KLASIFIKASI TANGGAPAN MASYARAKAT MENGENAI *CHILDFREE* MENGGUNAKAN *IMPROVED K-NEAREST NEIGHBOR* (STUDI KASUS: TWITTER)

SKRIPSI

Oleh:
NURUL IZZAH RAHMADHANI
NIM. 19650014

PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2023

KLASIFIKASI TANGGAPAN MASYARAKAT MENGENAI *CHILDFREE* MENGGUNAKAN *IMPROVED K-NEAREST NEIGHBOR* (STUDI KASUS: TWITTER)

SKRIPSI

Diajukan Kepada:
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang
Untuk Memenuhi Salah Satu Persyaratan Dalam
Memperoleh Gelar Sarjana Komputer (S.Kom)

Oleh:
NURUL IZZAH RAHMADHANI
NIM. 19650014

PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM MALANG
2023

HALAMAN PERSETUJUAN

KLASIFIKASI TANGGAPAN MASYARAKAT MENGENAI *CHILDFREE* MENGGUNAKAN *IMPROVED K-NEAREST NEIGHBOR* (STUDI KASUS: TWITTER)

SKRIPSI

Oleh :
NURUL IZZAH RAHMADHANI
NIM. 19650014

Telah Diperiksa dan Disetujui untuk Diuji:
Tanggal: 10 Desember 2023

Pembimbing I,	Pembimbing II,
Dr. Muhammad Faisal, M.T	Dr. Zainal Abidin, M.Kom
NIP. 19740510 200501 1 007	NIP. 19760613 200501 1 004

Mengetahui,
Ketua Program Studi Teknik Informatika
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang

Dr. Fachrul Kurniawan, M.MT, IPM
NIP. 19771020 200912 1 001

HALAMAN PENGESAHAN

KLASIFIKASI TANGGAPAN MASYARAKAT MENGENAI *CHILDFREE* MENGGUNAKAN *IMPROVED K-NEAREST NEIGHBOR* (STUDI KASUS: TWITTER)

SKRIPSI

Oleh :
NURUL IZZAH RAHMADHANI
NIM. 19650014

Telah dipertahankan di Depan Dewan Penguji Skripsi
Dan Dinyatakan Sebagai Salah Satu Persyaratan
Untuk Memperoleh Gelar Sarjana Komputer (S.Kom)

Susunan Dewan Penguji

Ketua Penguji	: Dr. Yunifa Miftachul Arif, M.T NIP. 19830616 201101 1 004	( )
Anggota Penguji I	: Ajib Hanani, M.T NIDT. 19840731 20160801 1 076	( )
Anggota Penguji II	: Dr. Muhammad Faisal, M.T NIP. 19740510 200501 1 007	( )
Anggota Penguji III	: Dr. Zainal Abidin, M.Kom NIP. 19760613 200501 1 004	( )

Mengetahui dan Mengesahkan,
Ketua Program Studi Teknik Informatika
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang

Dr. Fachrul Kurniawan, M.MT, IPM
NIP. 19771020 200912 1 001

PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan dibawah ini:

Nama	: Nurul Izzah Rahmadhani
NIM	: 19650014
Fakultas/Jurusan	: Sains dan Teknologi / Teknik Informatika
Judul Skripsi	: Klasifikasi Tanggapan Masyarakat Mengenai *Childfree* Menggunakan *Improved K-Nearest Neighbor* (Studi Kasus: Twitter)

Menyatakan dengan sebenarnya bahwa Skripsi yang saya tulis ini benar-benar merupakan hasil skripsi sendiri, bukan merupakan pengambilan data, tulisan atau pikiran orang lain yang saya akui sebagai hasil tulisan atau pikiran saya sendiri, kecuali dengan mencantumkan sumber cuplikan pada daftar pustaka.

Apabila dikemudian hari terbukti atau dapat dibuktikan Skripsi ini jiplakan, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Malang, 8 Desember 2023
Yang membuat Pernyataan,

Nurul Izzah Rahmadhani
NIM. 19650014

HALAMAN MOTTO

أَلَمۡ نَشۡرَحۡ لَكَ صَدۡرَكَ (١) وَوَضَعۡنَا عَنكَ وِزۡرَكَ (٢) ٱلَّذِيٓ أَنقَضَ ظَهۡرَكَ (٣) وَرَفَعۡنَا لَكَ ذِكۡرَكَ (٤)

فَإِنَّ مَعَ ٱلۡعُسۡرِ يُسۡرًا (٥) إِنَّ مَعَ ٱلۡعُسۡرِ يُسۡرٗا (٦) فَإِذَا فَرَغۡتَ فَٱنصَبۡ (٧) وَإِلَىٰ رَبِّكَ فَٱرۡغَب (٨)

Bukankah Kami telah melapangkan untukmu dadamu?, dan Kami telah menghilangkan daripadamu bebanmu, yang memberatkan punggungmu?, Dan Kami tinggikan bagimu sebutan (nama)mu, Karena sesungguhnya sesudah kesulitan itu ada kemudahan, sesungguhnya sesudah kesulitan itu ada kemudahan. Maka apabila kamu telah selesai (dari sesuatu urusan), kerjakanlah dengan sungguh-sungguh (urusan) yang lain, dan hanya kepada Tuhanmulah hendaknya kamu berharap.

(*QS. Al-Insyirah*: 1-8)

HALAMAN PERSEMBAHAN

Alhamdulillahirabbil'alamin, puji syukur kehadirat Allah *Subhanahu wa ta'ala*, Shalawat serta salam kepada Rasulullah *Shallallahu Alaihi Wasallam*. Dengan rahmat dan hidayah-Nya, penulis dapat menyelesaikan skripsi. Penulis menyadari bahwa penyelesaian skripsi ini tidak terlepas dari dukungan, bimbingan, serta doa restu dari berbagai pihak. Oleh karena itu, penulis ingin mengucapkan terima kasih yang tak terhingga kepada:

1. Orang tua tercinta, Bapak Samsul Huda dan Ibu Siti Zumaroh, terima kasih atas kepercayaan yang telah diberikan, terima kasih telah memberikan motivasi terbesar bagi penulis, yang selalu membimbing dan menuntun dengan sabar, yang selalu mendoakan, yang selalu memberikan dukungan penuh dan juga kasih sayang yang tak terhingga.
2. Kakak tersayang, Achmad Arif Rachman, yang senantiasa mendoakan serta memberikan dukungan moril maupun materiil selama penulis menempuh pendidikan hingga saat ini. Tak lupa keluarga dari Bani Islam dan Bani Hamid yang tidak bisa disebutkan satu persatu yang juga memberikan doa untuk penulis.
3. Bapak Dr. Muhammad Faisal, M.T., selaku dosen pembimbing 1 dan Bapak Dr. Zainal Abidin, M.Kom., selaku dosen pembimbing 2 yang senantiasa sabar dalam membimbing penulis untuk menyelesaikan skripsi. Tak lupa seluruh dosen dan staff program studi Teknik Informatika Universitas Islam Negeri

Maulana Malik Ibrahim Malang yang telah memberikan ilmu yang bermanfaat bagi penulis.

4. Teman-teman penulis, Ahmad Ahsanul Umam dan Dea Ariella, serta teman seperjuangan Teknik Informatika angkatan 2019, yang selalu memberikan energi positif dan memberikan dukungan terhadap penulis. Dan juga semua orang yang telah membantu dalam menyelesaikan pendidikan, penulis mengucapkan banyak terima kasih.

Kepada semua orang yang telah disebutkan, terima kasih telah menjadi bagian tak tergantikan dalam pencapaian ini. Semua kata-kata persembahan ini tidak dapat sepenuhnya menggambarkan rasa terima kasih dan penghargaan yang penulis rasakan. Semoga setiap capaian ini juga menjadi berkah bagi kita semua.

KATA PENGANTAR

Assalamu'alaikum Warahmatullahi Wabarakatuh

Puji dan syukur kehadirat Allah *Subhanahu wa ta'ala* atas berkat, rahmat, dan hidayah-Nya, sehingga peneliti diberikan kemudahan dan keberkahan dalam setiap menyelesaikan skripsi ini. Pada kesempatan ini, penulis ingin menyampaikan ucapan terima kasih yang sebesar-besarnya atas bimbingan dan dukungan selama penulis mempersiapkan, menyusun, dan menyelesaikan skripsi ini, kepada:

1. Prof. Dr. H. M. Zainuddin, M.A., selaku rektor Universitas Islam Negeri Maulana Malik Ibrahim Malang.
2. Prof. Dr. Hj. Sri Harini, M.Si., selaku dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang.
3. Dr. Fachrul Kurniawan, M. MT, IPM., selaku Ketua Program Studi Teknik Informatika Universitas Islam Negeri Maulana Malik Ibrahim Malang.
4. Dr. Muhammad Faisal, M.T., selaku dosen pembimbing 1 yang telah membimbing serta membantu penulis dalam menyelesaikan skripsi.
5. Dr. Zainal Abidin, M.Kom., selaku dosen pembimbing 2 yang telah membimbing serta membantu penulis dalam menyelesaikan skripsi.
6. Dr. Yunifa Miftachul Arif, M.T., selaku dosen penguji 1 dan Ajib Hanani, M.T., selaku dosen penguji 2 yang telah memberikan arahan dalam menyelesaikan skripsi.
7. H. Fatchurrochman, M.Kom., selaku dosen wali yang selalu senantiasa memberikan motivasi dan arahan serta saran selama pendidikan.

8. Seluruh Dosen Teknik Informatika yang telah mencurahkan ilmunya kepada penulis selama kuliah.
9. Teman-teman Teknik Informatika angkatan 2019, yang telah memberikan energi positif dalam menyelesaikan skripsi.
10. Sahabat serta teman-teman semua yang terlibat dan membantu secara langsung maupun tidak langsung dalam menyelesaikan skripsi.

Penulis sepenuhnya menyadari bahwa masih terdapat kekurangan dalam penyusunan skripsi ini. Penulis mengharapkan usulan, saran, dan ritik yang bersifat membangun sebagai upaya tindak lanjut dalam penelitian di skripsi ini. Penulis juga berharap terdapat manfaat yang bisa diambil dari skripsi penulis.

Wassalamu'alaikum Warahmatullahi Wabarakatuh

Malang, 10 Desember 2023

Penulis

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

ABSTRAK

Rahmadhani, Nurul Izzah. 2023. **Klasifikasi Tanggapan Masyarakat Mengenai *Childfree* Menggunakan *Improved K-Nearest Neighbor* (Studi Kasus: Twitter)**. Skripsi. Jurusan Teknik Informatika, Fakultas Sains dan Teknologi. Universitas Islam Negeri Maulana Malik Ibrahim Malang. Pembimbing: (I) Dr. Muhammad Faisal, M.T., (II) Dr. Zainal Abidin, M.Kom

Kata Kunci: *Childfree, Cosine Similarity, Google Colab, iKNN, Python, TF-IDF*

Kalimat perbincangan di tengah masyarakat terkadang mengandung unsur perundungan mengenai topik yang sedang dibahas. Topik perbincangan masyarakat sangat beragam, salah satunya ialah masalah gaya hidup. Ungkapan pendapat masyarakat tidak hanya sebatas pada komunikasi tatap muka, melainkan juga melibatkan ekspresi *online*, khususnya di *platform* media sosial Twitter. Penelitian ini bertujuan untuk mengklasifikasikan tanggapan masyarakat di Twitter terhadap topik *childfree* dengan menerapkan metode *Improved K-Nearest Neighbor* (*iKNN*). Metode tersebut menggunakan pembobotan kata dengan *TF-IDF* dan perhitungan jarak menggunakan *Cosine Similarity*. Penelitian dilakukan menggunakan bahasa pemrograman *Python* di *Google Colab*. Penelitian ini mengukur kinerja sistem menggunakan metrik *accuracy*, *precision*, dan *recall*. Pada uji coba ke-1 menggunakan data latih sebesar 90% dan data uji 10% menghasilkan nilai *accuracy* 42.6%, *precision* 47%, dan *recall* 70%. Uji coba ke-2 menggunakan data latih sebesar 80% dan data uji 20% menghasilkan nilai *accuracy* 51%, *precision* 52.5%, dan *recall* 83%. Uji coba ke-3 menggunakan data latih sebesar 70% dan data uji 30% menghasilkan nilai *accuracy* 56.6%, *precision* 56.7%, dan *recall* 78%. Uji coba ke-4 menggunakan data latih sebesar 60% dan data uji 40% menghasilkan nilai *accuracy* 52.5%, *precision* 53.7%, dan *recall* 77%. Hasil evaluasi menunjukkan bahwa sistem lebih cenderung mendeteksi secara benar pada teks positif yang ditandai nilai *recall* lebih tinggi.

ABSTRACT

Rahmadhani, Nurul Izzah. 2023. **Classification of Public Responses Regarding *Childfree* Using *Improved K-Nearest Neighbor* (Case Study: Twitter).** Thesis. Department of Informatics Engineering, Faculty of Science and Technology. Maulana Malik Ibrahim State Islamic University Malang. Supervisors: (I) Dr. Muhammad Faisal, M.T., (II) Dr. Zainal Abidin, M.Kom

Discussions within society sometimes contain elements of bullying related to the discussed topics. Public discourse covers a wide range of topics, one of which is lifestyle issues. The expression of public opinion goes beyond face-to-face communication but also involves online expressions, especially on the Twitter social media platform. This research aims to classify public responses on Twitter regarding the childfree topic using the *Improved K-Nearest Neighbor* (*iKNN*) method. The method employs word weighting with *TF-IDF* and distance calculation using *Cosine Similarity*. This study was conducted using the *Python* programming language on *Google Colab*. This research measures system performance using accuracy, precision and recall metrics. In the 1st trial using 90% training data and 10% test data resulting in accuracy values of 42.6%, precision 47%, and recall 70%. The second test used 80% training data and 20% test data resulting in an accuracy value of 51%, precision 52.5% and recall 83%. The 3rd trial used 70% training data and 30% test data resulting in accuracy values of 56.6%, precision 56.7%, and recall 78%. The 4th trial used 60% training data and 40% test data resulting in accuracy values of 52.5%, precision 53.7%, and recall 77%. The evaluation results show that the system is more likely to correctly detect positive text that is characterized by a higher recall value.

***Keywords**: Childfree, Cosine Similarity, Google Colab, iKNN, Python, TF-IDF*

الملخص

رحماني، نور العزة. 2023. تصنيف استجابات الجمهور بشأن "*Childfree*" باستخدام *Improved K-Nearest Neighbor* (دراسة حالة: تويتر). رسالة جامعية. قسم هندسة الحاسوب، كلية العلوم والتكنولوجيا. جامعة الدولة الإسلامية مولانا مالك إبراهيم مالانج. المشرفون: (I)الدكتور محمد فيصل، M.T.، (II) الدكتور زين العابدين، M.Kom

الكلمات الرئيسية: *Childfree*، *Cosine Similarity*، *Google Colab*، *iKNN*، *Python*، *TF-IDF*

تحتوي النقاشات داخل المجتمع أحيانًا على عناصر التنمر حول المواضيع المطروحة. تغطي الحوارات العامة مجموعة واسعة من المواضيع، واحدة منها هي قضايا أسلوب الحياة. التعبير عن رأي الجمهور يتعدى التواصل وجهًا لوجه بل يشمل أيضًا التعبير عبر الإنترنت، خاصة على منصة وسائل التواصل الاجتماعي تويتر. تهدف هذه الدراسة إلى تصنيف استجابات الجمهور على تويتر بشأن موضوع"*Childfree*" باستخدام طريقة(*Improved K-Nearest Neighbor* (*iKNN*) . تعتمد الطريقة على وزن الكلمات بواسطة *TF-IDF* وحساب المسافة باستخدام *Cosine Similarity* .تم إجراء هذه الدراسة باستخدام لغة البرمجة بايثون في *Google Colab* . يقيس هذا البحث أداء النظام باستخدام مقاييس الدقة والإحكام والاستدعاء. في التجربة الأولى، تم استخدام 90% من بيانات التدريب و10% من بيانات الاختبار مما أدى إلى قيم دقة تبلغ 42.6%، ودقة 47%، واستدعاء 70%. استخدم الاختبار الثاني 80% من بيانات التدريب و20% من بيانات الاختبار مما أدى إلى دقة تبلغ 51% ودقة 52.5% واسترجاع 83%. استخدمت التجربة الثالثة 70% من بيانات التدريب و30% من بيانات الاختبار مما أدى إلى قيم دقة 56.6%، ودقة 56.7%، واستدعاء 78%. استخدمت التجربة الرابعة 60% من بيانات التدريب و40% من بيانات الاختبار مما أدى إلى قيم دقة 52.5%، ودقة 53.7%، واستدعاء 77%. أظهرت نتائج التقييم أن النظام من المرجح أن يكتشف بشكل صحيح النص الإيجابي الذي يتميز بقيمة استدعاء أعلى.

BAB I

PENDAHULUAN

1.1 Latar Belakang Masalah

Kalimat perbincangan di tengah masyarakat terkadang mengandung unsur perundungan mengenai topik yang sedang dibahas. Perbincangan di tengah masyarakat mencakup berbagai topik, mulai dari masalah pendidikan, politik, kesehatan, isu ataupun masalah lingkungan, hingga gaya hidup yang saat ini tengah ramai dibicarakan. Di era digital ini, masyarakat cenderung menyuarakan pendapat mereka di media sosial mengenai topik-topik yang ramai dibahas atau topik yang sedang *trend*.

Media sosial menjadi alat komunikasi modern yang mendukung penyebaran informasi bagi masyarakat luas. Media sosial juga mampu mengubah obrolan satu arah menjadi komunikasi interaktif dengan memanfaatkan teknologi berbasis *web*. Seringkali, media sosial dimanfaatkan sebagai sarana untuk membangun hubungan sosial dengan individu lainnya serta berbagi aktivitas pribadi atau pengalaman kehidupan nyata (Kurniawan & Wibawa, 2021). Salah satu platform dalam dunia media sosial yang mendapat banyak data ungkapan atau tanggapan masyarakat mengenai topik pembahasan yang sedang nge-*trend* ialah Twitter. Twitter adalah jenis *platform* media sosial yang memungkinkan pengguna untuk menyampaikan informasi secara terbuka atau publik (Iwandini et al., 2023). Salah satu topik yang sedang ramai dibahas di Twitter saat ini ialah *childfree*.

Childfree merupakan gaya hidup yang mengacu pada keputusan atau pilihan hidup pasangan untuk tidak memiliki keturunan selama masa perkawinan (Meidina, 2023). *Childfree* sebenarnya bukan hal baru bagi masyarakat di luar negeri. Namun, di Indonesia, *childfree* menjadi topik pembahasan yang menimbulkan berbagai pro dan kontra. Istilah "*childfree*" mulai kenal lebih luas oleh masyarakat karena berawal dari adanya *influencer* dan selebriti yang secara terbuka mengungkapkan keputusan untuk tidak memiliki keturunan melalui sosial media mereka.

