ANALISIS SENTIMEN PADA TWEET DENGAN TAGAR #YANGCURANGGAKTENANG MENNGUNAKAN METODE DECISION TREE C4.5

SKRIPSI

Diajukan Untuk Memenuhi Salah Syarat Memperoleh Gelas Sarjana
Strata Satu Teknik Pada Fakultas Teknik
Universitas Islam Riau

FARAHDIVA ASSYFA ANDRIN
143510045

PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNIK
UNIVERSITAS ISLAM RIAU
PEKANBARU
2021

Perpustakaan Universitas Islam Riau
Dokumen ini adalah Arsip Milik :

KATA PENGANTAR

بسم الله الرحمن الرحيم

Assalamu'alaikum Warahmatullahi Wabarakatuh.

Dengan mengucapkan Allhamdulilah Robbil 'Alamin, berkat rahmat dan hidayah Allah SWT serta nikmat yang tak terhingga, penulis dapat menyelesaikan laporan skripsi ini dengan judul "ANALISIS SENTIMENT PADA TWEET DENGAN TAGAR #YANGCURANGGAKTENANG menggunakan metode Decision Tree C4.5" sebagai salah satu syarat untuk memperoleh gelar sarjana teknik pada fakultas Teknik Program Studi Teknik Informatika Universitas Islam Riau.

Dalam penyusunan laporan skripsi ini, Penulis sungguh sangat menyadari, bahwa penulisan ini tidak akan terwujud tanpa adanya dukungan dan bantuan dari berbagai pihak. Maka, dalam kesempatan ini penulis menghaturkan penghargaan dan ucapan terima kasih yang sebesar-besarnya kepada yang terhormat :

1. Bapak Ir. H. Abdul Kudus Zaini, MT selaku Dekan Fakultas Teknik dan selaku penasehat akademis.
2. Bapak Dr.Arbi Haza Nasution.,B.IT.,M.I.T selaku ketua Program Studi Teknik Informatika. Serta selaku dosen pembimbing saya yang telah ikhlas dan sabar memberikan bimbingan, saran dan motivasi yang bermanfaat dalam penyusunan laporan skripsi ini.

Perpustakaan Universitas Islam Riau
Dokumen ini adalah Arsip Milik :

3. Bapak dan Ibu Dosen Teknik UIR yang telah banyak memberikan ilmunya selama penulis berada di bangku pendidikan.
4. Teristimewa untuk kedua orang tua tercinta, adik serta saudara yang senantiasa selalu mendo'akan, serta memberikan dukungan baik moril maupun materil dalam menyelesaikan laporan skripsi ini.
5. Semua pihak yang telah membantu penulis dalam menyelesaikan maupun penyusunan dalam tuga akhir ini.

Semoga kebaikan dan do'a serta bantuan mereka dibalas oleh Allah SWT. Penyusunan laporan skripsi ini telah diusahakan semaksimal mungkin, namun penulis menyadari masih ada kekurangan, untuk itu penulis mengharapkan kritik dan saran yang membangun dari pembaca agar dapat disempurnakan pada kemudian hari.

Akhir kata penulis berharap penyusunan laporan skripsi ini dapat bermanfaat bagi dunia pendidikan dan dapat dikembangkan lebih lanjut, Amin.

Wassalamu'alaikum. Warahmatullahi Wabarakatuh.

Pekanbaru, 29 Maret 2021

Farahdiva Assyfa Andrin

Perpustakaan Universitas Islam Riau
Dokumen ini adalah Arsip Milik :

DAFTAR ISI

HALAMAN PENGESAHAN
HALAMAN PERNYATAAN
HALAMAN IDENTITAS
HALAMAN PERSEMBAHAN
KATAPENGANTAR ..... **i**
DAFTAR ISI ..... **iii**
DAFTAR TABEL ..... **vi**
DAFTAR GAMBAR ..... **vii**
ABSTRAK ..... **viii**
ABSTRACT ..... **ix**
BAB I PENDAHULUAN
1.1 Latar Belakang. 1
1.2 Identifikasi Masalah 3
1.3 Rumusan Masalah 4
1.4 Batasan Masalah 4
1.5 Tujuan Penelitian 4
1.6 Manfaat Penelitian 5

BAB II LANDASAN TEORI
2.1 Studi Kepustakaan 6
2.2 Dasar Teori. 8
2.2.1 Analisa Sentimen. 8
2.2.2 Twitter 8
2.2.3 Text Mining 9
2.2.4 Text Pre-Proccesing 9
2.2.5 Pembobotan TF-IDF 10
2.2.6 Klasifikasi 12
2.2.7 Decision Tree 12
2.2.8 Confusion Matrix 14

Perpustakaan Universitas Islam Riau
Dokumen ini adalah Arsip Milik :

2.2.9 Rapid Miner. 14
2.2.10 Python. 15
2.2.11 Jupyter Notebook. 16
2.2.12 Flowchart. 16

BAB III METODOLOGI PENELITIAN

3.1 Metodo Penelitian. 17
3.1.1 Metode Peneltian. 17
3.1.2 Spesifikasi Perangkat Keras (Hardware). 18
3.1.3 Spesifikasi Perangkat Lunak (Software). 18
3.2 Perancangan Sistem. 19
3.2.1 Analisis Sistem 19
3.2.2 Gambaran Umum. 19
3.2.3 Proses Dataset 20
3.2.4 Preprocessing. 21
3.2.5 Data Cleaning. 25
3.2.6 Term Frequency-Inverse Document Frequescy (TF-IDF). 25
3.2.7 Perhitungan Algoritma Decision Tree C4.5. 28

BAB IV HASIL DAN PEMBAHASAN

4.1 Model Analisis Sentimen. 36
4.1.1 Pelatihan. 36
4.1.2 Pengujian Dengan Data Uji 37
4.2 Confusion Matrix. 39
4.2.1 Accuracy 40
4.2.2 Tampilan Pohon Keputusan. 42
4.3 Antar muka Pada Analisis Sentimen. 43
4.4 Pengujian Kepada User 44
4.4.1 Implementasi User 44
4.4.2 Kesimpulan Implementasi Sistem 46

Perpustakaan Universitas Islam Riau
Dokumen ini adalah Arsip Milik :

BAB V PENUTUP

5.1 Kesimpulan..... 47

5.2 Saran..... 47

DAFTAR PUSTAKA..... 48

LAMPIRAN..... 51

Perpustakaan Universitas Islam Riau
Dokumen ini adalah Arsip Milik :

DAFTAR TABEL

Tabel 3.1 Case Folding. 22
Tabel 3.2 Remove URL. 22
Tabel 3.3 Remove Emoticon. 22
Tabel 3.4 Clean Number. 22
Tabel 3.5 Remove Punctuation. 23
Tabel 3.6 Remove Username. 23
Tabel 3.7 Remove Hastag 23
Tabel 3.8 Remove RT. 23
Tabel 3.9 Tokenezing. 24
Tabel 3.10 Filtering 24
Tabel 3.11 Contoh Dokumen. 25
Tabel 3.12 TF-IDF 27
Tabel 3.13 Contoh Training Data. 29
Tabel 3.14 Nilai Data Training. 29
Tabel 3.15 Perhitungan Gain Untuk Simpul Akar 32
Tabel 3.16 Nilai Data Training. 33
Tabel 3.17 Perhitungan Gain untuk Simpul Akar 33
Tabel 3.18 Nilai Data Training. 34
Tabel 3.19 Perhitungan Gain Simpul Kalah 34
Tabel 4.1 Tabel Confusion Matrix. 39
Tabel 4.2 Hasil Nilai Persentase Tiap Pertanyaan Kuisioner 46

Perpustakaan Universitas Islam Riau
Dokumen ini adalah Arsip Milik :

DAFTAR GAMBAR

Gambar 2.1 Logo Python 16
Gambar 3.1 Gambaran UmumAnalisis Sentimen 19
Gambar 3.2 DiagramAlur Dataset..... 20
Gambar 3.3 Hasil Perhitungan Nilai GainSimpulAkar 32
Gambar 3.4 Hasil Perhitungan Nilai GainSimpulKalah 34
Gambar 3.5 Hasil Perhitungan Nilai GainSimpulKalah 35
Gambar 4.1 TampilanData Latih..... 37
Gambar 4.2 Tampilan Form InputData Uji 38
Gambar 4.3 TampilanKetepatan Prediksi 38
Gambar 4.4 Perhitungan AccuracyPadaSistem..... 41
Gambar 4.5 SourcodePohon Keputusan..... 41
Gambar 4.6 FormInputUser 43
Gambar 4.7 FormOuputUser..... 43
Gambar 4.8 HasilKuisioner User 44

Perpustakaan Universitas Islam Riau
Dokumen ini adalah Arsip Milik :

ANALISIS SENTIMEN PADA TWEET DENGAN TAGAR #YANGCURANGGAKTENANG MENNGUNAKAN METODE DECISION TREE C4.5

ABSTRAK

Server Twitter menerima data tweet setiap hari dengan jumlah yang sangat besar. Dengan demikian, Perlu dilakukan pemanfaatan data dengan dikembangkan sistem untuk melakukan data mining dari banyaknya data tersebut yang akan di gunakan untuk kepentingan tertentu, Salah satunya yaitu untuk menganalisa anemo masyarakat terhadap isu kecurangan paska pilpers 2019. Teknik klasifikasi di terapkan untuk mengklasifikasikan data tweet, Sebelum di lakukan klasifikasi, data tweet memalui preprocessing dan pembobotan term frequency dan tf-idf. Dari pembobotan ini menghasilkan data training dan testing yang akan dilakukan klasifikasi menggunakan pohon keputusan dengan Algoritma C4.5 untuk menentukan Sentiment Analysis. Hasil pengujian di peroleh dengan akurasi yang cukup tinggi yaitu 80%. Analisa sentiment anemo masyarakat terhadap isu kecurangan paska pilpres 2019 tersebut fokusnya ada pada tingkat polaritas respon atau pendapat kedalam kategori positif dan negarif.

Kata Kunci : Twitter, Analisis Sentimen, Decision Tree C4.5

Perpustakaan Universitas Islam Riau
Dokumen ini adalah Arsip Milik :

SENTIMENT ANALYSIS ON TWEET WITH #YANGCURANGGAKTENANG FARNING USING METHOD DECISION TREE C4.5

ABSTRACT

Twitter servers receive an enormous amount of tweet data every day. Thus, it is necessary to use data by developing a system to conduct data mining from the amount of data that will be used for certain interests, one of which is to analyze public anemos against the issue of post-election fraud 2019. Classification techniques are applied to classify tweet data. Classified, tweet data through preprocessing and weighting terms frequency and tf-idf. This weighting results in training and testing data that will be classified using a decision tree with C4.5 Algorithm to determine Sentiment Analysis. The test results are obtained with a fairly high accuracy, namely 80%. Analysis of the public anemo sentiment on the issue of post-2019 presidential election fraud is the focus.

