

BAB II

TINJAUAN PUSTAKA

Sentimen Analisis juga dikenal sebagai *Opinion Mining* adalah proses otomatis menganalisis teks untuk menentukan apakah sentimen yang diungkapkan dalam teks itu *positive* atau *negative*. Penerapan Sentimen Analisis tidak terbatas dan telah menyebar ke hampir setiap domain yang memungkinkan, dari kesehatan (*healthcare*), jasa (*services*), konsumen produk (*product consumers*), hingga finansial (*financials*) dan politik (*politics*). Berikut ini adalah beberapa penelitian mengenai penerapan Sentimen Analisis dalam beberapa bidang dengan menggunakan berbagai teknik klasifikasi.

Menurut penelitian yang dilakukan oleh Jagdale, Shirsat, dan Deshmukh (2019) yaitu *Sentiment Analysis on Product Reviews Using Machine Learning Techniques*, menyatakan bahwa Sentimen Analisis adalah sebuah bidang yang sangat populer untuk menganalisis dan mendapatkan pengetahuan atau wawasan dari berbagai sumber data teks seperti, Facebook, Twitter, Amazon, dll. Sentimen Analisis memainkan peran vital dalam berbagai bisnis untuk dapat bekerja secara aktif dalam memperbaiki strategi dan memperoleh pengetahuan mendalam dari para pembeli mengenai produk mereka. Penelitian Sentimen Analisis tersebut membutuhkan studi komputasi dari perilaku individu mengenai minat beli dan kemudian menggali opini tentang entitas bisnis perusahaan. Entitas tersebut dapat digambarkan sebagai suatu kejadian (*event*), orang (*individual*), *blog post*, atau pengalaman produk (*product experience*). Penelitian tersebut mengambil *dataset* dari Amazon yang berisi ulasan dari produk kamera, laptop, ponsel, tablet, TV, dan CCTV. Penelitian tersebut mengaplikasikan algoritme pembelajaran mesin (*Machine Learning*) untuk melakukan klasifikasi pada ulasan, apakah *negative* atau *positive*. Penelitian tersebut menyimpulkan bahwa, teknik *Machine Learning* yang memberikan hasil terbaik

untuk mengklasifikasi ulasan produk adalah, *Naive Bayes* dengan nilai *Accuracy* 98.17% dan *Support Vector Machine* dengan *Accuracy* 93.54% untuk ulasan produk kamera. *Naive Bayes* menghasilkan nilai *Precision* 98.30% dan *F1-Score* 99.03%. Kemudian, *Support Vector Machine* menghasilkan nilai *Precision* 93.58% dan *F1-Score* 96.66% [12].

Penelitian selanjutnya dilakukan oleh Haja Sharieff, T. Sindhu, dan L. SaiRamesh (2020), yaitu *Comparison of Machine Learning Techniques for Sentimental Analysis on Restaurant Reviews*. Penelitian ini menganalisis reaksi dan pendapat orang terhadap restoran, dengan mengumpulkan dan menganalisis ulasan restoran. Analisis tersebut menghasilkan laporan analisis mengenai pendapat orang terhadap suatu restoran tertentu. Tujuan dari analisis tersebut adalah mengembangkan sebuah model untuk memprediksi apakah ulasan pada restoran tersebut *positive* atau *negative*. Peneliti melakukan analisis dengan menggunakan *dataset* ulasan restoran, menggunakan algoritme prediktif *Multinomial Naive Bayes*, *Bernoulli Naive Bayes*, dan *Logistic Regression*. Hasil dari model analisis tersebut adalah *Multinomial Naive Bayes* menghasilkan nilai *Accuracy* prediksi 77.67%, nilai *Precision* 78%, dan nilai *Recall* 77%. *Bernoulli Naive Bayes* dengan nilai *Accuracy* 77.0%, nilai *Precision* 76%, dan nilai *Recall* 78%. Kemudian *Logistic Regression* menghasilkan *Accuracy* 76,67%, nilai *Precision* 80%, dan nilai *Recall* 71%. Berdasarkan hasil tersebut, nilai *Accuracy* metode *Multinomial Naive Bayes* sedikit lebih baik dibandingkan dengan metode *Bernoulli Naive Bayes* dan *Logistic Regression*. Dengan demikian, model dengan *Accuracy* 77.67% dapat memberikan prediksi yang tepat [13].

Kemudian, topik penelitian selanjutnya adalah *Sentiment Analysis on Hotel Reviews Using Multinomial Naive Bayes Classifier*. Penelitian tersebut dilakukan oleh Farisi, Sibaroni, dan Faraby (2019). Berdasarkan penelitian tersebut di zaman modern saat ini di mana internet sedang berkembang dengan sangat pesat, keberadaannya dapat memudahkan wisatawan untuk mencari informasi apa pun. Di dalam bidang hotel pariwisata, internet sangat membantu dalam hal promosi hotel. Tersedianya ulasan di internet dalam jumlah yang banyak, membuat wisatawan sulit

untuk memahami semua ulasan apakah itu opini *positive* atau *negative*. Sentimen Analisis dibutuhkan untuk mendeteksi dengan cepat, apakah ulasan tersebut *positive* atau *negative*. Penelitian tersebut memberikan solusi dengan mengklasifikasi ulasan opini *positive* dan *negative* menggunakan metode *Multinomial Naive Bayes*, kemudian membandingkan kinerja model dengan menggunakan teknik *preprocessing*, *feature extraction*, dan *feature selection*. Berdasarkan dari hasil analisis dan pengujian, disimpulkan bahwa sistem klasifikasi berhasil dibangun dengan hasil eksperimen terbaik menggunakan *feature selection* dan *preprocessing* dengan *10 fold cross validation* dan memiliki hasil terbaik rata-rata *F1-Score* lebih dari 91% [14].