Di Indonesia, *childfree* dianggap sebagai sebuah pelanggaran norma budaya juga seringkali bertentangan dengan ajaran mayoritas agama. Namun, dari perspektif Hak Asasi Manusia (HAM), keputusan untuk tidak memiliki keturunan ialah hak pribadi setiap individu. Sedangkan pada sudut pandang agama Islam, jika ditinjau dalam ilmu Fiqih, keputusan *childfree* dilarang, hal ini disebabkan karena dalam penerapannya tidak didasarkan pada alasan yang berkaitan dengan kesehatan, tetapi menggunakan alasan yang berkaitan dengan hal-hal duniawi seperti karir, pekerjaan, ataupun keuangan (Hadi et al., 2022). Pandangan ini sejalan dengan ayat Al-Qur'an *Surah Al-Isra'* : 31

وَلَا تَقْتُلُوْٓا اَوْلَادَكُمْ خَشْيَةَ اِمْلَاقٍۗ نَحْنُ نَرْزُقُهُمْ وَاِيَّاكُمْۗ اِنَّ قَتْلَهُمْ كَانَ خِطْـًٔا كَبِيْرًا

"*Dan janganlah kamu membunuh anak-anakmu karena takut miskin. Kamilah yang memberi rezeki kepada mereka dan kepadamu. Membunuh mereka itu sungguh suatu dosa yang besar.*" (*Q.S. Al-Isra'* [17] : 31)

Surah Al-Isra' ayat 31 menegaskan pentingnya menjaga hak hidup anak-anak dengan keyakinan bahwa Allah-lah yang memberikan rezeki kepada mereka. Ayat ini juga mengingatkan bahwa membunuh anak-anak karena takut akan kemiskinan

adalah dosa besar. Hal ini menunjukkan bahwa dalam konteks agama, menjaga kehidupan anak-anak merupakan tugas yang sangat penting.

Pembahasan masyarakat mengenai topik *childfree* di Twitter memiliki implikasi positif dan negatif, tergantung pada cara pengguna menyampaikan pendapatnya. Jumlah respon masyarakat yang sangat besar di *platform* Twitter dapat menjadi sumber data yang baik. Untuk menghasilkan informasi yang lebih lanjut, data tanggapan masyarakat di Twitter dapat diolah dan diklasifikasikan menggunakan *machine learning*.

Machine Learning atau yang dikenal sebagai pembelajaran mesin adalah proses pemrograman komputer untuk mengoptimalkan kriteria kinerja menggunakan data contoh atau pengalaman masa lalu (Alpaydin, 2009). Data latih yang digunakan pada *machine learning* dalam bentuk set data. Pada penelitian ini, set data yang digunakan merupakan data berupa respon atau tanggapan masyarakat/*netizen* Indonesia mengenai topik *childfree* pada platform Twitter.

Terdapat berbagai metode pada *machine learning* untuk proses klasifikasi seperti metode *Support Vector Machines* (*SVM*), *Logistic Regression*, *Decision Trees*, *Random Forest*, *Naïve Bayes*, *K-Nearest Neighbor* (*KNN*), dan lain sebagainya. Penelitian ini akan berfokus pada klasifikasi dengan menggunakan metode *Improved K-Nearest Neighbor* karena memiliki performa yang baik dan nilai akurasi yang tinggi dari metode *KNN* (Mutrofin et al., 2014). Metode *Improved K-Nearest Neighbor* merupakan metode penyempurnaan dari *K-Nearest Neighbor*. Kelemahan metode *K-Nearest Neighbor* terletak pada tingkat akurasinya, karena

pada tiap kategori yang terdapat di data latih tidak diperhitungkan dan nilai *k* yang ditetapkan selalu sama.

Pada penelitian yang dilakukan oleh Shengyi Jiang, Guansong Pang, Meiling Wu, dan Limin Kuang (2012) menunjukkan bahwa menggunakan algoritma *Improved K-Nearest Neighbor* secara signifikan dapat mengungguli *KNN* dan lebih efektif serta tangguh daripada *Naïve Bayes* dan pengklasifikasi *Support Vector Machine* yang canggih. Pada proses perhitungan pertama menggunakan *Improved K-Nearest Neighbor* dan *KNN* mencapai hasil klasifikasi terbaiknya saat nilai *k* masing-masing adalah 45 dan 10. Dalam hal ini *Improved K-Nearest Neighbor Macro-F_1* mendapat hasil 88,50%, 2,54% lebih tinggi dari *KNN* biasa. Pada percobaan perhitungan selanjutnya, *Improved K-Nearest Neighbor* mengalami peningkatan kinerja rata rata sekitar 4,13%, dan mencapai hasil terbaik 75,12%, 3,27% lebih tinggi dari *KNN*. Berdasarkan kelebihan yang telah ditunjukkan oleh penelitian sebelumnya, maka pada penelitian ini penulis akan membuat program klasifikasi sentimen menggunakan metode *Improved K-Nearest Neighbor* dan dibantu menggunakan metode perhitungan lainnya. Penelitian ini diharapkan dapat berguna untuk membantu organisasi atau perusahaan dalam mengidentifikasi faktor-faktor yang mempengaruhi persepsi masyarakat dan dalam membuat kebijakan yang berkaitan dengan *childfree*.

1.2 Pernyataan Masalah

Berdasarkan latar belakang, pernyataan masalah adalah bagaimana mengukur tingkat *accuracy*, *precision*, dan *recall* algoritma *Improved K-Nearest Neighbor*

dalam mengklasifikasikan tanggapan positif dan negatif dari masyarakat mengenai *childfree* di Twitter.

1.3 Tujuan Penelitian

Tujuan dari penelitian ini adalah untuk mengukur tingkat *accuracy, precision,* dan *recall* algoritma *Improved K-Nearest Neighbor* dalam mengklasifikasikan tanggapan positif dan negatif dari masyarakat mengenai *childfree* di Twitter.

1.4 Manfaat Penelitian

Manfaat dari penelitian ini antara lain:

1. Memahami bagaimana respon masyarakat atau kelompok tertentu terhadap topik *childfree*. Hal ini dapat membantu mengidentifikasi faktor apa saja yang mempengaruhi persepsi juga opini mereka
2. Hasil dari penelitian ini dapat dimanfaatkan untuk menghasilkan keputusan yang lebih tepat pada pembuat kebijakan seperti organisasi atau perusahaan yang berkaitan dengan masalah *childfree*
3. Melalui penelitian ini, dapat diperoleh pemahaman yang lebih mendalam terkait *childfree*, seperti dampak terhadap keluarga, masyarakat, serta hubungan interpersonal

1.5 Batasan Masalah

Batasan masalah ditetapkan agar pembahasan tetap terfokus pada pokok bahasan materi. Beberapa batasan masalah antara lain:

1. Data yang digunakan diambil dari Twitter mulai tanggal 01 Januari 2023 sampai 30 Mei 2023 dengan kata kunci "*Childfree*"

2. Menggunakan data latih berupa teks dan disimpan dalam file berekstensi *csv*
3. Klasifikasi dilakukan pada tanggapan masyarakat yang menggunakan Bahasa Indonesia

BAB II

STUDI PUSTAKA

2.1 Penelitian Terkait

Sebagai acuan dan perbandingan untuk pembahasan dalam penelitian ini, maka dalam bab ini membahas beberapa penelitian terdahulu yang telah dilakukan.

Penelitian yang dilakukan oleh Siti Nur Arafah dan Fathoni (2022), untuk pengelompokan opini positif dan negatif pada pengguna jasa transportasi LRT Kota Palembang menggunakan metode *Improved K-Nearest Neighbor*, penelitian ini menggunakan metode pembobotan *TF-IDF* dengan hasil akurasi tertinggi sebesar sebesar 74.07% pada 90% data latih dan 10% data uji, dengan presisi 70%, *recall* 56%, dan *f-1 score* 59%. Sedangkan akurasi terendah dengan nilai akurasi 63.04% pada 50% data latih dan 50% data uji, dengan presisi 44%, *recall* 42%, dan *f-1 score* 42%.

Penelitian yang dilakukan oleh Wachid Darmawan, Muhammad Faizal Kurniawan, Wahyu Setianto, dan Wim Hapsoro (2013), untuk analisis sentimen tentang penerapan kurikulum merdeka menggunakan metode *K-Nearest Neighbor* dengan *Forward Selection* memberikan hasil bahwa metode *KNN* akurasinya sebesar 73.64%, sedangkan *KNN + FS* akurasinya meningkat menjadi 76.82%.

Penelitian oleh Dede Sandi, Ema Utami, dan Agus Fatkhurohman (2022) untuk klasifikasi opini pada berita vaksinasi menggunakan metode *K-Nearest Neighbor* mendapatkan hasil bahwa semakin banyak data *training* maka akurasi semakin bagus. Pengguna Twitter lebih banyak berpendapat bahwa berita vaksinasi

di Indonesia pada saat itu tergolong bagus.

Nurul Muslimah, Indriati, dan Randy Cahya Wihandika (2019) tentang klasifikasi film bersadaran sinopsis menggunakan metode *Improved K-Nearest Neighbor* dengan metode pembobotan *TF-IDF* memberikan hasil *precision* 1, *recall* 0.88, *f-measure* 0.936170213, dan akurasi sebesar 88%.

Indriya Dewi Onantya, Indriati, Putra Pandu Adikara (2019) tentang analisis sentimen pada ulasan aplikasi BCA Mobile menggunakan metode *Improved K-Nearest Neighbor* dengan metode pembobotan BM25 memberikan hasil *Precision* 0.946, *Recall* 0.934, *F-Measure* 0.939, *Accuracy* 0.942.

Dea Zakia Nathania, Indriati, Fitra Abdurrachma Bachtiar (2018) untuk klasifikasi spam pada Twitter menggunakan metode *Improved K-Nearest Neighbor* dan metode pembobotan *TF-IDF* memberikan hasil *Precision* 0.8946, *Recall* 0.9405, *F-Measure* 0.9155, *Accuracy* 89.57%.

Tabel 2.1 Penelitian Terkait

No.	Sitasi	Objek	Metode	Ekstraksi atau Seleksi Fitur	Hasil
1.	(Arafah & Fathoni, 2022)	Sentimen Analisis pada masyarakat terhadap LRT Kota Palembang	*Improved K-Nearest Neighbor*	*TF-IDF*	Akurasi tertinggi sebesar 74.07% pada 90% data latih dan 10% data uji, dengan presisi 70%, *recall* 56%, dan *f-1 score* 59%. Sedangkan akurasi terendah dengan nilai akurasi 63.04% pada 50% data latih dan 50% data uji, dengan presisi 44%, *recall* 42%, dan *f-1 score* 42%.

No.	Sitasi	Objek	Metode	Ekstraksi atau Seleksi Fitur	Hasil
2.	(Darmawan et al., 2023)	Analisis Sentimen Penerapan Kurikulum Merdeka	*K-Nearest Neighbor*	*Forward Selection*	Metode *K-NN* akurasinya sebesar 73.64%, sedangkan *K-NN + FS* akurasinya meningkat menjadi 76.82%
3.	(Sandi et al., 2022)	Klasifikasi Opini pada Berita Vaksinasi	*K-Nearest Neighbor*	-	Semakin banyak data training maka akurasi semakin bagus. Pengguna Twitter lebih banyak berpendapat bahwa berita vaksinasi di Indonesia pada saat itu tergolong bagus.
4.	(Muslimah et al., 2019)	Klasifikasi Film Berdasarkan Sinopsis	*Improved K-Nearest Neighbor*	*TF-IDF*	*Precision* 1, *recall* 0.88, *f-measure* 0.936170213, dan akurasi 88%
5.	(Onantya et al., 2019)	Analisis Sentimen pada Ulasan Aplikasi BCA Mobile	*Improved K-Nearest Neighbor*	Pembobotan BM25	*Precision* 0.946, *Recall* 0.934, *F-Measure* 0.939, *Accuracy* 0.942
6.	(Nathania et al., 2018)	Klasifikasi Spam	*Improved K-Nearest Neighbor*	*TF-IDF*	*Precision* 0.8946, *Recall* 0.9405, *F-Measure* 0.9155, *Accuracy* 89.57%

2.2 Landasan Teori

2.2.1 Analisis Sentimen

Analisis sentimen adalah disiplin ilmu komputasi tentang pendapat, sentimen, emosi, dan sikap yang dinyatakan oleh individu (Liu, 2015). B. Laurensz dan Eko Sediyono (dalam Pratama et al., 2023:531) menjelaskan bahwa analisis sentimen merupakan *sub*-bidang *Data Mining* yang biasanya digunakan untuk melakukan analisis teks dari opini yang mengandung polaritas, kemudian memproses informasi tekstual hingga menghasilkan data berupa informasi emosional dengan nilai positif

dan negatif. Liu, Bing (2015) juga menjelaskan analisis sentimen dapat dianggap sebagai permasalahan analisis semantik yang sangat terfokus dan terbatas. Sistem analisis sentimen tidak perlu sepenuhnya "memahami" setiap kalimat atau dokumen, melainkan hanya perlu memahamai beberapa aspek tertentu seperti opini positif dan negatif serta subjeknya.

Analisis sentimen Twitter adalah cara untuk menandai opini, penilaian, reaksi, evaluasi, hingga emosi seseorang mengenai sebuah topik bahasan di media sosial Twitter. Cara kerja analisis sentimen yaitu dengan menggabungkan teks atau data yang ada menjadi sebuah kalimat atau dokumen, kemudian mengklasifikan kalimat atau dokumen tersebut dalam bentuk positif atau negatif. Hasil dari analisis sentimen dapat digunakan sebagai pedoman untuk meningkatkan mutu suatu layanan atau produk (Suryono & Luthfi, 2021).

2.2.2 Twitter

Twitter adalah *platform microblogging* yang dapat digunakan untuk menyiarkan pembaruan status singkat (maksimum 140 karakter) (Russel, 2011). Novitasari (2020), mengungkapkan "*pada tanggal 7 November 2017 ditambah menjadi 280 karakter*". Twitter merupakan salah satu layanan untuk keluarga, teman, hingga rekan kerja dalam bidang komunikasi dan saling terhubung melalui pertukaran pesan. Twitter memungkinkan penggunanya memposting kicauan (biasa disebut *tweet*) serta berkirim dan membaca pesan yang berisikan foto, video, *link* atau tautan, dan teks. *Tweet* merupakan teks yang muncul di halaman profil pengguna dan telah diunggah sebelumnya. *Tweet* pada halaman profil pengguna dapat dilihat oleh publik, tetapi pengguna juga dapat menentukan siapa saja yang

dapat melihat *tweet* yang ada di halaman profil mereka, dan siapa saja yang mengikuti akun pengguna atau yang biasa disebut dengan pengikut atau *follower*. Twitter berhasil menarik perhatian pengguna internet karena dalam penggunaannya hanya butuh waktu yang singkat, namun informasi yang dikirimkan dapat tersebar secara luas.

2.2.3 *Text Mining*

Text Mining adalah penemuan komputer baru, informasi yang sebelumnya tidak diketahui, dengan secara otomatis mengekstraksi informasi dari sumber tertulis yang berbeda (Hearest, 2003). Tujuan dari *text mining* ialah untuk menemukan informasi yang sampai saat ini tidak diketahui, sesuatu yang belum diketahui siapa pun sehingga belum dapat dituliskan. Bidang *text mining* biasanya berurusan dengan teks yang fungsinya adalah komunikasi informasi faktual atau opini, dan motivasi untuk mencoba mengekstrak informasi dari teks semacam itu secara otomatis, bahkan jika keberhasilannya hanya sebagian (Witten, 2004).

Salah satu tema utama yang mendukung *text mining* ialah transformasi teks menjadi data numerik, jadi meskipun presentasi awalnya berbeda, pada beberapa tahap menengah atau tahap perantara, data berpindah ke pengkodean *data mining* (Weiss et al., 2015). Data yang tidak terstruktur menjadi terstruktur.

2.2.4 *Text Preprocessing*

Preprocessing merupakan tahap pertama dalam pengelolaan teks. *Preprocessing* adalah proses dimana data mentah disiapkan sebelum proses lainnya dilakukan. *Text Preprocessing* mengambil teks mentah sebagai input dan

menghasilkan token telah diolah secara bersih (Murugan et al., 2019). Token adalah kata tunggal atau sekelompok kata yang dihitung berdasarkan frekuensinya dan berperan sebagai fitur dalam analisis. *Preprocessing* data dilakukan dengan mengubah data ke dalam bentuk yang lebih mudah diproses oleh sistem seperti menghilangkan data atau kata yang tidak relevan. Pengeliminasian data atau kata pada tahap ini menghasilkan data *training* yang diterapkan sebagai input untuk algoritma klasifikasi.

2.2.5 Pembobotan Kata TF-IDF (*Term Frequency-Inverse Document Frequency*)

Pembobotan Kata atau *Term Weighting* adalah proses perhitungan bobot atau nilai di setiap kata dalam dokumen tertentu (Deolika et al., 2019). *Term Weighting* dilakukan dengan menghitung frekuensi kemunculan suatu *term* dalam dokumen. Frekuansi kemunculan *term* diasumsikan sebagai indikasi seberapa besar *term* tersebut merepresentasikan isi dari suatu dokumen.

Term Frequency-Inverse Document Frequency (*TF-IDF*) ialah metode pembobotan kata paling banyak diterapkan dalam *Text Mining*. *TF-IDF* adalah metode yang digunakan untuk mengukur sejauh mana *term* terkait dengan dokumen dengan memberikan bobot pada tiap-tiap kata. *TF-IDF* merupakan integrasi dari *Term Frequency* dan *Inverse Document Frequency* yang dapat digunakan untuk meng-*query corpus* dengan menghitung skor normalisasi yang menyatakan kepentingan suatu istilah dalam dokumen (Russel, 2011). *Term Frequency* mengukur berapa kali sebuah *term* muncul dalam sebuah dokumen, sedangkan *Inverse Document Frequency* mengukur seberapa sering *term* tersebut muncul di

seluruh dokumen dalam *corpus* (Tahmasebi & Risse, 2017).

