

BAB I

PENDAHULUAN

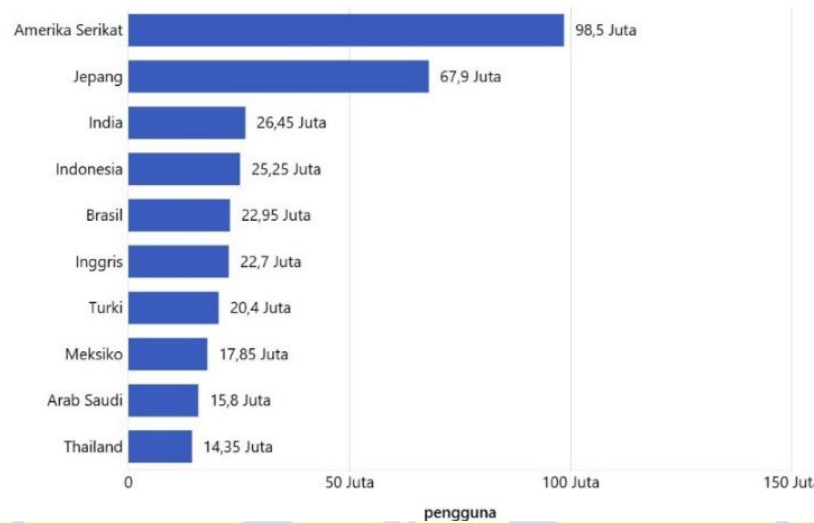
1.1 Latar Belakang

Sebagai negara demokrasi, Indonesia melaksanakan pemilihan umum atau yang biasa disebut dengan pemilu untuk memilih wakil rakyat. Pemilu merupakan pilar utama dan bentuk konstitusional dalam sistem sebuah negara demokrasi. Pada penjelasan atas Undang-Undang Republik Indonesia Nomor 3 Tahun 1999, dituliskan bahwa tujuan dari pemilu adalah untuk memilih wakil rakyat yang akan duduk di dalam lembaga permusyawaratan/perwakilan rakyat, membentuk pemerintahan, melanjutkan perjuangan mengisi kemerdekaan, dan mempertahankan keutuhan Negara Kesatuan Republik Indonesia. Pemilu dalam rangka pemilihan legislatif serta presiden dan wakil presiden dilaksanakan lima tahun sekali dan pemilu terakhir yaitu pemilu tahun 2024 merupakan pemilu Indonesia yang ke-13.

Pelaksanaan pemilu 2024 banyak diperbincangkan oleh masyarakat Indonesia tak hanya di lingkungan sekitar tetapi juga di media sosial seperti Instagram, X, TikTok, dan YouTube. Menurut Putra (2020), media-media sosial ini memberikan banyak manfaat dalam pemilu, seperti mempermudah akses informasi mengenai calon dan program politik, membentuk persepsi publik, serta meningkatkan partisipasi politik, terutama generasi muda. Platform ini juga menjadi sarana kampanye yang efektif dan hemat biaya, sekaligus memungkinkan interaksi langsung antara kandidat dan pemilih.

Dari berbagai media sosial tersebut diatas, media sosial Twitter atau yang sekarang disebut X memiliki keunggulan dalam pembahasan politik seperti pemilu dibandingkan dengan media sosial lainnya. Hal ini dikarenakan media sosial X yang memiliki sifat *real-time* dan kemampuan menjadi ruang diskusi publik. Selain itu, X sering menjadi media sosial utama bagi politisi, media, dan masyarakat untuk berinteraksi langsung, menjadikannya alat yang efektif untuk memantau dinamika politik selama pemilu. Keunggulan lainnya yaitu fakta bahwa media sosial X merupakan salah satu media sosial yang populer di berbagai kalangan masyarakat di seluruh dunia. Dilansir pada laman *databooks*, diketahui bahwa Indonesia menempati peringkat ke-4 sebagai negara dengan pengguna X terbanyak di dunia.

10 Negara dengan Jumlah Pengguna Twitter Terbanyak di Dunia (Juli 2023)



Gambar 1.1 10 Negara dengan Pengguna X Terbanyak di Dunia (Juli 2023)

Sumber: databoks.katadata.co.id

Pada Gambar 1.1, terlihat bahwa setidaknya 25,25 juta pengguna media sosial X merupakan Warga Negara Indonesia. Pada media sosial tersebut, masyarakat bebas mengunggah cuitan tentang hal apapun, berbincang dengan pengguna X lainnya atau mengutarakan tanggapan tentang suatu hal seperti pemilu. Melalui fitur seperti tagar (*hashtag*), pengguna dapat mengikuti tren, menyuarakan pendapat, dan menyebarkan informasi dengan cepat. Cuitan-cuitan pengguna X tersebut kemudian dapat memunculkan *trending topic*.