Keywords : Twitter, Analisis Sentimen, Decision Tree C4.5

Perpustakaan Universitas Islam Riau
Dokumen ini adalah Arsip Milik :

BAB I

PENDAHULUAN

1.1 Latar Belakang

Presiden dan Wakil Presiden yang menurut pasal 7 Undang-Undang Dasar Negara Republik Indonesia Tahun 1945, "Presiden dan Wakil Presiden memegang jabatan selama lima tahun, dan sesudahnya dapat dipilih kembali dalam jabatan yang sama, hanya untuk satu kali masa jabatan". Dengan demikian, Joko Widodo yang menjadi pemegang posisi Presiden Republik Indonesia untuk periode 2014-2019 berhak dan dapat mengajukan pencalonan kembali untuk pemilihan umum 2019 dengan masa jabatan 2019-2024. Di dampingi oleh Profesor Doktor KH Ma'ruf Amin sebagai calon wakil presiden bersaing dengan pasangan Prabawo Subiyanto dan Sandiago uno. Menjelang Pilpres 2019, fenomena haters dan hoaks bermunculan. Korbannya pun menyasar calon presiden dan calon wakil presiden yang akan bertarung di pesta demokrasi lima tahunan itu, baik Joko Widodo-Ma'ruf Amin maupun kubu Prabowo Subianto-Sandiaga Uno.

Pemilihan presiden dan wakil presiden 2019 telah selesai diselenggarakan pada hari Rabu 17 April 2019 yang diikuti oleh seluruh masyarakat indonesia, dari hasil perhitungan suara, pasangan Joko Widodo-Ma'ruf Amin menjadi presiden dan wakil presiden 2019, namun masih banyak terjadi pro dan kontra atas kemenangan yang diterima oleh pasangan tersebut, banyak pendapat dan anggapan baru bermunculan di masyarakat, sebagian kalangan juga menganggap

Perpustakaan Universitas Islam Riau
Dokumen ini adalah Arsip Milik :

pemilihan presiden dan wakil presiden ini tidak adil terutama dari kalangan pemenang pasangan prabowo-sandi, tidak hanya sekedar opini tapi juga didukung oleh beberapa bukti dan fakta yang terjadi.

Tim Hukum Badan Pemenangan Nasional (BPN) pasangan calon presiden dan wakil presiden nomor urut 02 Prabowo Subianto-Sandiaga Uno secara resmi telah mengajukan gugatan sengketa Perselisihan Hasil Pemilihan Umum (PHPU) 2019 ke Mahkamah Konstitusi (MK), Jumat (24/5/2019) malam. Tim Hukum BPN Prabowo-Sandi mencoba membuktikan bahwa penyelenggaraan Pilpres 2019 dilakukan penuh kecurangan yang terstruktur, sistematis, dan masis. Hal itu diukur dari penyalahgunaan APBN, ketidak netralan aparat, penyalahgunaan birokrasi, pembatasan media, dan diskriminasi perlakuan dan penyalahgunaan penegakan hukum, dari pemberitaan tersebut perang antar pendukung terus memanas. Pemberitaan mengenai kecurangan menjadi topik hangat di masyarakat dan kalangan pengguna sosial media, yang dirangkum dalam bentuk tagar, adapun tagar yang menempati posisi teratas adalah#yangcuranggaktenang sosial media yang menjadi wadah populer dalam menampumg anemo masyarakat pasca pilpres 2019 adalah sosial media twitter melalui tagar (#) yang beragam baik bernilai positif, netral maupun bernilai negatif. Dari beragam hastag tersebut akan dapat menjadi survei dalam bentuk analisa anemo masyarakat terhadap isu kecurangan paska pilpres 2019, maka dari itu untuk melihat antusias di kalangan masyarakat paska pilpres 2019 di nilai perlunya sebuah survei dalam bentuk sentimen analisis yang mampu menganalisa anemo masyarakat terhadap isu kecurangan paska pilpres 2019 berdasarkan tagar media sosial twitter.

Perpustakaan Universitas Islam Riau
Dokumen ini adalah Arsip Milik :

Analisis sentimenatau opinion mining merupakan proses memahami, mengekstrak dan mengolah data tekstual secara otomatis untuk mendapatkan informasi sentimen yang terkandung dalam suatu kalimat opini. Besarnya pengaruh dan manfaat dari analisis sentimen menyebabkan penelitian dan aplikasi berbasis analisis sentimen berkembang pesat. Bahkan di Amerika terdapat sekitar 20-30 perusahaan yang memfokuskan pada layanan analisis sentimen (Huang, 2009). dan penulis tertarik untuk mengangkat topik tersebut dalam tugas akhir dengan judul "Analisis Sentiment Pada Tweet Dengan Tagar #YangCurangGakTenang Menggunakan Metode Decision Tree C4.5".

1.2 Identifikasi Masalah

Adapun identifikasi masalah yang dapat diambil dari latar belakang tersebut yaitu:

1. Banyaknya informasi hoaks yang ada di kalangan masyarakat paska pilpres 2019.
2. Banyaknya pemberitaan dalam konteks yang bersifat positif maupun negatif di kalangan masyarakat paska pilpres 2019.
3. Sulitnya melakukan survey terhadap pandangan masyarakat terhadap presiden dan wakil presiden terpilih.

Perpustakaan Universitas Islam Riau
Dokumen ini adalah Arsip Milik :

1.3 Rumusan Masalah

Berdasarkan latar belakang masalah yang telah dikemukakan di atas, maka fokus permasalahan yang akan dibahas yaitu bagaimana anemo masyarakat terhadap isu kecurangan paska pilpres 2019 berdasarkan tagar media sosial twitter kemudian mengklasifikasikan sentimen data tweet tersebut.

1.4 Batasan Masalah

Adapun batasan masalah dalam penelitian ini yaitu sebagai berikut :

1. Sosial media yang menjadi sasaran penelitian adalah sosial media twitter.
2. Penelitian menggunakan data tweet yang diperoleh dari trending topik yang ada di kalangan masyarakat pengguna sosial media twitter.
3. Tagar yang digunakan adalah #yangcuranggaktenang.
4. Jumlah data tweet yang digunakan adalah 300 data yang diambil dari bulan Januari 2019 sampai dengan Februari 2019.
5. Metode yang digunakan untuk pengklasifikasikan dalam penelitian ini adalah metode *Decision Tree.*
6. Hanya menampilkan sentimen positif dan negatif..

1.5 Tujuan Penelitian

Adapun tujuan dari penelitian ini adalah untuk menganalisa anemo masyarakat terhadap isu kecurangan paska pilpres 2019 berdasarkan tagar media sosial twitter menggunakan metode *Decision Tree*.

Perpustakaan Universitas Islam Riau
Dokumen ini adalah Arsip Milik :

1.6 Manfaat Penelitian

Adapun manfaat dari penelitian ini, yaitu:

1. Penelitian ini dapat diharapkan dapat memberikan manfaat baik kepada masyarakat tentang gambaran sentimen para pengguna Twitter terhadap kasus anemo masyarakat paska pilpres 2019.
2. Memberikan kemudahan dalam melakukan evaluasi serta perbandingan terhadap pemberitaan, keberpihakan dan isu kecurangan paska pilpres 2019 yang ada di kalangan masyarakat.

Perpustakaan Universitas Islam Riau
Dokumen ini adalah Arsip Milik :

BAB II

LANDASAN TEORI

2.1 Studi Kepustakaan

Studi pustaka ini digunakan sebagai pembanding antara penelitian yang sudah dilakukan dan yang akan dilakukan peneliti. Penelitian yang dilakukan oleh Ghulam Asrofi Buntoro dalam judul "Analisis Sentimen Calon Gubernur DKI Jakarta 2017 Di Twitter" Pada penelitian ini Analisis sentimen dilakukan untuk melihat sebuah opini seseorang yang ditujukan kepada calon Gubernur DKI Jakarta 2017, opini itu bisa dimasukkan kategori opini positif, netral atau negatif. Besarnya sentimen yang ditujukan kepada calon Gubernur DKI Jakarta 2017 bisa dijadikan sebuah parameter kemenangan atau kekalahan seorang calon.

Selanjutnya penelitian yang dilakukan oleh Zamahsyari (2016) yang berjudul Analisis Sentimen Pada Berita Ekonomi Berbahasa Indonesia Menggunakan *Majority Vote Classifier*, dalam penelitian ini ditujukan untuk menganalisis sentimen berita ekonomi yang diperoleh dari berbagai media online di Indonesia. Penelitian ini akan menggunakan tiga algoritma yakni Decision tree, yaitu algoritma sederhana yang memiliki alur kerja menyerupai sebuah pohon keputusan yang dapat diterapkan untuk melakukan klasifikasi berita, *randomforest* yaitu algoritma pada dasarnya adalah kumpulan dari banyak pohon keputusan.

Support Vector Machine adalah algoritma klasifikasi yang menerapkan konsep pembuatan bidang pemisah antar kelas ditambah dengan algoritma

Perpustakaan Universitas Islam Riau
Dokumen ini adalah Arsip Milik :

penggabungan antara ketiganya menggunakan metode majority vote classifier Pada hasil Pengujian penelitian menggunakan 10-fold cross validation yang didapat dari berbagai media berita online menghasilkan nilai precision terbaik sebesar 75% oleh algoritma *majority vote classifier*, nilai recall terbaik sebesar 100% oleh algoritma *support vector machine*, dan nilai accuracy terbaik sebesar 72% yang diraih oleh algoritma support vector machine dan algoritma majority *vote classifier*.

Selanjutnya penelitian yang dilakukan oleh M. Fachrurrozi, M.T., dkk tentang Analisis Sentimen Pengguna Jejaring Sosial Menggunakan Metode Support Vector Machine. Pada penelitian ini membahas pesan yang akan disampaikan melalui jejaring sosial dapat di-post melalui antarmuka situs web jejaring sosial, pesan singkat (SMS), atau melalui berbagai aplikasi untuk perangkat seluler pesan pengguna jejaring sosial dapat digunakan untuk berbagai tujuan, seperti melakukan ulasan produk atau meninjau masalah politik dan sosial saat ini. Hal tersebut dapat dilakukan dengan menganalisis sentimen pengguna yang disampaikan melalui pesan di situs jejaring sosial. Metode support vector machine merupakan metode yang dapat digunakan untuk menganalisis sentimen. Analisis sentimen menggunakan metode support vector machine dilakukan dengan mengelompokkan emosi ke dalam kategori positif atau negatif.

Berdasarkan hasil ketiga penelitian terdahulu dapat dikemukakan sistem monitoring yang dibuat memiliki perbedaan studi kasus, maupun metode yang digunakan. Hasil penelitian diatas dapat dijadikan rujukan atau referensi dalam

Perpustakaan Universitas Islam Riau
Dokumen ini adalah Arsip Milik :

penelitian tentang analisis anemo masyarakat terhadap isu kecurangan paska pilpres 2019 berdasarkan sosial media twitter.

2.2 Dasar Teori

2.2.1 Analisa Sentimen

Anlisa sentimen adalah metode untuk menganlisa pembelajaran sentimen pada teks yang digunakan untuk mengolah opini yang diberikan oleh masyarakat melalui sosial media, produk, jasa baik itu dalam bentuk instansi. Dengan kata lain pengelompokan opini dari masyarakat, apakah pendapat dari masyarakat itu baik positif maupun negatif (Sianipar & Setiawan, 2015)

Analisa sentimen adalah metode untuk mengalisa pendapat seseorang, evaluasi seseorang, sikap seseorang, emosi seseorang, sentiment seseorang kedalam bahasa tulisan atau teks.Analisa sentimen yang dikeluarkan dalam bentuk dokumen atau bagian dari dokumen yang ditunjukan kepada objek atau subjek.Pada sentimen analis mempunyai 3 opini yaitu, positif, netral, negatif. Sehingaa dengan adanya sentimen analis bisa mengetahui pendapat seseorang melaui media teks (Sipayung et al., 2016)

2.2.2 Twitter

“Twitter merupakan layanan microblogging yang dapat digunakan untuk melakukan update dalam bentuk teks, dengan panjang maksimal 140 karakter. Update di Twitter disebut juga dengan tweet.. Twitter sendiri menggunakan istilah yang artinya apa yang sedang anda lakukan”(Maya, 2010).

