**ANALISIS URGENSI BERDIRINYA GERAI KOPI DI HALTE HARMONI BERDASARKAN PERSEPSI PENUMPANG DAN MASYARAKAT PENGGUNA MEDIA SOSIAL TWITTER**

**LADISA BUSAINA**

**222011349**

**PROGRAM STUDI: KOMPUTASI STATISTIK PROGRAM DIPLOMA IV**

**PEMINATAN: SAINS DATA**

****

**POLITEKNIK STATISTIKA STIS**

**JAKARTA**

**2022**

# **DAFTAR ISI**

[**DAFTAR ISI** i](#_Toc122076575)

[**DAFTAR GAMBAR** iii](#_Toc122076576)

[**DAFTAR TABEL** iv](#_Toc122076577)

[**BAB I** 1](#_Toc122076578)

[**PENDAHULUAN** 1](#_Toc122076579)

[**1.1** **Latar Belakang** 1](#_Toc122076580)

[**1.2** **Perumusan Masalah** 3](#_Toc122076581)

[**1.3** **Manfaat dan Tujuan Penelitian** 3](#_Toc122076582)

[**1.4** **Batasan Penelitian** 3](#_Toc122076583)

[**1.5** **Sistematika Penulisan** 3](#_Toc122076584)

[**BAB II** 5](#_Toc122076585)

[**KAJIAN PUSTAKA** 5](#_Toc122076586)

[**2.1** **Landasan Teori** 5](#_Toc122076587)

[**2.1.1** **Halte** 5](#_Toc122076588)

[**2.1.2** **Gerai Kopi** 6](#_Toc122076589)

[**2.1.3** **Twitter** 7](#_Toc122076590)

[**2.1.4** **Urgensi** 8](#_Toc122076591)

[**2.1.5** **Persepsi** 8](#_Toc122076592)

[**2.1.6** ***Text Mining*** 9](#_Toc122076593)

[**2.1.7** ***Web Scrapping*** 9](#_Toc122076594)

[**2.1.8** ***Text Preprocessing*** 10](#_Toc122076595)

[**2.1.9** ***Maximum Entropy Classification*** 12](#_Toc122076596)

[**2.1.10** **R** 12](#_Toc122076597)

[**2.1.11** ***Package* Rtweet** 13](#_Toc122076598)

[**2.2** **Penelitian Terkait** 14](#_Toc122076599)

[**2.2.1** **Analisis Sentimen *Online Review* Pengguna *E-Commerce* Menggunakan Metode *Support Vector Machine* dan *Maximum Entropy* (Studi Kasus: Buka Lapak pada Google Play)** 14](#_Toc122076600)

[**2.2.2** **Analisis Sentimen Pemilihan Presiden 2019 pada Twitter Menggunakan Metode *Maximum Entropy*** 15](#_Toc122076601)

[**2.2.3** **Metode Ekstraksi Fitur pada Analisis Sentimen Maskapai Penerbangan Menggunakan *Support Vector Machine* dan *Maximum Entropy*** 15](#_Toc122076602)

[**2.2.4** **Perbandingan Penelitian** 16](#_Toc122076603)

[**2.3** **Kerangka Pikir** 19](#_Toc122076604)

[**BAB III** 21](#_Toc122076605)

[**METODOLOGI** 21](#_Toc122076606)

[**3.1** **Ruang Lingkup Penelitian** 21](#_Toc122076607)

[**3.2** **Metode Pengumpulan Data** 21](#_Toc122076608)

[**3.3** **Metode Analisis** 22](#_Toc122076609)

[**DAFTAR PUSTAKA** 27](#_Toc122076613)

# **DAFTAR GAMBAR**

[Gambar 2. 1 Kerangka Pikir 19](#_Toc122076614)

[Gambar 3. 1 Diagram Alir Penelitian 22](#_Toc119863491)

# **DAFTAR TABEL**

[Table 2.1 Perbandingan Metode Klasifikasi Penelitian Sebelumnya 16](#_Toc119881843)

# **BAB I**

# **PENDAHULUAN**

## **Latar Belakang**

Menurut BPS DKI Jakarta, pada tahun 2021, jumlah penduduk Jakarta mencapai 10.609.681 jiwa. Jumlah kendaraan yang ada di DKI Jakarta pun tidak kalah banyak dengan jumlah penduduknya, yaitu sebanyak 21 juta unit secara keseluruhan. Dengan luas daerah yang hanya 664 kilometer persegi, penduduk sebanyak itu menjadikan DKI Jakarta menjadi kota terpadat di Indonesia. Padatnya DKI Jakarta menimbulkan banyak masalah bagi kota tersebut, di antaranya kemacetan. Kemacetan lalu lintas sudah menjadi keseharian penduduk DKI Jakarta. Agar mobilitas dan segala kegiatan yang bergantung dengan itu berlangsung lancar, Pemerintah DKI Jakarta melakukan berbagai cara untuk menyiasati masalah kemacetan lalu lintas tersebut. Salah satu caranya adalah dengan menyediakan transportasi umum bagi penduduknya.

Pada tahun 2004, Pemerintah DKI Jakarta mulai mengoperasikan sebuah transportasi umum dalam jumlah besar, yang dinamai dengan TransJakarta. TransJakarta merupakan transportasi umum berbasis bus dengan sistem *Bus Rapid Transit* (BRT) (Anggoro et al., 2020). TransJakarta berfokus untuk memberikan kenyamanan, keamanan, keselamatan, dan keterjangkauan bagi masyarakat. Biaya yang ditawarkan pun sesuai dengan kondisi ekonomi rata-rata masyarat menengah ke bawah di kota metropolitan Jakarta (Silaningsih, Gemina, & Yuningsih, 2015, dikutip dalam Anggoro et al., 2020). Program ini pun disambut gembira oleh masyarakat, khususnya penduduk Jakarta. Bahkan, pada tahun 2019, bus TransJakarta mencapai 5.843 unit dan jumlah halte sekitar 250 buah dengan 13 koridor.

Di samping semua keunggulan yang dijanjikan oleh TransJakarta, terdapat beberapa kekurangan saat pengoperasiannya. Salah satunya adalah halte yang tidak dapat menampung kapasitas penumpang yang ada. Hal ini tentu saja mengganggu kenyamanan dan keamanan penumpang yang telah melanggar tujuan dari dioperasikannya TransJakarta (Astuti et al., 2017). Umumnya, halte dengan kepadatan yang tinggi adalah halte transit yang dilalui banyak koridor, seperti Halte Harmoni. Keadaan Halte Harmoni menjadi semakin buruk dengan didirikannya sebuah gerai kopi milik Qopi Untuk Semua. Antrean pun tidak hanya berasal dari penumpang yang akan naik ataupun turun dari bus, tetapi juga dari pembeli kopi dari gerai kopi milik Qopi Untuk Semua. Penumpang yang merasakan ketidaknyamanannya karena hal tersebut satu per satu mulai menyuarakan keresahannya. Tidak hanya penumpang yang transit di Halte Harmoni, masyarakat, khususnya di daerah Jakarta, juga ikut menyuarakan pendapat mereka. Mereka menyuarakannya di berbagai media sosial, termasuk Twitter.