Kemudian pada penelitian *Sentiment Classification on Big Data Using Naive Bayes and Logistic Regression* oleh Anjuman Prabhat dan Vikas Khullar (2017), menyatakan bahwa ekspansi besar *World Wide Web* (WWW) telah melibatkan sebuah cara kontemporer untuk menyampaikan sikap atau pandangan manusia. *World Wide Web* adalah sebuah saluran di mana siapa saja dapat memvisualisasikan opini atau sentimen apa pun dari pelanggan yang berbeda. Sehingga, memungkinkan untuk mengklasifikasi opini yang diberikan terhadap produk ke dalam kategori yang berbeda. Informasi memainkan peran penting dalam tugas klasifikasi sentimen. Jumlah data besar yang tersimpan *online* dapat ditambang secara efektif, data tersebut dapat diekstrak untuk menghasilkan informasi yang berharga. Penelitian tersebut menggunakan data *real time* ulasan Twitter dan menggunakan algoritme *Machine Learning* yang berbeda, yaitu *Naive Bayes* dan *Logistic Regression* untuk mengklasifikasi ulasan Twitter. Kinerja dari kedua algoritme tersebut dibandingkan berdasarkan parameter yaitu, *Accuracy*, *Precision*, dan *Computation time*. *Naive Bayes* menghasilkan nilai *Accuracy* 66.667%, *Precision* 69.23%, dan *Computation time* 15732 mili-sekon. Sementara *Logistic Regression* menghasilkan nilai *Accuracy* 76.767%, *Precision* 73.575%, dan *Computation time* 3689 mili-sekon [15].

Untuk mempermudah pembaca, berikut adalah tabel perbandingan penelitian-penelitian sebelumnya yang telah dijelaskan :

Tabel 2.1 Perbandingan penelitian

No.	Peneliti	[12]	[13]	[14]	[15]	Hadiyana (2022)*
1.	Judul	<i>Sentiment Analysis on Product Reviews Using Machine Learning Techniques.</i>	<i>Comparison of Machine Learning Techniques for Sentimental Analysis on Restaurant Reviews.</i>	<i>Sentiment analysis on hotel reviews using Multinomial Naive Bayes classifier.</i>	<i>Sentiment classification on Big Data using Naive Bayes and Logistic Regression.</i>	Sentimen Analisis Keterbatasan Kesempatan Kerja Dalam Bidang yang Diminati Bagi Gen Z Menggunakan Algoritme Naive Bayes.
2.	Tujuan Penelitian	Mengembangkan model Sentimen Analisis yang digunakan untuk mengklasifikasi sentimen	Mengembangkan sebuah model <i>Machine Learning</i> yang dapat memprediksi	Mengembangkan model yang dapat melakukan klasifikasi sentimen ulasan publik terhadap	Mengembangkan sebuah model yang dapat melakukan klasifikasi sentimen dengan	Mengembangkan model yang dapat melakukan klasifikasi dan prediksi sentimen, yang digunakan

		pelanggan terhadap suatu produk, serta untuk mendapatkan pengetahuan dan memperbaiki strategi bisnis.	ulasan pada suatu restoran apakah <i>positive</i> atau <i>negative</i> .	suatu hotel.	menggunakan data cuitan Twitter (<i>tweet</i>).	untuk menganalisis studi kasus keterbatasan kesempatan kerja dalam bidang yang diminati bagi <i>Gen Z</i> .
3.	Algoritme	<i>Naive Bayes</i> dan <i>Support Vector Machine</i> .	<i>Multinomial Naive Bayes</i> , <i>Bernoulli Naive Bayes</i> , dan <i>Logistic Regression</i> .	<i>Multinomial Naive Bayes Classifier</i> .	<i>Naive Bayes Classifier</i> dan <i>Logistic Regression</i> .	<i>Naive Bayes</i> .
4.	Metrik Evaluasi	<i>Accuracy</i> , <i>Precision</i> , dan <i>F1-Score</i> .	<i>Accuracy</i> , <i>Precision</i> , dan <i>Recall</i> .	10 <i>fold cross validation</i> dan <i>F1-Score</i> .	<i>Accuracy</i> , <i>Precision</i> , <i>Computation time</i> .	<i>Precision</i> , <i>F1-Score</i> , <i>Recall</i> , <i>Accuracy</i> , dan <i>Confusion Matrix</i> .

*Penelitian yang dilakukan