah 238. Evaluasi menggunakan *confusion matrix* dengan skor terbaik antara dua skenario pengujian adalah rasio 80:20 memiliki data latih sebanyak 731 dan data uji sebanyak 183. Dari 183 data yang dijadikan data uji, terdapat 108 data yang bernilai *true*. Opini positif terhadap Qatar sebagai tuan rumah piala dunia FIFA 2022 adalah paling banyak diantara opini dari kelas lain yaitu opini positif sejumlah 67, opini negatif sejumlah 35, dan opini netral sejumlah 6.

5.2 Saran

Klasifikasi dalam penelitian analisis sentimen dengan algoritma C4.5 menggunakan data teks di Twitter yang telah dilakukan ini terdapat beberapa kekurangan. Oleh karena itu, untuk memberikan performa lebih tinggi pada sistem klasifikasi terdapat beberapa saran untuk dilakukan pada penelitian maupun implementasi selanjutnya. Berikut beberapa saran yang diusulkan:

1. Media sosial sangat berbeda dengan platform berita yang lebih memperhatikan kata yang baku. Oleh karena itu, perlu melakukan normalisasi kata yang salah ketik, kata gaul, bukan bahasa Indonesia, kata berlebih, singkatan, kata tidak baku, kata lain yang tidak termuat dalam kamus yang dimiliki *library* Sastrawi.
2. Membuat komparasi algoritma klasifikasi untuk menemukan model terbaik pada penelitian selanjutnya seperti Support Vector Machine (SVM), Naïve Bayes Classifier (NBC), Random Forest (RF), K-Nearest Neighbor (KNN), Logistic Regrresion (logit regression), dan lain-lain.
3. Menggunakan teknik pembelajaran mesin terbaru yaitu *reinforcement learning* yang dapat menggunakan algoritma State-action-reward-state-action, Q-Learning, Deep Q Network, dan lain-lain. Teknik ini juga disebut sebagai teknik yang semi *supervised learning*, namun lebih unggul. Teknik ini biasa dipakai untuk produk *artificial intelligence* dalam bidang apapun pada saat ini. Penggunaan data yang *multiclass* memiliki kemungkinan lebih mudah ditangani dengan teknik ini.
4. Mencari berbagai referensi *best practice* untuk data yang berjenis teks narasi berbahasa Indonesia.

DAFTAR PUSTAKA

- 2022 FIFA World Cup global opinion on Qatar as location. (2022). Statista. <https://www.statista.com/statistics/1298753/global-public-opinion-qatar-world-cup-different-country/>
- Agusta, L. (2009). Perbandingan algoritma stemming Porter dengan algoritma Nazief & Adriani untuk stemming dokumen teks bahasa indonesia. *Konferensi Nasional Sistem dan Informatika*, 196–201.
- Alrumaih, A., Al-Sabbagh, A., Alsabab, R., Kharrufa, H., & Baldwin, J. (2020). Sentiment analysis of comments in social media. *International Journal of Electrical and Computer Engineering (IJECE)*, 10(6), 5917. <https://doi.org/10.11591/ijece.v10i6.pp5917-5922>
- Ambarita, R., Azhar, A., & Abror, M. Y. (2022). *Analisis Kepentingan Qatar Dalam Penyelenggaraan Piala Dunia Fifa 2022* [PhD Thesis]. Sriwijaya University.
- Anggundia, P. A. (2017). *Implementasi Iman dan Taqwa (IMTAQ) dalam Membangun Ukhuwah Islamiyah di SMA Nahdlatul Ulama Palembang* [Diploma, UIN RADEN FATAH PALEMBANG]. <http://perpus.radenfatah.ac.id>
- Anshori, C. S. (2016). Ukhuwah Islamiyah sebagai Fondasi Terwujudnya Organisasi yang Mandiri dan Profesional. *Taklim : Jurnal Pendidikan Agama Islam*, 14(1), Article 1.
- Asir, A. (2014). Agama dan Fungsinya dalam Kehidupan Umat Manusia. *Al-Ulum Jurnal Pemikiran Dan Penelitian Ke Islaman*, 1(1), Article 1. <https://doi.org/10.31102/alulum.1.1.2014.50-58>
- Asmara, D. A., Khairani, D., & Masruroh, S. U. (2016). *Penerapan algoritma paice atau husk untuk stemming pada kamus bahasa inggris ke bahasa Indonesia*. <https://repository.uinjkt.ac.id/dspace/handle/123456789/31577>
- Asshiddiqi, M. F., & Lhaksana, K. M. (2020). Perbandingan Metode Decision Tree dan Support Vector Machine Untuk Analisis Sentimen Pada Instagram Mengenai Kinerja PSSI. *EProceedings of Engineering*, 7(3).
- Astiningrum, M., Hani'ah, M., & Pradana, Y. R. Y. (2020). Analisis Sentimen Tentang Opini Terhadap Performa Timnas Sepak Bola Indonesia Pada Twitter. *Seminar Informatika Aplikatif Polinema*, 35–39.

- Azzahra, S. A., & Wibowo, A. (2020). Analisis Sentimen Multi-Aspek Berbasis Konversi Ikon Emosi dengan Algoritme Naïve Bayes untuk Ulasan Wisata Kuliner Pada Web Tripadvisor. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 7(4), Article 4. <https://doi.org/10.25126/jtiik.2020731907>
- Baskoro, B. B., Susanto, I., & Khomsah, S. (2021). Analisis Sentimen Pelanggan Hotel di Purwokerto Menggunakan Metode Random Forest dan TF-IDF (Studi Kasus: Ulasan Pelanggan Pada Situs TRIPADVISOR). *Journal of Informatics Information System Software Engineering and Applications (INISTA)*, 3(2), Article 2. <https://doi.org/10.20895/inista.v3i2.218>
- Carvalho, F., & Guedes, G. P. (2020). *TF-IDFC-RF: A Novel Supervised Term Weighting Scheme*. <https://doi.org/10.48550/arXiv.2003.07193>
- Chamira, S. (2022). Implementasi Metode Text Mining Frequency-Invers Document Frequency (Tf-Idf) Untuk Monitoring Diskusi Online. *Journal of Informatics, Electrical and Electronics Engineering*, 1(3), Article 3.
- Databoks. (2017). *Apa Olahraga Paling Digemari di Dunia?* DataIndonesia.id. <https://dataindonesia.id/ragam/detail/apa-olahraga-paling-digemari-di-dunia>
- Djalil, B. (2014). *Ilmu Ushul Fiqih: Satu dan Dua*. Kencana Prenada Media Group.
- Elfariani, I. (2021). Prasangka dan Suudzon: Sebuah Analisa Komparatif Dari Perspektif Psikologi Barat dan Psikologi Islam. *Jurnal Psikologi Terapan (JPT)*, 2(1), Article 1. <https://doi.org/10.29103/jpt.v2i1.3621>
- Essra, A. (2016). Analisis Information Gain Attribute Evaluation untuk Klasifikasi Serangan Intrusi. *Journal Information System Development (ISD)*, 1(2). <https://ejournal-medan.uph.edu/index.php/isd/article/view/22>
- Fahrudin, T. M., Hartanto, M. H., Paramita, A. S., Aulia, A., Maulana, R. A., & Anniswa, I. R. (2022). Temu Kembali Informasi Berita Kegiatan Program Studi Menggunakan Algoritma Pembobotan TF-IDF dan Cosine Similarity. *Prosiding Seminar Nasional Teknologi Dan Sistem Informasi*, 2(1), Article 1.
- Feldman, R., & Sanger, J. (2007). *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*. Cambridge University Press.
- Filcha, A., & Hayaty, M. (2019). Implementasi Algoritma Rabin-Karp untuk Pendeteksi Plagiarisme pada Dokumen Tugas Mahasiswa. *JUITA : Jurnal Informatika*, 7(1), Article 1. <https://doi.org/10.30595/juita.v7i1.4063>

- Fiqri, K. G., Hanuranto, A. T., & Setianingsih, C. (2020). Analisis Perbandingan Klasifier Decision Tree, Random Forest, Dan Adaboost Dalam Mendeteksi Serangan Siber. *EProceedings of Engineering*, 7(1).
- Firdaus, S. A. (2020). *Prediksi Penyakit Diabetes Dengan Metode Hyperbolic Support Vector Machine* [Other, Universitas Komputer Indonesia]. https://doi.org/10/UNIKOM_Syahril%20Asidiq%20Firdaus_BAB%204.pdf
- Gymnastiar, A. (2003). *Mengatasi penyakit hati*. Penerbit Republika.
- Habibi, F. N., Setiawidayat, S., & Mukhsim, M. (2017). Alat Monitoring Pemakaian Energi Listrik Berbasis Android Menggunakan Modul PZEM-004T. *Prosiding Seminar Nasional Teknologi Elektro Terapan*, 1(01), Article 01.
- Haqie, A. B. (2013). *Sombong dan Penyembuhannya dalam Al-Qur'an (Perspektif Bimbingan dan Konseling Islam)* [IAIN Walisongo]. <https://eprints.walisongo.ac.id/id/eprint/167/>
- Hasanatuddaroini, U. (2020). *Konsep pendidikan karakter religius dan peduli sosial dalam Al-Qur'an Surat Luqman Ayat 13-19: Perspektif Tafsir Al-Munir dan Tafsir Al-Misbah* [Undergraduate, Universitas Islam Negeri Maulana Malik Ibrahim]. <http://etheses.uin-malang.ac.id/24724/>
- Hendrawan, A., Winarti, T., & Indriyawati, H. (2021). *Metode Stemming dalam Penerapan Artikel Berbahasa Indonesia*.
- Hidayat, D. R. (2013). Faktor-faktor Penyebab Kemunculan Prasangka Sosial (Social Prejudice) Pada Pelajar. *Jurnal Ilmiah Mimbar Demokrasi*, 12(2), Article 2. <https://doi.org/10.21009/jimd.v12i2.6284>
- Hidayatullah, M., Alam, S., & Jaelani, I. (2021). Sentiment Analysis of Police Performance On Twitter Users Using Naïve Bayes Method. *RISTEC : Research in Information Systems and Technology*, 2(2), Article 2.
- Ikhsan, M. A. (2017). Nilai-nilai Cinta Tanah Air dalam Perspektif al-Qur'an. *Jurnal Ilmiah Pendidikan Pancasila dan Kewarganegaraan*, 2(2), Article 2.
- Indraini, A. N., & Ernawati, I. (2022). Analisis Sentimen Terhadap Pembelajaran Daring Di Indonesia Menggunakan Support Vector Machine (SVM). *Jurnal Ilmiah FIFO*, 14(1), Article 1. <https://doi.org/10.22441/fifo.2022.v14i1.007>

