

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Perkembangan pengguna internet jumlahnya semakin banyak di Indonesia, menurut data dari internetworldstats.com memuat bahwa jumlah penduduk di Indonesia sebesar 278.820.685 jiwa, di mana jumlah pengguna internetnya mencapai 212.354.070 pengguna atau sekitar 76,3% dari populasi Indonesia per bulan desember tahun 2021 [1]. Dengan menggunakan internet Pengguna dapat dengan mudah dan cepat dalam berkomunikasi, mengakses berbagai informasi, memudahkan dalam berbisnis, hingga mengakses hiburan. Terdapat fakta bahwa pengguna internet di Indonesia lebih banyak menghabiskan waktu dalam mengakses media sosial, hal ini berdasarkan survei pada periode 2019-kuartal II/2020 yang dilakukan oleh Asosiasi Penyelenggara Jasa Internet Indonesia (AJII) menyatakan bahwa sebanyak 51,5% responden menggunakan jaringan internet untuk media sosial [2].

Media sosial yang paling banyak digunakan di Indonesia adalah Twitter, di mana pengguna Twitter di Indonesia mencapai 15,7 Juta per juli 2021 sehingga menempatkan Indonesia berada diperingkat 6 besar dunia pengguna Twitter [3]. Twitter merupakan media sosial di mana dengan *register* dan *login* akun ke dalam Twitter pengguna dapat lebih mudah dan cepat dalam mengakses informasi dikarenakan terdapat fitur *search* dan *trending topic*. Topik yang terdapat pada Twitter salah satunya adalah Formula E Jakarta mendapatkan berbagai tanggapan dari pengguna Twitter. Ada yang mendukung pergelaran Formula E Jakarta dan ada juga pihak yang kontra, serta terdapat juga pihak yang tidak suka polemik dari kubu pro dan kontra yang hanya ingin melihatnya sebagai kejuaraan otomotif [4].

Formula E merupakan sebuah kejuaraan balap mobil listrik yang masuk ke dalam olahraga bidang otomotif, mobil listrik tersebut dikemudikan oleh satu orang dan saat ini tahun 2022 diikuti oleh 22 pembalap dalam 11 tim. Ide seri balapan yang serba listrik bermula pada tanggal 3 maret 2011, yang digagas oleh presiden FIA (*Federasi Internationale de l'Automobile*) Jean Todt dan pengusaha Spanyol Alejandro Agag yang dalam pemikiran mereka akan membuat kejuaraan balapan mobil internasional yang serba listrik pertama di dunia. Misi pendiri Formula E adalah untuk mengadakan perlombaan balap mobil listrik melalui jalan-jalan perkotaan yang paling ikonik di dunia dengan pembalap dan tim terbaik yang akan berlomba dan dengan kendaraan listrik diharapkan kendaraan kedepan lebih baik dan lebih bersih. Formula E memulai balapan pertamanya pada tahun 2014 yang diadakan di

