

BAB I

PENDAHULUAN

1.1 Latar Belakang

Teknologi informasi (IT) telah menjadi bagian integral dari kehidupan modern, memengaruhi berbagai aspek mulai dari komunikasi hingga analisis data. Di era digital ini, penggunaan teknologi informasi tidak hanya terbatas pada sektor bisnis dan pemerintahan, tetapi juga merambah ke bidang penelitian dan analisis sosial. Salah satu penerapan teknologi informasi yang semakin berkembang adalah analisis sentimen, yang merupakan cabang dari pemrosesan bahasa alami. Analisis sentimen menjadi penting dalam memahami persepsi publik terhadap fenomena tertentu yang dapat mempengaruhi hasilnya. Analisis sentimen dapat membantu dalam pengambilan keputusan strategis dan komunikasi yang lebih efektif (Alhaqq et al., 2023).

Analisis sentimen menggunakan teknologi informasi memungkinkan peneliti untuk menggali pandangan, opini, dan emosi dari teks yang dihasilkan oleh pengguna media sosial dan platform digital lainnya. Dengan meningkatnya volume data yang tersedia di internet, metode analisis yang efisien dan akurat sangat diperlukan. *Machine learning* dan teknik optimasi merupakan solusi yang semakin populer untuk meningkatkan kinerja analisis sentimen. Generasi *Sandwich* merujuk pada kelompok individu yang berada di antara dua generasi yang memerlukan dukungan, yaitu orang tua yang sudah lanjut usia dan anak-anak yang masih memerlukan perhatian dan pengasuhan. Istilah ini menggambarkan beban ganda yang harus ditanggung oleh generasi tersebut, baik dari segi finansial, emosional, maupun waktu. Dalam konteks Indonesia, fenomena ini menjadi semakin relevan dengan perubahan demografis dan struktur keluarga yang semakin kompleks.

Seiring dengan meningkatnya penggunaan media sosial, sentimen publik terhadap fenomena "Generasi *Sandwich*" dapat dianalisis melalui data yang

tersedia di platform-platform tersebut. Analisis sentimen ini penting untuk memahami pandangan masyarakat dan dapat digunakan sebagai dasar untuk pengambilan keputusan dalam kebijakan sosial dan ekonomi (Irawan, 2024).

Metode analisis sentimen menggunakan *Machine Learning* telah berkembang pesat, salah satu metode yang sering digunakan adalah *K-Nearest Neighbor* (KNN), yang sederhana namun efektif untuk klasifikasi. Namun, kinerja KNN sangat bergantung pada pemilihan parameter yang tepat, seperti jumlah tetangga terdekat (k). Untuk mengatasi tantangan ini, *Particle Swarm Optimization* (PSO) dapat digunakan untuk mengoptimalkan parameter KNN. PSO adalah algoritma optimasi berbasis populasi yang terinspirasi oleh perilaku sosial kawanan burung atau ikan, dan telah terbukti efektif dalam berbagai aplikasi optimasi.

Metode *K-Nearest Neighbor* (KNN) yang diperbaiki dengan menggunakan *Particle Swarm Optimization* (PSO) untuk menganalisis sentiment fenomena *Generasi Sandwich*. *K-Nearest Neighbor* (KNN) adalah algoritma klasifikasi yang populer digunakan dalam analisis sentiment, sedangkan PSO adalah algoritma optimasi yang digunakan untuk menemukan solusi yang terbaik dalam berbagai aplikasi. *K-Nearest Neighbor* (KNN) telah digunakan secara luas dalam analisis sentimen untuk mengklasifikasikan sentimen dari teks. Namun, metode klasifikasi ini masih memiliki keterbatasan dalam menangani data yang kompleks dan besar (Rahman *et al.*, 2021).