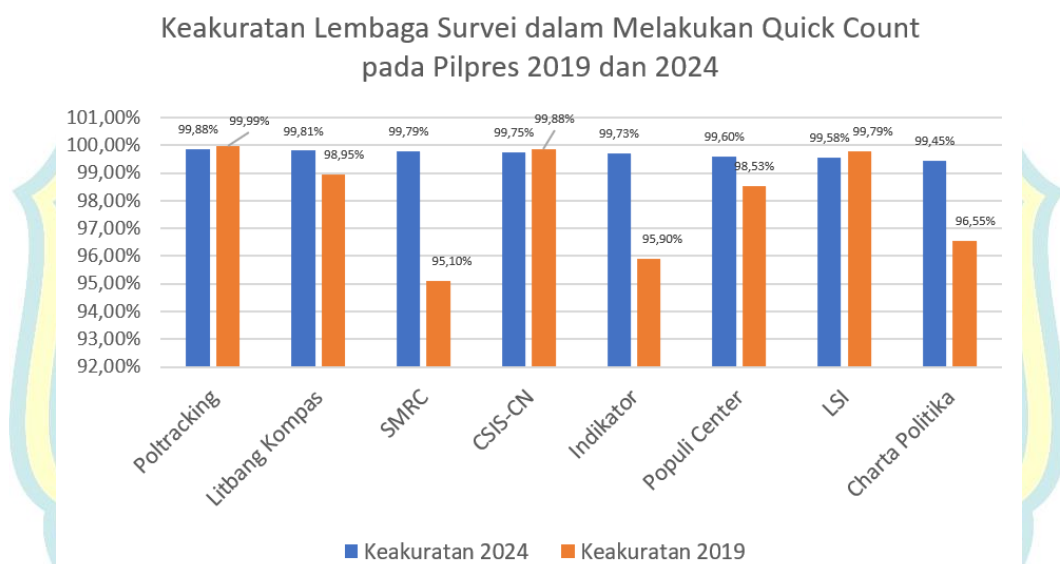
Menurut kamus internet *Zorraquino.com*, *trending topic* adalah topik yang mendominasi percakapan yang terjadi di media sosial pada waktu tertentu dan berbentuk frasa atau kata yang diawali dengan tagar. Selain itu, istilah ini juga digunakan untuk merujuk pada sebuah isu yang sedang menjadi perhatian utama di media sosial. Istilah *trending topic* pertama kali muncul di X, yang mencakup 10 topik dengan tren terbesar yang diletakan pada halaman pencarian media sosial tersebut, memberikan pengguna kesempatan untuk mempersonalisasikannya sesuai dengan lokasi mereka dan orang-orang yang mereka ikuti. *Trending topic* akan mendapatkan lebih banyak perhatian dari pengguna media sosial ketika terdapat suatu kejadian atau *event* besar baik itu terjadi di suatu negara ataupun seluruh dunia, salah satu contohnya adalah pelaksanaan pemilu di Indonesia yang dilaksanakan pada tahun 2024.

Salah satu tanggapan masyarakat terkait pelaksanaan Pemilu 2024 yang banyak dibicarakan di X adalah tentang hasil hitung cepat atau *quick count*. “*Quick Count*” menjadi *trending topic* di media sosial X pada hari pencoblosan Pemilu yaitu pada tanggal 14 Februari 2024. Menurut Pasal 1 angka 22 Peraturan KPU 9/2022 penghitungan cepat adalah kegiatan penghitungan suara hasil pemilu atau pemilihan secara cepat dengan menggunakan teknologi informasi atau berdasarkan metodologi tertentu. Perhitungan cepat dilaksanakan oleh berbagai lembaga survei dengan hasil yang berbeda-beda. Setidaknya terdapat 81 lembaga survei di Indonesia yang ditetapkan oleh KPU dan terdaftar atau bersertifikat untuk melakukan perhitungan cepat pemilu 2024. Selain *quick count* dari lembaga survei, masyarakat juga banyak membicarakan perhitungan dari KPU yang menampilkan angka *real count* pada laman Sirekap.