Twitter merupakan media sosial favorit masyarakat Indonesia. Hal ini terbukti dengan terus meningkatnya jumlah pengguna Twitter setiap tahunnya (Emeraldien et al., 2019). Karena sudah menjadi bagian dari kehidupan masyarakat, Twitter menampung berbagai cerita dan emosi penggunannya yang tertuang dalam cuitan. Ini menjadikan data yang berlalu lintas pada Twitter sangat besar dan dapat digunakan untuk berbagai keperluan, khususnya di bidang statistik.

Untuk menganalisis bagaimana persepsi penumpang dan masyarakat terhadap adanya gerai kopi Qopi Untuk Semua, digunakan metode *Maximum Entropy Classification*. Metode ini telah digunakan pada penelitian terhadap media sosial Twitter yang dilakukan oleh Ravi Parikh dan Matin Movassete pada tahun 2009 dan menunjukkan performa yang baik (Sunni & Widyantoro, 2012). Metode *Maximum Entropy Classification* digunakan untuk mengklasifikasikan data berdasarkan nilai *entropy*-nya (Sabily et al., 2019).

Dengan demikian, penelitian ini ingin melihat bagaimana persepsi masyarakat, khususnya penumpang TransJakarta yang transit di Halte Harmoni, mengenai urgensi gerai kopi Qopi Untuk Semua menggunakan data pada media sosial Twitter. Persepsi tersebut akan diklasifikasikan menggunakan metode *Maximum Entropy Classification*.

## **Perumusan Masalah**

1. Bagaimana persepsi masyarakat mengenai adanya gerai kopi Qopi Untuk Semua di Halte Harmoni?
2. Bagaimana harapan masyarakat terhadap Halte Harmoni agar kenyamanan dan keamanan tercapai?

## **Manfaat dan Tujuan Penelitian**

Penelitian ini untuk mengetahui bagaimana respons penumpang TransJakarta, khususnya penumpang yang menggunakan fasilitas halte harmoni, dan pengaruhnya terhadap kenyamanan dan keamanan penumpang. Respons ini diharapkan dapat menjadi masukan bagi pemerintah daerah DKI Jakarta untuk mengutamakan kenyamanan dan keamanan masyarakat yang menggunakan fasilitas publik dengan pertimbangan yang matang dalam memberikan izin menjalankan usaha di fasilitas publik terkait.

## **Batasan Penelitian**

Penelitian ini hanya akan menganalisis cuitanmengenai Halte Harmoni oleh akun yang berlokasi di daerah Jakarta pada Senin, 15 Agustus 2022 sebagai cuitanyang akan dilakukan analisis.

## **Sistematika Penulisan**

Bagian Awal memuat Halaman Sampul, Halaman Judul, Halaman Pernyataan, Halaman Pengesahan, Prakata, Abstrak, Daftar Isi, Daftar Tabel, Daftar Gamber, Daftar Lampiran. Bagian Isi terdiri dari 5 (lima) bab, yaitu Bab Pendahuluan, Bab Kajian Pustaka, Bab Metodologi, Bab Hasil dan Pembahasan, serta Bab Kesimpulan dan Saran. Bab Pendahuluan memuat 5 (lima) sub bab, yaitu Latar Belakang, Identifikasi Masalah, Tujuan Penelitian, Keterbatasan Penelitian (*optional*), dan Sistematika Penulisan. Bab Kajian Pustaka memuat 4 (empat) sub bab, yaitu Landasan Teori (bisa mengenai Terapan atau Metode Statistik, tergantung apa yang diteliti), Penelitian Terkait (bisa mengenai Terapan atau Metode Statistik, tergantung apa yang diteliti), Kerangka Pikir, dan Hipotesis Penelitian (jika ada). Bab Metodologi memuat 3 (tiga) sub bab, yaitu Ruang Lingkup Penelitian, Metode Pengumpulan Data, dan Metode Analisis. Bab Hasil dan Pembahasan memuat hasil pengolahan berupa tabel dan atau grafik yang sudah disesuaikan dengan format penulisan, disertai dengan pembahasan yang diarahkan untuk menjawab tujuan penelitian. Bab Kesimpulan dan Saran memuat 2 (dua) sub bab, yaitu Kesimpulan dan Saran. Bagian Akhir memuat Daftar Pustaka, Lampiran, dan Riwayat Hidup.

# **BAB II**

# **KAJIAN PUSTAKA**

## **Landasan Teori**

Berikut beberapa istilah penting yang digunakan dalam penelitian ini beserta definisinya.

### **Halte**

Halte adalah sebuah bangunan tempat pemberhentian untuk menaikkan atau menurunkan penumpang bagi transportasi umum (Sitohang & Ervin, 2019, dikutip dalam Pratama, 2021). Halte menjadi salah satu sarana pendukung transportasi umum yang menjadi bagian dari infrastruktur jalan. Jadi, dapat disimpulkan bahwa halte merupakan infrasturktur jalan berupa sebuah bangunan yang digunakan sebagai tempat pemberhentian transportasi umum untuk menaikkan atau menurunkan penumpang.

Karena TransJakarta merupakan bus yang memiliki jalur khusus, halte yang dibangun pun didesain khusus dengan ciri khas tertentu untuk membedakan halte TransJakarta dengan pelayanan transportasi umum lainnya. Agar desain khusus untuk halte TransJakarta terintegrasi dengan lingkungan dan memberikan pelayanan yang prima, halte yang dibangun harus memperhatikan keserasian lingkungan, berfungsi sebagai ornamen lingkungan, dan memperhatikan aksesibilitas bagi penyandang cacat (Darat, 2006, dikutip dalam Pratama, 2021).

Ada beberapa syarat umum yang harus dipenuhi agar halte dapat digolongkan sebagai tempat perhentian kendaraan umum, yaitu berada di sepanjang rute angkutan umum/bus, terletak pada jalur pejalan (kaki) dan dekat dengan fasilitas pejalan (kaki), disarankan dekat dengan pusat kegiatan atau pemukiman, dilengkapi dengan rambu petunjuk, dan tidak mengganggu kelancaran lalu lintas. Berdasarkan Pedomanan Perekayasa Tempat Perhentian Kendaraan Umum, fasilitas utama yang harus dimiliki oleh halte adalah identitas halte berupa nama dan/atau nomor, rambu petunjuk, papan informasi trayek, lampu penerangan, dan tempat duduk. Selain fasilitas utama, juga terdapat fasilitas tambahan, yaitu telepon umum, tempat sampah, pagar, dan papan iklan/pengumuman (Pratama, 2021).