Perpustakaan Universitas Islam Riau
Dokumen ini adalah Arsip Milik :

Twitter adalah suatu aplikasi yang sangat membantu untuk mempromosikan produk maupun data diri, twitter juga digunakan untuk perusahaan untuk mendapatkan layanan secara instan. Twitter juga digunakan sebagai sarana untuk menyebar informasi kepada semua orang melalui media sosial yaitu twitter(Putri, 2018).

2.2.3 Text Mining

Text mining adalah bagaimana teknik mempelajari informasi yang menarik, hal yang baru dimana pola yang belum diketahui sebelumnya atau menemukan kembali media informasi yang tersirat yang berasal dari sumber-sumber data teks yang terpercaya. Text mining mempelajari bagaimana mengekstrak informasi identifikasi dan eksplorasi dari pola yang menarik sehingga ditemukan pada *database record* akan tetapi dalam teks yang tidak terstruktur (Jumeilah, 2017)

2.2.4 Text Pre-Processing

Text preprocessing adalah suatu tahapan dimana untuk merubah data yang dimana data belum terstruktur menjadi data yang terstruktur. Text preprocessing ini bertujuan memudahkan proses untuk pengolahan data untuk memproses data yang selanjutnya (Puspitasari & Santoso, 2018).

1. *Cleaning*

   Cleaning adalah dimana tahapan ini semua karakter non alfabet akan dihapus dengan tujuan mengurangi *noise* sehingga karakter khusus, URL, *email*, angka, dan simbol akan dihapus.(Muttaqin& Bachtiar, 2015).

2. *Case Folding*

Perpustakaan Universitas Islam Riau
Dokumen ini adalah Arsip Milik :

Case folding adalah dimana tahapan *text preprocessing* terdapat beberapa hal yang dirubah,*case folding* hanya memproses huruf 'a' sampai dengan 'z' yang diterima (Muttaqin & Bachtiar, 2015).

3. *Tokenizing*

   Tokenizing adalah tahap pemisahan *string input* berdasarkan pada tiap kata yang menyusunnya (Muttaqin & Bachtiar, 2015)

4. Normalisasi

   Merupakan proses pengembalian kata-kata yang tidak baku ke dalam bahasa baku dalam kamus bahasa Indonesia (Muttaqin & Bachtiar, 2015)

5. *Stemming*

   Stemming adalah tahap mencari *root* kata dari tiap kata hasil *filtering*.Pada tahap ini dilakukan proses pengembalian berbagai bentukan kata kedalam suatu representasi yang sama (Muttaqin & Bachtiar, 2015)

2.2.5 Pembobotan TF-IDF

Pembobotan kata ini dilakukan untuk mendapatkan nilai dari kata (term). Data setelah preprocessing harus berbentuk numerik. Dalam mengubah data tersebut menjadi numerik yaitu menggunakan pembobotan TF-IDF. Proses *Term Frequency Invers Document Frequency* (TF-IDF) dimana tahapan yang digunakan untuk menentukan seberapa jauh keterhubungan kata (term) terhadap dokumen dengan memberikan bobot setiap kata. TF-IDF ini mengabungkan dua konsep yaitu frekuensi kemunculan sebuah kata didalam sebuah dokumen dan invers frekuensi

Perpustakaan Universitas Islam Riau
Dokumen ini adalah Arsip Milik :

dokumen yang mengandung kata tersebut. Dalam perhitungan bobot menggunakan TF-IDF.(Herwijayanti et al., 2018)

Berikut rumus untuk menghitung pembobotan TF-IDF adalah sebagai berikut.

$Idf = \log\left(\frac{N}{df}\right)$……..…….....…..**(2.1)**

W = tf * idf ……...**(2.2)**

Keterangan:

d = dokumen

t = kata pada dokumen

W = bobotdokumenke-d terhadap kata ke-t

Tf = banyaknya kemunculan term(kata) dalam dokumen

Ft,d = frekuensi kata pada d

IDF = Inversed Document Frequency.

dft = banyak kata yang mengandung kata i.

N = jumlah seluruh dokumen

Sistem klasifikasi yaitu nilai perbandingan antara jumlah kemunculan suatu kata dalam dokumen dibagi dengan jumlah keseluruhan kata yang ada dalam dokumen tersebut, sehingga jumlah dari semua tf dari kata yang ada di dalam dokumen tersebut sama dengan satu. Inverse dokument *frequency* (idf) atau kombinasi dari TF-IDF juga dapat digunakan.

Perpustakaan Universitas Islam Riau
Dokumen ini adalah Arsip Milik :

2.2.6 Klasifikasi

Klasifikasi adalah proses dimana mencari himpunan model yang dapat mendekskripsikan serta membedakan kelas-kelas data atau konsep-konsep, dengan klasifikasi ini bertujuan dapat menggunakan model untuk memprediksi kelas dari suatu objek yang dimana kelas belum diketahui sebelumnya (Rani, 2015).

Klasifikasi adalah dimana proses pengelompokan yang artinya mengumpulkan suatu benda atau entitas yang sama selanjutnya memisahkan benda atau entitas yang tidak sama tersebut (Syahid et al., 2016)

2.2.7 *Decision Tree*

Pohon keputusan adalah struktur yang digunakan untuk membagi kumpulan data yang besar menjadi sebuah himpunan yaitu *record* lebih kecil sehingga bisa diterapkan serangkai aturan keputusan. Pohon keputusan merupakan salah satu metode yang dapat mengubah fakta yang besar sehingga menjadi pohon keputusan yang memprediksi aturan, salah satu algoritma yang digunakan dalam pembentukan pohon keputusan adalah algoritma C4.5 (Mardi, 2017).

Algortima C4.5 adalah pohon keputusan ID3 (Iterative Dichotomiser 3).ID3 yang dikembangkan oleh J. Ross Quinlan. Dalam prosedur menggunaka algoritma ID3, dimana input berupa sampel training, label training dan atribut. Algoritma C4.5 merupakan pengembangan dari ID3 (Iterative Dichotomiser 3). Beberapa pengembangan yang dilakukan pada C4.5 adalah sebagai antara lain bisa mengatasi missing value, bisa mengatasi

Perpustakaan Universitas Islam Riau
Dokumen ini adalah Arsip Milik :

continue data dan pruning (Azwanti, 2018). Untuk mengetahui atribut harus berdasarkan dari nilai gain tertinggi dari atribut-atribut. Berikut rumus yang tertera dalam persamaan.

$$\textbf{Gain (S,A) = Entrophy(S)-} \sum_{i=1}^{n} \frac{|Si|}{|S|} * Entropy(Si) \text{ ..... } \textbf{(2.3)}$$

Berikut keterangan dari rumus:

S: himpunan kusus

A : atribut

N : jumlah partisi atribut A

|Si| : jumlah kasus pada partisi ke-i

|S| : jumlah kasus dalam S

Untuk mengetahui nilai entropi dapat dilihat berikut ini:

$$\textbf{Entropy(S)} = \sum_{i=1}^{n} -pi * log2pi \text{ ..... } \textbf{(2.4)}$$

Berikut keterangan dari rumus:

S: himpunan kasus

A : fitur

N : jumlah partisi *S*

Pi : proporsi dari *Si* terhadap *S*

Pada dasarnya algoritma C4.5 membangun pohon keputusan sebagai berikut:

1. Memilih atribut sebagai akar.
2. Membuat cabang untuk tiap-tiap nilai.
3. Bagi kasus dalam cabang.

Perpustakaan Universitas Islam Riau
Dokumen ini adalah Arsip Milik :

4. Mengulangi proses untuk setiap cabang sehingga semua kasus pada cabang memilih kelas yang sama.

2.2.8 *Confusion Matrix*

Confusion matrix adalah metode yang digunakan untuk melakukan perhitungan akurasi pada konsep data mining.*Confusion matrix* yang digambarkan dengan tabel yang menyatakan jumlah data uji yang benar diklasifikasikan dengan jumlah data uji yang salah diklasifikasikan (Rahman et al., 2017).

Confusion matrix adalah metode yang digunakan untuk menghitung akurasi, untuk menghitung kekuatan hasil dari pencarian yang akan dievaluasi nilai *accuracy*.

2.2.9 Rapid Miner

Rapid miner adalah aplikasi *opensource* yang digunakan untuk melakukan proses data mining. Rapid miner menyediakan prosedur data mining dan machine learning, dalam machine learning mempunyai bagian yaitu, ETL (extraction, transformation, learning) Rapid mainer mempunyai tools tipe nilai seperti(Imelda A.Muis & Muhammad Affandes, 2015).

1. Nominal
   Nominal adalah dimana tipe nilai yang berdasarkan nilai secara pertagori.
2. *Numeric*
   Numeric adalah mengambil secara umum

Perpustakaan Universitas Islam Riau
Dokumen ini adalah Arsip Milik :

3. Integer

   Integer adalah tipe nilai yang digunakan untuk bilangan, seperti bilangan bulat.

4. *Real*

   Real adalah dimana tipe nilai yang digunakan secara nyata

5. Text

   Text adalah dimana tipe nilai yang digunakan untuk teks bebas tanpa terstruktur.

6. Binomial

   Binomial adalah dimana tipe nilai yang digunakan terdiri dari dua nilai

7. Polynomial

   Polynomial adalah dimana nominal lebih dari dua nilai

8. Date-time

   Date-time adalah digunakannya untuk tanggal dan waktu.

2.2.10 Python

"Pyhton merupakan salah satu bahasa pemograman yang dinamis dan mempunyai sistem manajemen memori yang otomatis seperti bahasa pemograman yang dinamis lainnya, pyhton biasanya digunakan melalui *script.*Bahasa pemograman python ini bisa dipakai dan digunakan untuk segala macam kebutuhan dari pegembang software atayu perangkat lunak"(Awangga, 2019).

"Python adalah bahasa pemrograman yang dikembangkan pada tahun 1991 dan di publish dalam bentuk *open source*, sehingga python dapat berjalan di sistem operasi contohnya windows, linux, Mac/OS. Pyhton dapat mendukung bentuk model pemrograman seperti *struktural*, *functional*, *object oriented pemrograman*, prosedural" (Awangga, 2019).

Perpustakaan Universitas Islam Riau
Dokumen ini adalah Arsip Milik :

Gambar 2.1 Logo *Python* (*sumber*: python.org)

2.2.11 *Jupyter Notebook*

Jupyter notebook adalah dimana dokumen yang dapat menghasilkan oleh *jupyter notebook* dimana berisikan kode komputer seperti *paragraf*, persamaan matematik gambar dan tautan.*Jupyter notebook* bersifat *opensource* sehingga dapat mengunduh diinternet (Budidaya, 2015)

2.2.12 *Flowchart*

"*Flowchart* adalah dimana menggambarkan urutan logika dari suatu prosedur pemecahan masalah, sehingga *flowchart* merupakan langkah-langkah penyelesaian masalah yang dituliskan dalam simbol-simbol tertentu. Selain digunakan sebagai alat komunikasi, diagram alir ini juga perlu digunakan sebagai dokumen "(Sitorus, 2015).