Hasil hitung cepat pemilu yang paling banyak diperbincangkan pada media sosial X adalah hasil hitung cepat pemilihan presiden. Fenomena perbincangan tentang pemilihan presiden (pilpres) yang lebih banyak dibandingkan pemilihan legislatif (pileg) ini tidak terjadi pada media sosial X saja, tetapi juga terjadi di media sosial lainnya dan di lingkungan sekitar. Hal ini terjadi karena pilpres memiliki lebih sedikit calon dibandingkan pileg, yaitu hanya 3 pasangan calon presiden pada pemilu tahun ini. Pasangan calon presiden tersebut juga tentunya menyeluruh di seluruh daerah Indonesia. Sedangkan pileg memiliki lebih banyak calon, dan berbeda-beda di setiap daerah sehingga masyarakat cenderung tidak tertarik untuk mencari tau tentang calon legislatif yang jumlahnya terlalu banyak (Rahadian, 2019)

Keakuratan hasil *quick count* menunjukkan tingkat kedekatan hasil perhitungan cepat yang dilakukan lembaga survei dengan hasil rekapitulasi resmi yang ditetapkan oleh KPU. Secara umum, tingkat keakuratan diperoleh dengan membandingkan hasil persentase *quick count* terhadap hasil resmi KPU, semakin kecil selisih atau *error* maka semakin tinggi tingkat keakuratannya. Oleh karena itu, lembaga survei dengan nilai keakuratan yang tinggi cenderung memiliki tingkat kepercayaan masyarakat yang lebih baik. Perbandingan hasil *quick count* pilpres 2024 terbaik dari lembaga survei dengan rekapitulasi KPU telah dirangkum oleh *csis.or.id*. Lembaga survei tersebut diantaranya, yaitu Poltracking, Litbang Kompas,

SMRC, CSIS, Indikator Politik Indonesia, Populi Center, LSI, Charta Politika, PRC, dan Voxpol Center. Dilansir dari *CNNIndonesia.com* diketahui bahwa lembaga survei Poltracking juga menempati posisi pertama sebagai lembaga survei paling akurat dalam perhitungan cepat pilpres tahun 2019, diikuti CSIS dan LSI atau Lembaga Survei Indonesia. Pada Gambar 1.2 terlihat bahwa Poltracking memiliki hasil *quick count* paling akurat dan stabil dengan keakuratan sebesar 99,99% pada tahun 2019 dan 99,88% pada tahun 2024, diikuti oleh CSIS dengan keakuratan sebesar 99,88% pada tahun 2019 dan 99,75% pada tahun 2024 dan Lembaga Survei Indonesia dengan keakuratan sebesar 99,79% pada tahun 2019 dan 99,58% pada tahun 2024.



Gambar 1.2 Keakuratan Lembaga Survei dalam Melakukan *Quick Count* pada Pilpres 2019 dan 2024

Sumber: *csis.or.id* dan *cnnindonesia.com*

Hasil hitung cepat pemilihan presiden yang dilakukan oleh lembaga survei menuai berbagai tanggapan masyarakat di X dari tanggapan positif hingga negatif yang menarik untuk diteliti karena dapat dilihat bagaimana masyarakat menilai adanya *quick count* yang dilakukan oleh banyaknya lembaga survei. Tanggapan-tanggapan yang dikirimkan pada media sosial X merupakan tanggapan berbentuk teks sehingga untuk analisisnya digunakan metode *text mining* dan analisis sentimen. *Text mining* adalah proses ekstraksi informasi dari data teks melalui tahapan seperti pembersihan data (*preprocessing*), *feature selection*, *text*

representation, application of text mining text, dan visualisation of the result sehingga data tidak terstruktur dapat diolah menjadi informasi yang bermakna (Feldman & Sanger, 2007). Sementara analisis sentimen adalah langkah untuk memahami dan mengevaluasi sentimen yang terkandung dalam teks secara otomatis (Pang & Lee, 2008).

Sejumlah penelitian terkait analisis sentimen telah dilakukan pada berbagai cakupan masalah contohnya seperti kasus Omnibus Law Cipta Kerja (Pratama dkk., 2021), pandemi Covid-19 (Pamungkas & Kharisudin, 2020), perpanjangan PPKM atau Pemberlakuan Pembatasan Kegiatan Masyarakat (Februariyanti, 2022), dan RRU KUHP (Welmin dkk., 2020). Hasil penelitian tersebut menunjukkan bahwa analisis sentimen dapat diterapkan pada berbagai domain untuk mengidentifikasi opini masyarakat terhadap suatu isu.