Untuk meningkatkan efisiensi, metode *TF-IDF* mengeliminasi atau menghilangkan kata (*term*) tertentu serta kata imbuhan dengan memperhatikan frekuensi dan jumlah dokumen yang mengandung *term* yang bersangkutan. Kata yang beberapa kali muncul dalam dokumen, tetapi jarang muncul di semua koleksi dokumen, akan memiliki bobot atau nilai lebih tinggi daripada kata yang beberapa kali muncul di semua koleksi dokumen (Ni'mah & Arifin, 2020). Berikut rumus pembobotan kata *TF-IDF*:

$$w_{ij} = idf_j \times tf_{ij} = log\frac{N}{DF_i}(1 + logTFi) \quad (2.1)$$

Keterangan:
w_{ij} : bobot dokumen ke-*i* terhadap kata ke-*j*
idf_j : *inverse document frequency* (*log*(*N*/*df*)
tf_{ij} : banyaknya kata yang dicari
N : total teks
DF_i : jumlah dokumen yang mengandung kata yang dicari

2.2.6 *Cosine Similarity*

Sementara *TF-IDF* dapat menyediakan sarana untuk mempersempit *corpus* berdasarkan pencarian istilah (*term*), *cosine similarity* adalah salah satu teknik paling umum untuk membandingkan dokumen satu sama lain, yang merupakan inti dari menemukan dokumen serupa serta dalam analisis teks (Russel, 2011) (Han et al., 2012). *Cosine similarity* adalah metrik kesamaan antara dua vektor tidak nol yang didefinisikan dalam ruang produk dalam. Dalam hal ini, mengukur *cosinus* dari sudut antara dua vektor digunakan untuk menilai sejauh mana arah kedua vektor hampir sama (Han et al., 2012). Untuk menghitung *cosine similarity* antara dua vektor, dapat menggunakan rumus berikut:

$$cosSim(d_j, q_k) = \frac{\sum_{i=1}^{n}(td_{ij} \times tq_{ik})}{\sqrt{\sum_{i=1}^{n} td_{ij}^{\ 2} \times \sum_{i=1}^{n} tq_{ik}^{\ 2}}} \quad (2.2)$$

Keterangan:
$cosSim(d_j,q_k)$: tingkat kesamaan dokumen dengan *query* tertentu
td_{ij} : *term* ke-*i* dalam vektor untuk dokumen ke-*j*
tq_{ik} : *term* ke-*i* dalam vektor untuk *query* ke-*k*
n : jumlah *term* yang unik dalam *dataset*

2.2.7 *K-Nearest Neighbor*

K-Nearest Neighbor (*KNN*) merupakan analisis tetangga terdekat atau pencarian tetangga terdekat, yang merupakan algoritma untuk mengklasifikasikan titik data baru berdasarkan titik data terdekat (Parsian, 2015). Algoritma *KNN* merupakan sebuah *supervised learning classifier* yang banyak digunakan karena sederhana dan mudah diimplementasikan. Klasifikasi *KNN* adalah menemukan *k* titik data terdekat dalam kumpulan data ke data *query* yang diberikan, di mana *k* merupakan bilangan bulat lebih besar dari 0.

2.2.8 *Improved K-Nearest Neighbor*

Improved K-Nearest Neighbor merupakan metode pengembangan atau penyempurnaan dari *KNN* pada penetapan nilai *k-values*. Agar didapatkan nilai akurasi yang tinggi diperlukan penentuan nilai *k-values* yang tepat dalam tahap pengelompokan dokumen uji. Penentuan awal nilai *k-values* dilakukan pada masing-masing kategori sehingga setiap kategori memiliki nilai *k-values* yang berbeda. Perbedaan nilai *k-values* yang diberikan pada tiap kategori disesuaikan dengan jumlah data *training*/dokumen pada setiap kategori yang ada (Herdiawan, 2015). Setelah menghitung kemiripan dokumen menggubakan metode *Cosine*

Similarity, langkah berikutnya adalah menyusun hasil perhitungan tersebut secara menurun dalam setiap kategori, kemudian dilakukan perhitungan nilai *k-values* yang baru (*new_k_value*) (Herdiawan, 2015).

$$new\_k\_value = \left[\frac{k * N(c_m)}{maks\{N(c_m)|j = 1 \ldots . N_c\}}\right] \quad (2.3)$$

Keterangan:
new_k_value : nilai *k-values* baru
k : nilai *k-values* yang ditetapkan di awal
(c_m) : jumlah dokumen/data latih pada kategori *m*
$maks\{N(c_m)|j = 1 \ldots . N_c$: jumlah dokumen/data latih terbanyak pada semua kategori yang ada

Sejumlah *n* dokumen yang dipilih untuk setiap kategori adalah *top n* dokumen atau dokumen teratas dengan nilai kesamaan tertinggi di setiap kategori (Putri et al., 2013).

Untuk menentukan hasil ketegori, dilakukan perhitungan similaritas pada tiap kategori. Persamaan dibawah memberikan nilai maksimal untuk perbandingan kemiripan antara dokumen *X* dan data latih d_j dihitung untuk sejumlah tetangga teratas (*top n*) pada suatu kategori, baik itu dokumen *X* terhadap data latih d_j maupun sebaliknya pada *training set* (Baoli et al., 2003).

$$p(x, c_m) = argmax_m \frac{\Sigma d_j \in top\ n\ kNN(c_m) sim(x, d_j) y(d_j . c_m)}{\Sigma d_j \in top\ n\ kNN(c_m) sim(x, d_j)} \quad (2.4)$$

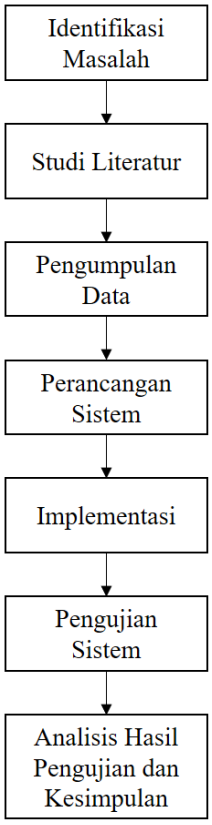
Keterangan:
$p(x, c_m)$: probabilitas dokumen *x* menjadi anggota kategori c_m
$sim(x, d_j)$: kemiripan antara dokumen *x* dengan dokumen latih d_j
top n kNN : *top n* tetangga
$y(d_j . c_m)$: fungsi atribut dari kategori

Dengan menghitung menggunakan persamaan diatas, selanjutnya akan dilakukan perbandingan probabilitas untuk setiap kategori. Kategori yang dipilih akan merujuk pada kategori dengan probabilitas tertinggi (Nathania et al., 2018).

BAB III
DESAIN DAN IMPLEMENTASI

3.1 Prosedur Penelitian

Prosedur penelitian merupakan serangkaian langkah atau tahapan yang dibuat untuk mencapai tujuan dalam penelitian. Gambar 3.1 adalah alur kegiatan yang dilakukan.



Gambar 3.1 Diagram Prosedur Penelitian

Berdasarkan gambar 3.1, tahapan penelitian yang dilakukan dalam penelitian ini ialah mengidentifikasi masalah dengan rumusan atau pernyataan masalah. Studi literatur dengan mengumpulkan teori-teori serta pengetahuan yang

berkaitan dengan pembahasan dalam penelitian. Pengumpulan data dengan memanfaatkan data yang didapat dari Twitter berupa *tweet* atau kicauan mengenai topik tertentu. Perancangan atau desain sistem menggunakan metode *Improved K-Nearest Neighbor* untuk memahami urutan atau rangkaian sistem. Tahap Implementasi dan Pengujian Sistem merupakan tahap di mana seluruh proses digabungkan menjadi kode, yang kemudian diuji. Tahap terakhir ialah Analisis Hasil, di mana akurasi yang dihasilkan oleh sistem dihitung atau dievaluasi kembali dengan menggunakan metode *Confusion Matrix*.

3.2 Pengumpulan dan Persiapan Data

Penelitian ini menggunakan data berupa *tweets* atau kicauan Twitter yang dikumpulkan kemudian diberi label sesuai dengan kategori positif dan negatif seperti pada contoh data di bawah. Tahap awal, pelabelan data dilakukan secara manual oleh 3 (tiga) guru bahasa Indonesia di SMA Al-Islam Krian Sidoarjo, yaitu Ibu Heri Widayati, S. Pd, Bapak Muhammad Nur, S. Pd, dan Ibu Siti Komariyah, S. Pd.

Tahap berikutnya, pelabelan otomatis dilakukan sistem untuk mengetahui tingkat *accuracy*, *precision*, dan *recall* dari metode *Improved K-Nearest Neighbor* yang digunakan. Hasil dari pelabelan data dengan cara manual digunakan untuk perbandingan dengan data yang diberi label secara otomatis oleh sistem. Berikut merupakan kriteria sebagai acuan dalam menetapkan apakah tanggapan atau *tweet* berkategori negatif, kriteria diambil dari beberapa pasal, diantaranya:

Pasal 27 ayat (3) UU ITE
Setiap orang dengan sengaja dan tanpa hak mendistribusikan dan/atau mentransmisikan dan/atau membuat dapat diaksesnya Informasi Elektronik

dan/atau Dokumen Elektronik yang memiliki muatan penghinaan dan/atau pencemaran nama baik. Ketentuan pada ayat ini mengacu pada ketentuan pencemaran nama baik dan/atau fitnah yang diatur dalam Kitab Undang-Undang Hukum Pidana ("KUHP").

Pasal 28 ayat (2) UU ITE
Setiap orang dengan sengaja dan tanpa hak menyebarkan informasi yang ditujukan untuk menimbulkan asa kebencian atau permusuhan individu dan/atau kelompok masyarakat tertentu berdasarkan atas suku, agama, ras, dan antargolongan (SARA).

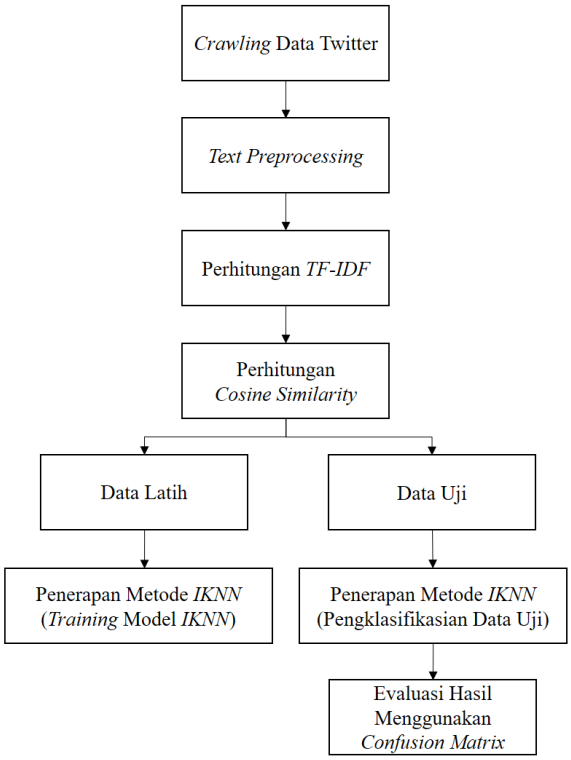
Kalimat yang tidak termasuk dalam kriteria diatas dianggap sebagai tanggapan atau *tweet* yang bersifat positif. Pelabelan yang dilakukan secara manual maupun otomatis oleh sistem pada dasarnya sama saja, ketika tanggapan memiliki bobot yang tinggi, dan mengandung banyak kata bersifat negatif, sistem akan menandai tanggapan tersebut ke dalam kategori negatif. Perbandingan yang dilakukan antara pelabelan dokumen secara manual dengan pelabelan otomatis oleh sistem guna untuk mengetahui seberapa tinggi kinerja sistem dapat mengenali tanggapan yang bersifat negatif.

Tabel 3.1 Contoh Dokumen

No.	*Tweet*	Label
1	Sorry but i have to say this. Mbak Gitasav, km di nyatakan gak bisa punya anaka kali ya sm dokter jadi berisik bgt bilang childfree childfree. Netijen udah pd gaada yg berisik soal keputusan u childfree. Km sendiri yg selalu terus menerus mention itu. HADEH pala batre	Negatif
2	gue ngikutin konten gitasav udh dari lama, suka liat youtube nya yg beropini, menurut gue terkait childfree itu emg udh jadi pilihannya ya its ok serah dia gw pikir semua org punya preferensi dan pilihan atau keputusan msing" mau punya anak atau childfree	Positif
3	jujur gpernah permasalahin keputusan seseorang utk childfree. tp liat reply dia ini kok agak kesel ya jatohnya kyk nyepelein orang yang bener2 mau punya anak. childfree = natural antiaging? no, mamahku anak 3 ttp awet muda tuh. kalo mau anti aging tu kuncinya adalah byk duit	Negatif
4	intinya hidup dibikin simple aja. semua orang yg punya keputusan childfree itu normal dan valid dan jg kaga bisa lu pada buat nyatuin isi pala lu dgn pala2 lainnya buat setuju/ga setuju.	Positif

3.3 Desain Sistem

Desain sistem dalam penelitian ini digunakan untuk menggambarkan perencanaan dalam mengatur struktur sistem dengan membuat alur dari beberapa proses dalam melakukan klasifikasi sentimen menggunakan metode *Improved K-Nearest Neighbor*. Penelitian ini melibatkan beberapa tahapan, diantaranya pengumpulan dan persiapan data yang dilakukan dengan *crawling* data di Twitter, kemudian melakukan *Text Preprocessing*, perhitungan *TF-IDF*, perhitungan *Cosine Similarity*, pembagian data, penerapan metode *Improved K-Nearest Neighbor*, dan evaluasi hasil menggunakan *Confusion Matrix*.



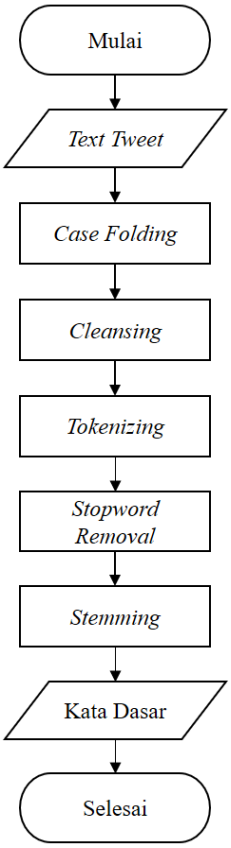
Gambar 3.2 Diagram Desain Sistem

Setelah pembagian data latih dan data uji, data latih diolah lebih lanjut kemudian digunakan untuk melatih model *IKNN*. Model *IKNN* yang telah dilatih

kemudian diterapkan pada data uji tanpa melibatkan kembali data latih dalam proses ini. Selanjutnya, data uji diolah menggunakan hasil latihan dari model *IKNN* untuk mengklasifikasikan sentimen. Performa model dievaluasi menggunakan metode *Confusion Matrix* berdasarkan hasil klasifikasi pada data uji.

3.4 *Text Preprocessing*

Preprocessing data dilakukan untuk membersihkan data dengan mengeliminasi kata yang tidak sesuai dan mengubahnya menjadi data bersih agar mudah diidentifikasi. Adapun beberapa tahap *preprocessing* pada penelitian ini, diantaranya: *case folding*, *cleansing*, *tokenizing*, *stopword removal*, dan *stemming*.



Gambar 3.3 *Flowchart* Tahap *Preprocessing*

3.4.1 *Case Folding*

Case Folding merupakan teknik pemrosesan teks dengan mengonversi huruf besar menjadi huruf kecil (*lowercase*). Karakter lain yang bukan huruf dan angka akan dihilangkan dan dianggap sebagai pemisah (*delimiter*). Berikut contoh tahap *Case Folding*:

Tabel 3.2 Contoh Proses *Case Folding*

Sebelum proses *Case Folding*	**Setelah proses *Case Folding***
Saling menghargai keputusan masing-masing bisa sih sebenernya. Mau childfree ataupun mau punya anak, setiap orang punya pertimbangannya masing-masing. Yang aku gak suka adalah mereka yang julid terhadap keputusan orang lain, yang merasa dirinya yang paling ideal.	saling menghargai keputusan masing-masing bisa sih sebenernya. mau childfree ataupun mau punya anak, setiap orang punya pertimbangannya masing-masing. yang aku gak suka adalah mereka yang julid terhadap keputusan orang lain, yang merasa dirinya yang paling ideal.

3.4.2 *Cleansing*

Cleansing merupakan proses pembersihan kalimat dari karakter-karakter yang tidak diperlukan guna mengurangi *noise* pada proses klasifikasi. Contoh penggunaan karakter yang banyak ditemukan pada *tweet* ialah karakter (‘@’) yang biasa digunakan untuk penyebutan, karakter *hashtag* yang dimulai dengan (‘#’), *link* atau tautan seperti (‘http’) dan (‘bit.ly’) serta *symbol* atau karakter lain seperti (~`! @ # $ % ^ & * () _ + {} |:,. [] <> ?). Karakter yang dihapus akan diganti dengan spasi.

Tabel 3.3 Contoh Proses *Cleansing*

Sebelum proses *Cleansing*	**Setelah proses *Cleansing***
@tanyakanrl Enggak sih, soal childfree menurut gue lebih ke pribadi. Siap enggak secara mental, financial, dll punya anak, bahkan trauma juga bisa mempengaruhi orang buat ambil keputusan childfree.	Enggak sih soal childfree menurut gue lebih ke pribadi Siap enggak secara mental financial dll punya anak bahkan trauma juga bisa mempengaruhi orang buat ambil keputusan childfree

3.4.3 *Tokenizing*

Tokenizing adalah proses pemotongan kalimat menjadi kata-kata dengan menganalisis *dataset* dengan mengekstraksi kata-kata dan menentukan struktur pada tiap kata. Berikut contoh penerapan *tokenizing*:

Tabel 3.4 Contoh Proses *Tokenizing*

Sebelum proses *Tokenizing*	**Setelah proses *Tokenizing***
I understand. Di indonesia kriteria keluarga sempurna tuh pasangan suami istri yang punya rumah, kendaraan dan yang paling penting keturunan	'i' 'understand' 'di' 'indonesia' 'kriteria' 'keluarga' 'sempurna' 'tuh' 'pasangan' 'suami' 'istri' 'yang' 'punya' 'rumah' 'kendaraan' 'dan' 'yang' 'paling' 'penting' 'keturunan'

3.4.4 *Stopword Removal*

Stopword adalah kumpulan kata yang sering muncul dalam dokumen. *Stopword* biasanya merupakan konjungsi atau kata hubung yang tidak bermakna sehingga *stopword* dapat tidak diikutsertakan dalam proses pengindeksan.

Tabel 3.5 Contoh Proses *Stopword Removal*

Sebelum *Stopword Removal*	**Setelah *Stopword Removal***
Enggak sih, soal childfree menurut gue lebih ke pribadi. Siap enggak secara mental, financial, dll punya anak, bahkan trauma juga bisa mempengaruhi orang buat ambil keputusan childfree.	childfree pribadi mental financial trauma mempengaruhi keputusan

3.4.5 *Stemming*

Stemming merupakan proses untuk menghilangkan infleksi suatu kata, atau kata yang mengandung *prefix* ataupun *suffix* menjadi ke dalam bentuk dasarnya, tetapi bukan berarti bentuk dasar sama dengan akar kata (*root word*).