Perpustakaan Universitas Islam Riau
Dokumen ini adalah Arsip Milik :

BAB III

METODOLOGI PENELITIAN

3.1 Metodologi Penelitian

3.1.1 Metode Penelitian

Metode penelitian adalah tahapan-tahapan yang dilalui oleh peneliti untuk memperoleh gambaran yang jelas mengenai penelitian. Tahapan yang dilalui dalam metode penelitian ini adalah sebagai berikut:

1. Pengumpulan Data

   Data yang dikumpulkan yaitu data yang diambil dari twitter, tweet yang memiliki tagar sesuai dengan topik yang diangkat yaitu tentang isu kecurangan paska pilpres 2019. Data tersebut diperoleh dengan cara melakukan pencarian data di twitter melalui rapid miner.

2. Studi Literatur

   Studi literatur dilakukan dengan cara mengumpulkan dan mempelajari segala macam informasi yang berhubungan dengan Analisis Sentimen Pada Twitter Menggunakan Metode Dicision Tree dengan algoritma C4.5.

3. Perancangan

   Pada tahap ini dilaksanakan perancangan terhadap perangkat lunak yang akan dibuat berdasarkan hasil studi literatur yang ada. Yaitu desain struktur data, desain aliran informasi, desain algoritma dan pemrograman.

Perpustakaan Universitas Islam Riau
Dokumen ini adalah Arsip Milik :

4. Pengujian dan Evaluasi

   Pada tahap ini dilakukan uji coba untuk mencari permasalahan yang mungkin terjadi, untuk mengevaluasi jalannya sistem dan melakukan perbaikan.

5. Penyusunan Laporan Penelitian

   Penyusunan laporan dilakukan pada tahap akhir sebagai dokumentasi. Dokumentasi ini dibuat untuk mempermudah orang lain dalam mempelajari dan mengembangkan sistem lebih lanjut.

3.1.2 Spesifikasi Perangkat Keras (*Hardware*)

Personal komputer atau laptop digunakan untuk perancangan dengan spesifikasi *hardware* sebagai berikut :

1.*Processor* : Intel(R) Celeron(R) CPU N2840 @ 2.16GHz 2.16 GHz

2.RAM : 2.00 GB

3.*System Type* : 64-bit *Operating Syatem*

3.1.3 Spesifikasi Perangkat Lunak (*Software*)

Perangkat lunak yang digunakan dalam pembuatan analisis sentimen pada twitter adalah sebagai berikut:

1. Sistem Operasi : Windows 7 Home
2. Bahasa Pemograman : Python
3. Mozilla Firefox dan Jupyter Notebook

Perpustakaan Universitas Islam Riau
Dokumen ini adalah Arsip Milik :

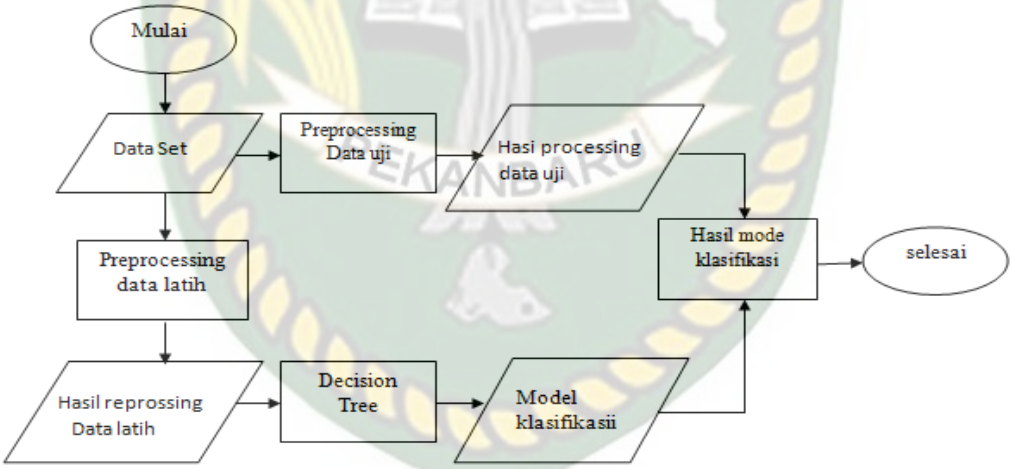
3.2 Perancangan Sistem

3.2.1 Analisis Sistem

Dalam proses pembuatan suatu sistem dilakukan analisis terhadap sistem yang akan dibangun. Analisis yang dilakukan untuk membangun aplikasi analisis sentimen pengguna twitter dengan tagar #yangcuranggaktenang menggunakan metode decession tree.

3.2.2 Gambaran Umum

Analisis Sentimen yang akan dibangun dapat digambarkan secara detail melalui perancangan sistem yang bisa dilihat pada gambar 3.1.



Gambar 3.1 Gambaran Umum *Analisis Sentimen*

Perpustakaan Universitas Islam Riau
Dokumen ini adalah Arsip Milik :

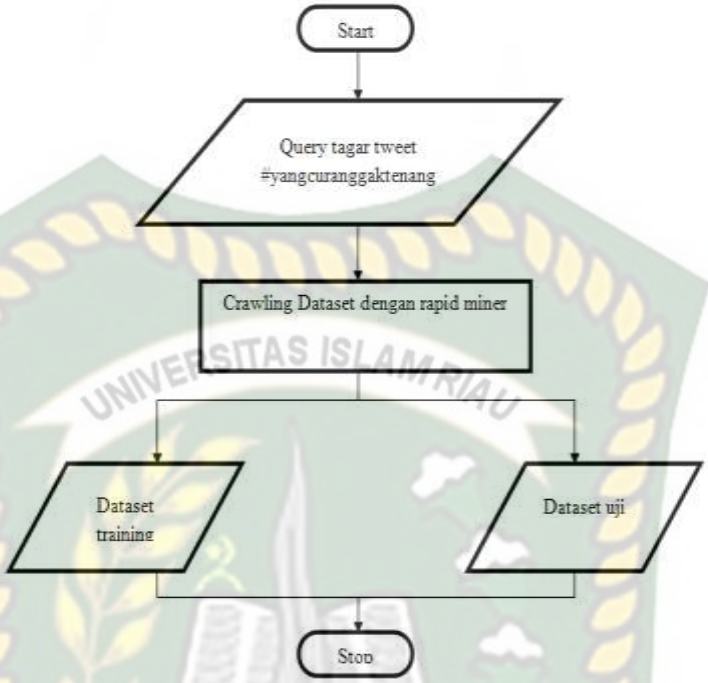
Berikut penjelasan dari gambar 3.1:

1. Mengambil data dari twitter menggunakan rapid miner dengan tagar #yangcuranggaktenang. Data yang diambil berupa data latih dan data uji.
2. Kemudian data set dipecah menjadi preprocessing data latih dan preprocessing data uji.
3. Setelah *preprocessing* didapatkan data bersih yang kemudian akan diterapkan metode *dicision tree* dengan algoritma C4.5.
4. Setelah menerapkan metode *dicision tree*, didalam metode tersebut akan melakukan klasifikasi.
5. Dan begitu juga dengan hasil data preprocessing uji di proses dalam model klasifikasi.
6. Selanjutnya adalah hasil evaluasi klasifikasi.

3.2.3 Proses Dataset

Dataset berupa teks berbahasa Indonesia yang diambil dari twitter. Data yang diambil untuk penelitian ini menggunakan tagar twitter #yangcuranggaktenang. Diagram alur proses pengambilan *dataset* seperti pada gambar 3.2.

Perpustakaan Universitas Islam Riau
Dokumen ini adalah Arsip Milik :



Gambar 3.2 Diagram Alur *Dataset*

Hasil dari dataset ini akan dibagi menjadi dua yaitu data training dan data uji yang di gambarkan pada gambar 3.2 Data training di klasifikasikan menggunakan Decision Tree dengan label negatif dan positif. Sedangkan data uji diklasifikasikan secara manual dengan label negatif dan positif. Data uji nantinya akan di gunakan pada saat evaluasi untuk menentukan ke akuratan data pada sistem.

3.2.4 Preprocessing

Proses *preprocessing* merupakan tahap selanjutnya, yaitu mengurangi atribut yang kurang berpengaruh terhadap proses klasifikasi. Data yang di masukkan pada tahap ini masih berupa data mentah, sehingga hasil dari

Perpustakaan Universitas Islam Riau
Dokumen ini adalah Arsip Milik :

proses ini adalah dokumen berkualitas yang harapan nya mempermudah dalam proses klasifikasi.

Tahap pertama yaitu Cleaning, pada proses ini akan di lakukan beberapa proses yaitu Case *Folding, Remove URL, Remove Emoticon, Clean Number, RemovePunctuation, Remove Username, Remove HastaG dan Remove RT* .Berikut proses preprocessing yang terjadi:

1. *Cleaning*

Yaitu semua karakter non alfabet akan di hapus dengan tujuan mengurangi noise sehingga karakter khusus, URL, emoticon, number, punctuation, username, hastag dan Rt akan di hapus. Adapun bagian-bagian dari cleaning yaitu :

- *Case folding*

Tabel 3.1 Case Folding

Data Awal	@MichelAdamNew: RAKYAT DIMISKINKAN, LALU DISURUH JADI KOMUNIS! Arahnya ke sana! LAWAN! JANGAN MAU JADI KOMUNIS!
Hasil	@michaeladamnew:rakyat dimiskinkan, lalu disuruh jadi komunis ! arahnya kesana ! lawan !jangan mau dijadikan komunis !

- *Remove URL*

Tabel 3.2 Remove URL

DataAwal	Sandiaga: 3 Hari Ini Banyak Berita yang Lucu-lucu #YangCurangGakTenang https://t.co/bTypOl7PSl
Hasil	Sandiaga 3 Hari Ini Banyak Berita yang Lucu-lucu #YangCurangGakTenang

Perpustakaan Universitas Islam Riau
Dokumen ini adalah Arsip Milik :

- *Remove Emoticon*

Tabel 3.3 Remove Emoticon

Data Awal	@amira_swandary Uye....☺ #YangCurangGakTenang Kitasupport hingga Top??
Hasil	Amira swandary uye .... YangCurangGakTenang Kita support hingga top

- *Clean Number*

Tabel 3.4 Clean Number

Data Awal	Selamat Bertugas Pak Presiden dan Wakil Presiden Terpilih untuk Periode 2019 - 2024. Semoga membawa IndonesiaLebih Maju.
Hasil	Selamat bertugas Pak Presiden dan Wakil Presiden Terpilih untuk periode. Semoga membawa indonesia lebih maju.

- *RemovePunctuation*

Tabel 3.5 Remove Punctuation

Data Awal	Tugas negara melayani rakyat, semoga amanah!
Hasil	Tugas negara melayani rakyat semoga amanah

- *Remove Username*

Tabel 3.6 Remove Username

Data Awal	@MichelAdamNew Revolusi mental ternyata.... Cebong tetap aja tutup mata... #YangCurangGakTenang
Hasil	Revolusi mental ternyata.... Cebong tetap aja tutup mata .. #YangCurangGakTenang

Perpustakaan Universitas Islam Riau
Dokumen ini adalah Arsip Milik :

- *Remove Hastag*

Tabel 3.7 Remove Hastag

Data Awal	@MichelAdamNew Revolusi mental ternyata.... Cebong tetap aja tutup mata... #YangCurangGakTenang
Hasil	MichelAdamNew Revolusi mental ternyata.... Cebong tetap aja tutup mata... YangCurangGakTenang

- *Remove RT*

Tabel 3.8 Remove RT

Data Awal	RT @CakKhum: Apakah ini Keberhasilan atau Kegagalan di Bidang Ekonomi?
Hasil	Apakah ini Keberhasilan atau Kegagalan di Bidang Ekonomi?