Particle Swarm Optimization (PSO) merupakan metode optimasi heuristik yang menginspirasi dari perilaku *swarm* (koleksi) dari partikel dalam alam. PSO telah digunakan dalam berbagai aplikasi, termasuk optimasi algoritma klasifikasi, untuk meningkatkan performa dan efisiensi (Shami *et al.*, 2022). Penerapan PSO pada KNN dapat meningkatkan akurasi klasifikasi dengan mencari solusi optimal dalam ruang fitur yang sangat besar. Ini sangat penting dalam analisis sentimen, di mana dataset dapat sangat besar dan kompleks (Setianingsih *et al.*, 2023).

Penerapan analisis sentimen salah satunya dilakukan untuk mengklasifikasikan ulasan *acne spot treatment* menjadi Positif atau negatif. Dalam penelitian ini digunakan metode klasifikasi *K-Nearest Neighbor* karena memiliki konsep sederhana yang mudah diaplikasikan dan dimengerti. Hasil sentimen didapatkan ulasan yang paling banyak mengandung sentimen positif adalah *Whitelab* sebanyak 1.190 ulasan dan yang paling banyak mengandung

sentimen negatif adalah *Skin Game* sejumlah 173 ulasan. Hasil klasifikasi menggunakan KNN akurasi terbaik adalah *Whitelab* sebesar 97%, kemudian *Skin Game* memperoleh akurasi 81%, dan nilai akurasi paling rendah adalah ERHA sebesar 75%. (Dyah Fritama et al., 2023)

Penelitian terdahulu yang membahas implementasi *particle swarm optimization* pada Algoritma *K- Nearest Neighbor* untuk optimasi penentuan klasifikasi penyakit Hepatitis C. Adanya implementasi tersebut diharapkan mampu meningkatkan nilai akurasi dalam klasifikasi dan mengatasi solusi untuk kelemahan pada Algoritma *K-Nearest Neighbor* tersebut. Dari hasil pengujian *K-Nearest Neighbor* diperoleh nilai akurasi 97,24% pada $K=5$ dan $K=3$. Adapun untuk hasil pengujian implementasi *Particle Swarm Optimization* pada *K-Nearest Neighbor* terjadi peningkatan nilai akurasi sebesar 2,07% menjadi 99,31%. Pengujian ini menunjukkan bahwa implementasi PSO mampu mengatasi kekurangan KNN dan model ini dapat dijadikan sebagai solusi terbaik untuk menentukan klasifikasi penyakit Hepatitis C (Setianingsih et al., 2023).

Analisis sentimen menggunakan *K-Nearest Neighbor* terhadap *New Normal* masa *Covid-19* di Indonesia adalah untuk memprediksi komentar ataupun opini masyarakat yang kecenderungan beropini positif maupun negatif. *Preprocessing* data menggunakan *cleaning*, *case folding*, normalisasi, *stemming*, *filtering*, dan *tokenizing*. Pada normalisasi kata bertujuan memperbaiki kesalahan penulisan kata berdasarkan KBBI dan TF-IDF sebagai metode pembobotan kata. Data yang digunakan terdiri dari 1000 tweet. Metode klasifikasi opini menggunakan metode *K - Nearest Neighbor* dan melakukan pengujian agar mendapatkan hasil akurasi yang paling terbaik serta mengevaluasi menggunakan *confusion matrix*. Hasil dari pelabelan untuk sentimen positif berjumlah 811 dan 189 untuk sentimen negatif. Klasifikasi K-NN dengan nilai $k = 1$ menghasilkan pengujian *use training set* dengan akurasi sebesar 100%, 92,60% untuk *10-fold cross-validation* dan 94,50% untuk *80% percentage split* (Furqan et al., 2022).

Penelitian lainnya juga menganalisa sentimen pada ulasan pengguna aplikasi investasi *online* yaitu *bibit* dan *bareksa*. Jumlah ulasan yang akan digunakan pada penelitian ini sebanyak 998 yang terdiri dari 484 sentimen positif dan 514 sentimen negatif untuk aplikasi *bareksa* sedangkan untuk aplikasi *bibit* menggunakan 1063 data yang terdiri dari 541 sentimen positif dan 522 sentimen negatif. Data tersebut juga melewati tahapan *preprocessing* dan *modelling*. Pada

penelitian ini menggunakan model *CRISP-DM (Cross Industry Standard Process for Data Mining)* dan algoritma yang digunakan pada penelitian ini adalah *K-Nearest Neighbors*. Berdasarkan hasil yang diperoleh dari tahapan *modelling* dengan menggunakan Algoritma *K-Nearest Neighbors* dan perbandingan 60:40 untuk data *training* dan data *testing*, maka nilai akurasi *precision* dan *recall* yang dihasilkan dari tiap aplikasi yaitu untuk bibit 85,14% , 91,91%, dan 76,44% sedangkan untuk bareksa yaitu 81,70% , 87,15%, 75,73%(Adhi Putra, 2021).