- Jamaluddin, T., Bijaksana, M. A., & Asror, I. (2020). Perbandingan Algoritma Sentencepiece Bpe Dan Unigram Pada Tokenisasi Artikel Bahasa Indonesia. *eProceedings of Engineering*, 7(2), Article 2. <https://doi.org/10.34818/eoe.v7i2.13126>
- Jondri, J. (2022). Prediksi Retweet Pada Twitter Menggunakan Metode Decision Tree. *CSRID (Computer Science Research and Its Development Journal)*, 14(2), Article 2. <https://doi.org/10.22303/csrid.14.2.2022.113-124>
- Khairunnisa, S., Adiwijaya, A., & Faraby, S. A. (2021). Pengaruh Text Preprocessing terhadap Analisis Sentimen Komentar Masyarakat pada Media Sosial Twitter (Studi Kasus Pandemi COVID-19). *JURNAL MEDIA INFORMATIKA BUDIDARMA*, 5(2), Article 2. <https://doi.org/10.30865/mib.v5i2.2835>
- Khan, B., Shah, T., Khan, M., Ali, M., & Idress, N. (2020). *CART, J-48graft, J48, ID3, Decision Stump and Random Forest: A comparative study*.
- Kim, J., Kim, H.-U., Adamowski, J., Hatami, S., & Jeong, H. (2022). Comparative study of term-weighting schemes for environmental big data using machine learning. *Environmental Modelling & Software*, 157, 105536. <https://doi.org/10.1016/j.envsoft.2022.105536>
- Lajnah Pentashihan Mushaf Al-Qur'an. (2010). *Hukum, Keadilan dan Hak Asasi Manusia* (1st ed.). Lajnah Pentashihan Mushaf Al-Qur'an Badan Litbang Dan Diklat Kementerian Agama RI. <https://pustakalajnah.kemenag.go.id/detail/53>
- Liu, B. (2012). *Sentiment Analysis and Opinion Mining*.
- Lubis, A. R., Prayudani, S., Lubis, M., & Nugroho, O. (2022). Sentiment Analysis on Online Learning During the Covid-19 Pandemic Based on Opinions on Twitter using KNN Method. *2022 1st International Conference on Information System & Information Technology (ICISIT)*, 106–111. <https://doi.org/10.1109/ICISIT54091.2022.9872926>
- Lusiana, S. (2021). *Pelaksanaan Takbir Keliling Dengan Iringan Musik DJ Pada Malam Hari Raya Idul Fitri Di Desa Sukodono Kecamatan Bonang Kabupaten Demak Perspektif QS. Al-Baqarah Ayat 42* [Skripsi, IAIN KUDUS]. <http://repository.iainkudus.ac.id/5869/>
- Mahawardana, P. P. O., Sasmita, G. A., & Pratama, I. P. A. E. (2022). *Analisis Sentimen Berdasarkan Opini dari Media Sosial Twitter terhadap "Figure Pemimpin" Menggunakan Python*. 3(1), 11.

- Mahfud, F. K. R., Mudawamah, N. S., & Hariyanto, W. (2020). Sentiment Analysis of Perpustakaan Nasional Republik Indonesia Through Social Media Twitter. *MATICS: Jurnal Ilmu Komputer Dan Teknologi Informasi (Journal of Computer Science and Information Technology)*, 12(1), Article 1. <https://doi.org/10.18860/mat.v12i1.8973>
- Mahfud, F. K. R., & Tjahyanto, A. (2017). Improving classification performance of public complaints with TF-IGM weighting: Case study : Media center E-wadul surabaya. *2017 International Conference on Sustainable Information Engineering and Technology (SIET)*, 220–225. <https://doi.org/10.1109/SIET.2017.8304138>
- Maulidah, M., Gata, W., Aulianita, R., & Agustyaningrum, C. I. (2020). Algoritma Klasifikasi Decision Tree Untuk Rekomendasi Buku Berdasarkan Kategori Buku. *E-Bisnis: Jurnal Ilmiah Ekonomi Dan Bisnis*, 13(2), 89–96.
- Muttaqin, F. A., & Bachtiar, A. M. (2016). Implementasi Teks Mining Pada Aplikasi Pengawasan Penggunaan Internet Anak ‘Dodo Kids Browser.’ *Jurnal Ilmiah Komputer Dan Informatika (KOMPUTA)*, 1–8.
- Nanda, M. R. I. (2015). *Implikasi Berburuk Sangka Terhadap Kehidupan Sosial: Telaah Surat al-Hujurat Ayat 12* [Undergraduate, UIN Sunan Ampel Surabaya]. <https://eu11.proxysite.com/process.php?d=sST51gV7ZdRCkobCatEaPnwWdCOAecvECQiyLw%3D%3D&b=1&f=norefer>
- Netton, I. R. (2001). *Dunia Spiritual Kaum Sufi* (1st ed.). PT. Raja Grafindo Persada.
- Norazizah, N., Dasuki, A., & Supriadi, A. (2021). Makna Istikbār dalam Al-Qur’an: Perspektif Semantik Toshihiko Izutsu. *Syams: Jurnal Studi Keislaman*, 2(2), Article 2.
- Nuechterlein, D. E. (1976). National interests and foreign policy: A conceptual framework for analysis and decision-making. *British Journal of International Studies*, 2(3), 246–266. <https://doi.org/10.1017/S0260210500116729>
- Patel, R., & Passi, K. (2020). Sentiment Analysis on Twitter Data of World Cup Soccer Tournament Using Machine Learning. *IoT*, 1(2), Article 2. <https://doi.org/10.3390/iot1020014>
- Perdana, A., Hermawan, A., & Avianto, D. (2022). Analisis Sentimen Terhadap Isu Penundaan Pemilu di Twitter Menggunakan Naive Bayes Classifier. *Jurnal*

- Sisfokom (Sistem Informasi dan Komputer)*, 11(2), Article 2.
<https://doi.org/10.32736/sisfokom.v11i2.1412>
- Pradana, M. G. (2020). Penggunaan Fitur Wordcloud Dan Document Term Matrix Dalam Text Mining. *Jurnal Ilmiah Informatika*, 8(01), Article 01.
<https://doi.org/10.33884/jif.v8i01.1838>
- Pratama, S. R., & Mirza, A. H. (2021). Penerapan Data Mining Untuk Memprediksi Tingkat Inflasi Menggunakan Metode Regresi Linier Berganda Pada Bps Kota Palembang. *Bina Darma Conference on Computer Science (BDCCS)*, 3(2), Article 2.
- Profile of the FIFA World Cup Qatar 2022™*. (2022). FIFA Publications.
<https://publications.fifa.com/en/sustainability-report/sustainability-at-the-fifa-world-cup/profile-of-the-fifa-world-cup-qatar-2022/>
- Putri, R. R. B. (2019). *Bimbingan dan Konseling Islam dengan terapi muhasabah untuk prasangka buruk seorang anak terhadap ibu kandungnya di Kelurahan Bulak Banteng Wetan Kenjeran Surabaya* [Undergraduate, UIN Sunan Ampel Surabaya].
<https://us19.proxysite.com/process.php?d=sST51gV7ZdRCkobCatEaPnwWdCOAecvEDA23NMY%3D&b=1>
- Rahutomo, F., Putra, Y. P., & Ali, M. H. (2019). Implementasi Manhattan Distance dan Dice Similarity pada Ujian Esai Daring Berbahasa Indonesia. *Seminar Informatika Aplikatif Polinema*, 171–174. N dan A sastrawi.
- Rais, A. N., Warjiyono, W., & Kurniawan, W. (2019). Analisa Akurasi dan F1 Score pada Algoritma Smote dan Naïve Bayes pada Dataset Bank Direct Marketing. *Journal Speed – Sentra Penelitian Engineering dan Edukasi*, 11(4).
- Ramdhani, M. A., & Rahim, O. N. (2014). Analisis Sentimen Untuk Mengukur Popularitas Tokoh Publik Berdasar Data Pada Media Sosial Twitter Menggunakan Algoritma Data Mining Dengan Teknik Klasifikasi. *Jurnal Informasi*, 6(2).
- Rico, F., & Fuady, M. E. (2014). Makna Emoticon dalam Media Sosial Path. *Prosiding Hubungan Masyarakat*, 0, Article 0.
<https://doi.org/10.29313/.v0i0.157>
- Ridwansyah, T. (2022). Implementasi Text Mining Terhadap Analisis Sentimen Masyarakat Dunia Di Twitter Terhadap Kota Medan Menggunakan K-Fold Cross Validation Dan Naïve Bayes Classifier. *KLIK: Kajian Ilmiah*

Informatika Dan Komputer, 2(5), Article 5.
<https://doi.org/10.30865/klik.v2i5.362>

