

BAB I PENDAHULUAN

1.1 Latar Belakang

Generasi Z, yang terdiri dari individu yang lahir antara pertengahan 1990-an hingga awal 2010-an. Generasi Z telah menjadi subjek perhatian yang signifikan dalam penelitian kesehatan mental karena mereka tumbuh dalam lingkungan yang dipengaruhi oleh teknologi digital dan media sosial, yang secara signifikan memengaruhi pola pikir, perilaku, dan kesehatan mental mereka [1]. Selain itu, pengaruh globalisasi dan dinamika sosial yang cepat telah menambah kompleksitas dalam pengalaman hidup mereka [2]. Hal ini menimbulkan tantangan baru dalam memahami dan mengatasi masalah kesehatan mental yang dihadapi oleh Generasi Z.

Perkembangan teknologi dan ketersediaan data dari media sosial telah memungkinkan peneliti untuk mengeksplorasi berbagai aspek kesehatan mental melalui analisis sentimen. Namun, tantangan utama dalam menggunakan algoritma seperti Naive Bayes adalah mencapai tingkat akurasi yang memadai dalam memahami sentimen yang kompleks dan seringkali ambigu dalam teks-tulisan yang diproduksi oleh generasi Z di platform seperti Twitter [3]. Teks yang diproduksi oleh Generasi Z di platform seperti Twitter seringkali mengandung sentimen yang kompleks dan ambigu terkait dengan kesehatan mental. Hal ini mempersulit pemrosesan dan klasifikasi sentimen secara akurat.

Algoritma Naive Bayes, meskipun sederhana dan mudah diimplementasikan, masih memiliki kelemahan dalam mencapai tingkat akurasi yang memadai ketika diterapkan pada data yang berkaitan dengan kesehatan mental dan menggunakan bahasa informal yang umumnya digunakan di media sosial. Ini menimbulkan kebutuhan untuk meningkatkan akurasi algoritma agar dapat menghasilkan prediksi yang lebih akurat terkait sentimen kesehatan mental Generasi Z di Twitter. Penelitian sebelumnya telah menunjukkan bahwa meskipun Naive Bayes merupakan metode klasifikasi yang sederhana dan mudah diimplementasikan, namun masih ada ruang untuk meningkatkan akurasinya,

terutama ketika diterapkan pada data yang berhubungan dengan kesehatan mental dan menggunakan bahasa informal yang umumnya digunakan di media sosial [3]–[5].

Untuk mengatasi kelemahan tersebut diusulkan pendekatan atau modifikasi tertentu pada algoritma Naive Bayes agar dapat memberikan hasil yang lebih akurat dalam menganalisis sentimen terkait kesehatan mental pada generasi Z di Twitter. Adapun modifikasi tersebut adalah menggunakan teknik ekstraksi fitur n-grams serta penyeimbang data Synthetic Minority Over-sampling Technique (SMOTE) [6], [7]. Data yang berkaitan dengan sentimen kesehatan mental pada Generasi Z mungkin mengalami ketidakseimbangan antara kelas positif dan negatif. Hal ini dapat menyebabkan kinerja yang bias terhadap kelas mayoritas dan kesalahan prediksi yang tinggi terhadap kelas minoritas. Oleh karena itu, diperlukan teknik penyeimbangan data seperti SMOTE untuk mengatasi masalah ini. Ekstraksi fitur n-grams bertujuan untuk mengekstrak fitur-fitur penting dari teks Twitter Generasi Z terkait dengan kesehatan mental. Tantangan di sini adalah untuk memilih dan mengekstrak fitur-fitur yang paling relevan dan informatif untuk meningkatkan kinerja algoritma klasifikasi sentimen.

Peningkatan akurasi algoritma Naive Bayes ini diharapkan dapat memberikan kontribusi yang signifikan dalam pemahaman dan deteksi dini masalah kesehatan mental pada generasi Z, serta membantu dalam pengembangan intervensi atau dukungan yang lebih efektif.

