**PROPOSAL TA**

A picture containing text, clipart

Description automatically generated PERBANDINGAN METODE *MACHINE LEARNING* UNTUK SENTIMEN ANALISIS *REVIEW* PENJUALAN PRODUK DI TOKOPEDIA

**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 “Pemanfaatan *Naives Bayes* untuk sentimen analisis *review* produk laptop di Shoppe” 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**

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 merupakan algoritma *supervised learning* yang bekerja seperti struktur pohon di setiap *node* atau simpul mewakili dari atribut yang dilatih, *leaf node*