### **Gerai Kopi**

Kedai kopi termasuk ke dalam golongan industri kreatif (Poernomo et al., 2019 dikutip dalam Pandu Adi Cakranegara, 2020). Salah satu cara untuk mempromosikannya adalah dengan menggunakan *branding*. Diharapkan dengan adanya *branding*, tercipta suatu identitas yang mudah dikenal konsumen.

Salah satu pendekatan yang digunakan kedai-kedai kopi, terutama untuk kedai kopi yang menyasar masyarakat dengan usia 20 hingga 40 tahun adalah dengan *sensory branding*. *Sensory branding* bertujuan untuk membuat kedai kopi menjadi menarik, yaitu dengan penggunaan desain ruangan yang menarik serta penataan produk dan visual (Aryani, 2019 dikutip dalam Pandu Adi Cakranegara, 2020).

Hal lain yang menentukan kesuksesan gerai kopi adalah harga, retensi pelanggan, dan kepuasan pegawai. Kedai kopi dengan tipe dan strategi seperti ini adalah kedai kopi yang disebut sebagai kedai kopi generasi ketiga. Generasi kopi ketiga ini kemudian akan berubah lagi menuju ke tren generasi kopi keempat yang disebut generasi “kopi *to go*”. Pada tahap terakhir, yaitu gelombang *in*, kopi sudah merupakan gaya hidup dan orang meminum kopi sambil beraktivitas (Reynold et al., 2017 dikutip dalam Pandu Adi Cakranegara, 2020).

### **Twitter**

Twitter adalah situs *microblog* yang memberikan fasilitas bagi pengguna untuk mengirimkan sebuah pesan teks dengan panjang maksimal 140 karakter melalui SMS, pengirim pesan instan, surat elektronik (Hadi, 2010, dikutip dalam Yunis Femilia Nugraini, Rd. Rohmat Saedudin, 2021). Twitter didirikan pada tahun 2006 oleh Jack Dorsey. Twitter menjadi salah satu media sosial yang banyak diminati dan digunakan oleh masyarakat dunia, termasuk masyarakat Indonesia. Banyaknya peminat media sosial Twitter dibuktikan dengan jumlah pengguna yang terus meningkat sejak didirikan. Pada tahun 2013, pengguna Twitter telah mencapai 550 juta pengguna. Sifat dan karakteristik yang sederhana dari Twitter menjadi alasan utama banyaknya peminat dari media sosial ini.

Twitter menyediakan berbagai fitur untuk penggunanya yang dapat diakses melalui perangkat seluler atau PC. Salah satu fitur yang disediakan oleh Twitter adalah fitur yang memungkinkan pengguna untuk mengirim dan melacak pesan yang dikenal sebagai *tweet*. Fitur lain yang ditawarkan oleh Twitter adalah *retweet*, yaitu fitur yang digunakan untuk mem-*posting* ulang status pengguna lain, dan juga fitur *replay*, yaitu fitur untuk merespons *posting*-an pengguna lain. Selain itu, Twitter memungkinkan penggunanya untuk berbagi tautan, seperti foto dari Instagram, lagu dari YouTube, atau berita dari saluran berita apa pun (Emeraldien et al., 2019).

Untuk terhubung dengan pengguna atau pemilik akun lain, pengguna dapat mengikuti akun Twitter pengguna lain tersebut. Cara lain untuk terhubung dengan pengguna lain adalah dengan menggunakan *mention*. Cukup mengetik nama akun Twitter pengguna lain di kolom status, dimulai dengan simbol "@". Selanjutnya, tagar (simbol "#" diikuti oleh kata kunci tertentu) pada Twitter dapat menyatukan orang dalam topik tertentu. Tagar sangat efektif dalam membangun komunitas aktual, seperti dalam menanggapi berita terkini (Burns dan Highfield, 2012, dikutip dalam Emeraldien et al., 2019).

### **Urgensi**

Manusia memiliki berbagai kepentingan dengan sifat yang beragam pula. Ada kepentingan yang bersifat sangat mendesak, mendesak, kurang mendesak dan tidak mendesak. Mendesak atau tidak mendesaknya kepentingan tersebut tergantung kepada kebutuhan dan akibat dari permintaan kebutuhan itu, atau yang dikenal dengan urgensi. Urgensi merupakan sesuatu yang sangat penting sehingga harus segera diberi perhatian. Dapat disimpulkan bahwa urgensi adalah suatu kepentingan yang harus segera diselesaikan, jika tidak akan berdampak pada aktivitas lainnya (Is Bukhari, 2021).

### **Persepsi**

Persepsi adalah kondisi saat kita memilih, mengatur, dan menerjemahkan masukan informasi untuk menciptakan gambaran dunia yang berarti (Kotler dan Keller, 2009, dikutip dalam Ilma Suryani Mardana, Junaidi, 2022). Persepsi adalah proses kognitif yang dipergunakan oleh individu untuk menafsirkan dan memahami dunia sekitarnya. Dengan kata lain, persepsi mencangkup penerimaan stimulus yang diorganisasikan, dan penerjemahan atau penafsiran stimulus yang diorganisasikan dengan cara yang dapat mempengaruhi perilaku dan pembentukan sikap (Haryo, 2004 dikutip dalam Supiani et al., 2021).

Jadi, dapat disimpulkan bahwa persepsi merupakan proses memakai sesuatu yang diterima melalui kelima indra sesuai dengan tingkat pemahaman dan karakter yang dimiliki agar seseorang dapat memilih, mengatur dan menafsirkan suatu informasi untuk memberikan pemahaman tentang lingkungan.

### ***Text Mining***

Feldman dan Sanger (2007) menyatakan *text mining* adalah sebuah proses pengetahuan intensif, yaitu pengguna berinteraksi dan bekerja dengan sekumpulan dokumen dengan menggunakan beberapa alat analisis (Feldman dan Sanger , 2007, dikutip dalam Yosmita Praptiwi, 2018).

*Text mining* dan *data mining* merupakan dua hal yang berbeda, *Text mining* atau *text analytics* adalah istilah yang mendeskripsikan sebuah teknologi yang mampu menganalisis data teks semi-terstruktur maupun tidak terstruktur. Hal tersebut menjadi pembeda *text mining* dengan *data mining*. Pada *data mining*, data yang diolah merupakan data yang sifatnya terstruktur (Jamil, 2017, dikutip dalam Yosmita Praptiwi, 2018).