Dengan memperbaiki akurasi algoritma Naive Bayes dalam menganalisis sentimen tentang kesehatan mental pada generasi Z di media sosial Twitter, penelitian ini dapat membuka pintu untuk pemahaman yang lebih dalam tentang permasalahan kesehatan mental dalam populasi ini dan menghasilkan rekomendasi yang lebih tepat sasaran untuk intervensi atau dukungan yang diperlukan. Selain itu, peningkatan akurasi algoritma Naive Bayes juga dapat memperluas aplikasi metode ini dalam berbagai konteks analisis sentimen lainnya di media sosial dan domain lain yang memanfaatkan data teks.

Meskipun dataset berasal dari Twitter tentang kesehatan mental Generasi Z,

belum tentu semua tweet berkaitan dengan topik yang sama. Oleh karena itu, penting untuk menggunakan metode seperti Latent Dirichlet Allocation (LDA) untuk mengidentifikasi topik-topik utama atau pembahasan utama dalam dataset. Hal ini akan membantu dalam memahami tren dan isu-isu kesehatan mental yang paling relevan bagi Generasi Z. Metode ini dipakai oleh beberapa penelitian sebelumnya dan berhasil menemukan topik-topik pembahasan utama pada dataset [8]–[10]. Kelebihan LDA dapat mengekstrak topik tanpa bergantung domain dan bahasa serta mendeskripsikan konten pada dokumen lebih komprehensif, akurat dan cepat [11]. Serta dapat mengekstrak topik secara akurat pada dokumen yang cukup besar [12].

1.2 Rumusan Masalah

1. Bagaimana algoritma naïve bayes dapat meningkatkan akurasi untuk analisis sentimen kesehatan mental Generasi Z di media sosial Twitter?
2. Bagaimana metode LDA dapat menemukan topik-topik utama pembahasan kesehatan mental Generasi Z di media sosial Twitter?

1.3 Batasan Masalah

1. Analisis sentimen menggunakan bahasa pemrograman python.
2. Algoritma yang digunakan adalah naïve bayes.
3. Teknik n-grams digunakan untuk ekstraksi fitur serta teknik SMOTE digunakan untuk menyeimbangkan dataset.
3. Metode Latent Dirichlet Allocation digunakan untuk menemukan topik-topik utama pada dataset.

1.4 Tujuan Penelitian

Tujuan penelitian ini adalah untuk mengetahui bagaimana algoritma naïve bayes dapat meningkatkan akurasi untuk analisis sentimen kesehatan mental Generasi Z di media sosial Twitter, serta menemukan topik-topik utama yang ada pada dataset dengan metode LDA.

1.5 Manfaat Penelitian

1. Memberikan pemahaman mendalam tentang cara Generasi Z berinteraksi

dengan isu-isu kesehatan mental di media sosial.

2. Memberikan wawasan bagi masyarakat untuk mengembangkan strategi penanggulangan yang lebih efektif atau mencari dukungan yang diperlukan ketika menghadapi isu kesehatan mental.
3. Membuka peluang untuk eksplorasi lebih lanjut terhadap topik-topik tertentu dalam kesehatan mental Generasi Z di media sosial, serta potensi intervensi yang dapat dikembangkan berdasarkan hasil penelitian ini.

1.6 Sistematika Penulisan

Sistematika penulisan dibuat untuk menjelaskan secara singkat pembahasan dalam penelitian ini melalui gambaran isi dari setiap bab. Penelitian ini akan terdiri dari 5 bab sebagai berikut:

BAB I PENDAHULUAN

Bagian ini menguraikan beberapa hal mengenai latar belakang masalah, perumusan masalah, tujuan penelitian, manfaat penelitian dan sistematika penulisan yang akan menjadi acuan dalam penelitian ini.

BAB II LANDASAN TEORI

Bagian ini menjelaskan mengenai penelitian terkait dan teori-teori yang mendasari menjadi acuan dalam penelitian ini secara terperinci.

BAB III METODE PENELITIAN

Bab ini menguraikan mengenai penyelesaian masalah berdasarkan metodologi yang telah dipilih serta berisi proses implementasi dari metode tersebut.

BAB IV HASIL DAN PEMBAHASAN

Bab ini berisi mengenai hasil yang didapat dari pengujian yang telah dilakukan pada penelitian ini serta evaluasi hasil dari penelitian tersebut.

BAB V PENUTUP

Bab ini berisikan kesimpulan dan hasil akhir dari penelitian yang telah dilakukan serta berisi saran untuk perbaikan jika nantinya akan dilakukan pengembangan.