**PROPOSAL TA**

A picture containing text, clipart

Description automatically generated ANALISIS SENTIMEN PUBLIK TWITTER MENGENAI PIALA DUNIA 2022 DI QATAR MENGGUNAKAN *NAIVES BAYES*

**Disusun Oleh:**

|  |  |  |
| --- | --- | --- |
| **Nama** | **:** | **Muhammad Reza** |
| **Nim** | **:** | **2019470055** |

**PROGRAM STUDI TEKNIK INFORMATIKA**

**FAKULTAS TEKNIK**

**UNIVERSITAS MUHAMMADIYAH JAKARTA**

**2023**

**ABSTRACT**

**ABSTRAK**

**KATA PENGANTAR**

*Alhamdulillaahirabbil‟aalamiin*, puji syukur penyusun panjatkan atas kehadirat Allah SWT yang telah melimpahkan rahmat dan hidayah-Nya, serta tidak lupa shalawat serta salam selalu tercurah kepada junjungan umat, yaitu Nabi Muhammad SAW sebagai suri tauladan umat, sehingga penyusunan tugas akhir yang berjudul “Analisis sentimen publik Twitter Mengenai Piala Dunia 2022 Di Qatar menggunakan *Naives Bayes*” sebagai syarat untuk kelulusan jenjang strata satu di Jurusan Teknik Informatika, Fakultas Teknik, Universitas Muhammadiyah Jakarta. Dalam penyusunan proposal tugas akhir penyusun banyak memperoleh petunjuk dan bimbingan dari berbagai pihak.

Untuk selanjutnya penyusun mengucapkan banyak terima kasih kepada pihak-pihak yang telah membantu dalam penyelesaianTA ini, yaitu:

1. Dekan Fakultas Teknik Bapak Irfan Purnawan, S.T., M.Chem.Eng.
2. Ketua Program Studi Teknik Informatika Ibu Popy Meilina, S.T., M. Kom
3. Dosen pembimbing Ibu Popy Meilina, S.T., M. Kom
4. Kedua orang tua penyusun yang selalu memberikan do’a dan motivasi

**DAFTAR ISI**

**DAFTAR TABEL**

**DAFTAR GAMBAR**

**DAFTAR LAMPIRAN**

**BAB I**

**PENDAHULUAN**

* 1. **Latar Belakang Masalah**

Piala dunia merupakan pergelaran turnamen sepakbola yang dilaksanakan 4 tahun sekali, setiap negara yang masuk ke pergelaran piala dunia harus lolos di babak kualifikasi, sampai ke tahap undian grup untuk bertanding melawan negara lain (Andersson et al., 2021).

Piala dunia pada tahun 2022 diadakan di negara Qatar, menurut Fifa pergelaran *event* piala dunia dilaksanakan 20 November 2022 s/d 18 Desember 2022. Pergelaran acara piala dunia ini tidak lepas dari opini publik netizen yang memberikan pendapat secara positif atau secara negatif .

Perkembangan teknologi yang pesat di zaman modern seperti ini mengubah kebiasaan manusia sebagai contoh membeli berbagai macam barang menggunakan toko *online* atau *electronic commerce* karena kemudahan serta praktis, banyaknya promosi yang menimbulkan minat beli barang di toko *online* menjadi konsumtif (Ricky et al., 2021).

Pembelian menggunakan toko *online* yang menjadi kebiasaan saat ini, akan tetapi calon konsumen membeli barang melihat dari fitur toko *online* yaitu *review* atau penilaian barang yang dijual oleh pihak toko *online* oleh konsumen yang sudah membeli barang ditoko itu apakah barang tersebut *positive* (bagus) atau *negative* (kurang bagus) (Kabiru & Sari, 2019).

Penilaian dari konsumen sebelumnya yang memberikan *review* barang yang sudah dibeli berdasarkan rating agar dapat menjadi parameter calon konsumen untuk selektif dalam membeli barang di toko *online* (Hariyanto & Trisunarno, 2021)*.* *Review* atau penilaian merupakan keadaan konsumen memberikan pendapat, kesan tentang barang atau produk yang dibeli (Harli et al., 2021).