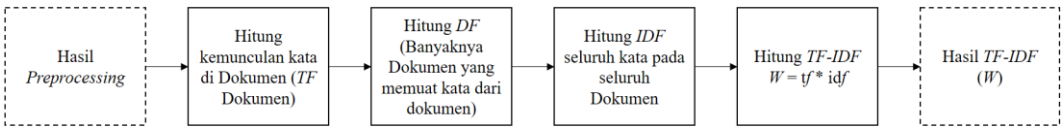
Tabel 3.6 Contoh Proses *Stemming*

Sebelum *Stemming*	**Setelah *Stemming***
setiap	tiap
perempuan	perempuan
BERHAK	hak
mengambil	ambil
keputusan	putus
dalam	dalam
hidupnya	hidup

3.5 *Improved K-Nearest Neighbor*

3.5.1 Pembobotan Kata *TF-IDF*

Proses pembobotan kata dalam penelitian ini dilakukan dengan menggunakan metode *TF-IDF* (*Term Frequency-Inverse Document Frequency*), *term* yang dihasilkan melalui proses *stemming* akan dihitung dengan metode ini. Metode *TF-IDF* digunakan untuk menghitung bobot pada tiap-tiap kata dalam dokumen. Berikut alur pembobotan *TF-IDF*:



Gambar 3.4 Blok Diagram Pembobotan *TF-IDF*

Tabel 3.7 dibawah merupakan data latih dan data uji yang telah dilakukan proses *preprocessing* yang kemudian dilakukan pembobotan kata *TF-IDF*

Tabel 3.7 Data Latih dan Data Uji untuk Pembobotan Kata *TF-IDF*

No.	**Dokumen**	***Tweet***	**Label**
1	D1	nyata berisik bilang childfree childfree udah berisik putus childfree terus batre	Negatif
2	D2	konten suka youtube opini kait childfree pilih serah preferensi pilih putus anak childfree	Positif
3	D3	jujur putus childfree orang anak childfree natural antiaging mamah anak awetmuda antiaging kunci duit	Negatif
4	D4	inti hidup bikin orang putus childfree normal valid isi tuju	Positif
5	D5	serah childfree anak childfree	?

Tabel 3.8 Pembobotan Kata *TF-IDF*

No.	Term	TF					DF	n/df	IDF	TF-IDF (W)				
		D1	D2	D3	D4	D5			(log n/df)	D1	D2	D3	D4	D5
1	nyata	1	0	0	0	0	1	5	0,7	0,7	0	0	0	0
2	berisik	2	0	0	0	0	1	5	0,7	1,4	0	0	0	0
3	bilang	1	0	0	0	0	1	5	0,7	0,7	0	0	0	0
4	childfree	3	2	2	1	2	5	1	0	0	0	0	0	0
5	udah	1	0	0	0	0	1	5	0,7	0,7	0	0	0	0
6	putus	1	1	1	1	0	4	1,3	0,1	0,1	0,1	0,1	0,1	0
7	terus	1	0	0	0	0	1	5	0,7	0,7	0	0	0	0
8	batre	1	0	0	0	0	1	5	0,7	0,7	0	0	0	0
9	konten	0	1	0	0	0	1	5	0,7	0	0,7	0	0	0
10	suka	0	1	0	0	0	1	5	0,7	0	0,7	0	0	0
11	youtube	0	1	0	0	0	1	5	0,7	0	0,7	0	0	0
12	opini	0	1	0	0	0	1	5	0,7	0	0,7	0	0	0
13	kait	0	1	0	0	0	1	5	0,7	0	0,7	0	0	0
14	pilih	0	2	0	0	0	1	5	0,7	0	1,4	0	0	0
15	serah	0	1	0	0	1	2	2,5	0,4	0	0,4	0	0	0,4
16	pikir	0	1	0	0	0	1	5	0,7	0	0,7	0	0	0
17	preferensi	0	1	0	0	0	1	5	0,7	0	0,7	0	0	0
18	anak	0	1	2	0	1	3	1,7	0,2	0	0,2	0,4	0	0,2
19	jujur	0	0	1	0	0	1	5	0,7	0	0	0,7	0	0
20	orang	0	0	1	1	0	2	2,5	0,4	0	0	0,4	0,4	0
21	natural	0	0	1	0	0	1	5	0,7	0	0	0,7	0	0
22	antiaging	0	0	2	0	0	1	5	0,7	0	0	1,4	0	0
23	mamah	0	0	1	0	0	1	5	0,7	0	0	0,7	0	0
24	awetmuda	0	0	1	0	0	1	5	0,7	0	0	0,7	0	0
25	kunci	0	0	1	0	0	1	5	0,7	0	0	0,7	0	0
26	duit	0	0	1	0	0	1	5	0,7	0	0	0,7	0	0
27	inti	0	0	1	0	0	1	5	0,7	0	0	0,7	0	0
28	hidup	0	0	1	0	0	1	5	0,7	0	0	0,7	0	0
29	bikin	0	0	1	0	0	1	5	0,7	0	0	0,7	0	0
30	normal	0	0	1	0	0	1	5	0,7	0	0	0,7	0	0
31	valid	0	0	1	0	0	1	5	0,7	0	0	0,7	0	0
32	isi	0	0	1	0	0	1	5	0,7	0	0	0,7	0	0
33	tuju	0	0	1	0	0	1	5	0,7	0	0	0,7	0	0

3.5.2 Perhitungan *Cosine Similarity*

Langkah awal dalam perhitungan *Cosine Similarity* ialah melakukan perkalian skalar antara D5 dengan 4 dokumen yang sudah terklasifikasi (D1, D2,

D3, dan D4).

Tabel 3.9 Perkalian Skalar WD5 dengan WD*i*

No.	WD5*WD*i*			
	D1	D2	D3	D4
1	0	0	0	0
2	0	0	0	0
3	0	0	0	0
4	0	0	0	0
5	0	0	0	0
6	0	0	0	0
7	0	0	0	0
8	0	0	0	0
9	0	0	0	0
10	0	0	0	0
11	0	0	0	0
12	0	0	0	0
13	0	0	0	0
14	0	0	0	0
15	0	0,158	0	0
16	0	0	0	0
17	0	0	0	0
18	0	0,049	0,098	0
19	0	0	0	0
20	0	0	0	0
21	0	0	0	0
22	0	0	0	0
23	0	0	0	0
24	0	0	0	0
25	0	0	0	0
26	0	0	0	0
27	0	0	0	0
28	0	0	0	0
29	0	0	0	0
30	0	0	0	0
31	0	0	0	0
32	0	0	0	0
33	0	0	0	0
Jumlah	0	0,208	0,098	0

Langkah kedua dalam perhitungan *Cosine Similarity* ialah menghitung panjang vektor setiap dokumen termasuk D5. Caranya yaitu dengan menguadratkan

setiap *term* (kata) dalam setiap dokumen, kemudian menjumlahkan nilai kuadrat tersebut lalu diakarkan.

Tabel 3.10 Perhitungan Panjang Vektor Seluruh Dokumen

No.	Panjang Vektor				
	D1	D2	D3	D4	D5
1	0,489	0	0	0	0
2	1,954	0	0	0	0
3	0,489	0	0	0	0
4	0	0	0	0	0
5	0,489	0	0	0	0
6	0,009	0,009	0,009	0,009	0
7	0,489	0	0	0	0
8	0,489	0	0	0	0
9	0	0,489	0	0	0
10	0	0,489	0	0	0
11	0	0,489	0	0	0
12	0	0,489	0	0	0
13	0	0,489	0	0	0
14	0	1,954	0	0	0
15	0	0,158	0	0	0,158
16	0	0,489	0	0	0
17	0	0,489	0	0	0
18	0	0,049	0,197	0	0,049
19	0	0	0,489	0	0
20	0	0	0,158	0,158	0
21	0	0	0,489	0	0
22	0	0	1,954	0	0
23	0	0	0,489	0	0
24	0	0	0,489	0	0
25	0	0	0,489	0	0
26	0	0	0,489	0	0
27	0	0	0,489	0	0
28	0	0	0,489	0	0
29	0	0	0,489	0	0
30	0	0	0,489	0	0
31	0	0	0,489	0	0
32	0	0	0,489	0	0
33	0	0	0,489	0	0
Jumlah	4,406	5,591	8,670	0,168	0,208
Akar	2,099	2,365	2,945	0,410	0,456

Langkah ketiga dalam perhitungan *Cosine Similarity* ialah menghitung kemiripan antara D5 dengan D1, D2, D3, dan D4 dengan menerapkan rumus *Cosine Similarity* (persamaan ke-2)

Cos(D5, D1) = 0 / (0,456 * 2,009) = 0

Cos(D5, D2) = 0,208 / (0,456 * 2,365) = 0,19287

Cos(D5, D3) = 0,098 / (0,456 * 2,945) = 0,07297

Cos(D5, D4) = 0 / (0,456 * 0,410) = 0

Tabel dibawah merupakan hasil perhitungan kemiripan antara D5 dengan D1, D2, D3, dan D4

Tabel 3.11 Hasil Pehitungan *Cosine Similarity*

D1	**D2**	**D3**	**D4**
0	0,19287	0,07297	0

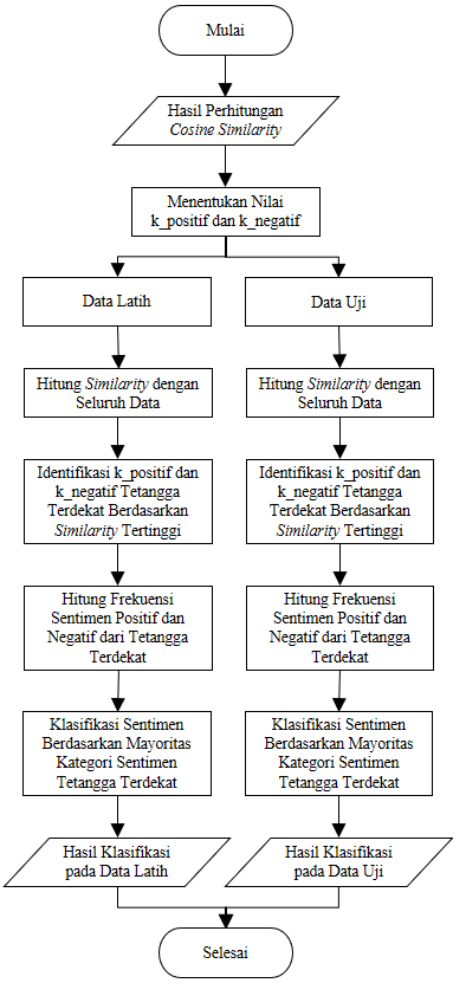
Langkah selanjutnya ialah mengurutkan hasil perhitungan *Cosine Similarity* dari nilai yang terbesar hingga terkecil.

Tabel 3.12 Pengurutan Nilai *Cosine SImilarity*

1	**2**	**3**	**4**
D2	D3	D1	D4
0,19287	0,07297	0	0

3.5.3 Implementasi Algoritma *Improved K-Nearest Neighbor*

Algoritma yang digunakan dalam klasifikasi tanggapan masyarakat mengenai *childfree* pada Twitter ialah algoritma *Improved K-Nearest Neighbor*. Penerapan algoritma *Improved K-Nearest Neighbor* dimulai dengan pembobotan kata menggunakan metode *TF-IDF*, yang selanjutnya digunakan untuk mengukur kemiripan dokumen menggunakan metode *Cosine Similarity*. Berikut merupakan alur dari proses klasifikasi menggunakan metode *Improved K-Nearest Neighbor*.



Gambar 3.5 *Flowchart* Proses Klasifikasi Metode *IKNN*

Langkah selanjutnya ialah menghitung nilai *n* (*k*-baru) untuk tiap kategori. Berikut merupakan perhitungan untuk nilai *k*-baru:

$$n\ positif = \left[\frac{3 \times 2}{2}\right] = 3$$

$$n\ negatif = \left[\frac{3 \times 2}{2}\right] = 3$$

Kemudian menghitung perbandingan kemiripan antara data latih dengan data uji pada setiap kategori. Berikut perhitungannya:

$$Cos\ positif = D2 + D4 = 0{,}07297 + 0 = 0{,}07297$$

$$Cos\ negatif = D1 + D3 = 0{,}19287 + 0 = 0{,}19287$$

Selanjutnya menghitung penjumlahan similaritas sebanyak top *n* tetangga pada data latih

$$Cosim\ Data\ Latih = D2 + D3 + D1 = 0{,}19287 + \ 0{,}07297 + 0 = 0{,}26584$$

Selanjutnya menghitung nilai maksimum perbandingan kemiripan antara dokumen D5 dengan data latih sebanyak top *n* tetangga

$$p(x, cm)\ positif = \frac{0{,}07297}{0{,}26584} = 0{,}27448$$

$$p(x, cm)\ negatif = \frac{0{,}19287}{0{,}26584} = 0{,}72551$$

>> Nilai maksimum merupakan kategori Negatif, jadi D5 terklasifikasi dalam kategori ***Negatif***.

3.6 Skenario Pengujian

Pengujian dalam penelitian ini menggunakan metode *Confusion Matrix* yang menunjukkan hasil dari evaluasi model yang digunakan dengan menggunakan tabel matrik, dalam penelitian ini *dataset* terdiri dari 2 kelas, yakni kelas positif dan kelas negatif. Evaluasi kinerja model menggunakan metode *Confusion Matrix* menghasilkan nilai *accuracy*, *precision*, dan *recall*. Mengukur kinerja sistem penting dilakukan untuk mengetahui akurasi yang dihasilkan oleh metode tersebut sehingga dapat diperoleh informasi tentang seberapa baik sistem untuk mengklasifikasikan data. Pada dasarnya, *Confusion Matrix* berisi informasi yang

membandingkan hasil klasifikasi yang dibuat oleh sistem dengan hasil klasifikasi yang seharusnya.

Tabel 3.13 *Confusion Matrix*

	Kelas Prediksi		
Kelas Sebenarnya		+	-
	+	*True Positive* (TP)	*False Negative* (FN)
	-	*False Positive* (FP)	*True Negative* (TN)

Confusion matrix menggunakan 4 (empat) istilah yang digunakan untuk merepresentasikan hasil proses klasifikasi, diantaranya *True Positive* (TP), *False Negative* (FN), *False Positive* (FP), dan *True Negative* (TN). Dimana *True Positive* (TP) merupakan jumlah data positif yang diklasifikasikan sebagai positif, *False Negative* (FN) merupakan jumlah data negatif yang diklasifikan sebagai positif, *False Positive* (FP) merupakan jumlah data positif yang diklasifikan sebagai negatif, *True Negative* (TN) merupakan jumlah data negatif yang diklasifikasikan sebagai negatif.

Selanjutnya memasukkan data uji dalam *Confusion Matrix*, nilai-nilai yang dimasukkan diperlukan untuk menghitung *accuracy*, *precision*, dan *recall*. Dimana *accuracy* dalam klasifikasi merupakan sebuah representasi ketepatan *record* data yang diklasifikasikan secara benar setelah dilakukannya pengujian pada hasil klasifikasi, *precision* merupakan proposikasi yang diprediksi positif yang pada data sebenarnya juga positif, sedangkan *recall* ialah proporsi data positif yang sebenarnya juga diprediksi positif secara benar. Berikut merupakan persamaan model *Confusion Matrix* (Putra & Wibowo, 2020):

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \quad (3.1)$$

Accuracy merupakan jumlah dari perbandingan data yang benar dari jumlah keseluruhan data

$$Precision = \frac{TP}{TP + FP} \quad (3.2)$$

Precision digunakan untuk mengukur proporsi kelas data positif yang diprediksi secara tepat dari hasil prediksi pada kelas positif secara keseluruhan

$$Recall = \frac{TP}{TP + FN} \quad (3.3)$$

Recall digunakan untuk mengukur presentase kelas data positif yang diprediksi secara tepat dari hasil prediksi pada kelas positif secara keseluruhan

BAB IV

UJI COBA DAN PEMBAHASAN

4.1 Skenario Uji Coba

Uji coba pada penelitian sistem klasifikasi tanggapan masyarakat menggunakan metode *Improved K-Nearest Neighbor* dilakukan dengan implementasi bahasa pemrograman *Python* dan menggunakan *Google Colab*. Uji coba dilakukan untuk mengukur performa sistem yang dibangun, termasuk *accuracy*, *precision*, dan *recall*. Tujuannya ialah untuk memastikan bahwa sistem dapat berjalan secara efektif dan memberikan hasil yang baik untuk pemahaman juga analisis terhadap penerapan algoritma yang digunakan.