Sistem analisis sentimen terhadap ulasan *Paris Van Java Resort Lifestyle Place* di Kota Bandung dengan menggunakan Algoritma KNN (*K-Nearest Neighbor*) berhasil mengklasifikasikan ulasan pada *Google Review* berupa sentimen positif dan sentimen negatif. Dari hasil pengujian pertama diperoleh bahwa hasil performa terbaik pada analisis sentimen terhadap ulasan *Paris Van Java Resort Lifestyle Place* di Kota Bandung dengan menggunakan Algoritma KNN (*K-Nearest Neighbor*), yaitu pada penggunaan *Term Frequency-Inverse Document Frequency (TF-IDF) Unigram* dengan seluruh nilai persentase rasio data yang digunakan dengan nilai perfoma terbaik sebesar 88.29%. Dapat dikatakan bahwa pengaruh N-gram sangat berpengaruh karena hasil yang diperoleh pada pengujian mempunyai perbedaan yang signifikan. Selain itu, dari hasil pengujian yang kedua, metode penghitungan jarak dengan seluruh nilai persentase rasio data yang digunakan dengan performa terbaik adalah *Euclidean Distance* yang dimana mencapai akurasi sebesar 88.29%. Penggunaan metode penghitungan jarak dapat mempengaruhi keakuratan Algoritma KNN (*K-Nearest Neighbor*) dalam analisis sentimen terhadap ulasan *Paris Van Java Resort Lifestyle Place* di Kota Bandung(Amardita et al., 2022).

Penelitian ini bertujuan untuk mengimplementasikan *K-Nearest Neighbor* berbasis *Particle Swarm Optimization* untuk analisis sentimen terkait fenomena "Generasi *Sandwich*". Diharapkan metode ini dapat meningkatkan akurasi klasifikasi sentimen dibandingkan dengan KNN konvensional. penelitian ini tidak hanya berkontribusi pada literatur akademik mengenai analisis sentimen, tetapi juga memberikan wawasan praktis bagi para pembuat kebijakan dan pengembang teknologi yang berfokus pada isu-isu sosial kontemporer seperti "Generasi *Sandwich*".

Berdasarkan latar belakang masalah diatas, penulis penulis akan melakukan penelitian tesis yang berjudul: **“IMPLEMENTASI K-NEAREST NEIGHBOR BERBASIS PARTICLE SWARM OPTIMIZATION UNTUK ANALISIS SENTIMEN FENOMENA GENERASI SANDWICH.**

1.2 Perumusan Masalah

Berdasarkan latar belakang masalah diatas, , terdapat beberapa masalah yang bisa diangkat. Penelitian ini akan difokuskan pada tiga perumusan masalah diantaranya adalah :

1. Bagaimana menerapkan Algoritma *K-Nearest Neighbor* untuk menganalisis sentiment fenomena *Generasi Sandwich*.
2. Bagaimana mengetahui kecenderungan opini publik terhadap fenomena *Generasi Sandwich* pada media sosial *Twitter*?
3. Bagaimana menguji tingkat akurasi penerapan *K-Nearest Neighbor* berbasis *Particle Swarm Optimization* dalam menganalisis sentimen fenomena *Genreasi Sandwich*.