Analisa terhadap penilaian toko *online* diperlukan, maka ilmu untuk mengolah data abstrak atau data teks diperlukan, ilmu untuk mengolah data teks adalah *Natural processing language,* karena *Natural processing language* merupakan cabang ilmu kecerdasan buatan yang dapat berinteraksi antara mesin dengan bahasa manusia (Nofiyanti & Oki Nur Haryanto, 2021).

Berdasarkan penjelasan *Natural processing language* yang sudah dijelaskan, didalam  *Natural processing language* ada metode analisis sentimen, analisis sentimen merupakan menambang opini atau analisa opini publik (Wankhade et al., 2022). Maka dari itu sangat tepat jika menggunakan analisis sentimen untuk menentukan positif atau negatif berdasarkan kata.

Penelitian terhadap analisis sentimen dalam ilmu *Natural processing language* menggunakan *machine learning* seperti melakukan analisis sentimen pada ulasan restoran menggunakan *Naives Bayes* sebagai *machine learning,* dengan memodelkan 1000 data, dilakukan pembagian kategori 1(positif) atau 0(negatif), dan menghasilkan hasil akurasi sebesar 77,5% (Reddy & Reddy, 2021).

Penelitian analisis sentimen menggunakan *decision tree* sebagai *machine learning* dengan fokus masalah ulasan mengenai hotel, kemudian dilakukan pengumpulan data, *preprocessing data,* pemodelan, sampai evaluasi akurasi menghasilkan sebesar 88,45% (Apriliani et al., 2020).

Penelitian berikutnya menggunakan *K - Nearest Neighbor* sebagai *machine learning* dengan fokus permasalahan analisis sentimen pengguna Twitter tentang topik Pilkada DKI (Daerah Khusus Ibukota) tahun 2017, melakukan pengambilan data dengan teknik *scraping*, kemudian melakukan *processing data*, menghitung bobot kata, kemudian pemodelan *K - Nearest Neighbor* menghasilkan hasil akurasi sebesar 67,2% dengan nilai K=5 (Deviyanto & Wahyudi, 2018).

Penyusun pada penelitian ini menganalisa sentimen di *e-commerce* Tokopedia, Tokopedia merupakan *e-commerce* yang berkembang dengan pengguna sebesar 153,46 juta (Handayani, 2021). Penelitian ini melakukan analisis sentimen *review* konsumen Tokopedia apakah penilaian konsumen bersifat positif atau negatif dengan komparasi *machine learning* untuk menentukan *machine learning* yang baik berdasarkan hasil akurasi.

* 1. **Identifikasi Masalah**

Berdasarkan permasalahan di latar belakang, permasalahan tersebut dilakukan. Analisa opini publik bersifat abstrak karena dalam mengolahnya berbentuk teks dengan rating, dan memerlukan ilmu untuk mengolah hal tersebut.

* 1. **Perumusan Masalah**

Berdasarkan permasalahan diatas, akan dilakukan perumusan atau kajian

sebagai berikut:

1. Bagaimana cara mengolah data teks untuk melakukan analisis sentimen, melakukan analisis sentimen dengan menghasilkan positif, dan nilai negatif, serta cara melakukan pelabelan sentimen berdasarkan rating bintang 1 sampai 5?
2. Bagaimana cara melakukan komparasi *machine learning* untuk analisis sentimen berdasarkan 2 kondisi positif dan negatif?
   1. **Batasan Masalah**

Proposal tugas akhir ini memiliki batasan agar lebih mengerucut lagi dan tidak

melebar, maka diberikan batasan-batasan sebagai berikut:

1. Mengolah data teks, yaitu penilaian atau ulasan konsumen berdasarkan rating apakah opini tersebut positif atau negatif.
2. Melakukan sentimen analisis ulasan produk, yaitu produk elektronik (*laptop*, *handphone*), produk pakaian(kemeja, kaos)
3. Melakukan komparasi *machine learning decision tree, naives bayes, K - Nearest Neighbor* untuk sentimen analisis ulasan konsumen Tokopedia di fitur ulasan dan *review*.
   1. **Tujuan dan Manfaat Penelitian**

Proposal tugas akhir ini memiliki tujuan penelitian, manfaat sebagai berikut:

1. Melakukan sentimen analisis atau klasifikasi opini konsumen di fitur ulasan *review* pembelian menggunakan algoritma *decision tree, naives bayes, K - Nearest Neighbor*
2. Membandingkan atau komparasi *machine learning* algoritma *decision tree, naives bayes, K - Nearest Neighbor* yang lebih baik berdasarkan tingkat akurasi.
3. Mengetahui bagaimana cara melakukan analisis sentimen dengan data teks ulasan *review* konsumen Tokopedia
4. Mengetahui bagaimana kata-kata yang mengandung positif atau negatif dari hasil analisis sentimen ulasan *review* konsumen.
   1. **Metodologi Penelitian**
5. **Data penelitian**

Data yang digunakan dalam proposal tugas akhir adalah data teks dari hasil *scrape* dari website Tokopedia yang data didalamnya terdapat konten berupa komentar dan rating berupa bintang 1 s/d 5, sedangkan data yang sudah didapat dari *scrape* dilakukan pelabelan berdasarkan rating bintang 1 s/d 5.

1. **Pengolahan data**

Pada proposal tugas akhir dilakukan proses agar dapat dilakukan ke pemodelan *machine learning decision tree, naives bayes, K-Nearest Neighbor*, dari *preprocessing text* dilakukan tahapan sebagai berikut:

1. *Case folding*

Merupakan tahap mengolah data teks jika memiliki huruf kapital atau *uppercase* makadiubah menjadi huruf kecil atau *lowercase* (KURNIAWAN & APRILIANI, 2020).

1. *Removal stopwords*

Merupakan tahap mengolah data teks untuk menghapus kata hubung seperti kata “atau”, ”dan” karena tersebut merupakan kata yang sering muncul dan tidak memiliki arti apapun (Pradana & Hayaty, 2019) (Deviyanto & Wahyudi, 2018).

1. *Stemming*

Merupakan tahap untuk mengurangi prefiks sebuah kata menjadi kata dasar (Pradana & Hayaty, 2019).

1. **Pembobotan kata**

Pada tahap ini setelah pengolahan data melakukan perhitungan kata dengan menggunakan metode *Term Frequency* *Inverse Document*, *Term Frequency* *Inverse Document* adalah metode perhitungan kata berdasarkan jumlah dokumen data dengan jumlah frekuensi kata yang muncul di setiap dokumen (Melita et al., 2018). *Term Frequency* *Inverse Document* mempunyai fungsi sebagai seleksi fitur untuk pemodelan *machine learning* klasifikasi (Prayoga et al., 2021).

1. **Pemodelan**

Pada tahap sesudah pembobotan kata melakukan pemodelan pada penelitian kali ini menggunakan *machine learning supervised learning*, *supervised learning* adalah pembelajaran dalam *machine learning* yang membutuhkan label untuk melakukan pelatihan (El Mohadab et al., 2019).

Pada tahap pemodelan melakukan komparasi berdasarkan nilai akurasi, antara *decision tree, naives bayes, k-nearest-neighbor* penjelasan mengenai ketiga *machine learning,* sebagai berikut:

1. *Decision tree*

*Decision* *tree* merupakan algoritma *supervised learning* yang bekerja seperti struktur pohon di setiap *node* atau simpul mewakili dari atribut yang dilatih (Panhalkar & Doye, 2022). *Decision tree* didalam klasifikasi memperlukan label berupa 2 variabel, setelah dilakukan pemilihan variabel dilakukan kalkulasi *informartion gain* untuk kalkulasi berapa informasi yang dihasilkan dari fitur *output* (fitur y) dan fitur *input* (fitur x) (Tangirala, 2020).