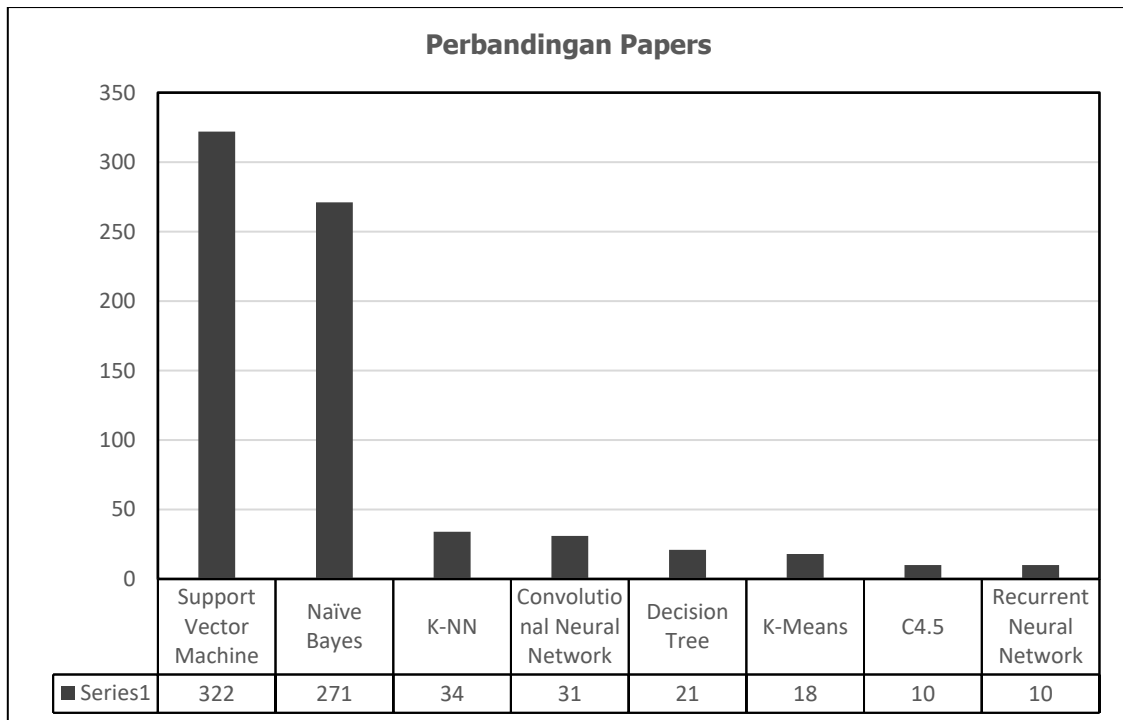
Olympic Park, Beijing dan untuk seri Indonesia digelar di Jakarta pada tahun 2022 di mana pada tahun tersebut kejuaraan Formula E ini menjadi balapan mobil listrik internasional pertama di Indonesia [5].

Dalam penyelenggaraan Formula E di Jakarta terdapat berbagai polemik mulai dari anggaran penyelenggaraan, lokasi balapan, proses pembuatan sirkuit, hingga kebermanfaatan bagi masyarakat. Polemik mengenai anggaran muncul pada tahun 2019-2020 di mana merujuk hasil audit Laporan Keuangan Pemerintah Daerah (LKPD) DKI Jakarta yang dilakukan oleh Badan Pemeriksa Keuangan (BPK) menyebutkan bahwa Pemprov DKI Jakarta menggelontorkan dana sebesar Rp 983,31 miliar, dengan rincian: *fee* yang telah dibayarkan pada tahun 2019 sebesar Rp 360 miliar, selanjutnya *fee* tahun 2020 sebesar Rp 200,31 miliar, dan bank garansi sebesar 423 miliar. Balapan Formula E pertama awalnya akan diagendakan pada juni 2020, namun pada tahun tersebut terjadi pandemi Covid 19 sehingga Pemrov DKI Jakarta menunda penyelenggaraan Formula E musim pertama. Atas penundaan tersebut maka dilakukan renegotiasi kepada pihak FEO Ltd selaku promotor dan pemegang lisensi Formula E yang menghasilkan persetujuan terkait penarikan bank garansi senilai Rp 423 miliar. Dalam dokumen resmi pemrov DKI Jakarta menyebutkan bahwa *commitment fee* Formula E berasal dari APBD tahun 2019 di mana pembiayaan tersebut dilakukan sebelum pandemi Covid 19 dengan total senilai Rp 560 miliar untuk balapan Formula E yang akan dilaksanakan pada tahun 2022, 2023, dan 2024. Dan dalam pelaksanaan Formula E nantinya pada tiap tahunnya akan mendapatkan pembiayaan dari *sponsorship* yakni PT Jakarta Propertindo (JakPro) senilai Rp 150 miliar [6]. Polemik selanjutnya mengenai tempat balapan diselenggarakan yang awalnya diagendakan dilaksanakan di area Monas sebagai sirkuit Formula E, namun hal tersebut tidak direstui oleh Sekretariat Negara dengan alasan bahwa area Monas termasuk kawasan ring 1 dan khawatir merusak bangunan cagar budaya, sehingga penyelenggara lokal mencari lokasi alternatif. Terdapat 5 lokasi alternatif yakni: Kawasan Indah Kapuk (PIK), Jalan Jenderal Soedirman, Kawasan Jakarta International Stadium, JIEXPO Kemayoran, dan kawasan Ancol. Lokasi alternatif tersebut kemudian dilakukan survei pihak penyelenggara Formula E dan memutuskan bahwa kawasan Ancol dipilih menjadi lokasi sirkuit formula E [7].