2. *Tokenizing*

Tokenizing yaitu setiap kata akan dipisahkan berdasarkan spasi yang di tentukan. Contoh Tokenizing sebagai berikut :

Tabel 3.9 Tokenizing

Data Awal	Jangan memulai kebohongan karna bangkai pasti akan tercium
Hasil	|jangan| |memulai| |kebohongan| |karna| |bangkai| |pasti| |akan| |tercium|

3. *Filtering*

Filtering yaitu tahap pembuangan kata-kata yang tidak penting dan daftar *stopwords* bahasa indonesia. Contoh *Filtering* sebagai berikut :

Perpustakaan Universitas Islam Riau
Dokumen ini adalah Arsip Milik :

Tabel 3.10 Filtering

Data Awal	|jangan| |memulai| |kebohongan| |karna| |bangkai| |pasti| |akan| |tercium|
Hasil	|jangan| |memulai| |kebohongan| |karna| |bangkai| |pasti| |tercium|

3.2.5 *Data Cleaning*

Pembersihan data adalah proses menghilangkan noise dan data yang tidak relevan. Pada umumnya data yang diperoleh, baik dari database memiliki isian-isian yang tidak sempurna seperti data yang hilang, data yang tidak valid atau juga hanya sekedar salah ketik. Data-data yang tidak relevan itu juga lebih baik dibuang. Pembersihan data juga akan mempengaruhi performasi dari teknik data mining karena data yang ditangani akan berkurang jumlah dan kompleksitasnya. Adapun perbedaan menggunakan data asli dan data cleaning yaitu terletak pada tinggi dan rendahnya akurasi. Data yang di cleaningkan akurasinya lebih tinggi di bandingkan data asli yang tidak di cleaning.

3.2.6 Term Frequency – Inverse Document Frequency (TF-IDF)

Langkah selanjutnya adalah pembobotan kata menggunakan perhitungan tf-idf. Tf-idf adalah cara pemberian bobot hubungan suatukata (*term*) terhadap kata.Untuk kata tunggal tiap kalimat dianggap sebagai dokumen. Berikut contoh perhitungan tf-idf.

Perpustakaan Universitas Islam Riau
Dokumen ini adalah Arsip Milik :

Tabel 3. 11 Contoh Dokumen

A	Kebohongan itu sementara, Kebenaran itu kekal.
B	Sudah curang pun masih kalah.
C	Dimenangkan itu pecundang dan yang curang gak pernah bahagia.
D	Yang curang gak tenang padahal sudah dimenangkan.
E	Janganlah memulai kecurangan karena pasti akan kalah.

1. Dari contoh dokumen pada tabel 3.11, hitung frekuensi kemunculan kata pada tiap dokumen sebagai langkah awal.
2. Setelah didapatkan frekuensi kata pada tiap dokumen, langkah selanjutnya adalah menghitung tf .
3. Selanjutnya untuk mencari nilai **IDF**dari kata “kebohongan” adalah dengan rumus **log (** $\frac{N}{dft}$**)** , maka log 5/1 = 0.698

Terakhir mencari nilai hasil bobot setiap dokumen berdasarkanhasil dari tf-idf yaitu dengan **tf×idf**, maka1*0.698 = 0.698. Berikut hasil dari perhitungan tf-idf.

Perpustakaan Universitas Islam Riau
Dokumen ini adalah Arsip Milik :

Tabel 3.12 TF-IDF

TERM	TF					Df	Idf=log(n/df)					Bobot(w)=tfxidf				
	A	B	C	D	E		A	B	C	D	E	A	B	C	D	E
Kebohongan	1	0	0	0	0	1	0.698	0	0	0	0	0.698	0	0	0	0
Itu	2	0	1	0	0	2	0.397	0	0.397	0	0	0.794	0	0.397	0	0
Sementara	1	0	0	0	0	1	0.698	0	0	0	0	0.698	0	0	0	0
Kebenaran	1	0	0	0	0	1	0.698	0	0	0	0	0.698	0	0	0	0
Kekal	1	0	0	0	0	1	0.698	0	0	0	0	0.698	0	0	0	0
Sudah	0	1	0	1	0	2	0	0.397	0	0.397	0	0	0.397	0	0.397	0
Curang	0	1	1	1	1	4	0	0.096	0.096	0.096	0.096	0	0.096	0.096	0.096	0.096
Pun	0	1	0	0	0	1	0	0.698	0	0	0	0	0.698	0	0	0
Masih	0	1	0	0	0	1	0	0.698	0	0	0	0	0.698	0	0	0
Kalah	0	1	0	0	1	2	0	0.397	0	0	0.397	0	0.397	0	0	0.397
Dimenangkan	0	0	1	1	0	2	0	0	0.397	0.397	0	0	0	0.397	0.397	0
Pecundang	0	0	1	0	0	1	0	0	0.698	0	0	0	0	0.698	0	0
Dan	0	0	1	0	0	1	0	0	0.698	0	0	0	0	0.698	0	0
Yang	0	0	1	1	0	2	0	0	0.397	0.397	0	0	0	0.397	0.397	0
Gak	0	0	1	1	0	2	0	0	0.397	0.397	0	0	0	0.397	0.397	0
Pernah	0	0	1	0	0	1	0	0	0.698	0	0	0	0	0.698	0	0
Bahagia	0	0	1	0	0	1	0	0	0.698	0	0	0	0	0.698	0	0
Tenang	0	0	0	1	0	1	0	0	0	0.698	0	0	0	0	0.698	0
Padahal	0	0	0	1	0	1	0	0	0	0.698	0	0	0	0	0.698	0
Janganlah	0	0	0	0	1	1	0	0	0	0	0.698	0	0	0	0	0.698
Memulai	0	0	0	0	1	1	0	0	0	0	0.698	0	0	0	0	0.698
Karena	0	0	0	0	1	1	0	0	0	0	0.698	0	0	0	0	0.698
Pasti	0	0	0	0	1	1	0	0	0	0	0.698	0	0	0	0	0.698
Akan	0	0	0	0	1	1	0	0	0	0	0.698	0	0	0	0	0.698

Perpustakaan Universitas Islam Riau
Dokumen ini adalah Arsip Milik :

3.2.7 Perhitungan Algoritma Decision Tree C4.5

Data hasil text processing selanjutnya di klasifikasikan menggunakan pohon keputusan yang dibangun algoritma C4.5. Sebelum pohon keputusan digunakan untuk mengklasifikasikan ke data sebenarnya, model pohon keputusan belajar dari data *training* yang telah mempunyai kelas, yaitu negatif dan positif. Yang akan di gunakan untuk membangun sebuah *tree.*

Berikut tabel perhitungan menggunakan algoritma C4.5 dengan rumus entropi sebagai berikut :

$$Entropi(\mathrm{S}) = \sum_{J=1}^{K} - \mathrm{Pj}\ log^2\ \mathrm{Pj} \quad \text{.....} \quad (3.1)$$

Keterangan :

S = Himpunan Partisi S

K = Banyaknya Partisi S

Pj = Probabilitas yang di dapat dari Sum (Ya dan Tidak)/ Total Kasus.

Rumus diatas juga digunakan untuk mencari entropi pada setiap kelas (positif dan negatif). Setelah mendapatkan nilai entropi, pemilihan atribut di lakukan dengan mencari nilai gain terbesar dengan rumus sebagai berikut :

$$Gain(A) = \mathrm{Entropi(S)} - \sum_{i=1}^{k} \left(a_n \cos \frac{|Si|}{|S|} x \mathrm{Entropi(Si)}\right) \text{.....} \ \mathbf{(3.2)}$$

Keterangan :

A = Atribut

S = Total nilai entropi

Perpustakaan Universitas Islam Riau
Dokumen ini adalah Arsip Milik :

$|Si|$ = Total jumlah (Ya dan Tidak)

|S| = Total jumlah pada kelas (Negatif dan Positif)

Pada dokumen *tweet* yang ada, beberapa kata di jadikan atribut untuk perhitungan entropy dan *gain*. Pada tabel 3.13 adalah contoh proses pembentukan pohon keputusan menggunakan algoritma C.45 menggunakan data training yang mempunyai kelas negatif dan positif.

Tabel 3. 13 Contoh Training Data

No	Data Training	Kelas
1	Kebohongan itu sementara kebenaran itu kekal	Positif
2	Sudah curang pun masih kalah	Negatif
3	Dimenangkan itu pecundang dan yang curang gak pernah bahagia	Negatif
4	Yang curang gak tenang padahal sudah dimenangkan	Positif
5	Janganlah memulai kecurangan karena pasti akan kalah	Positif

Selanjutnya setiap sampel dimasukan kedalam sebuah tabel dengan semua atribut, nilai dari atribut tersebut adalah biner (0 – 1), dimana nilai 0 mengartikan tidak munculnya kata tersebut pada tweet, dan nilai 1 mengartikan bahwa kata tersebut muncul pada tweet. Pada tabel 3.14 adalah tabel yang di hasilkan data training dan proses perhitungan gain terhadap semua atribut.

Tabel 3. 14 Nilai Data Training

D	Kebohongan	Kebenaran	curang	Kalah	Dimenangkan	pecundang	Tenang	Kelas
D1	1	1	0	0	0	0	0	+
D2	0	0	1	1	0	0	0	-
D3	0	0	1	0	1	1	0	-
D4	0	0	1	0	1	0	1	+
D5	0	0	0	1	0	0	0	-

Perpustakaan Universitas Islam Riau
Dokumen ini adalah Arsip Milik :

Perhitungan entropy untuk seluruh training data :

$entropy\ (total$

$$= \left(-\frac{2}{5}\right) xlog_2 \left(\frac{2}{5}\right) + \left(-\frac{2}{5}\right) xlog_2 \left(\frac{2}{5}\right) + \left(-\frac{3}{5}\right) xlog_2 \left(\frac{3}{5}\right)$$

$$+ \left(-\frac{3}{5}\right) xlog_2 \left(\frac{3}{5}\right) + \left(-\frac{1}{5}\right) xlog_2 \left(\frac{1}{5}\right) = 2{,}406$$

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

1. Kebohongan

Perhitungan entropy atribut "kebohongan" dengan nilai 0 dan 1 :

Entropy (0) = $\left(-\frac{3}{4}\right) xlog_2 \left(\frac{3}{4}\right) + \left(-\frac{1}{4}\right) xlog_2 \left(\frac{1}{4}\right) = 0{,}881$

Entropy (1) = 0

Gain (total, kebohongan) = $(2{,}406) - \left(\frac{4}{5}\right) x(0{,}811) + \left(\frac{1}{5}\right) x(0) = 1{,}757$

2. Kebenaran

Perhitungan entropy atribut "kebenaran" dengan nilai 0 dan 1 :

Entropy (0) = $\left(-\frac{3}{4}\right) xlog_2 \left(\frac{3}{4}\right) + \left(-\frac{1}{4}\right) xlog_2 \left(\frac{1}{4}\right) = 0{,}881$

Entropy (1) = 0

Gain (total, kebohongan) = $(2{,}406) - \left(\frac{4}{5}\right) x(0{,}811) + \left(\frac{1}{5}\right) x(0) = 1{,}757$

3. Curang

Perhitungan entropy atribut "curang" dengan nilai 0 dan 1 :

Entropy (0) = $\left(-\frac{1}{2}\right) xlog_2 \left(\frac{1}{2}\right) + \left(-\frac{1}{2}\right) xlog_2 \left(\frac{1}{2}\right) = 1$

