



Hak cipta dan penggunaan kembali:

Lisensi ini mengizinkan setiap orang untuk mengubah, memperbaiki, dan membuat ciptaan turunan bukan untuk kepentingan komersial, selama anda mencantumkan nama penulis dan melisensikan ciptaan turunan dengan syarat yang serupa dengan ciptaan asli.

Copyright and reuse:

This license lets you remix, tweak, and build upon work non-commercially, as long as you credit the origin creator and license it on your new creations under the identical terms.

BAB I

PENDAHULUAN

1.1 Latar Belakang

Pada saat ini media sosial tentunya bukan sesuatu hal yang asing untuk di perbincangkan, sosial media telah mengubah kehidupan manusia dan cara berinteraksi manusia karena sosial media sekarang ini digunakan seseorang sebagai sarana informasi dan media hiburan bagi penggunanya (Nasution & Hayaty, 2019). Sekarang ini, hampir bisa dipastikan setiap orang yang memiliki telepon pintar, juga memiliki akun sosial media, seperti *Facebook*, *Twitter*, *Instagram*, dan lain sebagainya (Mulawarman & Nurfitri, 2017). Pertumbuhan media sosial sendiri mengubah cara manusia untuk melakukan komunikasi terhadap sesamanya, masyarakat banyak menggunakan media sosial untuk mengekspresikan opini, pengalaman maupun hal lainnya yang menjadi perhatian mereka, dimana hal-hal seperti ini sering kali disebut dengan sentimen (Muchammad Shiddieqy Hadna et al., 2016). Dengan adanya konsep dari media sosial sekarang ini menjadi agenda utama dari para pelaku bisnis, baik itu para pengambil keputusan atau konsultan bisnis, semuanya berusaha untuk mengidentifikasi apakah perusahaannya dapat mendapatkan berbagai keuntungan (Pourkhani et al., 2019). Salah satu cabang riset yang kemudian berkembang dari situasi ledakan informasi di internet adalah *sentiment analysis* (Murnawan, 2017). Dengan melakukan sebuah analisis sentimen dapat diketahui opini publik mengenai suatu produk atau jasa yang ditawarkan atau untuk melakukan sebuah penelitian (Murphy et al., 2014).

Sentiment Analysis atau sering disebut juga *opinion mining* merupakan studi komputasional dari opini-opini orang, *appraisal*, dan emosi melalui entitas, *event* dan atribut yang dimiliki (Murnawan, 2017). Analisis sentimen sendiri belakangan ini menjadi topik yang populer untuk dijadikan sebuah penelitian dikarenakan analisis sentimen dapat diaplikasikan di banyak sektor industri (Ciric et al., 2013). Hampir seluruh industri, tidak terkecuali industri pariwisata merasakan manfaat dari adanya teknologi informasi (Murnawan, 2017). Sektor pariwisata di Indonesia sendiri selalu mengalami peningkatan tiap tahunnya dan sektor pariwisata ini berada didalam ranking 42 di dunia (Hermanto et al., 2018). Salah satu tujuan wisata yang terkenal baik di Indonesia maupun dunia adalah kota bali, tidak diragukan lagi kota bali memiliki kekayaan dan keanekaragaman warisan budaya dan keindahan panorama alami (Budiasa, 2011). Teknologi sendiri telah menjadi salah satu faktor penting dalam peningkatan sektor industri pariwisata di Indonesia, hal ini dikarenakan di era digital ini hampir seluruh masyarakat terhubung kedalam sosial media selama mereka berpegiان (Hermanto et al., 2018). Dalam beberapa tahun terakhir, *twitter* menjadi salah satu media sosial yang terkenal dan banyak digunakan oleh orang-orang (Setiawan, 2015). *Twitter* adalah salah satu media sosial dan merupakan layanan *microblogging* yang mengizinkan penggunanya untuk mengirimkan pesan secara *real time* (Muchammad Shiddieqy Hadna et al., 2016). Dengan adanya *twitter* sentimen atau opini publik terhadap pariwisata disuatu kota dapat diketahui karena didalam *tweet* seseorang sering mengandung informasi penting dari suatu peristiwa yang sangat berharga untuk digunakan sebagai alat untuk mengetahui opini publik terhadap objek wisata tersebut, selain