Beberapa polemik di atas ramai dibincangkan di media sosial Twitter sehingga dalam penelitian ini akan melakukan analisis sentimen Twitter pada topik Formula E Jakarta dengan menggunakan beberapa algoritma klasifikasi, yakni: algoritma *Support Vector Machine*, *Naïve Bayes*, dan *KNN*. Penelitian dengan perbandingan 3 algoritma pernah dilakukan oleh

Rahmatullah et al, (2021) dengan melakukan analisis sentimen Twitter tentang transportasi online, penelitian yang dilakukan tersebut menggunakan algoritma *Support Vector Machine*, *Naïve Bayes*, dan *KNN*. Hasil dari penelitian menunjukkan bahwa algoritma *Support Vector Machine* mendapatkan tingkat nilai *accuracy* tertinggi mencapai 98% dan AUC 0.988 [8]. Topik penelitian yang dilakukan tersebut terdapat keterbatasan karena tidak mengenai isu sensitif yang berkembang di masyarakat, isu yang dijadikan topik hanya sekedar sentimen terhadap jasa transportasi online sedangkan topik dalam penelitian yang akan dilakukan mengenai analisis sentimen Formula E Jakarta merupakan isu sensitif yang dibahas di Twitter yang dikaitkan dengan isu politik di Indonesia [9].

Penelitian ini akan melakukan perbandingan *accuracy* pada algoritma klasifikasi *Support Vector Machine*, *Naïve Bayes*, dan *KNN* dalam melakukan analisis sentimen dengan cara mengambil data dari Twitter dengan kata kunci “Formula E Jakarta”, kemudian dilakukan proses klasifikasi data menjadi 3 kategori sentimen, yaitu: sentimen positif, netral, dan negatif. Hasil akhir penelitian nantinya akan mengetahui tingkat *accuracy*, *precision*, dan *recall* dari tiga algoritma klasifikasi tersebut. Peneliti memilih algoritma klasifikasi *Support Vector Machine*, *Naïve Bayes*, dan *KNN* karena algoritma tersebut paling banyak digunakan dalam penelitian analisis sentimen di Indonesia. Data tersebut diperoleh peneliti berdasarkan *search* pada *Google Scholar* menggunakan aplikasi *Harzing’s Publish or Perish* pada bulan maret 2022 dengan kata kunci algoritma klasifikasi dalam analisis sentimen, seperti: “Support Vector Machine sentimen”, “Naïve Bayes sentimen”, “KNN sentimen”, “Convolutional Neural Network sentimen”, “Decision Tree sentimen”, “K-Means sentimen”, “C4.5 sentimen”, dan “Recurrent Neural Network sentimen”. Dan berikut grafik jumlah perbandingan seluruh *papers* algoritma klasifikasi sentimen dalam waktu 5 tahun terakhir (2017-2022):



Gambar 1.1 Perbandingan Jumlah Papers Algoritma Sentimen Terindeks Google Scholar

1.2 Perumusan Masalah

1.2.1 Permasalahan

Formula E merupakan sebuah kejuaraan balap mobil listrik yang masuk ke dalam olahraga bidang otomotif, Formula E ini diselenggarakan di Jakarta di mana gelaran tersebut merupakan kejuaraan balap mobil listrik internasional pertama kali di Indonesia. Dalam penyelenggaraan Formula E di Jakarta terdapat dukungan hingga terdapat polemik yang timbul di masyarakat, berbagai polemik yang dibicarakan oleh masyarakat melalui media sosial Twitter mulai dari anggaran yang dikeluarkan untuk gelaran Formula E, selanjutnya mengenai lokasi penyelenggaraan yang awalnya direncanakan di Kawasan Monas namun dipindah di Kawasan Ancol, hingga kebermanfaatan gelaran Formula E Jakarta bagi masyarakat.

Polemik yang timbul di atas perlu dilakukan analisis sentimen dengan mengkategorikan menjadi tiga kelas sentimen, yakni: sentimen positif, netral, dan negatif. Dalam melakukan analisis sentimen akan menggunakan tiga algoritma klasifikasi, yakni: *Support Vector Machine*, *Naïve Bayes*, dan *KNN*.