Entropy (1) = $\left(-\frac{2}{3}\right) xlog_2 \left(\frac{2}{3}\right) + \left(-\frac{1}{3}\right) xlog_2 \left(\frac{1}{3}\right) = 0{,}918$

Perpustakaan Universitas Islam Riau
Dokumen ini adalah Arsip Milik :

Gain (total, curang) = $(2{,}406) - \left(\frac{2}{5}\right) x(1) + \left(\frac{1}{5}\right) x(0{,}918) = 2{,}557$

4. Kalah

Perhitungan entropy atribut " kalah" dengan nilai 0 dan 1 :

Entropy (0) = $\left(-\frac{1}{3}\right) xlog_2 \left(\frac{1}{3}\right) + \left(-\frac{2}{3}\right) xlog_2 \left(\frac{2}{3}\right) = 0{,}918$

Entropy(1) = 0

Gain (total, kalah) = $(2{,}406) - \left(\frac{3}{5}\right) x(0{,}918) + \left(\frac{2}{5}\right) x(0) = 1{,}855$

5. Dimenangkan

Perhitungan entropy atribut "dimenangkan" dengan nilai 0 dan 1 :

Entropy (0) = $\left(-\frac{2}{3}\right) xlog_2 \left(\frac{2}{3}\right) + \left(-\frac{1}{3}\right) xlog_2 \left(\frac{1}{3}\right) = 0{,}918$

Entropy (1) = $\left(-\frac{1}{2}\right) xlog_2 \left(\frac{1}{2}\right) + \left(-\frac{1}{2}\right) xlog_2 \left(\frac{1}{2}\right) = 1$

Gain (total, dimenangkan) = $(2{,}406) - \left(\frac{3}{5}\right) x(0{,}918) + \left(\frac{2}{5}\right) x(1) = 2{,}553$

6. Pecundang

Perhitungan entropy atribut "pecundang" dengan nilai 0 dan 1 :

Entropy (0) = $\left(-\frac{2}{4}\right) xlog_2 \left(\frac{2}{4}\right) + \left(-\frac{2}{4}\right) xlog_2 \left(\frac{2}{4}\right) = 1$

Entropy (1) = 0

Gain (total, pecundang) = $(2{,}406) - \left(\frac{4}{5}\right) x(1) + \left(\frac{1}{5}\right) x(0) = 1{,}606$

7. Tenang

Perhitungan entropy atribut "tenang" dengan nilai 0 dan 1 :

Entropy (0) = $\left(-\frac{3}{4}\right) xlog_2 \left(\frac{3}{4}\right) + \left(-\frac{1}{4}\right) xlog_2 \left(\frac{1}{4}\right) = 0{,}811$

Entropy (1) = 0

Perpustakaan Universitas Islam Riau
Dokumen ini adalah Arsip Milik :

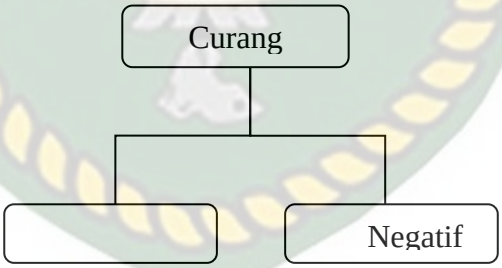
Gain (total, tenang) = $(2{,}406) - \left(\frac{4}{5}\right) x(0{,}811) + \left(\frac{1}{5}\right) x(0) = 1{,}757$

Proses perhitungan dilakukan sampai dengan ditentukannya simpul daun. Selanjutnya pada tabel 3.15 seluruh hasil perhitungan nilai gain untuk membentuk simpul akar pohon keputusan.

Tabel 3.15 Perhitungan Gain untuk Simpul Akar

Atribut	Neg(0)	Post(0)	Entropy(0)	Neg(1)	Post(1)	Entropy(1)	Gain
Kebohongan	3	1	0.811	0	1	0	1.757
Kebenaran	3	1	0.811	0	1	0	1.757
Curang	1	1	1	2	1	0.918	2.557
Kalah	1	2	0.918	2	0	0	1.855
Dimenangkan	2	1	0.918	1	1	1	2.255
Pecundang	2	2	1	1	0	0	1.606
Tenang	3	1	0.811	0	1	0	1.757
Gain Tertinggi : 2.557 Node : Curang							

Dari hasil perhitungan nilai gain untuk simpul akar didapatlah rule tree seperti pada gambar 3.3



Gambar 3.3 Hasil Perhitungan Nilai Gain Simpul Akar

Karena nilai entropy untuk cabang simpul “Curang” dengan nilai 0 masih tidak sama dengan 0, maka akan di lakukan perhitungan kembali untuk menentukan simpul selanjutnya.

Perpustakaan Universitas Islam Riau
Dokumen ini adalah Arsip Milik :

Tabel 3. 16 Nilai Data Training

D	Kebohongan	Kebenaran	Kalah	Dimenangkan	Pecundang	Tenang	Kelas
D1	1	1	0	0	0	0	Positif
D4	0	0	0	1	0	1	Positif
D5	0	0	1	0	0	0	Negatif

Perhitungan entropy total :

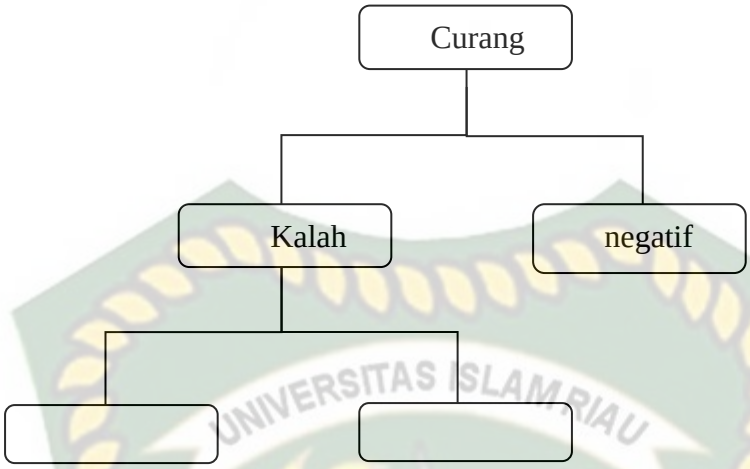
Entropy total = $\left(-\frac{2}{3}\right) x log_2 \left(\frac{2}{3}\right) + \left(-\frac{2}{3}\right) x log_2 \left(\frac{2}{3}\right) + \left(-\frac{1}{3}\right) x log_2 \left(\frac{1}{3}\right) =$ 1,308

Tabel 3.17 Perhitungan Gain untuk Simpul Akar

Atribut	Neg(0)	Post(0)	Entropy(0)	Neg(1)	Post(1)	Entropy(1)	Gain
Kebohongan	1	1	1	0	1	0	0.641
Kebenaran	1	1	1	0	1	0	0.641
Kalah	0	2	0	1	0	0	1.308
Dimenangkan	1	1	1	0	1	0	0.641
Pecundang	1	2	0.918	0	0	0	0.389
Tenang	1	1	1	0	1	0	0.641
Gain Tertinggi : 1.308 Node : Kalah							

Dari hasil perhitungan nilai gain untuk menentukan simpul kalah didapatlah rule tree seperti pada gambar 3.4

Perpustakaan Universitas Islam Riau
Dokumen ini adalah Arsip Milik :



Gambar 3.4 Hasil Perhitungan Nilai Gain Simpul kalah

Karena nilai entropy untuk cabang simpul "kalah" dengan nilai 1 masih tidak sama dengan 0, maka akan di lakukan perhitungan kembali untuk menentukan simpul selanjutnya.

Tabel 3. 18 Nilai Data Training

D	Kebohongan	Kebenaran	Dimenangkan	Pecundang	Tenang	Kelas
D1	1	1	0	0	0	Positif
D2	0	0	1	0	1	Positif

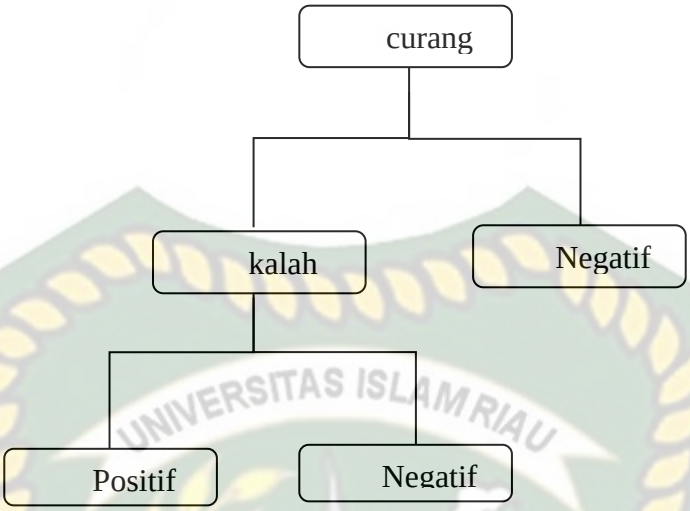
Perhitungan entropy total :

Entropy total = $\left(-\frac{2}{2}\right) x log_2 \left(\frac{2}{2}\right) + \left(-\frac{2}{2}\right) log_2 \left(\frac{2}{2}\right) = 0$

Tabel 3. 19 Perhitungan Gain simpul kalah

Atribut	Neg(0)	Post(1)	Entropy(0)	Neg(1)	Post(1)	Entropy(1)	Gain
Kebohongan	0	1	0	0	1	0	0
Kebenaran	0	1	0	0	1	0	0
Dimenangkan	0	1	0	0	1	0	0
Pecundang	0	2	0	0	0	0	0
Tenang	0	1	0	0	1	0	0

Perpustakaan Universitas Islam Riau
Dokumen ini adalah Arsip Milik :



Gambar 3.5 Hasil Perhitungan Nilai Gain Simpul kalah

Karena nilai entropy untuk cabang simpul "kalah" telah mencapai hasil sama dengan 0 maka kedua cabang atribut "kalah" berupa leaf node dan tidak dilakukan perhitungan gain untuk menentukan simpul selanjutnya.

Perpustakaan Universitas Islam Riau
Dokumen ini adalah Arsip Milik :

BAB IV

HASIL DAN PEMBAHASAN

4.1 Model Analisis Sentimen

Dalam penelitian ini, Penulis menggunakan jupyter notebook sebagai wadah dalam pembuatan analisis sentimen karena jupyter memiliki interface yang sederhana dan mudah untuk dipahami. Jupyter notebook memiliki kelebihan dapat memperlihatkan alur program yang berhasil dijalankan atau error pada alur program tertentu. Berikut adalah tahapan-tahapan analisis sentimen.

4.1.1 Pelatihan

Sub-bab ini menjelaskan proses pelatihan ini bertujuan untuk membentuk model prediksi dari data yang telah diketahui kelasnya. Data yang di gunakan dalam penelitian ini terdiri dari dua jenis, yaitu data latih dan data uji. Data latih yang digunakan di ambil dari kumpulan data *tweet* yang telah di labeli sesuai sentimennya secara manual, terdapat dua jenis label yang di bagi ke dalam sentimen yaitu positif dan negatif. Pada kasus ini penulis menggunakan data latih sebanyak 200 data dan menghasilkan akurasi 0.73. Berikut contoh tampilan data latih pada sistem pada gambar 4.1.