1. *Naïve Bayes*

*Naïve Bayes* merupakan algoritma klasifikasi probabilitas berdasarkan label data untuk memprediksi peluang masa depan dengan data sebelumnya (Watrianthos et al., 2019). *Naïve bayes* algoritma yang seringkali digunakan dalam sentimen analisis karena pembelajaran dari fitur untuk pengujian data untuk menghasilkan kemungkinan atau probabilitas (Watrianthos et al., 2019).

1. *K-Nearest-Neighbor*

Merupakan algoritma klasifikasi dengan menggunakan *input* fitur dan *output* fitur dengan melihat dari kelas atau fitur *Neighbor* terdekat (Cunningham & Delany, 2021). Menggunakan *K-Nearest-Neighbor* harus menentukan nilai K, dari bobot kata *Term Frequency* *Inverse Document* dikalkulasi untuk melihat kemiripan antar dokumen, kemudian dilakukan klasifikasi (Dwiki et al., 2021).

1. **Evaluasi**

Merupakan tahapan untuk mengukur bagaimana *machine learning* sudah melakukan tugas terhadap ketiga algoritma, yaitu *k-nearest-neighbor, decision tree, naïve bayes,* akan tetapi pada metode sentimen analisis yang sama seperti klasifikasi harus memperhatikan, sebagai berikut:

1. *False negative*

Merupakan hasil dari pengujian menggunakan data *testing* menunjukkan bahwa kelas yang harusnya prediksi benar tetapi prediksi salah (Romli et al., 2021).

1. *True negative*

Merupakan hasil dari pengujian menggunakan data testingmenunjukkan yang harusnya bernilai salah akan tetapi prediksi menunjukkan hasil salah (Syarifuddinn, 2020).

1. *False positive*

Merupakan hasil dari pengujian menggunakan data testingmenunjukkan bahwa kelas yang harusnya bernilai *negative* akan tetapi prediksi menghasilkan nilai *negative* (Syarifuddinn, 2020)*.*

1. *True positive*

Merupakan hasil dari pengujian menggunakan data testing menunjukkan bahwa kelas yang harusnya bernilai *positive* akan tetapi prediksi menghasilkan nilai *positive* (Syarifuddinn, 2020)*.*

Dari penjelasan mengenai *true positive, true negative, false positive, false negative* maka dilakukan perhitunagan *metrics* menggunakan 4 metode antara lain sebagai berikut:

1. Akurasi

Menghitung akurasi skor berdasarkan hasil prediksi dari data *testing*, dengan memperhatikan *true positive, true negative, false positive, false negative* (Romli et al., 2021). Berikut cara menghitung skor akurasi sebagai berikut:

1. *Recall*

Merupakan perhitungan dari hasil prediksi menggunakan data uji untuk menghasilkan skor nilai salah perhitungan *recall* dilakukan sebagi berikut:(Pintoko & Lhaksmana, 2018) (Romli et al., 2021).

1. *Precision*

Merupakan perhitungan dari hasil prediksi menggunakan data uji untuk mengukur prediksi nilai positif dari berapa banyak *true positive dengan false positive dilakukan* sebagai berikut: (Yun, 2021) (Romli et al., 2021).

1. *Confusion matrix*

Merupakan hasil dari evaluasi pemodelan *machine learning* yang berbentuk kotak, terdapat 2 kolom dan 2 baris yang didalamnya ada *false negative, true negative, true negative, false positive.* Berikut merupakan contoh *confusion matrix:* (Yun, 2021)(Hasnain et al., 2020)*.*

**DAFTAR PUSTAKA**

Andersson, S., Bengtsson, L., & Svensson, Å. (2021). Mega-sport football events’ influence on destination images: A study of the of 2016 UEFA European Football Championship in France, the 2018 FIFA World Cup in Russia, and the 2022 FIFA World Cup in Qatar. *Journal of Destination Marketing and Management*, *19*. https://doi.org/10.1016/j.jdmm.2020.100536

