

BAB I

PENDAHULUAN

1.1. Latar Belakang

Kanker adalah penyakit yang terjadi akibat adanya mutasi atau perubahan abnormal pada gen yang seharusnya mengatur pertumbuhan sel secara normal (Yussof et al., 2023). Penyakit ini menjadi penyebab kematian terbesar kedua di dunia, dengan kanker payudara sebagai salah satu jenis yang paling sering terjadi, terutama pada Wanita (Gumilang & Hernawan, 2023). Pada tahun 2022, sekitar 2,3 juta wanita didiagnosis menderita kanker payudara, dan sebanyak 670.000 di antaranya meninggal dunia akibat penyakit tersebut. Kanker payudara dapat menyerang wanita di berbagai negara dan usia, mulai dari masa setelah pubertas (World Health Organization, 2022). Berdasarkan penelitian (Yussof et al., 2023), gen di dalam inti setiap sel memiliki peran penting sebagai pengontrol utama aktivitas sel. Secara normal, tubuh kita meregenerasi sel melalui proses pertumbuhan yang teratur, di mana sel baru menggantikan sel yang lama. Namun, mutasi dapat menyebabkan gen tertentu aktif atau tidak aktif secara tidak wajar, sehingga sel-sel tersebut terus bertahan hidup dan membelah tanpa kendali. Hal ini dapat menghasilkan tumor. Oleh karena itu, deteksi dini dan pengobatan yang tepat menjadi kunci utama untuk meningkatkan peluang kesembuhan penderita kanker payudara.

Kesadaran dan edukasi tentang kanker payudara sangat penting untuk mendukung deteksi dini. Kampanye global terus dilakukan untuk mendorong masyarakat agar melakukan pemeriksaan rutin dan mengenali tanda-tanda awal penyakit ini. Media sosial, terutama *Twitter* yang saat ini berubah nama menjadi *X*, berperan besar dalam menyebarkan informasi terkait. Dengan pengguna yang sangat banyak, *X* memungkinkan penyebaran informasi kesehatan secara cepat dan luas. Analisis sentimen adalah teknik yang digunakan untuk memahami opini atau perasaan seseorang terhadap suatu isu melalui teks. Di media sosial, analisis sentimen membantu menilai respons publik terhadap kampanye tertentu, termasuk kesadaran kanker payudara. Salah satu metode analisis sentimen yang efektif adalah *Support Vector Machine (SVM)* karena hasil akurasi SVM memiliki akurasi yang tinggi dibanding *Naïve Bayes* (Madjid et al., 2023). Algoritma ini bekerja dengan

memisahkan data ke dalam kelas-kelas tertentu menggunakan hyperplane terbaik (Muntiari et al., 2023).

SVM sangat cocok untuk tugas klasifikasi, terutama dalam data berdimensi tinggi atau dengan margin antar kelas yang sempit (T. M. P. Aulia et al., 2021). Misalnya, dalam mengidentifikasi penyakit stroke, *SVM* berhasil mencapai akurasi rata-rata 76,52% (Azhar et al., 2022). Dalam analisis sentimen, *SVM* mengklasifikasikan teks ke dalam kategori sentimen positif atau negatif (Hafidz & Liliana, 2021). Selain *SVM*, algoritma seperti *Naive Bayes* juga sering digunakan untuk analisis sentimen serupa (Daniel & Martin, 2023). Metode *Naive Bayes* mengasumsikan independensi antar fitur, yang membuatnya cepat dan efisien dalam proses klasifikasi (Lestari & Sugiyono, 2024). Oleh karena itu, penelitian ini cocok jika membandingkan antara *SVM* dan *Naive Bayes* karena proses yang cepat dan efektif. Proses analisis ini melibatkan tahapan seperti preprocessing teks, ekstraksi fitur, dan pelatihan model menggunakan data berlabel sentimen. Setelah model dilatih, ia dapat mengklasifikasikan sentimen teks baru dengan akurat.