1.3 Batasan Masalah

Batasan masalah bertujuan untuk membantu dalam mengidentifikasi masalah, membatasi ruang lingkup dan menjadi gambaran terkait hal yang akan diteliti sehingga penelitian menjadi lebih terarah dan tidak menyimpang dari pokok bahasan dan permasalahan. Untuk membatasi ruang lingkup penelitian, maka batasan masalah dibatasi sebagai berikut:

1. Penelitian ini hanya akan menggunakan data opini publik *Generasi Sandwich* yang diambil dari platform media sosial *twitter* dari tanggal 1 agustus 2023 sampai 07 juli 2024.
2. *Tweet* yang digunakan hanya *tweet* yang menggunakan Bahasa Indonesia.
3. Metode yang digunakan dalam penelitian ini adalah *K-Nearest Neighbor* berbasis *Particle Swarm Optimization*.

1.4 Tujuan Penelitian

Berdasarkan perumusan masalah dan batasan masalah yang disampaikan, maka tujuan penulis dalam melakukan penelitian ini adalah sebagai berikut:

1. Memahami penerapan Algoritma *K-Nearest Neighbor* yang dioptimasi dengan *Particle Swarm Optimization* untuk menganalisis sentiment terhadap opini Generasi *Sandwich*.
2. Merancang metode *K-Nearest Neighbor* berbasis *Particle Swarm Optimization* kecenderungan opini publik terhadap fenomena *Generasi Sandwich* pada media sosial *twitter*.
3. Mengimplementasikan serta menguji tingkat akurasi analisis sentiment dengan menggunakan Algoritma *K-Nearest Neighbor* berbasis *Particle Swarm Optimization*.

1.5 Manfaat Penelitian

Hasil penelitian yang dilakukam diharapkan dapat dimanfaatkan baik bagi kepentingan peneliti, organisasi, maupun ilmu pengetahuan bagi peneliti selanjutnya. Adapun manfaat yang akan didapatkan melalui penelitian ini adalah sebagai berikut:

1. Hasil dari penelitian ini dapat digunakan untuk menganalisis sentimen yang lebih efektif dan akurat dalam memonitor dan menganalisis opini publik terkait Generasi *Sandwich*. Hal ini dapat membantu analis, dan peneliti dalam memahami persepsi masyarakat terhadap Generasi *Sandwich*
2. Hasil dari penelitian ini dapat memberikan gambaran bagaimana fenomena dan pembahasan mengenai *Generasi Sandwich* di indonesia.

1.6 Sistematika Penulisan

Sistematika penulisan dilakukan agar lebih mudah untuk dibaca dan dimengerti, maka penulis berusaha menyusun laporan penelitian ini dengan tata urutan secara sistematis. Berdasarkan hal itu, peneliti mengklasifikasikan

penelitian ini kedalam enam bab, antara bab satu dengan bab yang lain saling berhubungan.

BAB I PENDAHULUAN

Pada BAB I ini berisikan mengenai latar belakang, perumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian dan sistematika penulisan.

BAB II LANDASAN TEORI

BAB II ini membahas tentang beberapa hal yang mendasar tentang masalah yang akan dibahas, yang mana nantinya akan digunakan sebagai landasan dasar dalam melakukan pemecahan masalah pada penelitian yang dilakukan.

BAB III METODOLOGI PENELITIAN

BAB III ini membahas tentang langkah kerja dan prosedur penelitian serta metode yang akan dilakukan dalam penelitian.

BAB IV ANALISA DAN PERANCANGAN

BAB IV ini membahas tentang analisa data, pengolahan data *Twitter* mengenai femonema Generasi *Sandwich*. Selain itu bab ini juga memuat hasil dari penggunaan metode *K-Nearest Neighbor* berbasis *Particle Swarm Optimization*.

BAB V IMPLEMENTASI DAN HASIL

BAB V ini membahas tentang penerapan metode *K-Nearest Neighbor* berbasis *Particle Swarm Optimization* menggunakan bahasa pemrograman *Python* sebagai aplikasi pengklasifikasian data yang diambil dari *Twitter*.

BAB VI KESIMPULAN DAN SARAN

BAB VI ini membahas tentang kesimpulan dari hasil penelitian dan sara-saran untuk pihak lain dalam melakukan penelitian dan pengembangan selanjutnya.