- Rinandyaswara, R., Sari, Y. A., & Furqon, M. T. (2022). Pembentukan Daftar Stopword Menggunakan Term Based Random Sampling Pada Analisis Sentimen Dengan Metode Naïve Bayes (Studi Kasus: Kuliah Daring Di Masa Pandemi). *Jurnal Teknologi Informasi dan Ilmu Komputer*, 9(4), Article 4. <https://doi.org/10.25126/jtiik.2022934707>
- Roihan, A., Sunarya, P. A., & Rafika, A. S. (2020). Pemanfaatan Machine Learning dalam Berbagai Bidang. *Jurnal Khatulistiwa Informatika*, 5(1), 490845.
- Romadloni, N. T., Santoso, I., & Budilaksono, S. (2019). Perbandingan Metode Naive Bayes, Knn Dan Decision Tree Terhadap Analisis Sentimen Transportasi Krl Commuter Line. *Ikraith-Informatika*, 3(2), Article 2.
- Rosid, M. A., Fitriani, A. S., Astutik, I. R. I., Mulloh, N. I., & Gozali, H. A. (2020). Improving Text Preprocessing For Student Complaint Document Classification Using Sastrawi. *IOP Conference Series: Materials Science and Engineering*, 874(1), 012017. <https://doi.org/10.1088/1757-899X/874/1/012017>
- Rosid, M. A., Muharram, F., & Affandi, G. R. (2022). Comparison of the Performance of Machine Learning Algorithms for Sarcasm Detection in Bahasa: *Procedia of Social Sciences and Humanities*, 3, 1192–1195. <https://doi.org/10.21070/pssh.v3i.253>
- Rufiyanto, A., Rochcham, M., & Rohman, A. (2021). *Penerapan Algoritma C4.5 Untuk Prediksi Kepuasan Mahasiswa Tahun 2020*. Deepublish.
- Saadah, M. N., Atmagi, R. W., Rahayu, D. S., & Arifin, A. Z. (2013). Sistem Temu Kembali Dokumen Teks Dengan Pembobotan Tf-idf Dan Lcs. *JUTI: Jurnal Ilmiah Teknologi Informasi*.
<https://doi.org/DOI:10.12962/j24068535.v11i1.a16>
- Salim, S. S., & Mayary, J. (2020). Analisis Sentimen Pengguna Twitter Terhadap Dompot Elektronik Dengan Metode Lexicon Based Dan K – Nearest Neighbor. *Jurnal Ilmiah Informatika Komputer*, 25(1), Article 1. <https://doi.org/10.35760/ik.2020.v25i1.2411>
- Santoso, B., Azis, A. I. S., & Zohrahayaty. (2020). *Machine Learning & Reasoning Fuzzy Logic Algoritma, Manual, Matlab, & Rapid Miner*. Deepublish.

- Santoso, G. T. (2021). *Analisis Sentimen Pada Tweet Dengan Tagar #bpjsrasarentenir Menggunakan Metode Support Vectore Machine (Svm)*. Universitas Islam Riau Pekanbaru.
- Saputra, E., Saleh, N., & Santalia, I. (2022). Kualitas Pelayanan Komunikasi Perpustakaan Di Dinas Perpustakaan Provinsi Sulawesi Selatan. *Jurnal Mercusuar*, 3(2), Article 2. <https://journal3.uin-alauddin.ac.id/index.php/mercusuar/article/view/31086>
- Saragih, R. R. (2016). Pemrograman dan bahasa Pemrograman. *Praktek Otomasi Perkantoran*.
- Setiabudidaya, D. (2018). *Penggunaan Piranti Lunak Jupyter Notebook Dalam Upaya Mensosialisasikan Open Science*. INA-Rxiv. <https://doi.org/10.31227/osf.io/2h7q4>
- Shihab, M. Q. (2001). *Tafsir Al Mishbah* (Vol. 6). <http://archive.org/details/tafsiral-mishbah-prof-dr.-m.-quraish-shihab->
- Shihab, M. Q. (2016). *Yang Hilang Dari Kita: Akhlak*. Lentera Hati Group.
- Singgalen, Y. A. (2021). Pemilihan Metode dan Algoritma dalam Analisis Sentimen di Media Sosial: Systematic Literature Review. *Journal of Information Systems and Informatics*, 3(2), Article 2. <https://doi.org/10.33557/journalisi.v3i2.125>
- StatCounter Global Stats. (2022). *Social Media Stats Indonesia*. StatCounter Global Stats. <https://gs.statcounter.com/social-media-stats/all/indonesia>
- Stiawan, M. M., & Hidayat, R. (2019). Pengembangan Sistem Identifikasi Fakta Dan Tidak Fakta Berita di Media Informasi Berbahasa Indonesia. *Prosiding TAU SNAR-TEK Seminar Nasional Rekayasa Dan Teknologi*, 1(1), Article 1.
- Sudiantoro, A. V., & Zuliarso, E. (2018). Analisis Sentimen Twitter Menggunakan Text Mining Dengan Algoritma Naïve Bayes Classifier. *Dinamika Informatika: Jurnal Ilmiah Teknologi Informasi*, 10(2), Article 2. <https://doi.org/10.35315/informatika.v10i2.8135>
- Suprpto, A., & Yulianto, Y. (2023). Pandangan Islam Terhadap Pengembangan Dan Pemanfaatan Sains Dan Teknologi. *Es-Syajar*, 1(1), Article 1. <https://doi.org/10.18860/es.v1i1.20423>

- Taboada, M. (2016). Sentiment Analysis: An Overview from Linguistics. *Annual Review of Linguistics*. <https://doi.org/10.1146/annurev-linguistics-011415-040518>
- Taufik, A. (2018). Komparasi Algoritma Text Mining Untuk Klasifikasi Review Hotel. *Jurnal Teknik Komputer AMIK BSI*, 4(2), Article 2. <https://doi.org/10.31294/jtk.v4i2.3461>
- Taufikurrahman, T. (2017). *Sombong dalam al-Qur'an menurut al-Maragi* [PhD Thesis]. UIN Sunan Kalijaga.
- Taufikurrahman, T. (2020). Sombong dalam al-Qur'an Sebuah Kajian Tematik. *Jurnal Tafkere*, 8(1), Article 1. <https://journal.uin-alauddin.ac.id/index.php/tafsere/article/view/14802>
- Tavana, M., Shaabani, A., Raeesi Vanani, I., & Kumar Gangadhari, R. (2022). A Review of Digital Transformation on Supply Chain Process Management Using Text Mining. *Processes*, 10(5), Article 5. <https://doi.org/10.3390/pr10050842>
- Tiffani, I. E. (2020). Optimization of Naïve Bayes Classifier By Implemented Unigram, Bigram, Trigram for Sentiment Analysis of Hotel Review. *Journal of Soft Computing Exploration*, 1(1), Article 1. <https://doi.org/10.52465/josce.v1i1.4>
- Twitter: Number of users worldwide from 2019 to 2022*. (2022). Statista. <https://www.statista.com/statistics/303681/twitter-users-worldwide/>
- Wahyuni, R. T., Prastiyanto, D., & Supraptono, E. (2017). Penerapan Algoritma Cosine Similarity dan Pembobotan TF-IDF pada Sistem Klasifikasi Dokumen Skripsi. *Jurnal Teknik Elektro*, 9(1), Article 1. <https://doi.org/10.15294/jte.v9i1.10955>
- Wang, H., Can, D., Kazemzadeh, A., Bar, F., & Narayanan, S. (2012). A System for Real-time Twitter Sentiment Analysis of 2012 U.S. Presidential Election Cycle. *Proceedings of the ACL 2012 System Demonstrations*, 115–120. <https://aclanthology.org/P12-3020>
- Wardana. (2019). *Belajar Pemrograman dan Hacking Menggunakan Python*. Elex Media Komputindo.
- Widhiasih, N. P. N. D., Suarjaya, I. M. A. D., & Wiranatha, A. A. K. A. C. (2023). Rancang Bangun Engine Sentimen Analisis Tingkat Popularitas Kopi

Berdasarkan Pengolahannya Menggunakan Teknologi Big Data. *JITTER- Jurnal Ilmiah Teknologi dan Komputer*, 4(1).

World Atlas. (2022, October 20). *The Most Popular Sports In The World*. WorldAtlas. <https://www.worldatlas.com/articles/what-are-the-most-popular-sports-in-the-world.html>

Yan, H., Ma, M., Wu, Y., Fan, H., & Dong, C. (2022). Overview and analysis of the text mining applications in the construction industry. *Heliyon*, 8(12), e12088. <https://doi.org/10.1016/j.heliyon.2022.e12088>

Yulianto, Y. (2017). Mabadi' Asyroh Nalar Fikih Sosial Ali Yafie. *SHAHIH: Journal of Islamicate Multidisciplinary*, 2(1), 21. <https://doi.org/10.22515/shahih.v2i1.725>

Yulianto, Y. (2021). *Fikih Arsitektur Islam* (M. P. I. Yulianto, Ed.). UIN Maliki Press. <http://repository.uin-malang.ac.id/9591/>

Zaskya, M., Boham, A., & Lotulung, L. J. H. (2021). Twitter Sebagai Media Mengungkapkan Diri Pada Kalangan Milenial. *Acta Diurna Komunikasi*, 3(1).

Zuhad, F., & Wilantika, N. (2022). Perbandingan Penggunaan Kamus Normalisasi dalam Analisis Sentimen Berbahasa Indonesia. *Jurnal Linguistik Komputasional*, 5(1), Article 1. <https://doi.org/10.26418/jlk.v5i1.60>

LAMPIRAN

Lampiran 1.