Perpustakaan Universitas Islam Riau
Dokumen ini adalah Arsip Milik :

Out[3]:

	liked	txt
0	1	Gatot Nurmantyo: Saya Tak Takut Dibilang Provo...
1	1	@RezimTipuDaya @HukumDan @SaputraSuteja @na_di...
2	1	RT @Ponirahmujahid1: Yang masih perduli dengan...
3	1	RT @UllbimaNtb: Sandiaga: Wagub DKI Diserahkan...
4	1	@MichelAdamNew Revolusi mental ternyata.... \n...
5	1	Kebohongan itu sementara. Kebenaran itu kekal....
6	1	@joked2019 Sdh mulai terjebak\n#YangCurangGakT...
7	1	RT @geloraco: KPAI Menentang Ide Menghapus Pel...
8	1	Rasulullah shallallahu 'alaihi wa sallam bersa...
9	1	@33x_Dzikir Siap.... Semangat... \n\n#YangCura...
10	1	RT @ArulKha55306941: Kata pak Moeldoko...\nRek...

Gambar 4. 1 Tampilan Data Latih

Yang di maksud dengan liked yang terdapat pada gambar 4.1 adalah kelas dari kalimat 0 adalah negatif dan 1 adalah positif. Sedangkan txt adalah *tweet* yang telah diambil dari twitter. Data yang diinputkan berupa data dari excel dengan format csv. Gambar 4.1 adalah 10 contoh data yang ditampilkan dari data yang digunakan.

4.1.2 Pengujian Dengan Data Uji

Sub-bab ini penulis melakukan pengujian dengan data uji, ini adalah proses puncak atau mengetahui tingkat ke akuratan pada analisis. Pada proses ini *user* akan melakukan penginputan data dan memperoleh hasil berupa kelas dalam bentuk binner yaitu “0” sebagai “negatif” dan “1” sebagai “positif”. Hasil ini didapatkan oleh sistem melalui data latih yang sudah diinputkan. Untuk menginputkan data uji, dilakukan secara manual.

Perpustakaan Universitas Islam Riau
Dokumen ini adalah Arsip Milik :

Out[18]:

	txt
0	@Abeng_Ridwan81Jika pejuang oposisi jadi nyebo...
1	@ChasseurRouge7 Dont cheat#YangCurangGakTenang
2	@kaizer_1st Bro, can we just done it already?!...
3	@AsFairus Seperti inikah #YangCurangGakTenang ...
4	@utipahkiaDone ras and block...#YangCurangGakT...
5	@bmt_sabaryGo internasional brother#YangCurang...
6	@anggarini90 Wow The power of love
7	@f212w Hahahahahahahaaaaa....#LegitimasiDitang...
8	@PatpatPatrickSt which is never gonna happened...
9	bener kecurangan kecurangan sistematis dan ter...

Gambar 4. 2 Tampilan *Form* Input Data Uji

Hasil dari sistem ini di uji dengan hasil yang didapatkan oleh user dengan cara yang manual untuk mendapatkan ketepatan prediksi. Hasil dari uji manual yang sudah dilakukan oleh *user* diinputkan pada sistem secara manual. Disini penulis menggunakan 100 data. Untuk menghasilkan ketepatan prediksi dan menampilkan tabel ketepatan prediksi seperti pada gambar 4.3 di bawah ini.

Out[19]:

	Tweet	Sentiment	Uji Manual	Ketepatan Prediksi
0	@Abeng_Ridwan81Jika pejuang oposisi jadi nyebo...	1	1	Tepat
1	@ChasseurRouge7 Dont cheat#YangCurangGakTenang	0	0	Tepat
2	@kaizer_1st Bro, can we just done it already?!...	0	0	Tepat
3	@AsFairus Seperti inikah #YangCurangGakTenang ...	0	0	Tepat
4	@utipahkiaDone ras and block...#YangCurangGakT...	0	0	Tepat
5	@bmt_sabaryGo internasional brother#YangCurang...	0	0	Tepat
6	@anggarini90 Wow The power of love	1	1	Tepat
7	@f212w Hahahahahahahaaaaa....#LegitimasiDitang...	0	0	Tepat
8	@PatpatPatrickSt which is never gonna happened...	0	0	Tepat
9	bener kecurangan kecurangan sistematis dan ter...	1	1	Tepat

Gambar 4. 3 Tampilan Ketepatan Prediksi

Perpustakaan Universitas Islam Riau
Dokumen ini adalah Arsip Milik :

Dari tabel di atas menjelaskan *sentiment* merupakan hasil uji dari sistem sedangkan uji manual merupakan hasil uji dari *user*. Jika hasil uji dari sistem sama dengan hasil uji dari *user* maka prediksi nya adalah tepat, dan jika hasil uji tidak sama maka prediksi nya tidak tepat. Oleh karna itu dari pengujian yang penulis lakukan, dapat disimpulkan bahwa hasil pengujian ini menunjukkan hasil yang sesuai dengan yang diharapkan.

4.2 Confusion Matrix

Confusion Matrix dilakukan untuk menguji hasil klasifikasi dengan mengukur nilai kebenaran dari sistem. Jika dataset memiliki hanya dua kelas, yaitu yang sering di anggap positif dan yang lain negatif. Dalam kasus ini entri dalam dua baris dan kolom confusion matrix dirujuk sebagai *true and false positives* dan *true and false negatives*, seperti pada tabel 4.1 di bawah ini:

Tabel 4. 1 Tabel *Confusion Matrix*

	Predicted	
	Positive	Negative
Positive	True Positive	False negative
Negative	False Positive	True positive

True Positive (TP) adalah jumlah record positif yang diklasifikasikan sebagai positif, *false positive* (FP) adalah jumlah record negatif yang diklasifikasikan sebagai positif, *false negative* (FN) adalah jumlah record positif

Perpustakaan Universitas Islam Riau
Dokumen ini adalah Arsip Milik :

yang diklasifikasikan sebagai negatif, *true negative* (TN) adalah jumlah record negatif yang diklasifikasikan sebagai negatif. Dan untuk klasifikiasi text, biasanya pengukuran akurasinya menggunakan pengukuran *Accucacy*.

4.2.1 Accuracy

Accuracy yaitu persentase dokumen yang berhasil diklasifikasikan dengan tepat oleh sistem. *Accuracy* yang di hasilkan dari hasil perhitungan yang ditunjukkan pada persamaan di bawah ini.

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

Dari 100 data *testing* yang sudah diproses ketepatan prediksi nya, didapatkan hasil TP : 33, TN : 38, FP : 11, FN : 18 . Perhitungan *accuracy* nya adalah sebagai berikut:

$$Accuracy = \frac{26+54}{(26+6+14+54)} = \frac{80}{(100)} = 0.80$$

Disini penulis menggunakan bilangan bulat karna pada perhitungan ini penulis menentukan *range* untuk hasilnya dengan nilai 1-100 maka hasil nya dikalikan dengan 100, sehingga didapatkan hasil dari *accuracy* antara uji manual dan uji pada sistem adalah 80. Memiliki persentase angka yang cukup tinggi diklasifikasikan dengan tepat oleh sistem.

Pengujian yang dilakukan pada proses ini menggunakan pengujian *white box*, yaitu sebagai berikut:

Perpustakaan Universitas Islam Riau
Dokumen ini adalah Arsip Milik :

Output:Accuracy

1. read tp, tn, fp, fn
2. accuracy = ((tp + tn) * 100)/ float(tp + tn + fn + fp)
3. print “accuracy = “ , accuracy

Yang diperoleh pada *source code* adalah sebagai berikut :

```
In [33]: print('akurasi :', compute_akurasi(tp, tn, fp, fn))

         akurasi : 80.0
```

Gambar 4. 4 Perhitungan *Accuracy* Pada Sistem

Hasil yang di peroleh dari pengujian yang dilakukan ini dapat disimpulkan bahwa perhitungan accuracy sama dengan hasil yang dilakukan secara manual, pengujian ini mendapatkan hasil sesuai dengan yang diharapkan menghasilkan nilai yang cukup tinggi, karna data *testing* yang diinputkan memiliki kosa kata yang ada pada data *training*, sehingga sistem dapat mempelajari dari data *training* dan dapat memproses data *testing* dengan baik.

Perpustakaan Universitas Islam Riau
Dokumen ini adalah Arsip Milik :

4.2.2 Tampilan Pohon Keputusan

```
In [37]: from IPython.display import Image

         # Create DOT data
         dot_data = tree.export_graphviz(clf, label='all', rounded=True, filled=True)

          # Draw graph
         graph = pydotplus.graph_from_dot_data(dot_data)

         # Show graph
         Image(graph.create_png())
```

Gambar 4. 5 Sourcode Pohon Keputusan

Decision Tree Adalah metode pengambilan kesimpulan berdasarkan pohon keputusan, tersusun dari beberapa kondisi yang berbeda-beda dan semakin kebawah akan semakin spesifik hingga di dapatkanya kesimpulan tertentu. Adapun cara menampilkan graph untuk pohon keputusan ini menggunakan *graphviz* yang berfungsi sebagai memvisualisasikan data yang terstruktual seperti gambar jaringan yang berupa *PNG,GIF,PDF*. Ada beberapa fungsi dari pohon keputusan yaitu :

- *Gini* yaitu metrics yang menunjukan persentase sampels yang salah klasifikasi.
- *Sampels* yaitu jumlah data yang berujung di cabang keputusan tersebut.
- *Value* yaitu jumlah label yang harus berkelas positif atau negatif.

Dalam visualisasi ini, total yang ada di pohon keputusan ini adalah 21 level cabang keputusanya. Adapun tampilan dari pohon keputusan bisa dilihat pada lampiran hal 63.

Perpustakaan Universitas Islam Riau
Dokumen ini adalah Arsip Milik :

4.3 Antarmuka Pada Analisis Sentimen

Antarmuka ini berfungsi untuk pengguna umum bisa menggunakan analisis sentimen dan untuk melihat apakah sebuah tweet memiliki kelas positif atau negatif. Antarmuka yang disediakan ini menggunakan penginputan berupa text,yang nanti nya akan di proses dengan data training yang ada dan menghasilkan ouput berupa tabel yang berisi tweet yang diinputkan dan sentimen berupa positif atau negatif dari tweet tersebut. Berikut input antarmuka untuk pengguna umum bisa dilihat pada gambar 4.5 :

Gambar 4.5 Form *Input* User

Berikut output dari antarmuka tersebut dapat dilihat pada gambar 4.6 di bawah ini:

Perpustakaan Universitas Islam Riau
Dokumen ini adalah Arsip Milik :

Gambar 4.6 Form *Output* User

4.4 Pengujian Kepada User

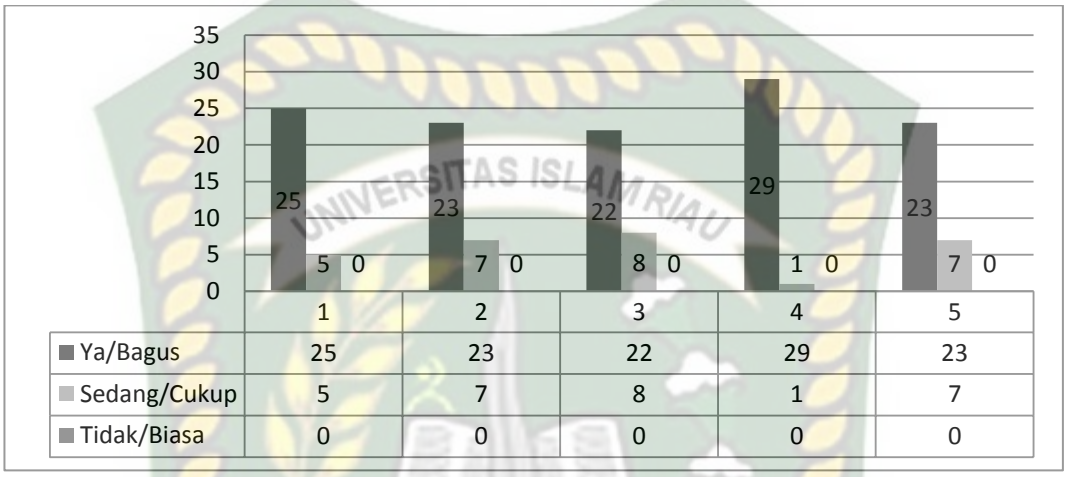
4.4.1 Implementasi User

Implementasi sistem yang di pakai adalah membuat kuisioner dengan menggunakan 5 pertanyaan dan 30 koresponden yang di tujukan kepada user yang ingin memakai sistem ini. Kepada 30 koresponden diajukan pertanyaan yang terkait dengan kinerja atau *performance* dari sistem. Adapun kelima pertanyaan yang dimaksud adalah sebagai berikut :

1. Apakah informasi yang ditampilkan mudah dimengerti oleh user?
2. Apakah bahasa yang digunakan pada analisis sentimen mudah dimengerti?
3. Bagaimana pendapat anda tentang tampilan pada analisis sentimen ini?
4. Apakah analisis sentimen ini cukup mudah digunakan (dioperasikan)?
5. Menurut anda, apakah analisis sentimen ini sudah layak dipublikasikan?