Salah satu tahapan penting dalam analisis sentimen adalah proses klasifikasi, yaitu mengelompokkan data ke dalam kategori sentimen tertentu. Umumnya, klasifikasi dilakukan ke dalam tiga kategori, yaitu positif, negatif, dan netral. Koppel dan Schler (2006) dalam penelitiannya menyatakan bahwa penyertaan kategori netral dalam proses klasifikasi dapat meningkatkan kinerja model dibandingkan hanya menggunakan dua kategori sentimen (positif dan negatif). Hasil penelitian mereka menunjukkan bahwa penggunaan tiga kategori menghasilkan nilai akurasi yang lebih tinggi dibandingkan dua kategori pada beberapa kasus pengujian. Oleh karena itu, penelitian ini menggunakan tiga kategori sentimen, yaitu positif, negatif, dan netral.

Dalam proses analisis sentimen, data latih memerlukan label sentimen sebagai acuan pada tahap klasifikasi. Salah satu pendekatan yang dapat digunakan untuk *labelling* secara otomatis adalah *lexicon-based*, yaitu pendekatan yang memanfaatkan kamus sentimen (*sentiment lexicon*) yang berisi daftar kata beserta nilai polaritasnya. Pada penelitian ini, pelabelan dilakukan menggunakan InSet (*Indonesian Sentiment Lexicon*), yaitu kamus sentimen berbahasa Indonesia untuk mengidentifikasi polaritas kata menjadi sentimen positif, negatif dan netral. Penggunaan InSet memungkinkan proses *labelling* otomatis sehingga dapat mengurangi subjektivitas dalam pelabelan manual serta mempermudah pengolahan data dalam jumlah besar (Koto & Rahmaningtyas, 2017).

Klasifikasi pada media sosial X dapat dilakukan dengan menggunakan berbagai algoritma seperti *Support Vector Machine* (SVM), *Naïve Bayes*, *K-Nearest Neighbor*, atau yang lainnya. *Naïve Bayes* adalah metode klasifikasi yang mengasumsikan setiap atribut dalam suatu kelas tidak saling memengaruhi satu sama lain (Junianto dkk., 2024). *K-Nearest Neighbor* (K-NN) adalah algoritma pembelajaran mesin yang digunakan untuk klasifikasi atau regresi, di mana data diklasifikasikan berdasarkan kedekatannya dengan K data terdekat yang sudah diberi label (Cover & Hart, 1952). Sedangkan *Support Vector Machine* (SVM) adalah teknik klasifikasi atau algoritma *machine learning* yang menggunakan fungsi kernel untuk memetakan ruang titik data ke ruang baru dengan minimnya klasifikasi *error*. Pada ruang titik data, data tidak dapat dipisahkan secara *linear* sedangkan pada ruang baru, data tersebut dapat dipisahkan secara *linear* (Mullen & Collier, 2004). *Support Vector Machine* juga dapat diartikan sebagai algoritma *machine learning* yang dirancang untuk melakukan klasifikasi dan regresi, dengan memanfaatkan kerangka kerja matematika yang kuat untuk *supervised learning*. Algoritma ini sangat kuat dalam menyelesaikan klasifikasi biner, tetapi juga dapat menangani klasifikasi multi-kelas (Wang, 2023).

Menurut Pamungkas dan Kharisudin (2020), metode *Support Vector Machine* merupakan metode yang paling cocok untuk mengklasifikasikan data tanggapan masyarakat di media sosial X terkait pandemi Covid-19 dengan akurasi sebesar 90,1% dibandingkan dengan metode *Naïve Bayes* dan *K-Nearest Neighbor* yang hanya memperoleh akurasi sebesar 79,2% dan 62,1%. Selain itu, pada penelitian yang dilakukan Rosyida, dkk. (2023) juga disimpulkan bahwa metode *Support Vector Machine* memiliki nilai akurasi yang lebih tinggi dibandingkan dengan metode *Naïve Bayes* dalam mengklasifikasi data sentimen masyarakat di media sosial X terhadap pemilihan presiden 2024. Hal ini juga diperkuat dengan hasil dari penelitian Sharma dan Dey (2012) yang juga menyimpulkan bahwa SVM memberikan performa klasifikasi yang lebih baik dibandingkan beberapa algoritma *machine learning* lainnya pada analisis sentimen. Oleh karena itu, metode SVM dipilih pada penelitian ini sebagai algoritma klasifikasi.