### ***Web Scrapping***

*Web scraping* adalah salah satu metode mengumpulkan data yang digunakan untuk mengekstraksi data dari suatu halaman web. Halaman web dibangun menggunakan bahasa berbasis teks (HTML dan XHTML) yang berisi banyak data dalam bentuk teks dan dapat digunakan untuk berbagai keperluan. Sebagian besar halaman web dirancang untuk kemudahan penggunaan oleh manusia. Oleh karena itu, alat dan perangkat lunak khusus telah dikembangkan untuk melakukan *web scraping* (Sahria, 2022, dikutip dalam Baskara & Rahma, 2022).

Secara hukum, *web scrapping* merupakan hal yang sah untuk dilakukan, dengan syarat, data yang dikumpulkan digunakan untuk kepentingan pribadi dan tidak melanggar undang-udang yang berlaku. Jika data akan dipublikasikan atau jika konten memiliki hak cipta dan melanggar persyaratan layanan, ada beberapa aturan dan hukum yang perlu diperhatikan (Sahria, 2022, dikutip dalam Baskara & Rahma, 2022).

*Web scraping* biasanya menargetkan situs web tertentu dengan tujuan mengumpulkan informasi spesifik di situs tersebut. *Scraper*, sebutan untuk alat yang digunakan pada *web scrapping*, dibuat untuk mengakses halaman spesifik dan perlu dimodifikasi jika situs berubah atau jika informasi yang ada di situs diubah (Sahria, 2022, dikutip dalam Baskara & Rahma, 2022).

Terdapat beberapa teknik *web scraping* yang umumnya digunakan. Yang pertama, menyalin data secara manual. *Web scraping* yang dilakukansecara manual merupakan bentuk paling sederhana yang dapat dilakukan. Ada suatu kondisi saat teknologi untuk melakukan *web scraping* terbaik sekalipun tidak dapat menggantikan pemeriksaan manual yang dilakukan oleh manusia. Teknik yang kedua merupakan *parsing* HTML. Dengan teknik ini, *web scraping* dapat dilakukan tidak hanya pada halaman website yang bersifat statis, tetapi juga dinamis. *Parsing* HTML juga memungkinkan untuk menyalin data dalam jumlah yang besar dalam waktu singkat (Sahria, 2022, dikutip dalam Baskara & Rahma, 2022).

### ***Text Preprocessing***

Dalam melakukan *text mining*, teks dokumen yang digunakan harus dipersiapkan terlebih dahulu, setelah itu baru dapat digunakan untuk proses utama. Proses mempersiapkan teks dokumen atau *dataset* mentah disebut juga dengan proses *text preprocessing*. *Text preprocessing* berfungsi untuk mengubah data teks yang tidak terstruktur atau tidak beraturan menjadi data yang terstruktur (Yosmita Praptiwi, 2018). Secara umum proses yang dilakukan dalam tahapan *preprocessing* adalah sebagai berikut:

1. *Tokenization* adalah proses memotong suatu teks menjadi bagian-bagian yang disebut sebagai token.
2. *Capitalization*/*Case-Folding*

*Capitalization* adalah proses mengubah teks menjadi seluruhnya huruf kapital (*uppercase*), sedangkan *case-folding* adalah proses mengubah teks menjadi seluruhnya huruf kecil (*lowercase*).

1. *Stopword Elimination*

*Stopword* merupakan kata-kata yang dianggap hanya sedikit berarti dalam pemrosesan teks, sehingga dapat dieliminasi. Setiap bahasa memiliki daftar *stopword* yang berbeda. Strategi lain penentuan *stopword* adalah dengan menghitung frekuensi kemunculan kata dan menganggap kata-kata yang jarang muncul dalam dokumen sebagai *stopword* yang dapat dieliminasi.

1. *Normalization* adalah mengelompokkan token yang penulisan hurufnya berbeda tetapi termasuk ke kata yang sama menjadi satu term yang sama. Contohnya yaitu kata anti vaksin dapat dituliskan dengan bebrapa variasi seperti berikut, antivaksin: {anti vaksin, antivaksin, anti-vaksin}. Selain itu, normalisasi dapat juga digunakan untuk mengelompokkan bahasa gaul ke bahasa bakunya, contohnya, saya: {gue, gua, aku, aq, sy}.
2. *Stemming* adalah mengubah kata-kata berimbuhan ke bentuk dasarnya.

### ***Maximum Entropy Classification***

Metode *Maximum Entropy Classification* merupakan metode yang digunakan untuk mengklasifikasikan data berdasarkan nilai *entropy*-nya (Sabily et al., 2019). Metode *Maximum Entropy* bertujuan untuk memaksimalkan *entropy* dalam sistem dengan memprediksi distribusi bersyarat label pada setiap kategori (Cindo, et al., 2019, dikutip dalam Mauludin Rohman & Adinugroho, 2021). Metode ini efektif digunakan dalam beberapa aplikasi pemrosesan bahasa alami (Htet & Myint, 2018, dikutip dalam Mauludin Rohman & Adinugroho, 2021).

Perhitungan rumus menghitung tingkat atau besaran *entropy* menggunakan persamaan berikut.

Keterangan:

= Peluang kelas 𝑋𝑖

Namun, dalam kasus untuk mengklasifikasikan teks, metode *Maximum Entropy* dapat menggunakan rumus berikut (Alroy, et al., 2019).

Keterangan:

c = kelas dari sebuah dokumen

d = setiap term yang terdapat pada dokumen

𝜆𝑖 = parameter atau nilai bobot fitur

𝑓𝑖 (𝑑, 𝑐) = peluang dari term d yang terdapat pada kelas c

### **R**

R adalah suatu software terintegrasi yang memiliki fasilitas untuk pemanipulasian data, perhitungan, dan penampilan grafik. Software ini terus dikembangkan dan diberikan secara gratis kepada masyarakat yang memerlukannya dengan cara mendown load melalui internet. R dapat dikatakan sebagai suatu implementasi bahasa S yang telah dikembangkan di Bell Laboratories oleh Rick Becker, John Chamber dan Allan Willks, orang yang juga membentuk dasar system S-Plus.

Walaupun lingkungan R (R *environment*) tidak menyebutkan statistik, namun banyak orang menggunakan R sebagai sistem statistik. Hal ini dikarenakan kemampuan R dalam mengolah dan menganalisis data cukup baik. R tidak hanya digunakan untuk keperluan statistik. Banyak orang menggunakan R untuk *machine learnig*.