Daftar 100 sampel dataset berlabel

No.	Tweets	Label
0	Qatar menginvestasikan 300 miliar dolar AS sebagai tuan rumah Piala Dunia 2022 https://t.co/rYluSaGzJc	positif
1	@idextratime @IExpressSports Qatar sebagai tuan rumah sudah sukses menyelenggarakan turnamen ini dengan baik, tapi FIFA sebagai perangkat pertandingan cukup kontroversial dan banyak menguntungkan pihak tertentu 🙄	positif
2	Piala Dunia 2022 Sudah Sampai Pertandingan Final, Qatar Disebut Sukses Sebagai Tuan Rumah https://t.co/BH6dy8LBvu	positif
3	Qatar memanfaatkan posisi sebagai tuan rumah Piala Dunia dan kehebatan ekonominya untuk memaksimalkan pengaruhnya, baik dalam politik regional maupun di panggung global. Baca Trias Kuncahyono di Opini Kompas. https://t.co/ZoO90rX814 @hariankompas @iaskunz	positif
4	Prancis pasti ada di Final. Pemilihan Qatar sebagai tuan rumah Piala Dunia lepas dari kontribusi Prancis. Hubungan ekonomi bilateral kedua negara juga sangat kuat.	positif
5	Sebagai tuan rumah dimana ini merupakan kali pertama Piala Dunia digelar di tanah Arab, pemerintah Qatar pun habis-habisan dalam mempersiapkannya. Mewujudkan kemegahan acara, dana sebesar USD 220 miliar pun digelontorkan. #FIFAWorldCup	positif
6	Bertemu teman lama, Chief Investment Officer Qatar Investment Authority untuk Asia, Sheikh Faisal Thani Al-Thani. Kerja sama di bidang olahraga dengan Qatar sebagai tuan rumah Piala Dunia, akan bermanfaat bagi kita dalam menggelar Piala Dunia U-20 tahun depan. https://t.co/4ycyUAwaP0	positif
7	Seri Paduka menzahirkan ucapan tahniah kepada Sheikh Tamim dan kerajaan Qatar sebagai tuan rumah, atas kejayaan menganjur kejohanan bola sepak Piala Dunia FIFA Qatar 2022. https://t.co/xlqSvddeMs	positif

No.	Tweets	Label
8	Tahun ini, gue dukung tim dari negara Maroco Gue optimis tim jagoan gue akan jadi juara karena semangat Qatar sebagai tuan rumah merasuk ke jiwa pemain Qatar.... Menang Kalah Pilih Main Bersih #WorldKahf	positif
9	Qatar, dapat pujian yang meluas sebagai tuan rumah #WorldCup2022 karena banyak hal yang baik dan berbeda dari piala dunia sebelum2nya.Wajar, jika banyak yg “mencari” kesalahannya-apalagi salah satu yg menonjol & keras melarang bendera pelangi	positif
10	Komite ini juga memastikan persiapan, perencanaan, dan operasional Qatar sebagai tuan rumah piala dunia	positif
11	Qatar sebagai tuan rumah tidak melarang semua suku, semua agama, semua warna kulit etnis dll. Untuk masuk bahkan yg LGBT pun mungkin ada yg ke Qatar....	positif
12	Jika memang kesuksesan Qatar sebagai tuan rumah dan kiprah Maroko di #PialaDunia2022 dianggap sebagai kegemilangan Muslim, itu tidaklah salah. Yang penting, setelah kompetisi berakhir, yuk kaum muslimin rajin baca buku, kembangkan sains dan tebar damai ke seluruh penjuru bumi.	positif
13	Itu juga alasan yang bikin Qatar (sebagai tuan rumah) mendukung banget Maroko. Setiap mereka main—apalagi menang—ada bendera Maroko di mana-mana. Para pendukung Maroko juga berasa ada di rumah/negara sendiri. https://t.co/qs3YeFrDNa	positif
14	diluar dari banyaknya kecaman dan kontroversi terhadap qatar sebagai tuan rumah. overall ini jadi salah satu gelaran piala dunia terseru..	positif
15	QATARSIS #tak cuma sukses jadi tuan rumah Piala Dunia, Qatar juga berhasil menggenapi sepakbola sebagai “agama baru” yang rahmatan lil umattan. Kini tak ada lagi neraka di lapangan bola. https://t.co/MXnkXNR9xY https://t.co/AVTVIqdqgn	positif
16	Pildun Qatar emang penuh dengan kejutan 😊😊 ((Even Qatar sebagai tuan rumah is also kinda...A suprise))	positif

No.	Tweets	Label
17	@LuqmanAwalludin @detikcom Kan ada videonya..Lihat dong perbandingan supporter Qatar..Sebagai tuan rumah juga,..Berkali-kali kalah telak dari tamu..Tidak satupun supporter nya marah, anarkis, banting-banting kursi apalagi merusak fasilitas umum..Tidak satupun ada perusakanMereka tenang,..Tertib..	positif
18	Piala Dunia 2022 bakal jadi Piala Dunia termahal sepanjang sejarah. Dana yang sudah dihabiskan tuan rumah Qatar mencetak rekor sebagai yang termahal. Hal itu juga membuat hadiah uang ikut bertambah!#StartingEleven #PialaDunia #TopListhttps://t.co/x9kfgBPwtn	positif
19	Orang kira pildun kali ini ga bakal seseru sebelum-sebelumnya tapi ternyata seru bgt. Apalagi Qatar jd tuan rumah wow megah betul. Banyak kisah-kisah menarik dari pildun kali ini, bisa diambil sebagai pelajaran juga	positif
20	Terpilihnya Qatar sebagai tuan rumah Piala Dunia 2022, membuat banyak orang mengunjungi negara tersebut untuk menyaksikan pertandingan sepak bola secara langsung. Selain menyaksikan pertandingan bola, para wisatawan juga bisa mencoba makanan khas Qatar.https://t.co/RglLggq8d	positif
21	@anonapostate @MichaelRainhard @dina_sulaeman Qatar sudah jelas menghargai event sepak bola terbesar di dunia justru kaum" lu yang gak menghargai qatar sebagai tuan rumah.	positif
22	@abdulrachim88 Nih model2 @nahdlatululama dan donaturnya mungkin kesel melihat indahnya qatar sebagai tuan rumah	positif
23	Saya ingin memulai dengan suara lantang menyampaikan selamat "mabruk dan mubarak" kepada negara Qatar sebagai negara Muslim dan Timur Tengah pertama, tidak saja menjadi tuan rumah sebuah perhalatan fenomenal, tapi juga telah membuktikan sebuah kemewahan... https://t.co/3sI0FeePIh	positif
24	Yang bikin seru piala dunia tahun ini tuh selain tim kuda hitam bikin kejutan, Qatar sebagai tuan rumah juga mengubah cara pandang dunia melihat Islam dengan disisipkan pesan ² dakwah lewat olahraga dan lewat moment ini	positif

No.	Tweets	Label
25	@radixwp Qatar menerapkan hukum negara ya wajar , jika qatar ke eropa juga mereka akan maklumi itu aturan merekakan itu tuan rumah nya piala dunia , ya harus patuhi lah :v anda saja kalau ke bali juga banyak melihat sesajen itu aturan sebagai warga bali dan kita sebagai wisatawan hormati	positif
26	@tvOneNews Kemenangan Qatar sebagai tuan Rumah... Tamu yg pecingkraan pulang duluan.	positif
27	Lanjut 4 Begitu juga dengan Qatar, sebagai tuan rumah dia mempunyai aturan aturan negara kedaulatannya. Jika pergi ke Barat, sebagai tamu ga mungkin juga kan larang orang pake baju tp ga pake BH 😊Dibalik aja ...	positif
28	Sobat MORA!Saat ini, seluruh mata dunia sedang fokus pada Qatar sebagai tuan rumah Piala Dunia 2022. Pada bulan Desember tahun 2006, dunia juga pernah fokus pada Qatar ketika kala itu Qatar menjadi tuan rumah event olahraga terbesar di Asia yaitu Asian Games XV. https://t.co/IGfjCjZCt	positif
29	@aheryawan Ternyata bukan sekedar sepak bola, Qatar juga menunjukkan sebagai tuan rumah yang ramah dan menghormati tamunya.	positif
30	Sejak ditunjuk FIFA menjadi tuan rumah Piala Dunia 2022 pada 2010, Qatar tak hanya menyiapkan infrastruktur, tetapi juga kemajuan sepak bolanya. Indonesia sebagai tuan rumah Piala Dunia U-20 2023 harus belajar. #Olahraga #AdadiKompas https://t.co/hkkKEmpv5M	positif
31	@6undul0h World cup qatar membuka pintu bagi ummat manusia utk mengenal dan mempelajari islam yg sesungguhnya dan mereka melakukan itu dgn hikmah dan mengedepankan sifat keramah tamahan sebagai tuan rumah..... insyaallah dunia lebih islami setelah piala dunia qatar.	positif
32	Menepis tuduhan buruk tentang Qatar sebagai tuan rumah piala dunia 2022#FIFAWorldCupQatar2022	positif

No.	Tweets	Label
33	Belakangan ini ramai pembicaraan tentang Qatar sebagai tuan rumah piala dunia 2022. Menariknya, piala dunia kali ini bernuansa penuh Islami. Nah, kira-kira bagaimana ya sejarah Islam di Qatar? Simak dalam artikel berikut, yuk! https://t.co/4hrjqHo3I#Qatar2022#FIFAWorldCup	positif
34	@CNNIndonesia Qatar negara paling sukses sebagai tuan rumah PD sepanjang sejarah	positif
35	Korsel harusnya mencontoh Qatar sebagai tuan rumah yang baik dan bersih. Qatar nggak pernah nyogok wasit dan sadar diri klw mereka masih di level Asia. Sedangkan Korsel, mereka jadi salah satu tuan rumah PD 2002 yang sangat curang dan nggak tau diri.	positif
36	Sebagai tuan rumah FIFA World Cup 2022 atau Piala Dunia, Qatar memiliki keunikan tersendiri hingga mencuri perhatian dunia saat ini. Berikut fakta seputar Qatar menarik untuk diketahui. MCP#iNews #Berita #FIFAWorldCup #PialaDunia2022 #Qatar https://t.co/RXD11bXfCJ	positif
37	@dina_sulaeman Toleransi itu ketika Qatar jadi tuan rumah melarang MIRAS dan LGBT/ONE LOVE yang kalian sebagai tamu harus menghormati.	positif
38	Mengucapkan, terima kasih sebesar-besarnya kpd.yth Tim Qatar, atas partisipasinya dalam gelaran turnamen sepakbola antar negara tahun 2022, dan juga atas kontribusinya dan perannya sebagai tuan rumah pada tahun ini.	positif
39	Betul lauk Qatar ni. Lauk sebijik macam Malaysia main Piala Asia 2007 layak sebagai tuan rumah	positif
40	Bikin Bangga! Ahli Tata Kota Indonesia Ikut Rancang Masterplan Qatar Sebagai Tuan Rumah Piala Dunia https://t.co/rdXxRiBxWU	positif
41	di sisi yang lain, sebagai tuan rumah, rules atau pengondisian yang disyaratkan sama Qatar ke FIFA itu harusnya diakomodasi dengan baik, lagian kan gaada hubungannya soal kampanye lgbtq+ sama bola anjiiiiirrrrrrrr 😭😭😭😭😭😭😭	positif