itu dengan menggunakan *twitter* dapat diketahui acara atau topik diskusi yang sedang populer terkait pariwisata dikota tersebut dengan menggunakan tagar (Hermanto et al., 2018). Untuk dapat melakukan sebuah analisis sentimen diperlukan penguasaan beberapa teknik seperti teknik untuk melakukan *text mining*, *machine learning* dan *natural language processing (NLP)* untuk dapat memproses data yang besar dan tidak terstruktur yang berasal dari *social media* (Alves et al., 2014). Pada sebuah penelitian mencatat algoritma *Support Vector Machine* dan *Naïve Bayes* adalah algoritma yang paling sering digunakan untuk melakukan analisis sentimen (Kaur et al., 2017). Namun beberapa penelitian yang telah dilakukan sebelumnya juga mengusulkan analisis sentimen dilakukan dengan menggunakan algoritma *K-Nearest Neighbors* dan *Decision Tree* (Mardiana et al., 2019). Algoritma *Support Vector Machine* adalah salah satu algoritma *machine learning (supervised learning)* yang dapat mengklasifikasikan opini dengan mencari bidang pemisah terbaik atau *hyperplane* yang dapat memisahkan teks dengan dimensi tinggi ke dalam beberapa kelas dengan sempurna (Kusumawati et al., 2019). Sedangkan algoritma *naïve bayes* adalah salah satu algoritma yang populer digunakan untuk keperluan data mining karna kemudahan penggunaanya, waktu pemrosesan yang cepat, mudah diimplementasikan dengan strukturnya yang cukup sederhana dan tingkat efektifitas yang tinggi (Muchammad Shiddieqy Hadna et al., 2016). Algoritma lain seperti *K-Nearest Neighbors* melakukan klasifikasi berdasarkan kemiripan suatu data dengan data yang lain, prinsip sederhana dari algoritma ini adalah “Jika suatu hewan seperti bebek, dan berpenampilan seperti bebek, dan bersuara kwek-kwek seperti bebek maka kemungkinan hewan itu adalah

bebek (Nasution & Hayaty, 2019). Berbeda dari ketiga algoritma sebelumnya, *Decision Tree* adalah sebuah metode klasifikasi representasi dari sebuah *tree* atau pohon keputusan, dimana sebuah attribute dipresentasikan sebagai *node*, dan sebagai nilainya adalah cabang pohon tersebut, serta kelas dipresentasikan sebagai kelas (Somantri & Dairoh, 2019). Penelitian yang dilakukan oleh (Mardiana et al., 2019) melakukan perbandingan metode untuk melakukan klasifikasi yang diantaranya adalah algoritma *Support Vector Machine*, *Naïve Bayes*, *K-Nearest Neighbors*, dan *Decision tree* menyebutkan bahwa akurasi tertinggi diperoleh dengan menggunakan algoritma SVM. Penelitian lain yang dilakukan oleh (Romadloni et al., 2019) melakukan perbandingan algoritma *Naïve Bayes*, *KNN*, dan *Decision Tree* menyebutkan bahwa akurasi tertinggi diperoleh dengan menggunakan algoritma *Decision Tree*. Sedangkan penelitian lain yang dilakukan oleh (Aulianita, 2016) yang membandingkan algoritma *SVM* dan *KNN* menyebutkan bahwa akurasi tertinggi diperoleh dengan menggunakan algoritma *KNN*.