4.1.1 Persiapan *Dataset*

Dataset dalam penelitian ini diperoleh dari *crawling* data Twitter yang disebut *tweets*, *tweets* dikumpulkan dengan alat bantu atau *library* bernama "*tweet-harvest*" dengan token akses '8c5f117d476c6cce1a7cf28cfbd6623d4a0c4dc8' yang merupakan bagian dari autentikasi dalam mengakses API Twitter. Semua *tweets* yang berhasil dikumpulkan dijadikan satu dalam file berekstensi *.csv*

Dataset yang digunakan sebanyak 1493 *tweets* yang telah diberi label sentimen positif dan negatif. 793 *tweets* berlabel positif, dan 700 *tweets* berlabel negatif. Berikut merupakan contoh *dataset* yang digunakan:

Tabel 4.1 Contoh *Dataset*

No.	*Tweet*	Label
1	@tanyakanrl Enggak sih, soal childfree menurut gue lebih ke pribadi. Siap enggak secara mental, financial, dll punya anak, bahkan trauma juga bisa mempengaruhi orang buat ambil keputusan childfree.	Positif

Lanjutan Tabel 4.1

No.	*Tweet*	Label
2	Sebelum gw pacaran gw udh blg ke papa gw kalo gw gamau nikah dan childfree, and he's totally fine with that! Tp stlh pny pacar gw jd pengen nikahin dia krn ternyata msh ada yg mau sm gw hehe. Gw cmn mau dia, klo gbs ya gw balik ke keputusan awal, gamau nikah.	Positif
3	@convomfs itu keputusan masing-masing, tapi klo childfree sampe tua ya keturunan bakal berenti di situ. at least klo gamau punya anak kandung adopsi aja walau adopsi anak yg udh gede, banyak soalnya di luar sana yang pengen ngerasain kasih sayang ortu	Positif
4	it's ok aja sih, asal ada keputusan dari kedua belah pihak, bukan hanya dari salah satunya. apalagi kalau alasan buat childfree karena ngga siap dari segi mental & fisik, karena kalau dipaksain buat punya anak, malah bisa ngerusak mental anaknya	Positif
5	@tanyakanrl Semoga hal tersebut gak ngebuatmu insecure, nder. Damaikan diri, jangan jadikan keputusan orang yg mau/enggak sama kamu jadi parameter nilai diri. Dan sepertinya, yang penting bisa saling terima. Nantinya mau edukasi dini ke anak/childfree putusin bersama. Dan banyakin duit aja😊	Positif
6	Terus mau lu ape, markonah? "Istri itu harusnya terserah dia dalam artian jadi wanita karier aja, terserah dia mau ngurus anaknya secara full time apa engga, semua pendapat dan keputusan terserah dia, bahkan gak usah nikah, childfree aja dari awal." Itu jawaban yang lo mau?	Negatif
7	real talk with my mom, keputusan buat childfree udah bulat, nyokap dukung-dukung aja.. jadi lebih nyaman jalanin hidup ☺☺ *tapi amal jariyahnya disuruh bikin sumur di afrika wkwk siaap nyonya https://t.co/CJql1Wpg3c	Positif
8	Sbnrnya gue ga nentang keinginan child free sendernya dim. Gue mah malah menghargai apapun keputusan org lain ttg anak, entah childfree atau engga, krn org punya pemikiran masing2. Keluarga deket gue jg banyak soalnya yg broken home dan korban kdrt Cuma rada aneh aja liat—	Positif
9	@indahpertiwii9 @crowdedrep @itssannia @txtdrimedia Indonesia lebih cocok mengatasi masalah ngewe dibawah umur/diluar nikah tuh masih masif, pendidikan yang belum merata, maka edukasi nya tergantung lingkungan. childfree bukan untuk di kampanye kan, lebih cocok dokter konsultasi yang ngomong jika masalah fisik dan mental	Positif
10	Kita semua tau dia ingin orang-orang menerima keputusan dia memilih childfree. But this is so wrong. Mending lo bantu anak-anak muda yang suka minum alkohol, ngerokok, makan makanan gak sehat, clubbing sampe pagi. Karena hal-hal itulah yang mempercepat penuaan dini. https://t.co/VGaQRur2SP	Negatif
11	Gw juga tim childfree tapi so sorry kalimat dia gak banget. Kalo mau awet muda, banyakin duit, Perawatan noh, kayak idol kpop sebulan abisin 2M buat wajah dan badan. Serah lu milih mau punya anak atau enggak, tapi HORMATI KEPUTUSAN ORANG LAIN YANG BEDA DARI LU	Negatif
12	—oleh setiap perempuan baik yang punya maupun tdk punya anak. Perempuan menua secara fisik bukan berarti dia stress dan tidak bahagia, itu memang normal dan alamiah terjadi" Alias lu marah kalo keputusan lu childfree dikomenin orang, tapi lu ngomongin perempuan yang punya anak—	Positif
13	gue gapeduli dia mau childfree or whatever, people freely choose kan ya asal jangan downgrade juga keputusan orang lain dan generalisir. let them be them. mbak gigi punya anak 2, masih awet muda, bisa tetep jalan jalan :P	Negatif

14	Jadikanlah dunia di tanganmu bkn di hatimu. Rasulullah SAW mengajak kita menikah untuk memperoleh keturunan jg. Yaudah kalo memang Anda childfree ya itu keputusan Anda. 2 kalimat teratas tidak ada dlm kehidupan Anda. Satu lagi, do not blame children for making you live unhappy.	Positif
15	sometimes i feel sorry for gita. org 'nonhijab' lain kalo bilang childfree blablabla dianggap itu keputusan individu tsb, biarin aja, dll. Pas dia yg ngomong (kebetulan berhijab) dan sekaligus ngebecandain emosi netizen, busetdah trending mulu.	Negatif
16	jujur gpernah permasalahin keputusan seseorang utk childfree. tp liat reply dia ini kok agak kesel ya jatohnya kyk nyepelein orang yang bener2 mau punya anak. childfree = natural antiaging? no, mamahku anak 3 ttp awet muda tuh. kalo mau anti aging tu kuncinya adalah byk duit	Negatif
17	@rengginang_p @shattletan harusnya dia fokus bahas soal pentingnya persiapan mental menjadi ibu atau edukasi orang-orang alasan kenapa orang memilih untuk childfree, cth : karena kesiapan diri. bukan malah cuap2 bilang anak pembawa masalah dan buat ibu-ibu merasa salah ambil keputusan buat punya anak.	Negatif
18	Poor you, gitasav. Aku sebenarnya respect dg keputusan org mau childfree atau apapun tapi kalo sdh ngomong negatif ttg anak tuh marah bgtt.	Negatif
19	on the other note, indonesian netizen jg pls normalize keputusan org yg childfree tanpa embel2 “liat ntar tua nya happy ga” KAYAK LEAVE HER ALONE JG (meski ga bs dikontrol jari2 bangsat, apalagi kl lo banyak followers). dah tau ni gitsav gampang kesulut & emosian.	Negatif
20	Kurasa keputusan dy childfree itu cm 1 fase aja dr proses recovery mental, ntar klo dah beres, bs aja kan berubah pikiran□. Cmn masalahnya, ndak perlu dikoar2in smp berantem trus ma netijen (yg memperparah mentalnya). Yg cf & anteng2 bae lho byk. Napa gt pilih jalan yg ruwet□	Negatif
21	Gue ada masalah sm keputusan childfree, tmn gue ada yg udh decide buat ga nikah dan ga punya anak. I respect her. Tp si gitasav ini knp sih gak ada respect sama sekali ke other women yg pgn punya anak dan jadi ibu.	Negatif
22	sebenernya keputusan mereka buat “childfree” itu agak ugly. banyak pasangan yg menanti hadirnya anak dalam hidupnya. cause having child is the most beautiful gift & very meaningful.	Negatif
23	@FaizaMardz Haha iya Mbak, bingung aku sama ybs, perihal keliatan muda di usia tertentu aja dikaitin sama keputusan punya anak apa ndak. Ngerti sih kalo mungkin ybs punya agenda promoting childfree. Tapi yo ora masuk, wong permasalahannya di wrinkles yg marginalized in beauty discourse. □	Negatif
24	sumpah ya gue ga peduli gita mau childfree atau apapun itu, tp bisa ga gausah merasa superior dengan keputusan childfree lo itu anjir???	Negatif
25	@chachathaib Ceunah kalo orang2 boleh posting ttg anak2nya, kenapa dia yg memilih childfree ga boleh posting ttg keputusan dia? Hahaha cape bgt liat opini2 dia	Negatif
26	Awalnya dulu bgt, alasan-alasan Gita utk childfree tuh masuk akal. Gw menghargai keputusan dia, ya siapa gw jg kan menghujat pilihan dia? Tapi comment dia ini beneran di luar nalar gw haha. Ngga mau punya anak, tp kok seakan menghujat orang lain yang punya anak?	Negatif
27	childfree or not, itu keputusan masing2 individu. krn hidup ya hidup mereka. yg merasa sanggup then good for them, wish them a lot of luck. yg merasa ga sanggup dan mutusin utk childfree they know what’s the best	Positif
28	@maririn_lie @themartinny mjb. Iyah karena kontennya itu cenderung 'menjulidkan' atau 'mengkonotasikan negatif' opini lain dengan menyebut 'early aging'. Seandainya dia netral atau hanya membicarakan hal positif dari keputusan childfree mungkin orang2 akan biasa aja.	Negatif
29	@littleangel48 Org tuamu nguliahin km di luar negeri jg bkn buat ngatain org lain dongo atau stunting hey! Kl happy sm keputusan childfree yaudah	Negatif

	gausah dikit2 nyinyir org yg memilih pilihan buat puny anak. Org yg pny anak jg ga pernah nyinyir keputusanmu bt childfree kan? Narrow minded ini mah	
30	In my opinion kalo nyokap gw milih childfree juga gw ga masalah si gaada di dunia ini, as long as she happy dan pasti untuk keputusan childfree itu butuh pertimbangan yang mendalam, pasti ada alasannya. Jadi mending aku gaada di dunia dripada maksain misal nykp blm/ga siap py ank	Positif

4.1.2 *Text Preprocessing*

Text Preprocessing dilakukan untuk membersihkan, mengubah, dan mempersiapkan data teks sebelum digunakan untuk tahap selanjutnya. Tujuan dari tahap ini ialah untuk meningkatkan kualitas data dan menghilangkan *noise* atau gangguan yang mungkin mempengaruhi atau bahkan mengganggu hasil kerja sistem. *Text preprocessing* membantu memastikan data yang digunakan untuk klasifikasi merupakan data bersih. *Text preprocessing* yang dilakukan dalam penelitian ini terdapat beberapa tahap, yaitu sebagai berikut:

a. *Case Folding*

Case folding dilakukan untuk mengubah semua huruf menjadi huruf kecil (*lowercase*). *Case folding* membantu menghindari perbedaan huruf besar dan huruf kecil dalam kata yang mungkin dapat menjadi kebingungan sistem dalam memproses dan memahami data. Penggunaan *text.lower* dalam fungsi bertujuan untuk mengubah *text* atau *string* menjadi huruf kecil. Gambar 4.1 merupakan hasil dari implementasi tahap *case folding*.

	Text	Label	Case Folding
0	@tanyakanrl Enggak sih, soal childfree menurut...	positif	@tanyakanrl enggak sih, soal childfree menurut...
1	Sebelum gw pacaran gw udh blg ke papa gw kalo ...	positif	sebelum gw pacaran gw udh blg ke papa gw kalo ...
2	@convomfs itu keputusan masing-masing, tapi kl...	positif	@convomfs itu keputusan masing-masing, tapi kl...
3	it's ok aja sih, asal ada keputusan dari kedua...	positif	it's ok aja sih, asal ada keputusan dari kedua...
4	@tanyakanrl Semoga hal tersebut gak ngebuatmu ...	positif	@tanyakanrl semoga hal tersebut gak ngebuatmu ...
5	Terus mau lu ape, markonah? "Istri itu harusny...	negatif	terus mau lu ape, markonah? "istri itu harusny...
6	real talk with my mom, keputusan buat childfre...	positif	real talk with my mom, keputusan buat childfre...
7	Sbnrnya gue ga nentang keinginan child free se...	positif	sbnrnya gue ga nentang keinginan child free se...
8	@indahpertiwii9 @crowdedrep @itssannia @txtdr...	positif	@indahpertiwii9 @crowdedrep @itssannia @txtdr...
9	Kita semua tau dia ingin orang-orang menerima ...	negatif	kita semua tau dia ingin orang-orang menerima ...
10	Gw juga tim childfree tapi so sorry kalimat d...	negatif	gw juga tim childfree tapi so sorry kalimat d...
11	—oleh setiap perempuan baik yang punya maupun ...	positif	—oleh setiap perempuan baik yang punya maupun ...
12	gue gapeduli dia mau childfree or whatever, pe...	positif	gue gapeduli dia mau childfree or whatever, pe...
13	Jadikanlah dunia di tanganmu bkn di hatimu. Ra...	positif	jadikanlah dunia di tanganmu bkn di hatimu. ra...
14	sometimes i feel sorry for gita. org 'nonhijab...	negatif	sometimes i feel sorry for gita. org 'nonhijab...
15	jujur gpernah permasalahin keputusan seseorang...	negatif	jujur gpernah permasalahin keputusan seseorang...
16	@rengginang_p @shattletan harusnya dia fokus ...	positif	@rengginang_p @shattletan harusnya dia fokus ...
17	Poor you, gitasav. Aku sebenarnya respect dg k...	negatif	poor you, gitasav. aku sebenarnya respect dg k...
18	on the other note, indonesian netizen jg pls n...	negatif	on the other note, indonesian netizen jg pls n...
19	Kurasa keputusan dy childfree itu cm 1 fase aj...	negatif	kurasa keputusan dy childfree itu cm 1 fase aj...
20	Gue ada masalah sm keputusan childfree, tmn gu...	negatif	gue ada masalah sm keputusan childfree, tmn gu...

Gambar 4.1 Hasil Implementasi Tahap *Case Folding*

b. *Cleansing*

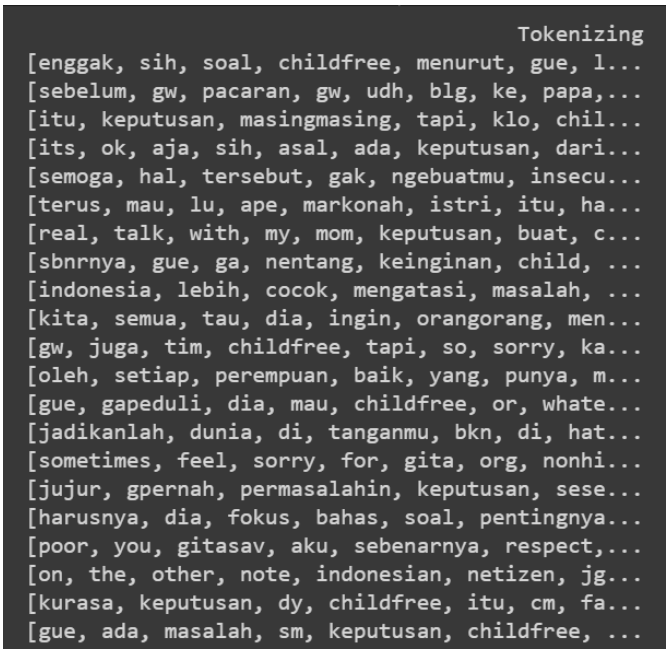
Cleansing dilakukan untuk membersihkan data dari elemen-elemen yang tidak diperlukan, seperti username Twitter, tautan atau *link*, tanda baca, *emoticon*, dan karakter khusus. Elemen-elemen yang dihapus, akan digantikan dengan spasi. Proses *cleansing* terkumpul dalam fungsi *clean_text*. Gambar 4.2 merupakan hasil dari implementasi tahap *cleansing*.

Cleansing
enggak sih soal childfree menurut gue lebih ke...
sebelum gw pacaran gw udh blg ke papa gw kalo ...
itu keputusan masingmasing tapi klo childfree ...
its ok aja sih asal ada keputusan dari kedua b...
semoga hal tersebut gak ngebuatmu insecure nde...
terus mau lu ape markonah istri itu harusnya t...
real talk with my mom keputusan buat childfree...
sbnrnya gue ga nentang keinginan child free se...
indonesia lebih cocok mengatasi masalah ngewe ...
kita semua tau dia ingin orangorang menerima k...
gw juga tim childfree tapi so sorry kalimat di...
oleh setiap perempuan baik yang punya maupun t...
gue gapeduli dia mau childfree or whatever peo...
jadikanlah dunia di tanganmu bkn di hatimu ras...
sometimes feel sorry for gita org nonhijab lai...
jujur gpernah permasalahin keputusan seseorang...
harusnya dia fokus bahas soal pentingnya persi...
poor you gitasav aku sebenarnya respect dg kep...
on the other note indonesian netizen jg pls no...
kurasa keputusan dy childfree itu cm fase aja ...
gue ada masalah sm keputusan childfree tmn gue...

Gambar 4.2 Hasil Implementasi Tahap *Cleansing*

c. *Tokenizing*

Tokenizing dilakukan untuk memisahkan teks atau kalimat menjadi token atau bagian-bagian yang lebih kecil, dalam hal ini ialah menjadi kata per-kata atau frasa sehingga lebih mudah dikelola dan dihitung oleh sistem. *Tokenizing* membantu sistem mengidentifikasi struktur teks dan membantu dalam mengenali entitas, kata, atau fitur yang relevan dalam teks. Implementasi tahap *tokenizing* dilakukan menggunakan modul bernama *RegexpTokenizer* dari pustaka *nltk.tokenize* yang merupakan fungsi untuk memecah teks menjadi token atau kata. Gambar 4.3 merupakan hasil dari implementasi tahap *tokenizing*.



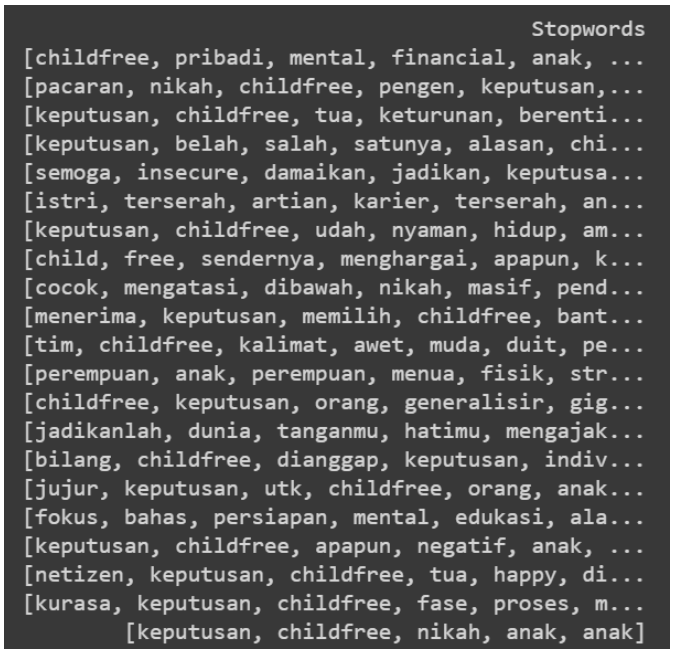
Tokenizing
[enggak, sih, soal, childfree, menurut, gue, l...
[sebelum, gw, pacaran, gw, udh, blg, ke, papa,...
[itu, keputusan, masingmasing, tapi, klo, chil...
[its, ok, aja, sih, asal, ada, keputusan, dari...
[semoga, hal, tersebut, gak, ngebuatmu, insecu...
[terus, mau, lu, ape, markonah, istri, itu, ha...
[real, talk, with, my, mom, keputusan, buat, c...
[sbnrnya, gue, ga, nentang, keinginan, child, ...
[indonesia, lebih, cocok, mengatasi, masalah, ...
[kita, semua, tau, dia, ingin, orangorang, men...
[gw, juga, tim, childfree, tapi, so, sorry, ka...
[oleh, setiap, perempuan, baik, yang, punya, m...
[gue, gapeduli, dia, mau, childfree, or, whate...
[jadikanlah, dunia, di, tanganmu, bkn, di, hat...
[sometimes, feel, sorry, for, gita, org, nonhi...
[jujur, gpernah, permasalahin, keputusan, sese...
[harusnya, dia, fokus, bahas, soal, pentingnya...
[poor, you, gitasav, aku, sebenarnya, respect,...
[on, the, other, note, indonesian, netizen, jg...
[kurasa, keputusan, dy, childfree, itu, cm, fa...
[gue, ada, masalah, sm, keputusan, childfree, ...