No.	Tweets	Label
42	Dalam persediaan Qatar sebagai tuan rumah FIFA 2022, Qatar telah membina 7 buah stadium baru beserta lain-lain fasiliti sukan. #FIFAWorldCup2022 #pialadunia2022 #pialaduniaqatar2022 https://t.co/zDIz5OXyD3	positif
43	Kenapa anda menolak menggunakan ban kapten LGBT?CR7 : Saya mengikuti aturan Qatar sebagai tuan rumah Piala Dunia 2022 dan menghormati ummat muslim.Misi kami datang hanya untuk menang.We Respect! https://t.co/kztfVcyXF	positif
44	Setiap negara memiliki peraturan tersendiri bagi para tamu yang akan datang. Kali ini, Qatar sebagai tuan rumah Piala Dunia 2022 membuat peraturan untuk berpakaian sopan dan tertutup. https://t.co/Oy0I7ojDcq	positif
45	Sebagai tuan rumah piala dunia 2022, ternyata Qatar menghabiskan Rp 3.344 triliun guys 🤯 Biarpun termasuk negara kecil, Qatar tuh kaya raya banget yah 🙌 #bisnismiracle #suksestanpabatas #qatar2022 #qatarworldcup2022 #pialadunia2022 #pialaduniaqatar2022 https://t.co/o9dG3mVP5n	positif
46	Dia mengatakan Qatar berhasil menggelar Piala Dunia secara lancar dan tertib tanpa ada pemberitaan tentang rusuh ataupun perkelahian.Sebelumnya Qatar sebagai Tuan Rumah Piala Dunia diprotes krn memiliki aturan yg berbeda dengan perhelatan sebelumnya.	positif
47	Qatar sebagai tuan rumah menjadi energi tambahan bagi negara tetangga wilayah Asia - Afrika	positif
48	Gua setuju. Tolong hargai culture Qatar sebagai tuan rumah. Gak semua open minded dapat diterapkan di berbagai negara. Gak semua negara setuju dengan one love. Hendaknya bersikaplah yg sopan kalo mau bertamu.	positif
49	Qatar tampil sebagai tuan rumah Piala Dunia pada edisi kali ini, mencatatkan debutnya di ajang yang paling bergengsi tersebut. Seluruh peserta Piala Dunia 2022 juga membawa kekuatan tempur	positif

No.	Tweets	Label
	berbahaya.#IDNTimesInfografis #PestaDiQatar #PialaDunia2022 #PialaDunia https://t.co/Jj7pENE838	
50	Tapi justru netijen bilang gini: 🗨️ Mereka ini debutan & tegangnnya luar biasa, wajar kalo kalah... 🗨️ Tim bolanya Qatar kalah tapi sebagai tuan rumah mereka menang... 🗨️ Permainan Qatar ada potensi mengarah bagus, cuma kualitas & mental pemain juga berpengaruh...(5/18) https://t.co/uKg2WVRX2l	positif
51	📷 Souvenir Qatar Buat Para Suporter Sebagai tuan rumah Piala Dunia... Qatar bagiin souvenir buat para suporter yang hadir di laga pembuka 😊 Isinya banyak banget ada kaos, stiker, minyak wangi, dan lainnya. (2/18) https://t.co/G10IyfbclT	positif
52	Sebagai tuan rumah Piala Dunia 2022, Qatar dan Doha menjadi lebih terkenal. Di Doha, tepatnya di Al Courniche terdapat satu monumen yang unik, yakni The Pearl Monument. Monumen ini berupa sebuah kerang raksasa dengan mutiara di dalamnya. #Headline https://t.co/PcTfnSD6HE	positif
53	Masalah lgbt, qatar dah bener, dia sebagai tuan rumah mengajukan hal tersebut kan berdasar ya. Kalo dari kacamata agama karena jelas salah, dan qatar dengan mayoritas muslim, menjadi bukan budaya mereka, artinya ada norma sosial disitu. Ya, istilahnya menghargai yg punya rumah.	positif
54	Fakta Kehidupan di Qatar, Tuan Rumah Piala Dunia 2022 yang Dinobatkan sebagai Negara Teraman – Skor – https://t.co/oOfCt7QcS5 – #Arenanews #OmKamidi- https://t.co/YjAvUgBZnZ	positif
55	@sukma_rio @SiaranBolaLive Ok lah kalau alasannya krn menang. Minimal itu sebagai ucapan terima kasih kepada tuan rumah krn di Qatar tim negaranya memetik kemenangan. Masa hal simple gitu aja gak bisa nangkep?	positif
56	@zuriat_fadil Ini yg ngelarang lgbt aturan dinegara penyelenggara, bukan fansKlo mau lgbt diterima waktu ajang world cup yaa knpa	positif

No.	Tweets	Label
	ga dari dulu aja demo, tolak qatar sebagai tuan rumah world cup piala duniaa	
57	Pembangunan Infrastruktur Qatar sebagai Tuan Rumah Piala Dunia Banyak Menarik Perhatian Berbagai Negara di Dunia https://t.co/ybpdf3ONmZ	positif
58	Konfederasi Sepak Bola Asia (AFC) menepis kritik yang menyebut Qatar mempersiapkan diri sebagai tuan rumah Piala Dunia 2022 sambil melakukan pelanggaran Hak Asasi Manusia (HAM). Justru sebaliknya, AFC malah memuji Qatar.#HAM #AFC #pialadunia #Lampungpost https://t.co/KodLm02e8n	positif
59	@saef_sumarno @CNNIndonesia Perancis melarang cadar di negaranya sendiri. Timnas Perancis menghormati aturan Qatar sebagai tuan rumah World Cup.	positif
60	Gapapa, setidaknya sebagai tuan rumah sudah bikin sejarahSukses Sebagai Tuan Rumah Tak Berimbas ke Timnas Qatar, Kembali Telan Kekalahan Usai Dibekuk Ekuador 0-2 https://t.co/hysqLmEHWC	positif
61	Jadi pakar bola satu kali lagi:Terima kasih juga buat Qatar sebagai tuan rumah yang sudah meramaikan piala dunia haha. Emang secara permainan dan basic sepak bola dari passing dan control pun masih ada gep dengan negara lain. Gapapa sekarang udah bisa fokus jadi tuan rumah 😊	positif
62	Tuan rumah Qatar sebagai Tim pertama yang tersingkir karena telah kalah dua kali yakni dari Ekuador dan dari Senegal...nasib nasib... ngga apa-apa yg penting nanti tercatat sebagai penyelenggara Piala Dunia yg sukses ...edisi menghibur diri...#Qatar2022	positif
63	Hazard menilai aksi Der Panzer tersebut sarat akan muatan politik. Sebagai pemain sepak bola, tidak seharusnya mereka mencampurkan urusan politik ke lapangan hijau.Hazard lebih memilih menghormati tuan rumah Qatar yang dengan jelas melarang penggunaan atribut LGBTQ.	positif

No.	Tweets	Label
64	@CNNIndonesia Respect Bapak Lloris terkait statement dia yg hormati peraturan qatar sebagai tuan rumah 😊	positif
65	@wptlzkmoonb @azizaha2z2ih @tanyarlifefes fyi: dalam sepakbola gaboleh ada politik sih ye setau gue sih sebagai penikmat bola... dan balik lagi qatar negara islam, so ga ada salahnya mereka melarang karena "main di tanah mereka(tuan rumah) serta mempunyai aturan dalam negaranya" wajar kok menurut gue. #cmiiw	positif
66	Ribet amat sih kaum pelangi, piala dunia itu ajang olahraga kompetisi bola bukan ajang kampanye kaum kalian. Kebetulan Qatar sebagai tuan rumah yg notabenenya menolak keras keberadaan LGBTJika kalian ingin di hargai coba menghargai juga peraturan serta budaya negara Qatar.	positif
67	Qatar akhirnya memberi gol pertama di Piala Dunia sebagai tuan rumah! Namun, Senegal bermain lebih dominan dan memenangkan pertandingan.#Senegal #Qatar #FIFAWorldCup #PialaDunia2022 #Sepakbola #PialaDunia #Indosport https://t.co/er0DhDXuPD	positif
68	Qatar ada 12 tahun untuk sediakan squad sebagai tuan rumah piala dunia. Pentas ini terlalu besar untuk mereka	positif
69	QATAR 🇪🇲 sebagai tuan rumah maka aku mendukungmu. Tapi apalah dayamu saat ini	positif
70	Konfederasi Sepak Bola Asia (AFC) menepis kritik yang menyebut Qatar mempersiapkan diri sebagai tuan rumah Piala Dunia 2022 sambil melakukan pelanggaran Hak Asasi Manusia (HAM). Justru sebaliknya, AFC malah memuji Qatar#MedcomDemamBola https://t.co/PhQNCCcImi via @medcom_id	positif
71	Qatar sebagai Tuan Rumah Piala Dunia 2022, Tawarkan Ragam Pertunjukan Festival Musik – Tribunnews – https://t.co/Y2fKDBLc97 – #Arenanews #OmKamidi- https://t.co/JwLbObRN4E	positif
72	Ayok Qatar QAQAQAKamu pasti bisa comebackDan menang dipertandingan keduaMasak sebagai tuan rumah gak bisaTim Asia lainnya bisa loh #Qatar	positif