### ***Package* Rtweet**

Ketertarikan pada data Twitter terus meningkat, tetapi mengumpulkan dan menganalisis data melalui API Twitter masih sulit untuk dilakukan. Misalnya, untuk berinteraksi dengan pengguna API Twitter, selain mengidentifikasi dan mencerna informasi yang relevan dari dokumentasi pengembang Twitter, juga harus membuat/mengirim/menerima permintaan, mengelola batas kecepatan, dan memperdebatkan objek respons bersarang dan *real-time* ke dalam struktur data yang mudah analisis. Untungnya, *package* ‘rtweet’ pada R dirancang untuk menyederhanakan proses ini. *Package* ‘rtweet’ membuat interaksi dengan API Twitter lebih mudah diakses oleh lebih banyak pengguna.

Tujuan utama dari *package* ‘rtweet’ ada dua. Tujuan pertama adalah membuat interaksi dengan API Twitter lebih mudah didekati dan disederhanakan untuk pengguna yang kurang berpengalaman secara komputasi. Tujuan kedua adalah untuk membantu analisis data Twitter melalui konversi informasi yang dikembalikan oleh API Twitter ke dalam struktur data tabular dan menyediakan beberapa fungsi kemudahan untuk teknik analitis umum seperti memeriksa jaringan Twitter atau frekuensi cuitan lembur. Singkatnya, meskipun sangat mungkin bagi pengguna untuk menulis API Twitter mereka sendiri sebagai fungsi pembungkus, pekerjaan berat dapat dilakukan oleh ‘rtweet’, yaitu untuk (a) merampingkan pembuatan, otorisasi, dan pengiriman permintaan API, (b) memperdebatkan data JSON yang bersarang dalam ke dalam struktur tabel, dan (c) menyediakan fungsi kemudahan untuk pembuatan teknik analitis yang relevan dan populer yang berharga di bidang pengumpulan dan analisis data Twitter.

## **Penelitian Terkait**

Penelitian terkait sangatlah penting bagi penulis sebagai kajian untuk mengetahui keterkaitan antara penelitian terdahulu dengan penelitian yang akan dilakukan, untuk menghindari terjadinya tindakan duplikasi yang dilakukan oleh penulis. Tujuan dari kajian pustaka ini adalah untuk menunjukkan bahwa penelitian yang dilakukan penulis mempunyai manfaat dan kotribusi terhadap ilmu pengetahuan. Berikut beberapa ulasan tentang penelitian-penelitian yang pernah dilakukan sebelumnya berkenaan dengan data dan metode yang digunakan.

### **Analisis Sentimen *Online Review* Pengguna *E-Commerce* Menggunakan Metode *Support Vector Machine* dan *Maximum Entropy* (Studi Kasus: Buka Lapak pada Google Play)**

Yosmita Praptiwi, D (2018) menganalisis sentimen *online review* pengguna *e-commerce* dengan menggunakan metode *Support Vector Machine* (SVM) dan *Maximum Entropy* (ME). Metodologi yang digunakan untuk melakukan analisis sentimen dimulai dari *data collecting*, analisis deskriptif, *preprocessing*, klasifikasi, dan visualisasi. Klasifikasi sentimen menggunakan metode SVM menghasilkan tingkat akurasi sebesar 91,95%. Artinya, dari 385 data ulasan yang diujikan, terdapat 354 ulasan yang diklasifikasikan dengan benar oleh metode SVM. Dengan menggunakan metode ME, tingkat akurasi yang dihasilkan lebih tinggi daripada metode SVM, yaitu sebesar 92,98%. Artinya, dari 385 data ulasan yang diujikan, terdapat 358 ulasan yang diklasifikasikan dengan benar oleh metode *Maxent*.

### **Analisis Sentimen Pemilihan Presiden 2019 pada Twitter Menggunakan Metode *Maximum Entropy***

Dalam penelitian yang dilakukan Sabily, dkk (2019) tentang analisis sentimen pemilihan presiden 2019 pada twitter, digunakan metode *Maximum Entropy* (ME) dengan metode evaluasinya adalah *Confusion Matrix* yang menghitung nilai *Macro* dan *Micro*. Metodologi yang digunakan untuk melakukan analisis sentimen dimulai dari *data collecting*, *preprocessing*, pembobotan, klasifikasi, dan analisis. Dari hasil penelitiannya, Sabily, dkk menyatakan bahwa metode ME menghasilkan akurasi yang cukup tinggi, yaitu sebesar 89,16%, dengan nilai *Presicion* 100%, *Recall* 89,16% dan *F-measure* sebesar 94,27%. Hasil evaluasi metode *Maximum Entropy* menggunakan metode *Confusion Matrix* menghasilkan nilai *Macro* dan *Micro* yang sama.

### **Metode Ekstraksi Fitur pada Analisis Sentimen Maskapai Penerbangan Menggunakan *Support Vector Machine* dan *Maximum Entropy***

Penelitian lain juga dilakukan oleh Cindo, dkk (2019) tentang analisis sentimen maskapai penerbangan. Metodologi yang digunakan untuk melakukan analisis sentimen dimulai dari *data collecting*, *preprocessing*, ekstraksi fitur, klasifikasi, dan evaluasi. Metode klasifikasi *Support Vector Machine* (SVM) dan *Maximum Entropy* (ME) digunakan untuk membandingkan hasil prediksi klasifikasi yang terbaik. Penelitian ini menggunakan 5 ekstraksi fitur, yaitu fitur *pragmatic*, *lexical n-grams*, *lexical sentiment*, *POS-grams*, dan LDA topik yang diharapkan dapat meningkatkan akurasi dari analisis sentimen. Hasil eksperimen pada semua fitur menunjukkan metode ME memiliki tingkat akurasi yang lebih tinggi, yakni sebesar 92,7%, dibandingkan dengan metode SVM, yakni sebesar 89,2%.