Gambar 4.3 Hasil Implementasi Tahap *Tokenizing*

d. *Stopword Removal*

Stopword removal dilakukan untuk menghilangkan kata hubung atau kata yang tidak bermakna. Tujuannya ialah untuk membersihkan kalimat atau teks dari

kata yang sering muncul dan tidak memiliki makna seperti konjungsi. Menghapus *stopwords* dapat memfokuskan sistem pada kata-kata penting atau kata kunci yang lebih informatif. Hal ini juga dapat membantu mengurangi ukuran data dan meningkatkan efisiensi analisis. Proses *stopword removal* menggunakan *library nltk*, kamus *stopwords* bahasa Indonesia diunduh dari *nltk* dengan perintah '*nltk.download("stopwords")*'. Peneliti juga menambah kata-kata *stop* yang diunduh dari *github.com/irfandythalib/python-indonesia-stopwords-remover*, juga menambahkan kata-kata *stop* yang tidak terdapat dalam kamus, seperti "*kl*", "*knp*", "*yh*", dan lain sebagainya. Gambar 4.4 merupakan hasil dari implementasi tahap *stopword removal*.

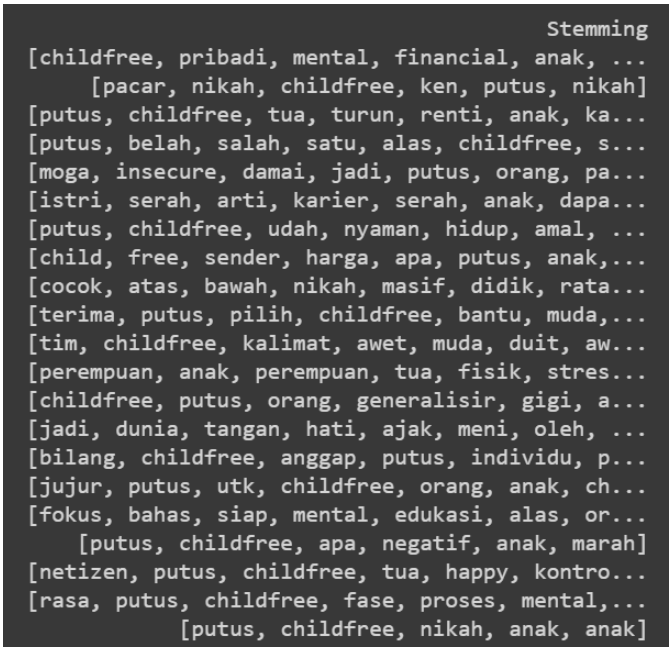


Gambar 4.4 Hasil Implementasi Tahap *Stopword Removal*

e. *Stemming*

Stemming dilakukan untuk mengubah kata dalam teks atau kalimat ke dalam bentuk dasarnya sehingga kata-kata yang memiliki akar kata yang sama akan

dianggap sama. Tujuannya ialah untuk mengurangi variasi kata dalam teks dan mengidentifikasi kata-kata yang memiliki makna yang serupa. *Stemming* membantu sistem dalam perhitungan frekuensi dan membantu dalam pengurangan dimensi data juga meningkatkan efisiensi analisis. Implementasi *stemming* dilakukan menggunakan modul *StemmerFactory* dari *library Sastrawi*. Gambar 4.5 merupakan hasil dari implementasi tahap *stemming*.



Stemming
[childfree, pribadi, mental, financial, anak, ...
[pacar, nikah, childfree, ken, putus, nikah]
[putus, childfree, tua, turun, renti, anak, ka...
[putus, belah, salah, satu, alas, childfree, s...
[moga, insecure, damai, jadi, putus, orang, pa...
[istri, serah, arti, karier, serah, anak, dapa...
[putus, childfree, udah, nyaman, hidup, amal, ...
[child, free, sender, harga, apa, putus, anak,...
[cocok, atas, bawah, nikah, masif, didik, rata...
[terima, putus, pilih, childfree, bantu, muda,...
[tim, childfree, kalimat, awet, muda, duit, aw...
[perempuan, anak, perempuan, tua, fisik, stres...
[childfree, putus, orang, generalisir, gigi, a...
[jadi, dunia, tangan, hati, ajak, meni, oleh, ...
[bilang, childfree, anggap, putus, individu, p...
[jujur, putus, utk, childfree, orang, anak, ch...
[fokus, bahas, siap, mental, edukasi, alas, or...
[putus, childfree, apa, negatif, anak, marah]
[netizen, putus, childfree, tua, happy, kontro...
[rasa, putus, childfree, fase, proses, mental,...
[putus, childfree, nikah, anak, anak]

Gambar 4.5 Hasil Implementasi Tahap *Stemming*

4.1.3 Pembagian *Dataset*

Dataset dibagi menjadi dua bagian, yakni: data pelatihan (*train* set) dan data pengujian (*test* set). Data pelatihan merupakan kelompok data yang digunakan untuk melatih atau mengembangkan sistem. Data pelatihan membantu sistem dalam memahami dan mempelajari pola data yang digunakan. Sedangkan data pengujian merupakan kelompok data yang digunakan untuk mencoba atau menguji sistem klasifikasi yang telah dilatih sebelumnya. Dalam hal ini, kedua jenis data digunakan

untuk melatih dan menguji model yang dapat mengklasifikasikan tanggapan masyarakat sebagai kategori positif atau negatif.

Pembagian dilakukan secara acak oleh sistem menggunakan *library* atau pustaka bernama *train test split* dengan tujuan untuk memastikan bahwa kedua set data memiliki variasi yang terdapat dalam *dataset* secara adil dan tidak bias, sehingga representasi yang didapatkan dapat lebih akurat. Berikut merupakan pembagian *dataset* yang dilakukan:

Tabel 4.2 Pembagian *Dataset*

Uji Coba	Data Latih	Data Uji
1	90%	10%
2	80%	20%
3	70%	30%
4	60%	40%

4.1.4 Pelatihan Model *Improved K-Nearest Neighbor*

Pada tahap pelatihan model, pembagian *dataset* disesuaikan dengan skenario yang telah dirancang sebelumnya. Percobaan dilakukan sebanyak 4 kali yaitu dengan menggunakan data latih sebanyak 90%, 80%, 70%, dan 60%.

a. Ekstraksi Fitur

Peneliti menggunakan metode *TF-IDF* (*Term Frequency-Inverse Document Frequency*) sebagai cara untuk memberikan bobot kata dalam data pelatihan. Metode ini mengubah teks menjadi representasi vektor numerik yang akan digunakan dalam analisis selanjutnya. Pada implementasinya, peneliti menggunakan *library* bernama *TfidfVectorizer*. Tabel 4.3 merupakan contoh hasil dari nilai *TF-IDF* yang dapat dibaca dengan, kalimat pada *index* ke-16, bobot dari kata ‘anak’ bernilai 0, bobot kata ‘*childfree’* bernilai 0,145791, bobot kata ‘pilih’ bernilai 0,477658, dan seterusnya.

Tabel 4.3 Contoh Hasil Pembobotan Kata *TF-IDF*

No.	Kalimat (*index* ke-)	Bobot (dari kata:)				
		anak	*childfree*	pilih	pribadi	putus
1	16	0	0,145791	0,477658	0	0
2	17	0,16759	0,090202	0	0	0
3	18	0	0,053186	0,174254	0,276124	0,134776
4	19	0,189162	0,101813	0,333572	0	0,258
5	20	0	0,09099	0	0	0
6	21	0,395191	0,106352	0	0,552145	0,269502
7	22	0	0,087885	0	0	0
8	23	0	0,046792	0,153305	0	0
9	24	0,260841	0,140392	0	0	0
10	25	0,091691	0,049351	0	0,256214	0,250116

b. Perhitungan Kemiripan Dokumen

Tahap ini melibatkan perhitungan *similarity* antar dokumen menggunakan *cosine similarity*. Ditahap awal, perhitungan dilakukan pada keseluruhan dataset dengan menggunakan representasi *TF-IDF*. Hasil dari perhitungan ini akan menciptakan matriks *cosine similarity* untuk seluruh data. Tabel 4.4 merupakan contoh hasil dari perhitungan *cosine similarity* pada keseluruhan data, yang dapat dibaca dengan, kemiripan antara kalimat *index* ke-0 dengan kalimat *index* ke-0 nilai kemiripannya sebesar 1, kalimat *index* ke-0 dengan kalimat *index* ke-1 nilai kemiripannya sebesar 0,0607.

Tabel 4.4 Hasil Perhitungan Kemiripan Dokumen pada Keseluruhan Data

index	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
0	1	0,0607	0,0583	0,2143	0,0512	0,0428	0,0412	0,0498	0,0550
1	0,0607	1	0,1215	0,0463	0,0358	0,1440	0,0515	0,0439	0,0916
2	0,0583	0,1215	1	0,0700	0,0188	0,0385	0,0270	0,0447	0,0022
3	0,2143	0,0463	0,0700	1	0,0219	0,0449	0,0315	0,0521	0,1626
4	0,0512	0,0358	0,0188	0,0219	1	0,0179	0,0243	0,0207	0,0842
5	0,0428	0,1440	0,0385	0,0449	0,0179	1	0,0257	0,0322	0,0239
6	0,0412	0,0515	0,0270	0,0315	0,0243	0,0257	1	0,0298	0,0028
7	0,0498	0,0439	0,0447	0,0521	0,0207	0,0322	0,0298	1	0,0024
8	0,0550	0,0916	0,0022	0,1626	0,0842	0,0239	0,0028	0,0024	1

Selanjutnya, dilakukan perhitungan *cosine similarity* khusus untuk data latih dan data uji guna membangun matriks *cosine similarity* yang akan digunakan oleh model *Improved K-Nearest Neighbor*. Matriks *cosine similarity* pada data latih dan data uji dapat memberikan informasi untuk mengidentifikasi dokumen-dokumen yang memiliki kemiripan tinggi dengan dokumen-dokumen yang telah dilihat sebelumnya (data latih). Dengan demikian, memahami kemiripan antar dokumen di kedua dataset membantu model dalam membuat prediksi yang lebih akurat, terutama saat menghadapi data yang belum pernah dilihat sebelumnya.

c. Pengaturan Parameter

Tahap ini peneliti menetapkan nilai *k* yang berbeda untuk tiap kategori sentimen, dimana nilai *k* pada kelas minoritas akan lebih besar dari kelas mayoritas, yakni *k*=11 untuk kategori positif, dan *k*=15 untuk kategori negatif. Hal ini dikarenakan ketidakseimbangan jumlah data dari kedua kelas, jumlah data berlabel positif lebih besar dari jumlah data berlabel negatif. Nilai *k* menentukan jumlah tetangga terdekat yang akan digunakan oleh model *Improved K-Nearest Neighbor* dalam mengklasifikasikan teks.

d. Pelatihan Model

Class ImprovedKNN dibuat untuk mengimplementasikan model *Improved K-Nearest Neighbor* dalam mengklasifikasikan sentimen. Dalam inisialisasinya, kelas menerima parameter *k_value_positif* dan *k_value_negatif* yang merepresentasikan nilai *k* yang berbeda untuk kategori positif dan negatif. Selain itu, terdapat opsi untuk menyertakan bobot (*weight_positif* dan *weight_negatif*) yang akan

diaplikasikan pada skor positif dan negatif.

Method hitung_bobot digunakan untuk menghitung bobot berdasarkan rasio jumlah dokumen positif terhadap dokumen negatif dalam dataset. Selanjutnya, terdapat dua *Method*, yaitu *hitung_k_value_positif* dan *hitung_k_value_negatif* yang melakukan perhitungan skor untuk masing-masing nilai *k* pada kategori positif dan negatif. Proses ini melibatkan pemilihan dokumen berdasarkan urutan skor tertinggi.

Method hitung_probabilitas mengintegrasikan hasil perhitungan kategori positif dan negatif untuk setiap fitur pada dataset. Metode ini memberikan output berupa *DataFrame* yang berisi hasil perhitungan. Terakhir, *Method compare_result* digunakan untuk membandingkan hasil dari kategori positif dan negatif untuk setiap fitur dan memberikan prediksi akhir berupa kategori sentimen, yaitu 'positif' atau 'negatif'. Metode ini berguna untuk mengevaluasi dan membandingkan hasil kategori dari setiap fitur dalam dataset.

4.1.5 Pengujian Model *Improved K-Nearest Neighbor*

Pengujian model dimulai dengan pembuatan sebuah *instance* dari *class ImprovedKNN* tanpa menetapkan bobot positif dan negatif. Selanjutnya, bobot dihitung menggunakan data latih yang terdiri dari fitur-fitur sentimen (*X_train*), dan label sentimen yang sesuai (*y_train*), melalui pemanggilan *method hitung_bobot* pada objek *calculator*.

Proses selanjutnya ialah menghitung probabilitas sentimen dihitung menggunakan data latih dan bobot yang telah dihitung sebelumnya. Proses yang sama dilakukan pada data uji, dimana probabilitas sentimen dihitung dan dibandingkan

menggunakan metode *hitung_probabilitas* pada objek *calculator*. Hal ini untuk memastikan konsistensi tipe data, kolom ‘Label’ pada dataset awal dikonversi kembali ke tipe data kategori.

Hasil dari perbandingan *Improved K-Nearest Neighbor* antara data latih dan data uji kemudian dicetak, memberikan gambaran tentang keputusan model pada kedua set data. Prediksi sentimen untuk data latih dan data uji disimpan untuk evaluasi performa model secara keseluruhan.

4.1.6 Evaluasi Model *Improved K-Nearest Neighbor*

Evaluasi model dalam penelitian dilakukan dengan mengukur sejumlah metrik untuk menggambarkan performa model. Langkah awal melibatkan penyamaan indeks untuk *DataFrame y_test* dan *predicted_labels_test* guna memastikan evaluasi yang akurat. Selanjutnya, metrik kinerja model diukur menggunakan *Confusion Matrix*, yang menghasilkan *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN). Penggunaan metode ini, evaluasi model memberikan gambaran tentang kemampuan model dalam mengklasifikasikan sentimen pada dataset uji, termasuk *accuracy*, *precision*, dan *recall*.

4.2 Hasil Uji Coba

Peneliti melakukan beberapa uji coba dengan variasi pasangan nilai *k* pada kelas positif dan negatif untuk mengamati bagaimana perubahan nilai *k* dapat mempengaruhi kinerja model. Tujuan dari uji coba ini ialah untuk menemukan nilai *k* yang memberikan hasil yang optimal dalam mengatasi ketidakseimbangan data

dan meningkatkan akurasi serta performa klasifikasi secara keseluruhan. Uji coba pada nilai *k* dapat dilihat pada tabel dibawah:

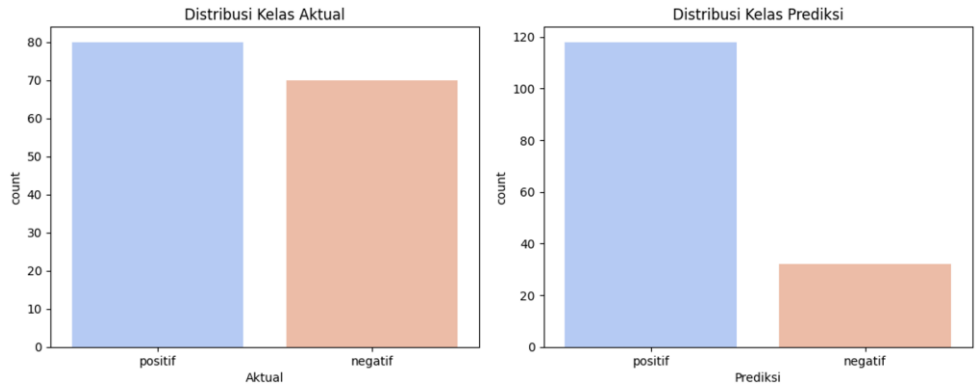
Tabel 4.5 Percobaan Nilai *k*

No	**Percobaan Nilai *k***				
1	***k_positif* = 7, *k_negatif*=9**				
	Data Latih	Data Uji	*Accuracy*	*Precision*	*Recall*
	90%	10%	53%	54%	76%
	80%	20%	53%	54%	69%
	70%	30%	52%	54%	68%
	60%	40%	49%	51%	66%
2	***k_positif* = 7, *k_negatif*=11**				
	Data Latih	Data Uji	*Accuracy*	*Precision*	*Recall*
	90%	10%	52%	53%	76%
	80%	20%	53%	54%	71%
	70%	30%	52%	54%	69%
	60%	40%	49%	51%	66%
3	***k_positif* = 9, *k_negatif*=11**				
	Data Latih	Data Uji	*Accuracy*	*Precision*	*Recall*
	90%	10%	50%	52%	76%
	80%	20%	54%	55%	73%
	70%	30%	53%	54%	71%
	60%	40%	49%	52%	66%
4	***k_positif* = 9, *k_negatif*=15**				
	Data Latih	Data Uji	*Accuracy*	*Precision*	*Recall*
	90%	10%	52%	53%	80%
	80%	20%	54%	55%	77%
	70%	30%	52%	53%	71%
	60%	40%	48%	51%	67%
5	***k_positif* = 11, *k_negatif*=15**				
	Data Latih	Data Uji	*Accuracy*	*Precision*	*Recall*
	90%	10%	51%	52%	78%
	80%	20%	54%	55%	77%
	70%	30%	52%	53%	73%
	60%	40%	49%	51%	68%
6	***k_positif* = 11, *k_negatif*=15**				
	Data Latih	Data Uji	*Accuracy*	*Precision*	*Recall*
	90%	10%	42%	47%	70%
	80%	20%	51%	52%	83%
	70%	30%	56%	57%	78%
	60%	40%	52%	53%	77%
7	***k_positif* = 11, *k_negatif*=17**				
	Data Latih	Data Uji	*Accuracy*	*Precision*	*Recall*
	90%	10%	50%	52%	81%
	80%	20%	54%	55%	78%
	70%	30%	52%	53%	71%
	60%	40%	48%	51%	66%

Hasil uji coba model *Improved K-Nearest Neighbor* memberikan hasil klasifikasi yang dilakukan oleh sistem dengan membandingkan sentimen aktual dengan sentimen prediksi. Evaluasi model mencakup *accuracy*, *precision*, dan *recall*. Tahap ini peneliti menjelaskan uji coba ke-6 pada percobaan nilai *k* menggunakan nilai *k* pada kelas positif dan kelas negatif masing-masing *k*=11 dan *k*=15.