Apriliani, D., Abidin, T., Sutanta, E., Hamzah, A., & Somantri, O. (2020). Sentiment analysis for assessment of hotel services review using feature selection approach based-on decision tree. *International Journal of Advanced Computer Science and Applications*, *11*(4), 240–245. https://doi.org/10.14569/IJACSA.2020.0110432

Cunningham, P., & Delany, S. J. (2021). K-Nearest Neighbour Classifiers-A Tutorial. In *ACM Computing Surveys* (Vol. 54, Issue 6). Association for Computing Machinery. https://doi.org/10.1145/3459665

Deviyanto, A., & Wahyudi, M. D. R. (2018). PENERAPAN ANALISIS SENTIMEN PADA PENGGUNA TWITTER MENGGUNAKAN METODE K-NEAREST NEIGHBOR. *JISKA (Jurnal Informatika Sunan Kalijaga)*, *3*(1), 1. https://doi.org/10.14421/jiska.2018.31-01

Dwiki, A., Putra, A., & Juanita, S. (2021). Analisis Sentimen pada Ulasan pengguna Aplikasi Bibit Dan Bareksa dengan Algoritma KNN. *JATISI (Jurnal Teknik Informatika Dan Sistem Informasi)*, *8*(2), 636–646. https://doi.org/10.35957/JATISI.V8I2.962

El Mohadab, M., Bouikhalene, B., & Safi, S. (2019). Predicting rank for scientific research papers using supervised learning. *Applied Computing and Informatics*, *15*(2), 182–190. https://doi.org/10.1016/j.aci.2018.02.002

Handayani, R. N. (2021). Optimasi Algoritma Support Vector Machine untuk Analisis Sentimen pada Ulasan Produk Tokopedia Menggunakan PSO. *Media Informatika*, *20*(2), 97–108. https://doi.org/10.37595/MEDIAINFO.V20I2.59

Hariyanto, H. T., & Trisunarno, L. (2021). Analisis Pengaruh Online Customer Review, Online Customer Rating, dan Star Seller terhadap Kepercayaan Pelanggan Hingga Keputusan Pembelian pada Toko Online di Shopee. *Jurnal Teknik ITS*, *9*(2). https://doi.org/10.12962/j23373539.v9i2.56728

Harli, I. I., Mutasowifin, A., & Andrianto, M. S. (2021). Pengaruh Online Consumer Review dan Rating terhadap Minat Beli Produk Kesehatan pada E-Marketplace Shopee Selama Masa Pandemi COVID-19. *INOBIS: Jurnal Inovasi Bisnis Dan Manajemen Indonesia*, *4*(4). https://doi.org/10.31842/jurnalinobis.v4i4.205

Hasnain, M., Pasha, M. F., Ghani, I., Imran, M., Alzahrani, M. Y., & Budiarto, R. (2020). Evaluating Trust Prediction and Confusion Matrix Measures for Web Services Ranking. *IEEE Access*, *8*, 90847–90861. https://doi.org/10.1109/ACCESS.2020.2994222

Kabiru, I. N., & Sari, P. K. (2019). Analisa Konten Media Sosial E-commerce Pada Instagram Menggunakan Metode Sentiment Analysis Dan Lda-based Topic Modeling (studi Kasus: Shopee Indonesia). *EProceedings of Management*, *6*(1).