### **Perbandingan Penelitian**

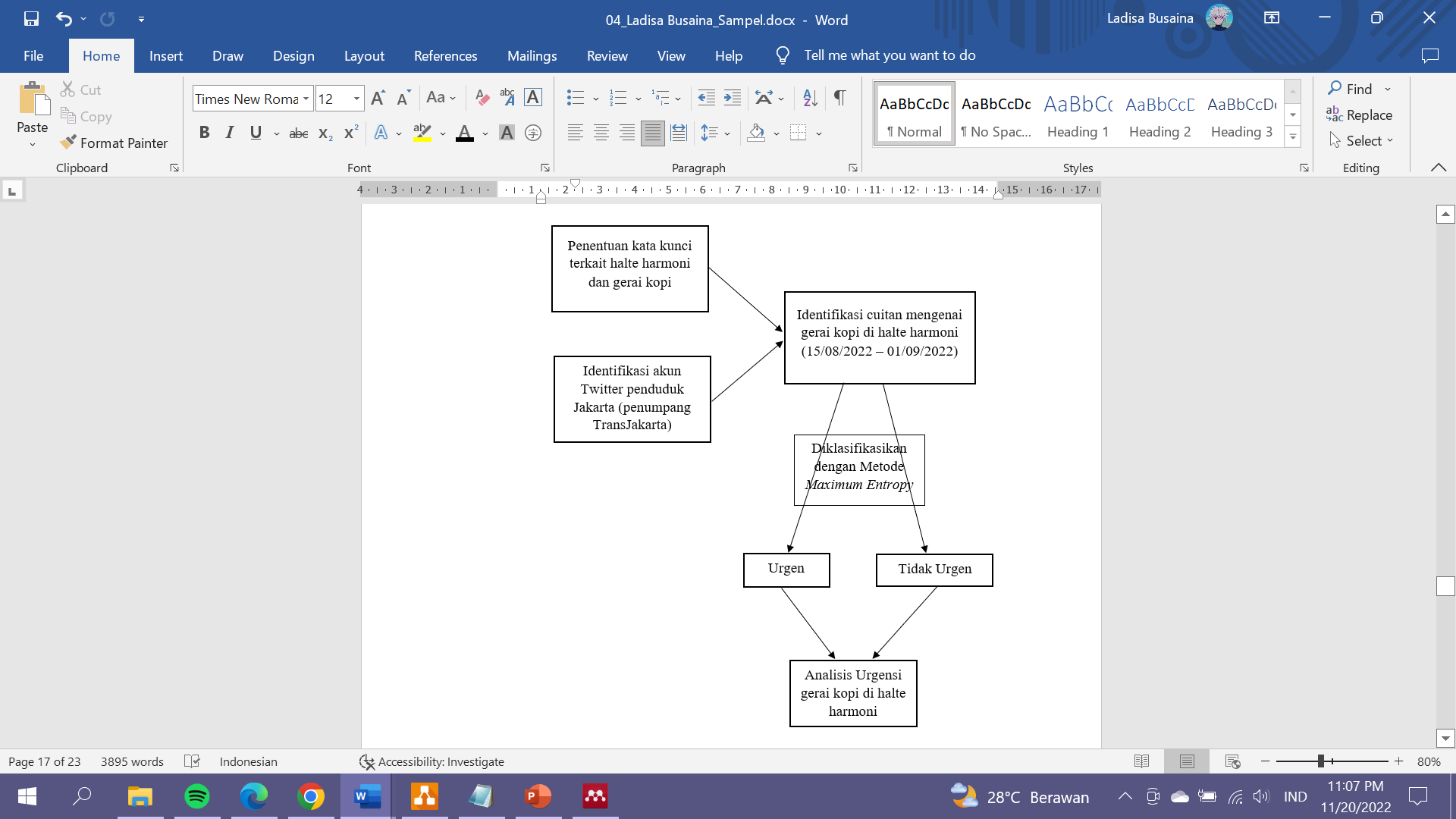
Berikut tabel perbandingan metode-metode klasifikasi yang digunakan pada penelitian-penelitian sebelumnya.

Table 2.1 Perbandingan Metode Klasifikasi Penelitian Sebelumnya

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Penulis, tahun** | **Tujuan** | **Metode** | **Sampel** | **Temuan** | **Kesamaan** | **Perbedaan** |
| Yosmita Praptiwi, D (2018) | Mengetahui seberapa tepat machine learning dengan menggunakan metode Support Vector Machine (SVM) dan *Maximum Entropy* (*Maxent*) dalam mengklasifikasikan teks berbahasa Indonesia mengenaiulasan pengguna Bukalapak berdasarkan situs Google Play. | *Support Vector Machine* (SVM) dan *Maximum Entropy* (ME) | Data ulasan tentang Bukalapak | Metode ME memiliki tingkat akurasi yang lebih tinggi, yakni sebesar 92,98%, dibandingkan dengan metode SVM, yakni sebesar 91,95%. | Menggunakan metode *Maximum Entropy* dalam pengklasifikasian. | Menggunakan metode SVM sebagai perbandingan dengan metode *Maximum Entropy* |
| Sabily dkk (2019) | Menganalisis opini yang berasal dari ungkapan seseorang melalui media sosial Twitter mengenai Pemilihan Presiden 2019. | *Maximum Entropy* (ME) | Cuitan di media sosial Twitter mengenai Pemilihan Presiden 2019. | Metode ME menghasilkan akurasi yang cukup tinggi, yaitu sebesar 89,16%, dengan hasil evaluasi berupa nilai *Macro* dan *Micro* yang sama. | Menggunakan metode *Maximum Entropy* dalam pengklasifikasian. | Hasil dari analisis sentimen tidak dianalis lebih lanjut, hanya dievaluasi. |
| Cindo dkk (2019) | Menganalisis kepuasan pelanggan perusahaan untuk meningkatkan layanannya. | *Support Vector Machine* (SVM) dan *Maximum Entropy* (ME) | Data Twitter tentang opini maskapai penerbangan yang tersedia di internet. | Pada semua fitur, metode ME memiliki tingkat akurasi yang lebih tinggi, yakni sebesar 92,7%, dibandingkan dengan metode SVM, yakni sebesar 89,2%. | Menggunakan metode *Maximum Entropy* dalam pengklasifikasian. | Menggunakan metode SVM sebagai perbandingan dengan metode *Maximum Entropy*. |

## **Kerangka Pikir**

Penelitian ini akan membahas tentang pengklasifikasian data Twitter dalam menentukan urgensi didirikannya gerai kopi di halte harmoni. Untuk mendapatkan data dari Twitter, ditentukan dahulu *keyword* yang berkaitan dengan gerai kopi dan halte harmoni. Akun Twitter yang dihitung *tweet* atau *reply*-nya adalah akun Twitter penduduk Jakarta, terutama akun milik penumpang TransJakarta.



Gambar 2. 1 Kerangka Pikir

Setelah kata kunci dan akun Twitter ditentukan, lakukan identifikasi *tweet* mengenai gerai kopi di halte harmoni pada tanggal 15 Agustus 2022 sampai dengan 1 September 2022. *Tweet* yang terjaring akan diklasifikasikan dengan Metode *Maximum Entropy* yang berdasarkan penelitian-penelitian sebelumnya menghasilkan akurasi yang tinggi. *Tweet* diklasifikasikan menjadi *tweet* yang menyatakan bahwa gerai kopi di halte harmoni urgen dan *tweet* yang menyatakan bahwa gerai kopi di halte harmoni tidak urgen. Terakhir, analisis hasil klasifikasi yang didapatkan.

# **BAB III**

# **METODOLOGI**

## **Ruang Lingkup Penelitian**

Data yang digunakan pada penelitian ini merupakan data dari media sosial Twitter. Data tersebut berupa cuitan pada media sosial tersebut. Cuitan yang digunakan akan dibatasi hanya untuk cuitan pada tanggal 15 Agustus 2022 sampai dengan 1 September 2022. Pembatasan periode cuitan ini dilakukan agar data cuitan terjaring secara maksimal.