Sebelum proses klasifikasi menggunakan SVM, data teks perlu diubah ke dalam bentuk numerik melalui proses pembobotan kata. Salah satu metode

pembobotan yang umum digunakan adalah *Term Frequency-Inverse Document Frequency* (TF-IDF). Metode ini memberikan bobot berdasarkan frekuensi kemunculan suatu kata pada sebuah dokumen serta tingkat keunikan kata tersebut dalam keseluruhan dokumen, sehingga kata-kata yang lebih representatif terhadap isi dokumen akan memiliki bobot yang lebih besar (Jumeilah, 2017). Penggunaan TF-IDF telah banyak diterapkan pada penelitian analisis sentimen karena mampu menghasilkan representasi fitur yang efektif untuk proses klasifikasi.

Berdasarkan pemaparan tersebut, peneliti tertarik untuk melakukan penelitian analisis sentimen masyarakat pada media sosial X terhadap lembaga survei pada pemilihan presiden dengan menggunakan algoritma *Support Vector Machine*. Pada penelitian ini akan dilakukan klasifikasi sentimen dimana kata yang tersusun pada *tweet* akan dilakukan pembobotan kata menggunakan metode TF-IDF sehingga memiliki skala pengukuran rasio sebagai variabel bebas (X) dan data kategorik dari sentimen masyarakat yang dikategorikan menjadi kategori positif, negatif, dan netral sebagai variabel terikat (Y). Penelitian ini dilakukan untuk mengetahui pendapat masyarakat Indonesia di media sosial X terhadap lembaga survei yang melakukan *quick count* pada pilpres dalam klasifikasi sentimen positif, netral dan negatif, serta untuk melihat tingkat keakuratan *Support Vector Machine* dalam mengklasifikasikan data tersebut.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dipaparkan, maka penelitian ini memiliki beberapa permasalahan yang dapat dirumuskan sebagai berikut:

1. Bagaimana performa *Support Vector Machine* dalam melakukan pengklasifikasian pada data sentimen X terkait lembaga survei pada pemilihan presiden?
2. Bagaimana gambaran umum sentimen masyarakat terhadap lembaga survei pada pemilihan presiden di media sosial X?

1.3 Batasan Masalah

Batasan masalah penelitian ini adalah sebagai berikut:

1. Data yang digunakan dalam penelitian ini adalah data *tweet* atau sentimen dari masyarakat Indonesia di media sosial X terhadap lembaga survei pada pemilihan presiden dengan rentang waktu 14 Februari - 20

Maret 2024 serta akan diklasifikasikan dalam tiga kategori yaitu positif, negatif, dan netral.

2. Data dikumpulkan dengan metode *crawling text*. *Crawling text* dilakukan langsung pada media sosial X dengan menggunakan *tools* pada python yaitu *Tweet Harvest* terhadap pencarian nama-nama lembaga survei terpilih dengan Bahasa Indonesia dan rentang waktu yang sudah ditentukan.
3. Lembaga survei terpilih dibatasi menjadi 3 nama lembaga berdasarkan tingkat kestabilan dan keakuratan hasil *quick count* yang dilakukan lembaga pada pilpres tahun 2019 dan 2024, yaitu Poltracking, CSIS dan Lembaga Survei Indonesia atau LSI.
4. Pada penelitian ini akan digunakan metode SVM dengan pendekatan One-Against-All atau OAA dan menggunakan fungsi *Radial Basis Function* (RBF) dan *Linear Kernel* untuk kernelitas yang parameternya akan dipilih dengan menggunakan *Grid Search*.

1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah sebagai berikut:

1. Mengetahui seberapa baik performa *Support Vector Machine* dalam melakukan pengklasifikasian data sentimen X terkait lembaga survei pada pemilihan presiden.
2. Mengetahui gambaran umum sentimen masyarakat terhadap lembaga survei pada pemilihan presiden di media sosial X.

1.5 Manfaat Penelitian

Diharapkan penelitian ini mampu memberikan manfaat diantaranya:

1. Memberikan pemahaman bagaimana dan seberapa baik performa *Support Vector Machine* dalam melakukan pengklasifikasian terhadap pendapat atau opini masyarakat tentang lembaga survei pada pemilihan presiden di media sosial X, sehingga dapat dijadikan rujukan atau acuan pada penelitian selanjutnya.
2. Menjadi referensi bagi masyarakat di kemudian hari untuk dapat mengetahui tanggapan masyarakat Indonesia pengguna media sosial X terhadap lembaga survei yang ada.