1.2.2 Metode Usulan

Dalam penelitian ini mengusulkan analisis sentimen mengenai topik Formula E Jakarta di media sosial Twitter. Sentimen akan dikategorikan menjadi tiga kelas, yaitu: sentimen positif, netral, dan negatif. Analisis sentimen akan menggunakan algoritma klasifikasi *Support Vector Machine*, *Naïve Bayes*, dan *KNN*. Model klasifikasi pada tiap-tiap algoritma akan dilakukan evaluasi dengan menggunakan *10 Fold-Cross Validation* sehingga akan diperoleh nilai tingkat *accuracy*, *precision*, dan *recall* dari masing-masing algoritma klasifikasi.

1.2.3 Pertanyaan Penelitian

Berdasarkan latar belakang yang sudah dipaparkan sebelumnya, maka terdapat rumusan masalah yang diangkat dalam penelitian, yakni: berapa persen tingkat *accuracy*, *precision*, dan *recall* yang diperoleh dari analisis sentimen Formula E Jakarta pada Twitter dengan menggunakan algoritma klasifikasi *Support Vector Machine*, *Naïve Bayes*, dan *KNN* serta apa algoritma klasifikasi terbaik yang mempunyai tingkat *accuracy* tertinggi dalam penelitian ini?

1.3 Tujuan dan Manfaat

1.3.1 Tujuan

Tujuan yang diperoleh dalam penelitian ini adalah mengetahui tingkat *accuracy* dari algoritma klasifikasi *Support Vector Machine*, *Naïve Bayes*, dan *KNN* dalam analisis sentimen Formula E Jakarta pada Twitter.

1.3.2 Manfaat

Manfaat yang diperoleh dalam penelitian analisis sentimen, yaitu:

1. Memberikan pengetahuan mengenai analisis sentimen Formula E Jakarta dengan menggunakan algoritma klasifikasi *Support Vector Machine*, *Naïve Bayes*, dan *KNN*.
2. Melihat dan melakukan analisis sentimen secara cepat ke dalam kelas positif, netral, dan negatif.
3. Menjadi referensi penelitian mendatang mengenai analisis sentimen menggunakan algoritma klasifikasi *Support Vector Machine*, *Naïve Bayes*, dan *KNN*.

1.4 Batasan Masalah

Terdapat batasan masalah agar penelitian dalam tugas akhir ini tidak keluar dari pokok permasalahan maka ruang lingkup penelitian dibatasi sebagai berikut:

1. Sumber data analisis sentimen berasal dari media sosial Twitter dengan kata kunci “formula e jakarta”, data diambil pada saat balapan Formula E Jakarta berlangsung yakni pada tanggal 4 Juni 2022 pada *tweet* yang menggunakan Bahasa Indonesia.
2. Analisis sentimen menggunakan algoritma klasifikasi *Support Vector Machine*, *Naïve Bayes*, dan *KNN* dengan menggunakan aplikasi *RapidMiner*.
3. Analisis sentimen mencakup kelas positif, netral, dan negatif serta perbandingan tingkat *accuracy* dari algoritma klasifikasi *Support Vector Machine*, *Naïve Bayes*, dan *KNN*.

1.5 Sistematika Penulisan

Penulisan skripsi ini disusun secara sistematis dan dibagi menjadi beberapa BAB, seperti berikut:

1. BAB I Pendahuluan

Pada BAB ini berisi latar belakang, rumusan masalah, tujuan, manfaat, batasan masalah, serta sistematika penulisan

2. BAB II Tinjauan Pustaka

Pada BAB ini membahas mengenai landasan teori yang berkaitan dengan penelitian.

3. BAB III Metodologi Penelitian

Pada BAB ini menguraikan mengenai waktu dan tempat penelitian, serta berisi prosedur dalam melakukan analisis sentimen.

4. BAB IV Hasil Penelitian dan Analisis

Pada BAB ini mengimplementasikan dan menjelaskan prosedur yang dilakukan dalam analisis sentimen serta berisi hasil dari tingkat *accuracy* algoritma klasifikasi *Support Vector Machine*, *Naïve Bayes*, dan *KNN*.

3. BAB V Penutup

Pada BAB ini berisi kesimpulan dan saran dari hasil penelitian.