Untuk pemilihan tanggalnya, peneliti menentukan berdasarkan tanggal dibuatnya cuitan pertama mengenai gerai kopi di halte harmoni dan diakhiri dengan tanggal dibongkarnya gerai kopi di halte harmoni. Cuitan yang dianalisis juga akan dibatasi untuk cuitan dari akun-akun penduduk Jakarta. Hal ini dikarenakan letak halte harmoni yang berada di Provinsi Jakarta sehingga akan lebih sesuai jika penduduk Jakarta sebagai pengguna fasilitas yang disediakan oleh Pemerintah Jakarta untuk menilai urgensi dari didirikannya gerai kopi di halte harmoni.

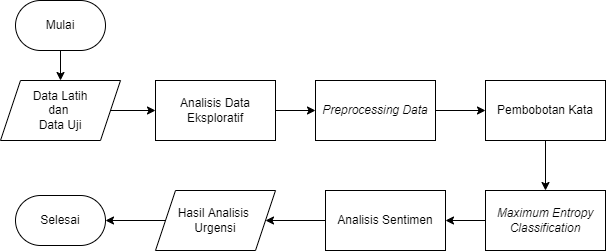
## **Metode Pengumpulan Data**

Dalam mengumpukan data respon dan persepsi penduduk Jakarta mengenai didirikannya gerai kopi Qopi Untuk Semua di halte harmoni dalam bentuk cuitan di media sosial Twitter, digunakan teknik *web scrapping*. Kegiatan *web scrapping* pada penelitian ini dilakukan menggunakan aplikasi R dengan bantuan *package* ‘rtweet’.

Terdapat 2 kata kunci yang digunakan pada penelitian ini, yaitu “gerai kopi” dan “halte harmoni”. Kedua kata kunci tersebut dinilai dapat menjaring semua data cuitan penduduk Jakarta mengenai gerai kopi Qopi Untuk Semua di halte harmoni yang mengganggu kenyamanan para penumpang TransJakarta dari tanggal 15 Agustus 2022 sampai dengan 1 September 2022.

## **Metode Analisis**

Terdapat tujuh tahapan yang dilakukan pada penelitian ini, yaitu pengumpulan data yang terdiri dari data latih dan data uji, *preprocessing data*, analisis data eksploratif, pembobotan kata, *maximum entropy classification*, analisis sentimen, dan hasil analisis urgensi. Gambar berikut menunjukkan tahapan yang dilakukan pada penelitian ini.



Gambar 3. 1 Diagram Alir Penelitian

2. Data Latih dan Data Uji

*Dataset* yang diperoleh dari Twitter dan digunakan untuk menganalisis urgensi didirikannya gerai kopi Qopi Untuk Semua di halte harmoni diperkirakan memiliki jumlah yang banyak. Agar pengolahan data menjadi efektif, peneliti akan membagi data tersebut dibagi menjadi dua jenis, yaitu data latih dan data uji. Dari keseluruhan data yang diperoleh, data latih memiliki porsi 20% dan data uji 80%.

Data latih yang ada akan diklasifikan menjadi dua kelas, yaitu data dengan sentimen positif dan data dengan sentimen negatif. Pengklasifikasian data ini didasarkan pada kamus positif-negatif yang dibuat Liu, Hu, & Cheng yang telah diterjemahkan ke dalam Bahasa Indonesia dan umum digunakan dalam pengklasifikasian sentimen.

1. Analisis Data Eksploratif

Analisis data eksploratif digunakan untuk memvisualisasikan data yang diperoleh dengan diagram yang sesuai, seperti *barchart* ataupun *piechart*. Visualisasi ini dilakukan untuk melihat sebaran data yang diperoleh. Selain visualisasi, data yang diperoleh juga dilakukan pengecekan *missing value*. Pengecekan *missing value* dilakukan untuk memastikan *preprocessing* yang dilakukan sebelumnya telah berjalan dengan baik dan benar. Terakhir, dilakukan perhitungan jumlah kata dari data yang ada.

1. *Preprocessing Data*

Untuk kemudahan pengolahan data, data tidak terstruktur yang telah dikumpulkan terlebih dahulu dirapikan sehingga menjadi data yang terstruktur atau yang dikenal dengan proses *preprocessing data*. *Preprocessing data* yang dilakukan pada penelitian ini adalah mengubah seluruh huruf menjadi huruf kecil (*lowecase*), menghapus tanda baca, menghapus karakter-karakter lain, menghapus tanda pagar (*hashtag*) di depan suatu kata, menghapus HTML *tags*, menghapus emoji, menghapus kata yang tidak penting, dan melakukan normalisasi kata. Pada tahap normalisasi kata, semua kata ditransformasi menjadi kata formal dan baku sesuai dengan KBBI. Penghapusan kata yang tidak penting dilakukan untuk mengefisienkan waktu me-*running* data dan pengolahan data.

1. Pembobotan Kata

Pembobotan kata dilakukan melalui beberapa tahap. Pertama, setiap kata pada data uji dihitung kemunculannya terhadap data uji atau yang dikenal dengan *term frequency*. Data uji yang dimaksud di sini adalah 80% dari semua cuitan yang berhasil dijaring pada tahap pengumpulan data. Kemudian, dihitung *document frequency* dari data uji tersebut, yaitu banyaknya cuitan dimana suatu kata muncul. Dari *document frequency*, akan dihitung *invers document frequency* yang merupakan ukuran keinformatifan dari suatu kata atau *term*. Terakhir, dilakukan pembobotan. Pemberian bobot pada semua kata pada data uji dilakukan dengan mengkombinasikan *term frequency* dan *invers document frequency* yang telah didapatkan sebelumnya. Berikut rumus dari pembobotan menggunakan *term frequency* dan *invers document frequency*.

1. *Maximum Entropy Classification*

Setelah pembobotan kata dilakukan, semua kata pada data uji akan diklasifikasikan dengan metode *Maximum Entropy*. Rumus yang digunakan untuk peluang menggunakan metode *Maximum Entropy* adalah sebagai berikut.

Jika diuraikan, c pada rumus berarti kelas dari cuitan pada data, yang pada penelitian ini adalah data uji, d merupakan kata atau *term* yang terdapat pada setiap cuitan, lambda merupakan bobot dari kata, dan f merupakan nilai peluang dari kata yang terdapat pada kelas c. Rumus tersebut akan menghasilkan nilai peluang dari kelas c dengan kondisi munculnya suatu kata pada cuitan tertentu. Jika data uji memiliki data yang heterogen, nilai *entropy* yang dihasilkan akan semakin besar. Begitu juga sebaliknya, jika data uji memiliki data yang homogen, nilai *entropy* yang dihasilkan akan semakin kecil.