No.	Tweets	Label
73	@springblo_om Padahal ya, sebagai tamu apa salahnya menghargai tuan rumah, tuan rumah juga punya aturan. Itu tempat2 mereka gt, terserah mereka mau ngijinin apa engga,Jujur aku kesel sama org2 yg masih memperlakukan di Qatar gak boleh ini itu.Kayak lu tamu tapi lu mau seenak Dewe gt	positif
74	terlepas dari agama dan pose foto jerman, tiap negara tuh punya culture masing masing masa mau dipaksain? qatar sebagai tuan rumah punya privilege buat ngasi rules yang sudah seharusnya dihormati negara lain yang masuk sebagai tamu 😊	positif
75	@hotuyee @bigwinlla @tanyarlfees Padahal kayak swiss, Prancis yang notabene nya dukung lgbt aja ga macem macem dan tetep menghargai Qatar sebagai tuan rumah. Dahal event nya juga sebentar cuman setahun doang itu juga kalau lolos ya kalau ngga tinggal caw. Gitu aja ko ribet	positif
76	@mannschaft89 @raihanalthff @ilhamzada Anda gagal paham,sebagai tuan rumah Qatar punya aturan,punya kedaulatan yg harusnya dihargai,dihormati oleh siapapun Tamu yg hadir.Ibarat ada orang masuk rumahlu,trus berbuat seenaknya.Apa lu akan diam aja?kan nggak!!!	positif
77	AFC Bela Tuan Rumah Piala Dunia 2022 Qatar, Tuding Kritik Sebagai Sikap Munafik https://t.co/WBxcs49H2a	positif
78	@Donutbanana73 @tanyarlfees Gak perlu memuskanahkan. Itu hak asasi mereka. Terserah mereka...Tapi yo please lah, gausah mengaitkan sepak bola dengan politik dan kampanye2 seperti ini. Cukup hargai Qatar sebagai tuan rumah yang berpegang teguh pada nilai-nilai norma disana. Simple..Saling menghargai.	positif
79	@hotuyee @tanyarlfees yg bikin keputusan & berwenang juga FIFA atas penentuan negara tuan rumah. Qatar nya tu melarang campaign & penggunaan atribut yg identik dg lgbt karena ngga sesuai dengan aturan & latar belakang negaranya. Qatar memiliki hak karena ia sebagai tuan rumah.	positif
80	Meskipun banyak pihak yang kejang-kejang, protes dengan dalih Hak Asasi Manusia. Kalian itu tamu, jadi hormati aturan Qatar	positif

No.	Tweets	Label
	sebagai tuan rumah dong. Dasar kaum wabah terkutuk, baik pelaku dan pendukungnya NORAK!Ada pepatah "Dimana bumi dipijak, disitu langit dijunjung" https://t.co/QsouAINNbo	
81	Qatar sebagai tuan rumah @FIFAWorldCup 2022 sangat mengerti bagaimana esensi dari Piala Dunia adalah ajang untuk silaturahmi antar suku & bangsa agar saling mengenal dan dapat menciptakan ketertiban dunia, melalui olahraga. 🏆	positif
82	@CatHoolicisme @gfrx_ @memefess lah suka2 qatar lah, kan dia tuan rumahnya. sebagai tamu ya harusnya ngikutin peraturan tuan rumah.	positif
83	@hotuyee @bigwinlla @tanyarlfe disini tuh posisinya qatar sebagai TUAN RUMAH dan negara lain sebagai TAMU. udah seharusnya menghormati qatar yg tidak mendukung lgbt. world cup di qatar sebulan berasa seumur hidup.	positif
84	@eccejentesiast @Askrlfess Iyaaa, padahal negara2 lain yg membebaskan lgbt pun kaya jepang, kanada, prancis gitu mereka masi respect qatar sebagai tuan rumah dan ga neko2	positif
85	@crutskii @AdeKikiSumarna1 @bgdiuuuuuu @idextratime Gini lo boti, di Perancis dilarang untuk mengenakan jilbab, sama seperti di Qatar, ada aturan tersendiri dari negara tuan rumah. Hak ini adl privilege sebagai penyelenggara, FIFA dan negara Eropa dgn hal ini, mendiskreditkan Qatar di rumahnya sendiri. Kulit putih menjjjikan.	positif
86	@tanyarlfe Selain itu, Qatar ini sebagai tuan rumah & negara yg mayoritasnya muslim meminta untuk dihormati karena culture di sana yg kontra dengan lgbt serta ada larangan alkohol.The guest must respect the host. Basic manner.	positif
87	Keberkahan Qatar sebagai Negeri Muslim menjadi Tuan Rumah Piala Dunia 2022 yakni sebagian Muslim di sana Berdakwah dan Alhamdulillah banyak yg bersyahadat orang-orang yg datang dari Negeri ² Kufar yg ingin mendukung Timnas kesayangan nya... https://t.co/Tv0GgZgOwr	positif

No.	Tweets	Label
88	5. Qatar adalah negara Timur Tengah pertama yang menjadi tuan rumah Piala Dunia sekaligus menjadi negara kedua di Benua Asia yang mendapatkan kehormatan tersebut. Piala Dunia pertama kali mendaratangi Asia pada tahun 2002 lalu, dengan Korea-Jepang sebagai tuan rumah bersama.	positif
89	Segala hal soal Qatar sebagai tuan rumah membuat Piala Dunia kali ini lebih mencuri perhatian. 1. Piala Dunia 2022 ini merupakan Piala Dunia paling mahal. Diperkirakan Qatar sudah mengeluarkan sekitar 200 miliar USD untuk berbagai keperluan infrastruktur.	positif
90	@yedamlife @misrror @tanyarlifes Konteksnya kan fifa, yg mau bunuh saha anjir? Orang pada main bola. Lagian qatar cuma minta dihormati sebagai tuan rumah karena culture disana kontra lgbt.	positif
91	@nooooookia @aulinatsya @cermateguh @tanyakanrl Apa salahnya qatar teguh pendirian? Validasi kaum pelangi tidak diterima di negara qatar, dan qatar sebagai tuan rumah punya hak akan itu, dan qatar sendiri bilang kalau kalian mau diakuin ya di lingkungan kalian aja tapi ga di negara mereka.	positif
92	Terkait atribut LGBTQ di FIFA ini, menurut aku ya seharusnya mematuhi policy Qatar sebagai negara tuan rumah. Konsepnya sederhana si, bukannya melanggar HAM or whatever you name it, tapi lebih ke menghargai tuan rumah. Masa ga paham ya sesederhana itu? Tiap negara beda aturan.	positif
93	@SiaranBolaLive Apa salahnya kita menghargai Qatar sebagai tuan rumah? Mereka menghabiskan begitu banyak uang, tenaga, bahkan waktu yg tidak sedikit untuk event ini. Dimana letak otakmu yg katanya manusia paling punya adab?	positif
94	@itsmenayaah @nukunaru @Askrlfess Jadi, Qatar kan tuan rumah World Cup 2022 trus ngelarang hal yang berbau LGBT harusnya mereka sebagai TAMU hormat dong sama TUAN RUMAH di mana bumi dipijak di situ langit dijunjung tp mereka ngeyel ngotot kampanye kaum mereka gitu krn ajang pildun kan gede jd sorotan org di dunia	positif

No.	Tweets	Label
95	@idextratime Dihargai aja tuan rumah, perhelatan ga lama juga masih maksain nilai sendiri. Sebagai tuan rumah Qatar berhak pegang teguh nilai mereka, ga usah tersinggung nih orang2 putih.	positif
96	MasyaAllah TabarakaAllah.. sepanjang Qatar menjadi Tuan rumah perhalatan sepak bola,, sebanyak 558 supporter masuk Agama Islam 😊 Qatar bukan saja sengaja bermewah” mengadakan sepak bola akan tetapi menggunakan sepak bola juga sebagai ladang Dakwah mensyiarkan Agama Islam 👍 https://t.co/MditLPiuRL	positif
97	@mannschaft89 @raihanalthff @ilhamzada Point nya simpel aja bang, menghargai Qatar sebagai tuan rumah. Pngen dihargai tapi kok gak mau menghargai wkwk	positif
98	@martinhtpea_ @veliputt @Askrfess logikanya kalo ada orang bertamu ke rumah lo pasti lo mengharuskan dia buat buka sandal/sepatu karna buat ngehargai lo sebagai tuan rumah. ya begitu pula sama Qatar ada aturan2 tertentu yg mereka buat sebagai tuan rumah. hadehh gini aja gapaham	positif
99	@Askrfess kalo gamau di rasisin, tolong jangan rasis, qatar punya budayanya sendiri dan agama yang mereka junjung, hargai tuan rumah kalau mau dihargai sebagai tamu	positif
100	Yaudah anjir banyak bacot, kalo ga bisa ikutin aturan di suatu negara yaudah, ribet amat lo. Udah jelas jelas Qatar ditunjuk sebagai tuan rumah,ya mau gak mau ikutin lah aturan dan budaya disana. Ribet amat anjrit, acara skala dunia nih yang “ditunggangi” buat cari pembelaan	positif

Lampiran 2.