Gambar 4.6 merupakan hasil perbandingan klasifikasi sentimen aktual dan prediksi dari uji coba ke-1 dengan menggunakan 90% data latih dan 10% data uji yang dibagi acak oleh sistem, dimana:

Jumlah data uji : 150
Jumlah data latih : 1343
Jumlah data uji dengan sentimen positif : 80
Jumlah data uji dengan sentimen negatif : 70
Jumlah data latih dengan sentimen positif : 713
Jumlah data latih dengan sentimen negatif : 630



Gambar 4.6 Perbandingan Klasifikasi Sentimen dari Uji Coba ke-1

Berdasarkan hasil klasifikasi sentimen pada Gambar 4.6, maka *confusion matrix* ialah sebagai berikut:

Tabel 4.6 *Confusion Matrix* dari Uji Coba ke-1

	Kelas Prediksi		
Kelas Sebenarnya		+	-
	+	56	24
	-	62	8

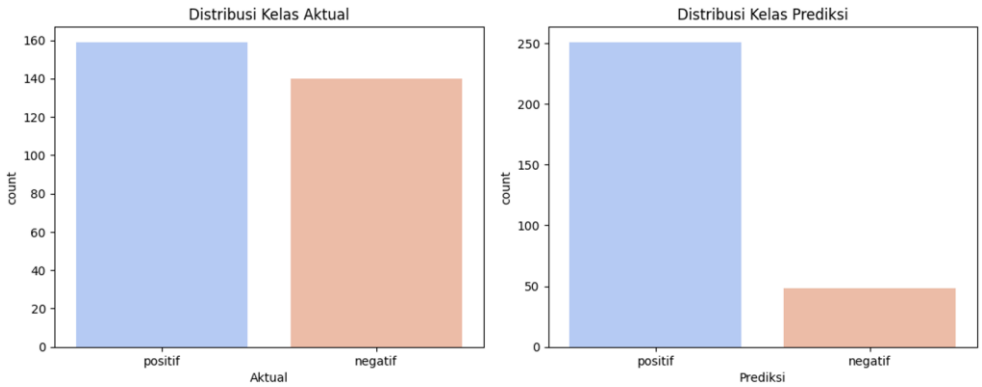
Dimana,
True Negative (TN) : 8
False Negative (FN) : 24
True Positive (TP) : 56
False Positive (FP) : 62

Berdasarkan pada Tabel 4.6, jika dihitung menggunakan persamaan 3.1, 3.2, dan 3.3 maka menghasilkan hasil sebagai berikut:

Accuracy : 42.6%
Precision : 47%
Recall : 70%

Gambar 4.7 merupakan hasil perbandingan klasifikasi sentimen aktual dan prediksi dari uji coba ke-2 dengan menggunakan 80% data latih dan 20% data uji yang dibagi acak oleh sistem, dimana:

Jumlah data uji : 299
Jumlah data latih : 1194
Jumlah data uji dengan sentimen positif : 159
Jumlah data uji dengan sentimen negatif : 140
Jumlah data latih dengan sentimen positif : 634
Jumlah data latih dengan sentimen negatif : 560



Gambar 4.7 Perbandingan Klasifikasi Sentimen dari Uji Coba ke-2

Berdasarkan hasil klasifikasi sentimen pada Gambar 4.7, maka *confusion matrix* ialah sebagai berikut:

Tabel 4.7 *Confusion Matrix* dari Uji Coba ke-2

	Kelas Prediksi		
Kelas Sebenarnya		+	-
	+	132	27
	-	119	21

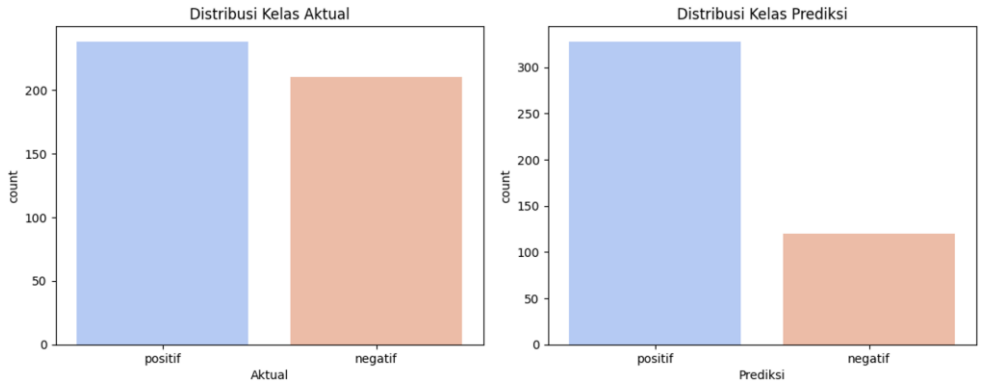
Dimana,
True Negative (TN) : 21
False Negative (FN) : 27
True Positive (TP) : 132
False Positive (FP) : 119

Berdasarkan pada Tabel 4.7, jika dihitung menggunakan persamaan 3.1, 3.2, dan 3.3 maka menghasilkan hasil sebagai berikut:

Accuracy : 51%
Precision : 52.5%
Recall : 83%

Gambar 4.8 merupakan hasil perbandingan klasifikasi sentimen aktual dan prediksi dari uji coba ke-3 dengan menggunakan 70% data latih dan 30% data uji yang dibagi acak oleh sistem, dimana:

Jumlah data uji : 448
Jumlah data latih : 1045
Jumlah data uji dengan sentimen positif : 238
Jumlah data uji dengan sentimen negatif : 210
Jumlah data latih dengan sentimen positif : 555
Jumlah data latih dengan sentimen negatif : 490



Gambar 4.8 Perbandingan Klasifikasi Sentimen dari Uji Coba ke-3

Berdasarkan hasil klasifikasi sentimen pada Gambar 4.8, maka *confusion matrix* ialah sebagai berikut:

Tabel 4.8 *Confusion Matrix* dari Uji Coba ke-3

	Kelas Prediksi		
Kelas Sebenarnya		+	-
	+	186	52
	-	142	68

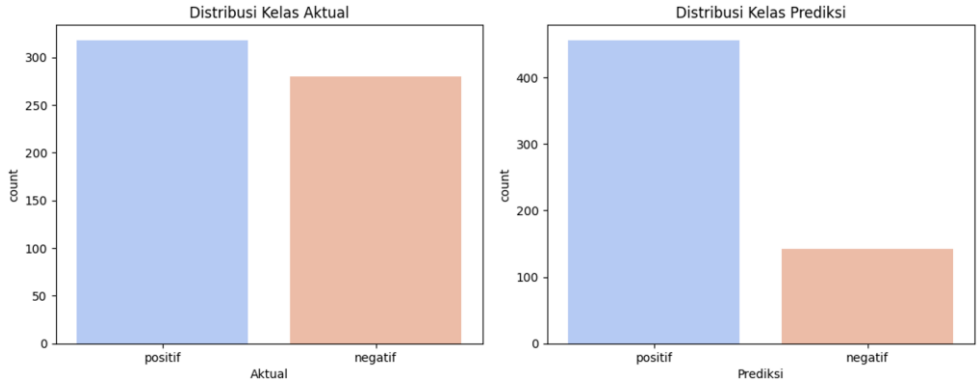
Dimana,
True Negative (TN) : 68
False Negative (FN) : 52
True Positive (TP) : 186
False Positive (FP) : 142

Berdasarkan pada Tabel 4.8, jika dihitung menggunakan persamaan 3.1, 3.2, dan 3.3 maka menghasilkan hasil sebagai berikut:

Accuracy : 56.6%
Precision : 56.7%
Recall : 78%

Gambar 4.9 merupakan hasil perbandingan klasifikasi sentimen aktual dan prediksi dari uji coba ke-4 dengan menggunakan 60% data latih dan 40% data uji yang dibagi acak oleh sistem, dimana:

Jumlah data uji : 598
Jumlah data latih : 895
Jumlah data uji dengan sentimen positif : 318
Jumlah data uji dengan sentimen negatif : 280
Jumlah data latih dengan sentimen positif : 475
Jumlah data latih dengan sentimen negatif : 420



Gambar 4.9 Perbandingan Klasifikasi Sentimen dari Uji Coba ke-4

Berdasarkan hasil klasifikasi sentimen pada Gambar 4.9, maka *confusion matrix* ialah sebagai berikut:

Tabel 4.9 *Confusion Matrix* dari Uji Coba ke-4

	Kelas Prediksi		
Kelas Sebenarnya		+	-
	+	245	73
	-	211	69

Dimana,
True Negative (TN) : 69
False Negative (FN) : 73
True Positive (TP) : 245
False Positive (FP) : 211

Berdasarkan pada Tabel 4.9, jika dihitung menggunakan persamaan 3.1, 3.2, dan 3.3 maka menghasilkan hasil sebagai berikut

Accuracy : 52.5%
Precision : 53.7%
Recall : 77%

Uji coba kembali dilakukan dengan jumlah data latih tetap yakni 60% dari keseluruhan data dengan jumlah data uji yang bervariasi mulai dari 10%, 20%, 30%, dan 40%. Tabel 4.10 berikut merupakan hasil dari percobaan:

Tabel 4.10 Uji Coba dengan Jumlah Data Latih 60%

Uji Coba	Data Latih	Data Uji	*Accuracy*	*Precision*	*Recall*
1	60%	10%	75%	68.9%	100%
2	60%	20%	68.8%	66.6%	98%
3	60%	30%	68.7%	71%	98%
4	60%	40%	72%	71%	98%

4.3 Pembahasan

Hasil uji coba model *Improved K-Nearest Neighbor* (*iKNN*) pada empat percobaan memberikan gambaran terkait klasifikasi sentimen terhadap gaya hidup *childfree*. Penggunaan *confusion matrix* untuk perhitungan *accuracy*, *precision*, dan *recall* guna memahami sejauh mana model mampu memberikan prediksi yang akurat.

Uji coba pertama menggunakan pembagian data latih 90% dan data uji 10%, didapatkan akurasi sebesar 42.6%. Berdasarkan *confusion matrix*, sistem menunjukkan kemampuan yang baik dalam mengidentifikasi sentimen positif, sebagaimana terindikasi oleh nilai *True Positive* (TP) yang mencapai 56. Meskipun *True Negative* (TN) bernilai 8, yang menunjukkan kemampuan sistem dalam mengklasifikasikan sentimen negatif, terdapat kekhawatiran yang muncul dari nilai *False Positive* (FP) yang tinggi yakni 62 dan *False Negatif* (FN) mencapai 24. Kedua nilai FP dan FN yang cukup tinggi menggambarkan bahwa sistem memiliki kecenderungan memberikan lebih banyak *False Positive* (FP) dan tidak mengenali beberapa sentimen positif. *Precision* sebesar 47%, lebih rendah dari pada nilai *recall* 70%, mengindikasikan bahwa sistem cenderung mengidentifikasi teks positif secara lebih baik daripada mengidentifikasi teks yang diklasifikasikan sebagai positif benar-benar positif.

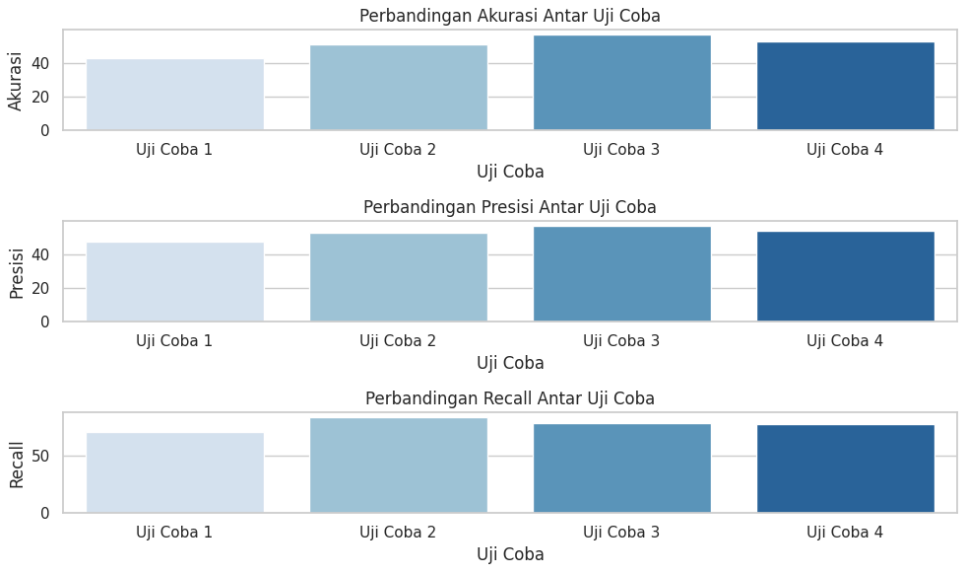
Uji coba kedua menunjukkan perbaikan yang signifikan, dengan akurasi 51%

ketika data latih dan data uji dibagi menjadi 80% dan 20%, terlihat adanya peningkatan dalam kemampuan sistem dengan *True Negative* (TN) sebanyak 21 dan *True Positive* (TP) sebanyak 132, menunjukkan kemampuan sistem dalam mengidentifikasi sentimen positif lebih baik. Meskipun demikian, nilai *False Positive* (FP) masih tinggi, mencapai 119, menunjukkan kecenderungan memberikan positif palsu (FP) yang perlu diperhatikan. Selain itu, terdapat *False Negative* (FN) sebanyak 27, mengindikasikan bahwa beberapa sentimen positif mungkin masih tidak dikenali/tidak terdeteksi. Nilai *precison* 52.5%, lebih rendah dari nilai *recall* 83%, mengindikasikan bahwa sistem memiliki kecenderungan memberikan klasifikasi positif yang lebih luas. Namun, sistem juga mengklasifikasikan beberapa teks negatif sebagai positif.

Uji coba ketiga menggunakan pembagian data latih 70% dan data uji 30% dengan tingkat akurasi 56%, terlihat adanya peningkatan yang baik dalam *True Negative* (TN) sebanyak 68 dan *True Positive* (TP) sebanyak 186. Meskipun demikian, nilai *False Positive* (FP) masih tinggi, mencapai 142, menunjukkan adanya kemungkinan positif palsu yang tinggi. Meskipun nilai *False Negative* (FN) menurun, evaluasi lebih lanjut tetap diperlukan untuk memahami penyebab *False Positive* yang tinggi dan memastikan bahwa peningkatan performa tidak diiringi dengan peningkatan yang signifikan dalam *False Negative* (FN). *Precision* 56.7%, mengindikasikan bahwa sistem memiliki kecenderungan memberikan klasifikasi positif lebih baik, hal ini ditandai dengan nilai *recall* yang tinggi yakni 78%.

Uji coba keempat memberi nilai akurasi sebesar 52.5% saat menggunakan pembagian data latih 60% dan data uji 40%, terjadi peningkatan nilai pada *True*

Negative (TN) dan *True Positive* (TP), menunjukkan kemampuan yang baik dalam mengidentifikasi sentimen positif dan negatif. Nilai *False Positif* (FP) mencapai 211, menunjukkan adanya permasalahan dalam memberikan positif palsu. Selain itu, *False Negative* (FN) juga meningkat menjadi 73, menandakan kemungkinan tidak mengenali beberapa sentimen positif. Nilai *recall* 77% yang lebih tinggi dari nilai *precision* 53.7% menunjukkan bahwa sistem lebih cenderung mengidentifikasi teks positif secara lebih baik.



Gambar 4.10 Perbandingan Hasil Antar Uji Coba

Gambar 4.14 diatas memberikan perbandingan metrik kinerja antar empat uji coba yang dilakukan pada model *iKNN*. Setiap *bar* pada grafik merepresentasikan nilai *accuracy*, *precision*, dan *recall* dari setiap uji coba. Sub-*plot* pertama menunjukkan perbandingan *accuracy*, terlihat bahwa Uji Coba ke-3 memiliki nilai tertinggi, sementara Uji Coba ke-1 memiliki performa yang lebih rendah. Sub-*plot* kedua menunjukkan perbandingan *precision*, terlihat bahwa Uji Coba ke-3 juga memiliki nilai tertinggi meskipun perbedaan nilai di antara uji coba lainnya tidak

terlalu signifikan. Sub-*plot* ketiga menunjukkan perbandingan nilai *recall*, menunjukkan bahwa Uji Coba ke-2 memiliki nilai tertinggi, sementara Uji coba ke-1 memiliki nilai yang lebih rendah.

Uji coba pertama hingga keempat, sistem menunjukkan kemajuan dalam mengidentifikasi sentimen positif dan negatif. Uji coba pertama, sistem memiliki kemampuan yang baik dalam mengenali sentimen positif, namun kekhawatiran muncul akibat tingginya nilai *False Positive* (FP) dan *False Negative* (FN). Uji coba kedua, terjadi peningkatan yang signifikan dalam kemampuan mengidentifikasi sentimen positif dan negatif, meskipun *False Positif* (FP) masih tinggi, dan beberapa sentimen positif tidak terdeteksi (*False Negative*/FN). Uji coba ketiga mengalami peningkatan yang baik dalam *True Negative* (TN) dan *True Positive* (TP), tetapi tingginya *False Positive* (FP) menandakan potensi positif palsu yang perlu diperbaiki. Uji coba keempat menunjukkan kemampuan yang baik dalam mengenali sentimen positif dan negatif. Tetapi nilai *False Positive* yang tinggi dan peningkatan nilai *False Negative* menandakan bahwa perlu perbaikan lebih lanjut untuk meningkatkan keandalan sistem.

Masing-masing nilai *accuracy*, *precision*, dan *recall* pada keempat uji coba memiliki pola yang sama, yakni nilai *accuracy* selalu lebih rendah dibandingkan dengan nilai *precision* dan *recall*. Mengindikasikan bahwa sistem memiliki kecenderungan memberikan klasifikasi positif yang lebih luas, hal ini ditandai dengan nilai *recall* yang tinggi, namun sistem juga mengklasifikasikan beberapa teks negatif sebagai positif. Selanjutnya, nilai *recall* yang lebih tinggi dari nilai *precision* menunjukkan bahwa sistem lebih cenderung mengidentifikasi teks positif

secara lebih baik daripada mengidentifikasi teks yang diklasifikasikan sebagai positif benar-benar positif. Meskipun pola nilai pada keempat uji coba selalu konsisten, rendahnya nilai *accuracy* menggambarkan bahwa sistem masih memerlukan perbaikan untuk mengurangi kesalahan dalam klasifikasi, terutama terkait dengan teks negatif yang mungkin keliru diklasifikasikan sebagai positif.

Hasil pada uji coba terakhir (Tabel 4.10) menunjukkan variasi pada performa model ketika jumlah data uji divariasikan. Uji coba 1 dengan 10% data uji menghasilkan nilai *accuracy* 75%, yang menunjukkan tingkat ketepatan klasifikasi model. Namun, *recall* mencapai 100% yang menunjukkan model memiliki kemampuan yang sangat baik dalam mengidentifikasi seluruh teks positif yang sebenarnya. Peningkatakan jumlah data uji pada uji coba ke-2 hingga ke-4 terlihat adanya perubahan pada nilai *accuracy*, *precision*, dan *recall*. Hal ini mengindikasikan bahwa proporsi data uji terhadap keseluruhan dataset dapat mempengaruhi performa model. Uji coba 2 dan 3 akurasi sedikit menurun, tetapi *precision* dan *recall* masih dalam rentang yang baik.