Dari pembahasan diatas ditemukan bahwa sosial media telah digunakan oleh lebih dari setengah penduduk dunia dan sosial media sendiri memiliki banyak manfaat. Salah satu keuntungan yang didapat dari adanya sosial media adalah dapat dilakukannya analisa sentimen atau opini publik terhadap suatu hal karena sosial media sekarang ini digunakan masyarakat untuk mengekspresikan opini atau pengalaman yang menjadi perhatian mereka. Beberapa penelitian mengenai analisa sentimen menggunakan twitter sebagai sumber data, karena twitter menyediakan API yang mengizinkan penggunanya menarik data dari aplikasi twitter. Selain itu,

twitter menjadi sosial media yang dijadikan tempat penggunaanya untuk menyampaikan opini ataupun berbagi pengalaman. Analisa sentimen sendiri sendiri dapat diaplikasikan ke banyak sektor industri, termasuk pariwisata. Untuk dapat melakukan analisis sentimen terdapat beberapa metode yang bertujuan untuk mengetahui sentimen atau opini publik terhadap suatu hal, apakah itu mengarah ke hal yang positif, negatif atau bersifat netral. Terdapat beberapa algoritma yang biasanya digunakan dalam melakukan analisis sentimen, yaitu algoritma *Naive Bayes*, *Support Vector Machine*, *K-Nearest Neighbors*, dan *Decision Tree*. Tentunya setiap metode atau algoritma yang digunakan untuk menganalisa sentimen akan memerlukan cara yang berbeda dan mengeluarkan hasil yang berbeda. Oleh karena hal tersebut, penelitian kali ini akan membandingkan keempat algoritma tersebut dalam melakukan analisis sentimen pariwisata di Indonesia, khususnya kota Bali dengan menggunakan sosial media *twitter* sebagai sumber data.

1.2 Rumusan Masalah

Dari latar belakang yang telah diuraikan, terdapat beberapa rumusan masalah yang harus dijawab pada penelitian ini, antara lain:

1. Bagaimana sentimen masyarakat terhadap pariwisata di Indonesia (studi kasus: kota Bali) dengan menggunakan social media *Twitter* sebagai sumber data?
2. Bagaimana hasil perbandingan analisis sentimen pariwisata di Indonesia (studi kasus: kota Bali) menggunakan algoritma *Naive Bayes*, *Support Vector Machine*, *K-Nearest Neighbors*, dan *Decision Tree*?

1.3 Batasan Masalah

Beberapa batasan masalah pada penelitian ini:

1. Sosial media yang akan dijadikan sumber data pada penelitian kali ini adalah *Twitter*.
2. Algoritma klasifikasi sentimen yang akan dibandingkan pada penelitian ini adalah algoritma *Naive Bayes*, *Support Vector Machine*, *K-Nearest Neighbors*, dan *Decision Tree*.
3. Nilai *Area Under Curve* akan menjadi acuan algoritma mana yang tepat untuk digunakan dalam menganalisis sentimen masyarakat terhadap kota Bali.
4. Studi kasus dalam melakukan analisis sentimen pada penelitian kali ini adalah sentimen publik terhadap kota Bali yang terkait dengan pariwisatanya, baik itu keindahan alamnya ataupun suasana di kota bali.
5. Pengumpulan data dimulai sejak tanggal 21 Maret 2020 dan data yang dapat dikumpulkan maksimal 7 hari kebelakang terhitung pada tanggal pengumpulan data dilakukan.

1.4 Tujuan Penelitian

Tujuan dari penelitian ini antara lain:

1. Mengetahui sentimen atau opini publik terhadap pariwisata Indonesia (Studi kasus: Kota Bali) dengan menggunakan media sosial *Twitter* sebagai sumber data.
2. Membandingkan algoritma *Naive Bayes*, *Support Vector Machine*, *K-Nearest Neighbors*, dan *Decision Tree* untuk mengetahui algoritma mana yang paling tepat digunakan untuk menganalisis sentimen masyarakat terhadap kota Bali.