Daftar data latih dan data uji

Rasio 70:30	
Data	Nomor Dokumen
Latih	0, 1, 2, 6, 7, 8, 9, 11, 12, 13, 15, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 29, 32, 33, 34, 35, 36, 37, 38, 40, 42, 43, 44, 47, 49, 50, 52, 53, 54, 55, 56, 57, 59, 60, 62, 63, 66, 67, 68, 69, 70, 71, 72, 74, 75, 76, 77, 79, 82, 84, 87, 88, 89, 90, 93, 94, 96, 98, 100, 104, 106, 107, 110, 113, 114, 115, 116, 117, 119, 120, 122, 125, 129, 130, 131, 133, 135, 138, 139, 140, 141, 142, 147, 148, 149, 150, 151, 152, 153, 154, 156, 158, 160, 161, 162, 165, 166, 167, 168, 169, 171, 172, 173, 174, 176, 178, 179, 181, 182, 183, 184, 185, 187, 188, 191, 193, 196, 198, 200, 202, 203, 204, 205, 207, 208, 210, 212, 213, 215, 217, 218, 219, 220, 221, 222, 223, 225, 226, 228, 229, 230, 231, 233, 234, 236, 240, 241, 242, 244, 245, 248, 249, 250, 251, 252, 253, 254, 256, 257, 258, 259, 260, 262, 265, 267, 268, 271, 272, 273, 274, 275, 276, 277, 278, 280, 281, 282, 283, 284, 285, 286, 289, 291, 292, 295, 296, 297, 298, 299, 300, 302, 304, 305, 306, 307, 308, 309, 310, 311, 312, 313, 315, 316, 317, 318, 320, 321, 322, 325, 326, 327, 329, 331, 332, 334, 335, 337, 338, 340, 341, 342, 343, 344, 345, 348, 351, 353, 355, 356, 359, 361, 363, 364, 365, 366, 367, 368, 369, 370, 372, 373, 374, 375, 377, 378, 379, 380, 382, 383, 384, 389, 390, 392, 394, 396, 397, 398, 399, 400, 402, 403, 404, 405, 406, 408, 410, 412, 413, 414, 416, 417, 418, 419, 420, 421, 422, 423, 424, 425, 426, 427, 429, 430, 432, 433, 434, 435, 436, 438, 441, 442, 443, 444, 445, 447, 449, 451, 452, 453, 454, 455, 456, 457, 459, 460, 462, 463, 467, 468, 470, 471, 472, 473, 474, 476, 477, 479, 480, 481, 482, 483, 484, 485, 486, 489, 490, 491, 492, 493, 494, 495, 496, 497, 498, 499, 500, 502, 503, 504, 505, 506, 507, 508, 509, 510, 511, 512, 513, 514, 515, 516, 517, 519, 520, 521, 524, 526, 527, 528, 529, 531, 533, 537, 538, 540, 541, 543, 544, 545, 546, 547, 548, 549, 551, 553, 555, 557, 558, 559, 560, 561, 563, 564, 566, 568, 569, 572, 573, 575, 576, 577, 579, 580, 581, 582, 583, 584, 586, 587, 588, 590, 591, 594, 596, 597, 598, 599, 601, 602, 603, 604, 605, 606, 607, 608, 609, 611, 616, 618, 619, 621, 622, 623, 625, 626, 628, 630, 634, 635, 636, 637, 639, 640, 641, 643, 645, 647, 650, 653, 654, 655, 656, 657, 658, 659, 660, 661, 662, 663, 664, 666, 667, 668, 670, 671, 672, 673, 676, 677, 678, 679, 681, 682, 683, 684, 685, 686, 687, 688, 690, 692, 693, 694, 695, 696, 699, 700, 701, 702, 703, 704, 708, 709, 711, 712, 714, 716, 717, 718, 719, 720, 721, 723, 725, 726, 727, 728, 729, 731, 732, 733, 735, 736, 738, 739, 741, 744, 747, 749, 751, 752, 754, 755, 756, 758, 761, 763, 764, 765, 766, 767, 769, 770, 771, 773, 774, 777, 779, 781, 782, 783, 784, 785, 786, 788, 789, 790, 793, 794, 796, 797, 798, 799, 800, 802, 804, 805, 808, 809, 810, 811, 812, 814, 817, 818, 819, 821, 823, 824, 826, 827, 828, 829, 830, 831, 832, 833, 834, 835, 837, 838, 839, 840, 841, 842, 843, 844, 845, 850, 851, 853, 854, 855, 856, 857, 858, 859, 860, 862, 863, 864, 866, 867, 868, 869, 870, 871, 872, 873, 874, 875, 876, 877, 880, 881, 882, 883, 884, 885, 886, 888, 889, 891, 892, 893, 894, 895, 896, 897, 898, 899, 900, 904, 905, 907, 908, 909, 912, 913

Rasio 70:30	
Data	Nomor Dokumen
Uji	3, 4, 5, 10, 14, 16, 27, 28, 30, 31, 39, 41, 45, 46, 48, 51, 58, 61, 64, 65, 73, 78, 80, 81, 83, 85, 86, 91, 92, 95, 97, 99, 101, 102, 103, 105, 108, 109, 111, 112, 118, 121, 123, 124, 126, 127, 128, 132, 134, 136, 137, 143, 144, 145, 146, 155, 157, 159, 163, 164, 170, 175, 177, 180, 186, 189, 190, 192, 194, 195, 197, 199, 201, 206, 209, 211, 214, 216, 224, 227, 232, 235, 237, 238, 239, 243, 246, 247, 255, 261, 263, 264, 266, 269, 270, 279, 287, 288, 290, 293, 294, 301, 303, 314, 319, 323, 324, 328, 330, 333, 336, 339, 346, 347, 349, 350, 352, 354, 357, 358, 360, 362, 371, 376, 381, 385, 386, 387, 388, 391, 393, 395, 401, 407, 409, 411, 415, 428, 431, 437, 439, 440, 446, 448, 450, 458, 461, 464, 465, 466, 469, 475, 478, 487, 488, 501, 518, 522, 523, 525, 530, 532, 534, 535, 536, 539, 542, 550, 552, 554, 556, 562, 565, 567, 570, 571, 574, 578, 585, 589, 592, 593, 595, 600, 610, 612, 613, 614, 615, 617, 620, 624, 627, 629, 631, 632, 633, 638, 642, 644, 646, 648, 649, 651, 652, 665, 669, 674, 675, 680, 689, 691, 697, 698, 705, 706, 707, 710, 713, 715, 722, 724, 730, 734, 737, 740, 742, 743, 745, 746, 748, 750, 753, 757, 759, 760, 762, 768, 772, 775, 776, 778, 780, 787, 791, 792, 795, 801, 803, 806, 807, 813, 815, 816, 820, 822, 825, 836, 846, 847, 848, 849, 852, 861, 865, 878, 879, 887, 890, 901, 902, 903, 906, 910, 911

Rasio 80:20	
Data	Nomor Dokumen
Train	0, 1, 2, 3, 5, 6, 7, 8, 9, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 29, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 42, 43, 44, 47, 49, 50, 52, 53, 54, 55, 56, 57, 59, 60, 62, 63, 64, 66, 67, 68, 69, 70, 71, 72, 74, 75, 76, 77, 79, 82, 84, 87, 88, 89, 90, 93, 94, 95, 96, 97, 98, 100, 104, 105, 106, 107, 110, 111, 112, 113, 114, 115, 116, 117, 119, 120, 122, 125, 127, 128, 129, 130, 131, 132, 133, 134, 135, 137, 138, 139, 140, 141, 142, 146, 147, 148, 149, 150, 151, 152, 153, 154, 156, 158, 159, 160, 161, 162, 163, 165, 166, 167, 168, 169, 170, 171, 172, 173, 174, 176, 178, 179, 181, 182, 183, 184, 185, 186, 187, 188, 190, 191, 193, 196, 198, 199, 200, 202, 203, 204, 205, 207, 208, 209, 210, 212, 213, 215, 217, 218, 219, 220, 221, 222, 223, 224, 225, 226, 228, 229, 230, 231, 233, 234, 236, 237, 240, 241, 242, 244, 245, 247, 248, 249, 250, 251, 252, 253, 254, 256, 257, 258, 259, 260, 261, 262, 265, 267, 268, 269, 271, 272, 273, 274, 275, 276, 277, 278, 279, 280, 281, 282, 283, 284, 285, 286, 288, 289, 291, 292, 294, 295, 296, 297, 298, 299, 300, 301, 302, 303, 304, 305, 306, 307, 308, 309, 310, 311, 312, 313, 314, 315, 316, 317, 318, 320, 321, 322, 325, 326, 327, 328, 329, 330, 331, 332, 334, 335, 337, 338, 340, 341, 342, 343, 344, 345, 347, 348, 349, 351, 353, 354, 355, 356, 357, 359, 360, 361, 362, 363, 364, 365, 366, 367, 368, 369, 370, 372, 373, 374, 375, 377, 378, 379, 380, 382, 383, 384, 387, 388, 389, 390, 392, 394, 396, 397, 398, 399, 400, 402, 403, 404, 405, 406, 408, 409, 410, 411, 412, 413, 414, 415, 416, 417, 418, 419, 420, 421, 422, 423, 424, 425, 426, 427, 429, 430, 432, 433, 434, 435, 436, 438, 441, 442, 443, 444, 445, 447, 448, 449, 451, 452, 453, 454, 455, 456, 457, 458, 459, 460, 461, 462, 463, 464, 467, 468, 470, 471, 472, 473, 474, 476, 477, 479, 480, 481, 482, 483, 484, 485, 486, 488, 489, 490, 491, 492, 493, 494, 495, 496, 497, 498, 499, 500, 501, 502, 503, 504, 505, 506, 507, 508, 509, 510, 511, 512, 513, 514, 515, 516, 517, 519, 520, 521, 524, 526, 527, 528, 529, 531, 532, 533, 535,