Penelitian ini bertujuan untuk menganalisis sentimen komentar di X tentang kampanye kesadaran kanker payudara menggunakan metode *SVM* dan *naive bayes* untuk mengetahui seberapa baik performa metode *SVM* dan *naive bayes*. Penelitian ini memberikan wawasan yang bermanfaat bagi organisasi kesehatan dan pihak-pihak yang melakukan kampanye kesadaran. Akibatnya, pendekatan komunikasi yang lebih efisien dapat dibuat untuk meningkatkan kesadaran dan pemahaman masyarakat tentang kanker payudara.

1.2. Rumusan Masalah

1. Bagaimana persepsi publik dan pengukuran efektivitas kampanye kanker payudara di *Twitter* dalam meningkatkan kesadaran akan bahaya kanker payudara berdasarkan hasil sentimen analisis?
2. Seberapa baik performa metode *Support Vector Machine* dan *Naive Bayes* dalam mengklasifikasikan sentimen publik terhadap kampanye kesadaran kanker payudara di *Twitter* berdasarkan evaluasi akurasi, presisi, recall, dan F1-score?

1.3. Tujuan Penelitian

1. Mengidentifikasi dan menganalisis persepsi publik serta mengukur efektivitas kampanye kanker payudara di *Twitter* dalam meningkatkan kesadaran akan bahaya kanker payudara.
2. Menilai evaluasi model akurasi, presisi, recall, dan FI- score metode *Support Vector Machine* dan *Naive Bayes* dalam mengklasifikasikan sentimen publik terhadap kampanye kesadaran kanker payudara di *Twitter*

1.4. Batasan dan Asumsi Penelitian

1. Data yang digunakan dalam penelitian ini hanya mencakup tweet yang berbahasa Indonesia yang terkait dengan kampanye kesadaran kanker payudara dengan menggunakan kata kunci kanker payudara dan data dikumpulkan mulai dari bulan Oktober 2022 – Oktober 2024 untuk menghindari data yang melebar bisa mengakibatkan munculnya ambiguitas dan tidak seimbang
2. Penelitian ini diharapkan menghasilkan output dengan hasil metode *Support Vector Machine (SVM)* lebih baik daripada *Naive Bayes*.

1.5. Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat yang luas bagi berbagai pihak. Studi ini mengkaji dan membandingkan dua metode populer, yaitu *Support Vector Machine (SVM)* dan *Naive Bayes*, dalam konteks akademis untuk memperkaya pemahaman tentang analisis sentimen dan pengolahan bahasa alami (*NLP*). Hasil penelitian ini diharapkan dapat menjadi referensi bagi peneliti lain

yang tertarik untuk mengeksplorasi kelebihan dan kekurangan kedua metode tersebut dalam menganalisis sentimen di platform media sosial. Selain itu, penelitian ini juga bertujuan untuk mengidentifikasi teknik analisis yang lebih presisi dan efisien guna mengevaluasi keberhasilan kampanye kesadaran kanker payudara di media sosial. Bagi masyarakat luas, penelitian ini diharapkan dapat meningkatkan pemahaman akan pentingnya kampanye yang lebih terarah untuk mendukung deteksi dini kanker payudara.

Dengan demikian, penelitian ini dapat menjadi rujukan bagi peneliti lain yang ingin memahami kelebihan dan kekurangan kedua metode tersebut dalam konteks analisis sentimen di media sosial. Studi ini juga membantu organisasi kesehatan dalam memahami persepsi masyarakat terhadap kampanye kesadaran kanker payudara yang dijalankan di media sosial. Tujuan utama dari penelitian ini adalah untuk meningkatkan kesadaran akan pentingnya deteksi dini kanker payudara melalui kampanye yang lebih efektif, yang dirancang berdasarkan umpan balik publik. Selain itu, penelitian ini dapat membantu para pengembang teknologi dalam memilih pendekatan terbaik untuk menciptakan aplikasi analisis sentimen. Aplikasi tersebut dapat memantau dan meningkatkan efektivitas kampanye media sosial secara real-time, serta menghasilkan data yang dapat memberikan wawasan berharga tentang sejauh mana kampanye kesadaran kanker payudara berhasil mencapai tujuannya