KURNIAWAN, R., & APRILIANI, A. (2020). ANALISIS SENTIMEN MASYARAKAT TERHADAP VIRUS CORONA BERDASARKAN OPINI DARI TWITTER BERBASIS WEB SCRAPER. *Jurnal INSTEK (Informatika Sains Dan Teknologi)*, *5*(1), 67. https://doi.org/10.24252/instek.v5i1.13686

Melita, R., Amrizal, V., Suseno, H. B., & Dirjam, T. (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

Nofiyanti, E., & Oki Nur Haryanto, E. M. (2021). Analisis Sentimen terhadap Penanggulangan Bencana di Indonesia. *Jurnal Ilmiah SINUS*, *19*(2), 17. https://doi.org/10.30646/sinus.v19i2.563

Panhalkar, A. R., & Doye, D. D. (2022). Optimization of decision trees using modified African buffalo algorithm. *Journal of King Saud University - Computer and Information Sciences*, *34*(8), 4763–4772. https://doi.org/10.1016/j.jksuci.2021.01.011

Pintoko, B. M., & Lhaksmana, K. M. (2018). Analisis Sentimen Jasa Transportasi Online Pada Twitter Menggunakan Metode NaÃ¯ve Bayes Classifier. *EProceedings of Engineering*, *5*(3). https://openlibrarypublications.telkomuniversity.ac.id/index.php/engineering/article/view/7447

Pradana, A. W., & Hayaty, M. (2019). The Effect of Stemming and Removal of Stopwords on the Accuracy of Sentiment Analysis on Indonesian-language Texts. *Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control*, 375–380. https://doi.org/10.22219/kinetik.v4i4.912

Prayoga, A. Y., Hadiana, A. I., & Umbara, F. R. (2021). Deteksi Hoax pada Berita Online Bahasa Inggris Menggunakan Bernoulli Naïve Bayes dengan Ekstraksi Fitur Tf-Idf. ., *2*(10), 1808–1823. https://doi.org/10.46799/JSA.V2I10.327

Reddy, K. N., & Reddy, D. B. I. (2021). Restaurant Review Classification Using Naives Bayes Model. *Journal of University of Shanghai for Science and Technology*, *23*(08), 646–656. https://doi.org/10.51201/JUSST/21/08443

Ricky, R. D. M., Kawung, E., & Goni, S. Y. V. . (2021). Dampak Aplikasi Belanja Online (Online Shop) di Masa Pandemi Covid-19 Terhadap Minat Belanja Masyarakat di Kelurahan Girian Weru Ii Kecamatan Girian Kota Bitung Provinsi Sulawesi Utara. *Jurnal Ilmiah*, *1*(ilmiah).

Romli, I., Prameswari R, S., & Kamalia, A. Z. (2021). Sentiment Analysis about Large-Scale Social Restrictions in Social Media Twitter Using Algoritm K-Nearest Neighbor. *Jurnal Online Informatika*, *6*(1), 96. https://doi.org/10.15575/join.v6i1.670

Syarifuddinn, M. (2020). ANALISIS SENTIMEN OPINI PUBLIK TERHADAP EFEK PSBB PADA TWITTER DENGAN ALGORITMA DECISION TREE,KNN, DAN NAÏVE BAYES. *INTI Nusa Mandiri*, *15*(1), 87–94. https://doi.org/10.33480/INTI.V15I1.1433

Tangirala, S. (2020). Evaluating the impact of GINI index and information gain on classification using decision tree classifier algorithm. *International Journal of Advanced Computer Science and Applications*, *2*, 612–619. https://doi.org/10.14569/ijacsa.2020.0110277

Wankhade, M., Rao, A. C. S., & Kulkarni, C. (2022). A survey on sentiment analysis methods, applications, and challenges. *Artificial Intelligence Review*, *55*(7), 5731–5780. https://doi.org/10.1007/s10462-022-10144-1

Watrianthos, R., Suryadi, S., Irmayani, D., Nasution, M., & Simanjorang, E. F. S. (2019). Sentiment Analysis Of Traveloka App Using Naïve Bayes Classifier Method. *INTERNATIONAL JOURNAL OF SCIENTIFIC & TECHNOLOGY RESEARCH*, *8*, 7. www.ijstr.org

Yun, H. (2021). Prediction model of algal blooms using logistic regression and confusion matrix. *International Journal of Electrical and Computer Engineering*, *11*(3), 2407–2413. https://doi.org/10.11591/ijece.v11i3.pp2407-2413