Rasio 80:20	
Data	Nomor Dokumen
	537, 538, 539, 540, 541, 542, 543, 544, 545, 546, 547, 548, 549, 551, 553, 555, 557, 558, 559, 560, 561, 563, 564, 566, 568, 569, 570, 571, 572, 573, 575, 576, 577, 578, 579, 580, 581, 582, 583, 584, 586, 587, 588, 589, 590, 591, 593, 594, 595, 596, 597, 598, 599, 600, 601, 602, 603, 604, 605, 606, 607, 608, 609, 611, 616, 617, 618, 619, 621, 622, 623, 625, 626, 628, 629, 630, 634, 635, 636, 637, 639, 640, 641, 643, 644, 645, 647, 649, 650, 651, 653, 654, 655, 656, 657, 658, 659, 660, 661, 662, 663, 664, 666, 667, 668, 670, 671, 672, 673, 676, 677, 678, 679, 681, 682, 683, 684, 685, 686, 687, 688, 690, 692, 693, 694, 695, 696, 699, 700, 701, 702, 703, 704, 707, 708, 709, 711, 712, 713, 714, 716, 717, 718, 719, 720, 721, 723, 725, 726, 727, 728, 729, 730, 731, 732, 733, 734, 735, 736, 738, 739, 740, 741, 742, 744, 747, 749, 750, 751, 752, 754, 755, 756, 758, 759, 761, 763, 764, 765, 766, 767, 769, 770, 771, 773, 774, 777, 779, 781, 782, 783, 784, 785, 786, 787, 788, 789, 790, 793, 794, 796, 797, 798, 799, 800, 802, 804, 805, 808, 809, 810, 811, 812, 813, 814, 816, 817, 818, 819, 821, 823, 824, 826, 827, 828, 829, 830, 831, 832, 833, 834, 835, 836, 837, 838, 839, 840, 841, 842, 843, 844, 845, 848, 849, 850, 851, 852, 853, 854, 855, 856, 857, 858, 859, 860, 861, 862, 863, 864, 866, 867, 868, 869, 870, 871, 872, 873, 874, 875, 876, 877, 878, 879, 880, 881, 882, 883, 884, 885, 886, 887, 888, 889, 890, 891, 892, 893, 894, 895, 896, 897, 898, 899, 900, 904, 905, 907, 908, 909, 910, 912, 913
Test	4, 10, 27, 28, 30, 41, 45, 46, 48, 51, 58, 61, 65, 73, 78, 80, 81, 83, 85, 86, 91, 92, 99, 101, 102, 103, 108, 109, 118, 121, 123, 124, 126, 136, 143, 144, 145, 155, 157, 164, 175, 177, 180, 189, 192, 194, 195, 197, 201, 206, 211, 214, 216, 227, 232, 235, 238, 239, 243, 246, 255, 263, 264, 266, 270, 287, 290, 293, 319, 323, 324, 333, 336, 339, 346, 350, 352, 358, 371, 376, 381, 385, 386, 391, 393, 395, 401, 407, 428, 431, 437, 439, 440, 446, 450, 465, 466, 469, 475, 478, 487, 518, 522, 523, 525, 530, 534, 536, 550, 552, 554, 556, 562, 565, 567, 574, 585, 592, 610, 612, 613, 614, 615, 620, 624, 627, 631, 632, 633, 638, 642, 646, 648, 652, 665, 669, 674, 675, 680, 689, 691, 697, 698, 705, 706, 710, 715, 722, 724, 737, 743, 745, 746, 748, 753, 757, 760, 762, 768, 772, 775, 776, 778, 780, 791, 792, 795, 801, 803, 806, 807, 815, 820, 822, 825, 846, 847, 865, 901, 902, 903, 906, 911

Lampiran 3.**Hasil evaluasi *Confusion Matrix* model *machine learning*****Rasio 70:30**

	precision	recall	f1-score	support
negatif	0.54	0.63	0.58	71
netral	0.08	0.36	0.12	14
positif	0.83	0.55	0.66	190
accuracy			0.56	275
macro avg	0.48	0.51	0.46	275
weighted avg	0.72	0.56	0.61	275

Rasio 80:20

	precision	recall	f1-score	support
negatif	0.59	0.64	0.61	55
netral	0.14	0.50	0.22	12
positif	0.83	0.58	0.68	116
accuracy			0.59	183
macro avg	0.52	0.57	0.50	183
weighted avg	0.71	0.59	0.63	183

Lampiran 4. 5 sampel hasil Pre-Processing

Out[11]:

	KONTEN	LABEL	clean	Stopwords	Stemmed	Tokenized
0	Qatar menginvestasikan 300 miliar dolar AS sebagai tuan rumah Piala Dunia 2022 https://t.co/rYUSaGzLc	positif	qatar menginvestasikan miliar dolar as sebagai tuan rumah piala dunia	qatar menginvestasikan miliar dolar as tuan rumah piala dunia	qatar investasi miliar dolar as tuan rumah piala dunia	[qatar, investasi, miliar, dolar, as, tuan, rumah, piala, dunia]
1	@dextratime @ExpressSports Qatar sebagai tuan rumah sudah sukses menyelenggarakan turnamen ini dengan baik, tapi FIFA sebagai perangkat pertandingan cukup kontroversial dan banyak menguntungkan pihak tertentu	positif	qatar sebagai tuan rumah sudah sukses menyelenggarakan turnamen ini dengan baik tapi fifa sebagai perangkat pertandingan cukup kontroversial dan banyak menguntungkan pihak tertentu	qatar tuan rumah sukses menyelenggarakan turnamen fifa perangkat pertandingan kontroversial menguntungkan	qatar, tuan, rumah, sukses, selenggara, turnamen, fifa, perangkat, tanding, kontroversial, untung	[qatar, tuan, rumah, sukses, selenggara, turnamen, fifa, perangkat, tanding, kontroversial, untung]
2	Piala Dunia 2022 Sudah Sampai Pertandingan Final, Qatar Disebut Sukses Sebagai Tuan Rumah https://t.co/BHdydLByu	positif	piala dunia sudah sampai pertandingan final qatar disebut sukses sebagai tuan rumah	piala dunia pertandingan final qatar sukses tuan rumah	piala dunia tanding final qatar sukses tuan rumah	[piala, dunia, tanding, final, qatar, sukses, tuan, rumah]
3	Qatar memanfaatkan posisi sebagai tuan rumah Piala Dunia dan kelebihan ekonominya untuk memaksimalkan pengaturannya baik dalam politik regional maupun di panggung global. \Baca Trias Kuncalyono di Opini Kompas \n @jaskunz ompas	positif	qatar memanfaatkan posisi sebagai tuan rumah piala dunia dan kelebihan ekonominya untuk memaksimalkan pengaturannya baik dalam politik regional maupun di panggung global baca trias kuncalyono di opini kompas	qatar memanfaatkan posisi tuan rumah piala dunia, kelebihan ekonominya memaksimalkan pengaturannya politik regional panggung global baca trias kuncalyono opini kompas	qatar manfaat, posisi, tuan, piala, dunia, hebat, ekonomi, maksimal, pengatur, politik, regional, panggung, global, baca, trias, kuncalyono, opini, kompas	[qatar, manfaat, posisi, tuan, rumah, piala, dunia, hebat, ekonomi, maksimal, pengatur, politik, regional, panggung, global, baca, trias, kuncalyono, opini, kompas]
4	Prancis pasti ada di Final \nPemilihan Qatar sebagai tuan rumah piala dunia gk lepas dari kontribusi Prancis. Hubungan ekonomi bilateral kedua negara juga sangat kuat.	positif	prancis pasti ada di final pemilihan qatar sebagai tuan rumah piala dunia gk lepas dari kontribusi prancis hubungan ekonomi bilateral kedua negara juga sangat kuat	prancis final pemilihan qatar tuan rumah piala dunia gk lepas kontribusi prancis hubungan ekonomi bilateral negara kuat	prancis final pilih qatar, tuan rumah piala dunia gk lepas kontribusi prancis hubungan ekonomi bilateral negara kuat	[prancis, final, pilih, qatar, tuan, rumah, piala, dunia, gk, lepas, kontribusi, prancis, hubungan, ekonomi, bilateral, negara, kuat]

Lampiran 5.
 Hasil Pembobotan oleh Tfidf Vectorizer dari library Sklearn

abadi	abai	abis	absen	acara	adab	adat	adli	agama	agen	...	wisata	wisatawa	wisatawan	wisawatan	workers	world	worldcup	wujud	zona	
0.0	0.0	0.0	0.0	0.0000000	0.0	0.0	0.0	0.0000000	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0000000	0.0
0.0	0.0	0.0	0.0	0.0000000	0.0	0.0	0.0	0.0000000	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0000000	0.0
0.0	0.0	0.0	0.0	0.0000000	0.0	0.0	0.0	0.0000000	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0000000	0.0
0.0	0.0	0.0	0.0	0.0000000	0.0	0.0	0.0	0.0000000	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0000000	0.0
0.0	0.0	0.0	0.0	0.0000000	0.0	0.0	0.0	0.0000000	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0000000	0.0
0.0	0.0	0.0	0.0	0.0000000	0.0	0.0	0.0	0.0000000	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0000000	0.0
0.0	0.0	0.0	0.0	0.0000000	0.0	0.0	0.0	0.0000000	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0000000	0.0
0.0	0.0	0.0	0.0	0.0000000	0.0	0.0	0.0	0.0000000	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0000000	0.0
0.0	0.0	0.0	0.0	0.0000000	0.0	0.0	0.0	0.0000000	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0000000	0.0
0.0	0.0	0.0	0.0	0.0000000	0.0	0.0	0.0	0.0000000	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0000000	0.0
0.0	0.0	0.0	0.0	0.0000000	0.0	0.0	0.0	0.302968	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0000000	0.0
0.0	0.0	0.0	0.0	0.0000000	0.0	0.0	0.0	0.0000000	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0000000	0.0
0.0	0.0	0.0	0.0	0.0000000	0.0	0.0	0.0	0.0000000	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0000000	0.0
0.0	0.0	0.0	0.0	0.0000000	0.0	0.0	0.0	0.0000000	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0000000	0.0

rows × 1493 columns

Lampiran 6.

Hasil Cek Plagiasi Turnitin

Analisis Sentimen terhadap Qatar sebagai Tuan Rumah Piala Dunia FIFA 2022 pada Media Sosial Twitter

ORIGINALITY REPORT

8%	4%	1%	5%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Universiti Teknologi MARA Student Paper	3%
2	repository.uin-suska.ac.id Internet Source	1%
3	Submitted to Kaplan College Student Paper	1%
4	id.123dok.com Internet Source	1%
5	repository.uir.ac.id Internet Source	1%
6	issuu.com Internet Source	1%
7	ejournal.unsrat.ac.id Internet Source	1%

Exclude quotes Off
Exclude bibliography Off

Exclude matches < 1%