1.6. Sistematika Penulisan

Bab I **Pendahuluan** menjelaskan latar belakang penelitian yang berkaitan dengan analisis sentimen kampanye kesadaran kanker payudara di *Twitter*. Dalam bab ini, dibahas pentingnya deteksi dini kanker payudara serta peran media sosial, khususnya *Twitter*, sebagai sarana penyebaran informasi kesehatan. Bab ini juga memuat rumusan masalah, tujuan penelitian, serta manfaat penelitian yang bertujuan memberikan wawasan mengenai efektivitas kampanye kesadaran melalui analisis sentimen menggunakan algoritma *Support Vector Machine (SVM)* dan *Naïve Bayes*.

Bab II **Landasan Teori** memuat teori-teori yang mendasari penelitian ini. Pembahasan meliputi penjelasan tentang kanker payudara, gejala, dampaknya, dan

pentingnya deteksi dini, serta peran media sosial dalam kampanye kesehatan. Selain itu, bab ini menjelaskan konsep analisis sentimen, proses ekstraksi fitur, dan preprocessing teks. Algoritma *Support Vector Machine (SVM)* dan Naïve Bayes juga dijelaskan sebagai teknik utama untuk klasifikasi.

Bab III **Metodologi Penelitian** menjelaskan metode yang digunakan dalam penelitian, termasuk tahapan implementasi algoritma *SVM* dan Naïve Bayes untuk analisis sentimen. Teknik pengumpulan data dijabarkan melalui pemanfaatan tweet berbahasa Indonesia terkait kampanye kanker payudara pada periode Oktober 2022–Oktober 2024. Proses preprocessing data, seperti pembersihan data, tokenisasi, stemming, dan ekstraksi fitur, dijelaskan secara rinci. Bab ini juga menjelaskan objek penelitian berupa dataset komentar di *Twitter* dengan label sentimen, serta metode evaluasi model yang menggunakan metrik akurasi, presisi, recall, dan F1-score.

Bab IV **Pengumpulan dan Pengolahan Data** menguraikan proses pengumpulan data dari *Twitter* dengan menggunakan kata kunci terkait kampanye kanker payudara. Proses ini meliputi preprocessing teks, seperti normalisasi data, penghapusan noise, tokenisasi, serta pembagian data menjadi data latih dan data uji. Selanjutnya, algoritma *SVM* dan Naïve Bayes diterapkan untuk mengklasifikasikan sentimen publik. Hasil pengujian dievaluasi menggunakan metrik tertentu untuk mengukur performa model.

Bab V **Analisis dan Pembahasan** berfokus pada hasil pengujian algoritma *SVM* dan Naïve Bayes dalam mengklasifikasikan sentimen publik terhadap kampanye kanker payudara. Perbandingan performa kedua algoritma didasarkan pada metrik evaluasi, seperti akurasi, presisi, recall, dan F1-score. Selain itu, pola sentimen publik yang muncul terhadap kampanye ini dianalisis, dan kelebihan serta kekurangan masing-masing algoritma dalam tugas analisis sentimen dibahas secara mendalam.

Bab VI **Kesimpulan dan Saran** memuat kesimpulan dari hasil penelitian, seperti performa terbaik dari algoritma yang digunakan serta temuan utama mengenai efektivitas kampanye kanker payudara berdasarkan analisis sentimen. Bab ini juga memberikan saran untuk penelitian selanjutnya, seperti penggunaan dataset yang

lebih besar, implementasi algoritma machine learning lainnya, atau pengujian di platform media sosial yang berbeda.