1. Analisis Sentimen

Dari pengklasifikasian menggunakan metode *Maximum Entropy*, akan dilakukan analisis sentimen terdahap data uji. Analisis yang dilakukan berupa evaluasi terhadap pengklasifikasian menggunakan metode *Maximum Entropy*. Evaluasi pada penelitian ini menggunakan *confusion matrix*.

*Confusion matrix* merupakan tabel yang berisi 4 kombinasi berbeda dari nilai prediksi dan nilai aktual. Nilai prediksi merupakan keluaran dari program yang nilainya positif dan negatif, nilai aktual merupakan nilai sebenarnya yang nilai *true* dan *false*, sedangkan 4 kombinasi dari nilai prediksi dan aktual adalah *true positive*, *true negative*, *false positive*, dan *false negative*.

Dari *confusion matrix* dapat dihasilkan evaluasi berupa nilai *precision*, *recall*, akurasi, dan *f-measure*. *Precision* merupakan porsi dari dokumen yang diambil yang relevan dengan kebutuhan pengguna. Berikut rumus dari *precision*.

*Recall* merupakan porsi dari dokumen yang relevan dalam koleksi yang diambil. Berikut rumus dari *recall*.

Akurasi merupakan ukuran ketepatan pengklasifikasian dari suatu mesin pencari. Berikut rumus dari akurasi.

*F-measure* merupakan ukuran kombinasi yang menggabungkan *precision* dan *recall*. Berikut rumus dari *f-measure*.

1. Analisis Urgensi

Terakhir, setelah data diklasifikasikan menjadi beberapa sentimen, dan dievaluasi hasil pengklasifikasiannya, dilakukan analisis urgensi. Analisis urgensi merupakan tahapan untuk menilai urgensi didirikannya gerai kopi Qopi Untuk Semua di halte harmoni berdasarkan proporsi dari sentimen positif dan negatif dari data yang dikumpulkan. Jika proporsi sentimen negatif lebih banyak dibandingkan dengan proporsi sentimen positif, dapat disimpulkan bahwa gerai kopi di halte harmoni tidak urgen untuk didirikan dan pemerintah harus mempertimbangkan kembali izin didirikannya gerai kopi tersebut. Namun, jika proporsi sentimen positif lebih banyak dibandingkan dengan proporsi sentimen negatif, dapat disimpulkan pendirian gerai kopi di halte harmoni tidak mengganggu kenyamanan mayoritas penumpang. Kemungkinan ada faktor lain selain adanya gerai kopi di halte harmoni yang menyebabkan antrean di halte harmoni menjadi tidak teratur dan berdesak-desakan.

# **DAFTAR PUSTAKA**

Anggoro, A. G., Hadi, W., & Daryati, D. (2020). Analisis Tingkat Pelayanan Terminal Transjakarta Pada Stasiun Harmoni Central Busway Berdasarkan Persepsi Penumpang. *Menara: Jurnal Teknik Sipil*, *15*(1), 30–36. https://doi.org/10.21009/jmenara.v15i1.18063

Astuti, S. I., Arso, S. P., & Wigati, P. A. (2017). Analisis Standar Pelayanan Minimum Transjakarta Berdasarkan Observasi Dan Persepsi Penumpang. *Analisis Standar Pelayanan Minimal Pada Instalasi Rawat Jalan Di RSUD Kota Semarang*, *3*(November), 103–111.

Baskara, R., & Rahma, F. (2022). Implementasi Web Scraping Pada Media Sosial Instagram. *Automata*, *3*, 1–3.

Emeraldien, F. Z., Sunarsono, R. J., & Alit, R. (2019). Twitter Sebagai Platform Komunikasi Politik Di Indonesia. *SCAN: Jurnal Teknologi Informasi Dan Komunikasi*, *14*(1), 21–30. http://ejournal.upnjatim.ac.id/index.php/scan/article/view/1457/1207

Ilma Suryani Mardana, Junaidi, dan S. A. A. (2022). Persepsi Mahasiswa Akuntansi terhadap Etika Bisnis dan Etika Profesi Dipandang dari Segi Gender (Studi Kasus Mahasiswa Akuntansi Unisma). *E-JRA Vol. 11 No. 09 Februari 2022 Fakultas Ekonomi Dan Bisnis Universitas Islam Malang*, *11*(09), 84–92.

Is Bukhari, P. R. (2021). Urgensi Pengkajian Dan Penelitian Bagi Insan Akademik. *Agama Dan Sains*, *V*(2), 17.

Mauludin Rohman, M., & Adinugroho, S. (2021). *Analisis Sentimen pada Ulasan Aplikasi Mobile JKN Menggunakan Metode Maximum Entropy dan Seleksi Fitur Gini Index Text*. *5*(6), 2646–2654. http://j-ptiik.ub.ac.id

Pandu Adi Cakranegara. (2020). Gerai Kopi di Jakarta dalam Adaptasi Kehidupan Baru. *EDUTOURISM Journal Of Tourism Research*, *2*(02), 1–7. https://doi.org/10.53050/ejtr.v2i02.133

Pratama, M. R. (2021). *Evaluasi Halte Bus Trans Metro Deli Koridor 5 Medan Lapangan Merdeka - Tembung terhadap Naik Turun Penumpang Trans Metro Deli*. 6.

Sabily, A. F., Adikara, P. P., & Fauzi, M. A. (2019). Analisis Sentimen Pemilihan Presiden 2019 pada Twitter menggunakan Metode Maximum Entropy. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, *3*(5), 4204–4209.

Sunni, I., & Widyantoro, D. H. (2012). Analisis Sentimen dan Ekstraksi Topik Penentu Sentimen pada Opini Terhadap Tokoh Publik. *Jurnal Sarjana Institut Teknologi Bandung Bidang Teknik Elektro Dan Informatika*, *1*(2), 200–206.

Supiani, S., Rahmat, F., & Budiman, F. (2021). Pengaruh Budaya dan Persepsi Masyarakat terhadap Keputusan Menabung di Bank Syariah. *Al-Bank: Journal of Islamic Banking and Finance*, *1*(1), 49. https://doi.org/10.31958/ab.v1i1.2618

Yosmita Praptiwi, D. (2018). *Analisis Sentimen Online Review Pengguna E-Commerce Menggunakan Metode Support Vector Machine Dan Maximum Entropy*.

Yunis Femilia Nugraini, Rd. Rohmat Saedudin, R. A. (2021). Implementasi Data Mining Dalam Kasus Mental Health Pada Sosial Media Twitter Menggunakan Metode Naive Bayes. *EProceedings …*, *8*(5), 9260–9265. https://openlibrarypublications.telkomuniversity.ac.id/index.php/engineering/article/view/15827