Perpustakaan Universitas Islam Riau
Dokumen ini adalah Arsip Milik :

Dari pertanyaan-pertanyaan diatas, maka hasil jawaban atau tanggapan dari koresponden terhadap kinerja atau *performance* dari sistem berdasarkan pertanyaan yang diajukan adalah sebagai berikut:



	1	2	3	4	5
Ya/Bagus	25	23	22	29	23
Sedang/Cukup	5	7	8	1	7
Tidak/Biasa	0	0	0	0	0

Gambar 4.7 Hasil Kuisioner

Keterangan:

1. Apakah informasi yang ditampilkan mudah dimengerti oleh user memiiki nilai YA : 25 koresponden, SEDANG : 5 koresponden, TIDAK : 0 koresponden.
2. Apakah bahasa yang digunakan pada analisis sentimen mudah dimengerti memiliki nilai YA : 23 koresponden, SEDANG : 7 koresponden, TIDAK : 0 koresponden.
3. Bagaimana pendapat anda mengenai tampilan pada analisis sentimen ini memiliki nilai BAGUS : 22 koresponden, CUKUP : 8 koresponden, BIASA : 0 koresponden.
4. Apakah analisis sentimen ini cukup mdah digunakan (dioperasikan) memiliki nilai YA : 29 koresponden, SEDANG : 1 koresponden, TIDAK : 0 koresponden.

Perpustakaan Universitas Islam Riau
Dokumen ini adalah Arsip Milik :

5. Menurut anda, apakah analisis sentimen ini sudah layak dipublikasikan memiliki nilai YA : 23 koresponden, CUKUP : 7 koresponden, TIDAK : 0 koresponden.

4.4.2 Kesimpulan Implementasi Sistem

Berdasarkan hasil kuisioner tersebut maka dapat disimpulkan bahwa aplikasi manajemen jurnal online ini memiliki persentase sebagai berikut:

Tabel 4. 2 Hasil Nilai Persentase Tiap Pertanyaan Kuisioner

No	Pertanyaan	Jumlah Persentase Koresponden		
		Ya / Bagus	Sedang / Cukup	Tidak / Biasa
1	Apakah informasi yaditampilkan mudah dimengerti oleh user?	83%	17%	0%
2	Apakah bahasa yang digunakan pada analisis sentimen mudah dimengerti?	77%	23%	0%
3	Bagaimana pendapat anda mengenai tampilan pada analisis sentimen ini?	73%	27%	0%
4	Apakah analisis sentimen ini cukup mudah digunakan (dioperasikan)	97%	3%	0%
5	Menurut anda, apakah analisis sentimen ini sudah layak dipublikasikan	77%	23%	0%

Dari hasil persentase tabel 4.2 nilai presentase tiap pertanyaan kuisioner, analisis sentimen pada twitter dengan tagar #yangcuranggaktenang memiliki *performance* baik dengan nilai persentase rata-rata sebesar 81.4%, sehingga aplikasi ini dapat diimplementasikan ke publik.

Perpustakaan Universitas Islam Riau
Dokumen ini adalah Arsip Milik :

BAB V

PENUTUP

5.1 Kesimpulan

Berdasarkan penelitian yang telah dilakukan, mengenai analisis sentimen pengguna Twitter pada tagar #yangcuranggaktenang menggunakan metode Decision Tree C4.5 telah berhasil dilakukan dengan hasil akurasi yang cukup tinggi yaitu 80%. Analisa sentiment untuk anemo masyarakat terhadap isu kecurangan paska pilpres 2019 tersebut fokusnya ada pada tingkat polaritas respon atau pendapat kedalam kategori positif dan negatif.

5.2 Saran

Penelitian yang di lakukan ini sudah tentu tidak terlepas dari kekurangan dan kelemahan. Oleh karena itu, maka ada beberapa saran yang perlu dilakukan dalam penelitian berikutnya, diantaranya :

1. Penelitian selanjutnya diharapkan menggunakan dua atau lebih metode atau algoritma untuk meningkatkan keakuratan sistem.
2. Sistem belum dapat mengidentifikasi penyingkatan kata-kata yang berpengaruh pada keakuratan sehingga diperlukan adanya fitur untuk mengidentifikasi masalah tersebut.

Sistem yang dibangun masih terpisah antara modul untuk mengambil data dari Twitter dan modul untuk perhitungan sentimen analisis. Pada penelitian selanjutnya diharapkan dapat dibuat sistem yang lebih terintegrasi.

Perpustakaan Universitas Islam Riau
Dokumen ini adalah Arsip Milik :

DAFTAR PUSTAKA

Amrizal, V. (2018). Penerapan Metode *Term Frequency Inverse* Document Frequency (Tf-Idf) Dan *Cosine Similarity* Pada Sistem Temu Kembali Informasi Untuk Mengetahui Syarah Hadits Berbasis Web (Studi Kasus: Hadits Shahih Bukhari-Muslim). Jurnal Teknik Informatika, *11*(2), 149–164. https://doi.org/10.15408/jti.v11i2.8623

Awangga, R. M. (2019). pengantar sistem informasi geografis (M.yusri helmi setyawanl (ed.)).

Azwanti, N. (2018). Analisa Algoritma C4.5 Untuk Memprediksi Penjualan Motor Pada Pt. Capella Dinamik Nusantara Cabang Muka Kuning. Informatika Mulawarman*: Jurnal Ilmiah Ilmu Komputer*, *13*(1), 33. https://doi.org/10.30872/jim.v13i1.629

Budidaya, dedi setia. (2015). teknologi pembelajaran fisika berbasis web, *151*, 10–17. https://doi.org/10.1145/3132847.3132886

Buntoro,Ghulam A, 2018. Analisis Sentimen Calon Gubernur DKI Jakarta 2017 Di TwitterTeknik Informatika, Fakultas Teknik, Universitas Muhammadiyah Ponorogo

Herwijayanti, B., Ratnawati, D. E., & Muflikhah, L. (2018). Klasifikasi Berita Online dengan menggunakan Pembobotan TF-IDF dan *Cosine Similarity.* Pengembangan Teknologi Informasi Dan Ilmu Komputer, *2*(1), 306–312.

Imelda A.Muis & Muhammad Affandes, M. . (2015). Penerapan Metode Support Vector Machine (SVM) Menggunakan Kernel Radial Basis Function (RBF) Pada Klasifikasi Tweet. Sains, Teknologi Dan Industri.UIN Sultan Syarif Kasim Riau, *12*(2), 189–197.

Perpustakaan Universitas Islam Riau
Dokumen ini adalah Arsip Milik :

Jumeilah, F. S. (2017). Penerapan *Support Vector Machine* (SVM) untuk Pengkategorian Penelitian. Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi), *1*(1), 19. https://doi.org/10.29207/resti.v1i1.11

Mardi, Y. (2017). Data Mining : Klasifikasi Menggunakan Algoritma C4.5. Jurnal Edik Informatika, *2*(2), 213–219.

M. Fachrurrozi, M.T., dkk, 2016. Analisis Sentimen Pengguna Jejaring Sosial MenggunakanMetode Support Vector Machine. Teknik Informatika Fakultas IlmuKomputer, Universitas Sriwijaya

Maya, M. (2010). *Teknik Remote PC* (A. Magabella (ed.)). PT. Elex Media Komputindo.

Muttaqin, F. A., & Bachtiar, A. M. (2015). Implementasi Teks Mining Pada Aplikasi Pengawasan Penggunaan Internet Anak "Dodo Kids Browser." Jurnal Ilmiah Komputer Dan Informatika (KOMPUTA).

Puspitasari, A. A., & Santoso, E. (2018). Klasifikasi Dokumen Tumbuhan Obat Menggunakan Metode Improved k-Nearest Neighbor. Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer, *2*(2), 486–492.

Putri, K. D. (2018). *e-ISSN: 2623-1212* Jurnal diakom. *1*(2).

Rahman, M. F., Alamsah, D., Darmawidjadja, M. I., & Nurma, I. (2017). Klasifikasi Untuk Diagnosa Diabetes Menggunakan Metode *Bayesian Regularization Neural Network* (RBNN). *Jurnal Informatika*, *11*(1), 36. https://doi.org/10.26555/jifo.v11i1.a5452

Rani, L. N. (2015). Klasifikasi Nasabah Menggunakan Algoritma. Jurnal KomTekInfo Fakultas Ilmu Komputer, *2*(2), 33–38.

Sianipar, R., & Setiawan, E. B. (2015). Pendeteksian Kekuatan Sentimen Pada Teks Tweet BerbahasaIndonesiaMenggunakanSentistrength.*Openlibrary. Telkomuniversity.Ac.Id*,*2*(3),79227928.https://openlibrary.telkomuniversity.a

Perpustakaan Universitas Islam Riau
Dokumen ini adalah Arsip Milik :

c.id/pustaka/files/104179/jurnal_eproc/jurnal_eproc.pdf

Sipayung, E. M., Maharani, H., & Zefanya, I. (2016). Perancangan Sistem Analisis Sentimen Komentar Pelanggan Menggunakan Metode Naive Bayes Classifier. Jurnal Sistem Informasi, *8*(1), 958–965.

Sitorus, L. (2015). algoritma dan pemograman (A. Pramesta (ed.)).

Syahid, D., Jumadi, J., & Nursantika, D. (2016). Sistem Klasifikasi Jenis Tanaman Hias Daun Philodendron Menggunakan Metode K-Nearest Neighboor (KNN) Berdasarkan Nilai Hue, Saturation, Value (HSV). Jurnal Online Informatika, *1*(1), 20. https://doi.org/10.15575/join.v1i1.6

T, Maulana.(2015). Analisis Sentiment Pengguna Twitter Terhadap Partai Politik Peserta Pemilu 2014 Menggunakan Algoritma C.45 Decision Tree Learning, Unikom.

Zamahsyari, 2016.Analisis Sentimen Pada Berita Ekonomi Berbahasa Indonesia MenggunakanMajority Vote Classifier. Universitas Gadjah Mada

Perpustakaan Universitas Islam Riau
Dokumen ini adalah Arsip Milik :