4.4 Perspektif Klasifikasi Menggunakan Algoritma *Improved K-Nearest Neighbor*

Dalam agama Islam, tidak dibahas secara langsung tentang klasifikasi, namun Allah *SWT* menjelaskan dalam firman-Nya tentang keberagaman atau pengelompokan yang kemudian dapat menjadi dasar untuk konsep klasifikasi. Salah satu ayat Al-Qur'an yang membahas tentang pengelompokan ialah ayat 13 dalam *Surah Al-Hujurat* yang berbunyi:

يٰٓاَيُّهَا النَّاسُ اِنَّا خَلَقْنٰكُمْ مِّنْ ذَكَرٍ وَّاُنْثٰى وَجَعَلْنٰكُمْ شُعُوْبًا وَّقَبَاۤىِٕلَ لِتَعَارَفُوْاۚ اِنَّ اَكْرَمَكُمْ عِنْدَ اللّٰهِ اَتْقٰىكُمْۗ

اِنَّ اللّٰهَ عَلِيْمٌ خَبِيْرٌ ۝

Wahai manusia, sesungguhnya Kami telah menciptakan kamu dari seorang laki-laki dan perempuan. Kemudian, Kami menjadikan kamu berbangsa-bangsa dan bersuku-suku agar kamu saling mengenal. Sesungguhnya yang paling mulia di antara kamu di sisi Allah adalah orang yang paling bertakwa. Sesungguhnya Allah Maha Mengetahui lagi Maha Teliti.(QS. *Al-Hujurat*:13)

Firman Allah dalam ayat tersebut, menurut penafsiran Hasbie Ash-Shiddieqiy, mengajarkan bahwa manusia berasal dari keturunan yang sama, yakni Adam dan Hawa. Hasbie menjelaskan bahwa Allah menciptakan keberagaman berbangsa-bangsa dan bersuku-suku tidak dimaksudkan untuk menciptakan permusuhan di antara manusia. Sebaliknya, keberagaman tersebut bertujuan agar manusia saling mengenal satu sama lain. Allah menciptakan keragaman suku bangsa dan warna kulit agar manusia lebih tertarik untuk saling mengenal. Menurut Hasbie Ash-Shiddieqiy, kedudukan dan kemuliaan seseorang ditentukan oleh tingkat ketakwaannya kepada Allah, dan ayat tersebut mencerminkan prinsip demokrasi dalam Islam yang menolak konsep kasta dan perbedaan bangsa. Hasbie Ash-Shiddieqiy menyebutkan penolakan terhadap rasialisme dan *apartheid*, serta mengingatkan bahwa Allah Maha Mengetahui segala hal, mendorong manusia untuk hidup dengan takwa sebagai bekal terbaik dalam kehidupan (Adju & Imran, 2022).

Secara keseluruhan penafsiran Hasbie Ash-Shiddieqiy tentang keragaman manusia dalam hal keberagaman suku bangsa dipandang sebagai cerminan demokrasi yang sehat dalam Islam. Dari perspektif ini, konsep kasta dan sikap rasis dihilangkan, sementara perbedaan suku atau warna kulit dianggap sebagai faktor yang dapat memunculkan ketertarikan antar manusia untuk saling mengenal.

Hasbie menekankan bahwa manusia seharusnya tidak bermusuhan karena perbedaan suku, golongan, juga warna kulit, karena pada hakikatnya, mereka berasal dari satu keturunan yang sama. Hasbie menegaskan bahwa nilai kemuliaan manusia dihadapan Allah ditentukan oleh tingkat takwanya, bukan oleh faktor fisik atau materi. Oleh karena itu, Hasbie menekankan bahwa takwa harus dijadikan bekal terbaik dalam kehidupan.

Setelah mempelajari pandangan Hasbie Ash-Shiddieqiy mengenai keberagaman manusia dan pentingnya takwa sebagai bekal terbaik, terdapat keterkaitan yang dapat dijelaskan mengenai kepedulian terhadap orang-orang terdekat. Sebagai implementasi nyata dari takwa, Islam menganjurkan umatnya untuk menyebar kebaikan, salah satunya melalui sedekah kepada orang-orang terdekat.

وَعَنْهُ قَالَ: قَالَ رَسُولُ اَللَّهِ صلى الله عليه وسلم (" تَصَدَّقُوا " فَقَالَ رَجُلٌ: يَا رَسُولَ اَللَّهِ, عِنْدِي دِينَارٌ? قَالَ: " تَصَدَّقْ بِهِ عَلَى نَفْسِكَ " قَالَ: عِنْدِي آخَرُ, قَالَ: " تَصَدَّقْ بِهِ عَلَى وَلَدِكَ " قَالَ: عِنْدِي آخَرُ, قَالَ: " تَصَدَّقْ بِهِ عَلَى خَادِمِكَ " قَالَ: عِنْدِي آخَرُ, قَالَ: " أَنْتَ أَبْصَرُ ".) رَوَاهُ أَبُو دَاوُدَ وَالنَّسَائِيُّ, وَصَحَّحَهُ اِبْنُ حِبَّانَ وَالْحَاكِمُ

Rasulullah SAW bersabda: "Bersedekahlah kalian". Lalu seorang laki-laki berkata:Wahai Rasulullah, saya mempunyai satu dinar? Beliau bersabda: "Bersedekahlah pada dirimu sendiri." Lalu laki-laki tersebut bertanya kembali: Lantas ketika saya mempunyai satu dinar lagi? Beliau bersabda: "Sedekahkan kepada anakmu", orang itu bertanya kembali: ketika satu dinar lagi? Beliau bersabda: "Sedekahkan kepada pelayanmu", lalu bertanya kembali: ketika satu dinar lagi? Beliau bersabda: "Kamu lebih mengetahui siapa yang akan kamu beri". (HR. Abu Dawud & Nasa'i, dan dinilai shahih oleh Ibnu Hibban dan Hakim)

Dalam hal klasifikasi dan pengelompokan, hadis diatas memberikan pandangan tambahan tentang keutamaan sedekah. Rasulullah *SAW* menekankan pentingnya bersedekah, namun dengan pendekatan yang mencerminkan prinsip klasifikasi. Hadis tersebut menggambarkan bahwa dalam memberikan sedekah

prioritas diberikan pada kelompok yang lebih dekat dan memiliki keterkaitan yang lebih erat, seiring dengan metode *K-Nearest Neighbor* (*KNN*) yang menitikberatkan pada tetangga terdekat.

Berdasarkan hadis tersebut, Rasulullah *SAW* memberikan arahan untuk memberikan sedekah kepada diri sendiri, anak, dan pelayan, yang merujuk pada pengelompokan prioritas berdasarkan kedekatan hubungan. Hal ini sejalan dengan konsep *KNN*, di mana hal-hal yang memiliki kesamaan atau kedekatan akan dikelompokkan bersama. Dengan demikian, hadis ini tidak hanya mengajarkan tentang nilai sedekah, tetapi juga mencerminkan prinsip klasifikasi dalam memberikan prioritas pada kelompok yang lebih terdekat dan relevan.

Dengan memahami konsep tersebut, manusia diingatkan untuk tidak hanya memberikan sedekah secara umum, tetapi juga melakukan klasifikasi atau pengelompokan dalam memberikan bantuan, sesuai dengan prinsip kebersamaan dan kedekatan yang ditekankan oleh Rasulullah.

BAB V

PENUTUP

5.1 Kesimpulan

Hasil evaluasi menggunakan *Confusion Matrix* dalam penelitian menggunakan model *Improved K-Nearest Neighbor* (*iKNN*) dalam empat uji coba menunjukkan kemajuan dalam pengenalan sentimen positif dan negatif, tetapi perlu perbaikan, terutama dalam mengurangi nilai *False Positive* (FP) dan *False Negative* (FN). Pada uji coba ke-1 menggunakan data latih sebesar 90% dan data uji 10% menghasilkan nilai *accuracy* 42.6%, *precision* 47%, dan *recall* 70%. Uji coba ke-2 menggunakan data latih sebesar 80% dan data uji 20% menghasilkan nilai *accuracy* 51%, *precision* 52.5%, dan *recall* 83%. Uji coba ke-3 menggunakan data latih sebesar 70% dan data uji 30% menghasilkan nilai *accuracy* 56.6%, *precision* 56.7%, dan *recall* 78%. Uji coba ke-4 menggunakan data latih sebesar 60% dan data uji 40% menghasilkan nilai *accuracy* 52.5%, *precision* 53.7%, dan *recall* 77%. Pola yang konsisten dalam nilai *accuracy*, *precision*, dan *recall* menunjukkan kecenderungan sistem memberikan klasifikasi positif yang lebih luas, namun rendahnya nilai *accuracy* menunjukkan perlunya perbaikan untuk mengurangi kesalahan klasifikasi, terutama terkait dengan teks negatif yang mungkin keliru diklasifikasikan sebagai positif.

5.2 Saran

Berdasarkan hasil penelitian yang telah dilakukan, beberapa saran dapat diajukan untuk perbaikan dan pengembangan model *iKNN* pada penelitian selanjutnya:

1. Optimasi Parameter

   Melakukan penyetelan ulang terhadap parameter pada model *iKNN* untuk meningkatkan performa dan mengurangi tingkat kesalahan, terutama pada kasus *False Positive* (FP) dan *False Negative* (FN)

2. Analisis Fitur Tambahan

   Mempertimbangakan penambahan fitur tambahan, seperti analisis emosi, kata kunci seperti '*childfree*', atau analisis konteks kalimat, untuk meningkatkan keakuratan sistem dalam klasifikasi sentimen

3. Peningkatan Dataset

   Memperluas dan menyempurnakan dataset dengan menambahkan data yang lebih representatif dan beragam untuk meningkatkan model

4. Pengembangan Metode Lain

   Mengeksplorasi dan menguji metode klasifikasi lainnya yang mungkin lebih cocok untuk kasus klasifikasi sentiment terhadap gaya hidup childfree

5. Validasi Data yang Lebih Mendalam

   Melakukan validasi data yang lebih mendalam dengan memeriksa keberagaman, keseimbangan kelas, dan potensi bias dalam dataset.

Dengan penerapan saran-saran tersebut, diharapkan model *iKNN* dapat ditingkatkan keandalannya dalam mengklasifikasikan sentimen terhadap gaya hidup *childfree*.

DAFTAR PUSTAKA

Adju, A. M., & Imran, M. (2022). Keragaman Manusia dalam Tafsir An-Nur Karya Hasbie Ash-Shiddieqiy. *Al-Mustafid: Journal of Quran and Hadith Studies*, *1*(2), 49–62. https://doi.org/10.30984/mustafid.v1i2.407

Al-Hafidz Ibnu Hajar Al-Ashqolani. 2020. *Bulughul Maram Min Adillatil Ahkam*. Semarang. Toha Putra. Halaman 120

Al-Quran Online Al-Hujurat Terjemah dan Tafsir Bahasa Indonesia: NU Online Diakses pada 28 November 2023 dari https://quran.nu.or.id/al-hujurat/13#

Al-Quran Online Al-Isra' Terjemah dan Tafsir Bahasa Indonesia: NU Online Diakses pada 19 Oktober 2023 dari https://quran.nu.or.id/al-isra'/31#

Alpaydin, E. (2009). Introduction to Machine Learning Second Edition. In *The MIT Press*.

Arafah, S. N., & Fathoni. (2022). Sentiment Analysis Pada Masyarakat Terhadap LRT Kota Palembang Menggunakan Metode Improved K-Nearest Neighbor. *Jurnal Media Informatika Budidarma*, *6*(3), 1554–1561. https://doi.org/10.30865/mib.v6i3.4434

Baoli, L., Shiwen, Y., & Qin, L. (2003). An Improved k -Nearest Neighbor Algorithm for Text Categorization. *Reading*, 678.

Darmawan, W., Kurniawan, M. F., Setianto, W., & Hapsoro, W. (2023). Analisis Sentimen Penerapan Kurikulum Merdeka Pada Pengguna Twitter Menggunakan Metode K-Nearest Neighbor Dengan Forward Selection. *Smart Comp: Jurnalnya Orang Pintar Komputer*, *12*(1), 245–253. http://ejournal.poltektegal.ac.id/index.php/smartcomp/article/view/4634

Deolika, A., Kusrini, K., & Luthfi, E. T. (2019). Analisis Pembobotan Kata Pada Klasifikasi Text Mining. *Jurnal Teknologi Informasi*, *3*(2), 179. https://doi.org/10.36294/jurti.v3i2.1077

Hadi, A., Khotimah, H., & Sadari. (2022). CHILDFREE DAN CHILDLESS DITINJAU DALAM ILMU FIQIH DAN PERSPEKTIF PENDIDIKAN ISLAM. *Journal of Educational and Language Research*, *8721*(Muksalmina 2020), 647–652.

Han, J., Kamber, M., & Pei, J. (2012). Third Edition : Data Mining Concepts and Techniques. In *University of Illinois at Urbana-Champaign Micheline Kamber Jian Pei Simon Fraser University*. http://library.books24x7.com/toc.aspx?bkid=44712

Hearest, M. (2003). What Is Text Mining. In *SIMS, UC Berkeley* (p. 5).

Herdiawan. (2015). ANALISIS SENTIMEN TERHADAP TELKOM INDIHOME BERDASARKAN OPINI PUBLIK MENGGUNAKAN METODE IMPROVED K-NEAREST NEIGHBOR. *Jurnal Ilmiah Komputer Dan*

Informatika (KOMPUTA).

Iwandini, I., Triayudi, A., & Soepriyono, G. (2023). Analisa Sentimen Pengguna Transportasi Jakarta Terhadap Transjakarta Menggunakan Metode Naives Bayes dan K-Nearest Neighbor. *Journal of Information Systems Research (JOSH)*, *4*(2), 543–550. https://doi.org/10.47065/josh.v4i2.2937

Jiang, S., Pang, G., Wu, M., & Kuang, L. (2012). An improved K -nearest-neighbor algorithm for text categorization. *Expert Systems With Applications*, *39*(1), 1503–1509. https://doi.org/10.1016/j.eswa.2011.08.040

Kurniawan, F., & Wibawa, A. P. (2021). Text Mining Techniques for Identify Islamophobic Conversation Language by Selecting Preprocessing Feature. *Research Square*, 1–9. https://www.researchsquare.com/article/rs-1105114/latest.pdf

Liu, B. (2015). SENTIMENT ANALYSIS: Mining Opinions, Sentiments, and Emotions. In *Cambridge University Press*. https://doi.org/10.1109/WI-IAT.2010.63

Meidina, A. R. (2023). Childfree PPractices in Indonesia (Study on the Response of Islamic Community Organizations in Kebumen Distric). *Indonesian Journal of Multidisciplinary Islamic Studies*, *7*(1), 17–32.

Murugan, A., Chelsey, H., & Thomas, N. (2019). Practical Text Analytics. In *Springer* (Vol. 2).

Muslimah, N., Indriati, & Wihandika, R. . (2019). Klasifikasi Film Berdasarkan Sinopsis dengan Menggunakan Improved K-Nearest Neighbor (K-NN). *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, *3*(1), 196–204.

Mutrofin, S., Izzah, A., Kurniawardhani, A., & Masrur, M. (2014). Optimasi Teknik Klasifikasi Modified K Nearest Neighbor Menggunakan Algoritma Genetika. *Gamma*, *10*(1), 130–134.

Nathania, D. Z., Indriati, & Bachtiar, F. A. (2018). Klasifikasi Spam Pada Twitter Menggunakan Metode Improved K-Nearest Neighbor. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, *2*(10), 3948–3956. http://j-ptiik.ub.ac.id

Ni'mah, A. T., & Arifin, A. Z. (2020). Perbandingan Metode Term Weighting terhadap Hasil Klasifikasi Teks pada Dataset Terjemahan Kitab Hadis. *Rekayasa*, *13*(2), 172–180. https://doi.org/10.21107/rekayasa.v13i2.6412

Novitasari, S. (2020). Pengaruh Media Sosial Twitter @Womanfeeds_Id Terhadap Perilaku Konsumtif Followers. *Jurnal Online Mahasiswa (JOM) Bidang Ilmu Sosial Dan Ilmu Politik*, *7*, 5–24.

Onantya, I. D., Indriati, & Adikara, P. P. (2019). Analisis Sentimen Pada Ulasan Aplikasi BCA Mobile Menggunakan BM25 Dan Improved K-Nearest

Neighbor. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, *3*(3), 2575–2580. http://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/4754

Parsian, M. (2015). *Data Algorithms*. O'Reilly Media.

Pratama, Y., Murdiansyah, D. T., & Lhaksmana, K. M. (2023). Analisis Sentimen Kendaraan Listrik Pada Media Sosial Twitter Menggunakan Algoritma Logistic Regression dan Principal Component Analysis. *Jurnal Media Informatika Budidarma*, *7*, 529–535. https://doi.org/10.30865/mib.v7i1.5575

Putra, D., & Wibowo, A. (2020). Prediksi Keputusan Minat Penjurusan Siswa SMA Yadika 5 Menggunakan Algoritma Naïve Bayes. *Prosiding Seminar Nasional Riset Dan Information Science (SENARIS)*, *2*, 84–92.

Putri, P. A., Ridok, A., & Indriati. (2013). IMPLEMENTASI METODE IMPROVED K-NEAREST NEIGHBOR PADA ANALISIS SENTIMEN TWITTER BERBAHASA INDONESIA. *Repositori Jurnal Mahasiswa PTIIK UB*, *2*, 1–8.

Russel, M. A. (2011). *Mining the Social Web: Analyzing Data from Facebook, Twitter, LinkedIn, and Other Social Media Sites*.

Sandi, D., Utami, E., & Fatkhurohman, A. (2022). Klasifikasi Opini Dengan Menggunakan Algoritma K-Nearest Neighbor Pada Berita Vaksinasi Di Twitter. *Nuansa Informatika*, *16*(1), 156–160. https://doi.org/10.25134/nuansa.v16i1.5343

Suryono, S., & Luthfi, E. T. (2021). Analisis sentimen pada Twitter dengan menggunakan metode Naïve Bayes Classifier. *Jnanaloka*, *1*(2), 81–86. https://doi.org/10.36802/jnanaloka.2020.v1-no2-81-86

Tahmasebi, N., & Risse, T. (2017). On the uses of word sense change for research in the digital humanities. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *10450 LNCS*, 246–257. https://doi.org/10.1007/978-3-319-67008-9_20

Weiss, S. M., Indurkhya, N., & Zhang, T. (2015). Fundamentals of Predictive Text Mining. In *Springer*. http://www.ncbi.nlm.nih.gov/pubmed/20549881

Witten, I. H. (2004). *Text mining*. https://doi.org/10.1201/9780203507223