

BAB II

KERANGKA TEORI

2.1. Analisa Sentimen

Analisa Sentimen merupakan sebuah teknik yang berorientasi untuk mempelajari mengenai persepsi dan karakteristik terhadap suatu populasi atau kelompok pasar, dan untuk menjelaskan kredibilitas dari konten dan ulasan *postingan* (Ribeiro, Araújo, G. Pollyana, & G. M. André, 2016).

Menurut (Alaei, Becken, & Stantic, 2017), Analisa Sentimen berhubungan dengan ulasan-ulasan konsumen, informasi yang diberikan mengandung dasar pemikiran atau premis melalui kalimat (seperti sebuah ulasan) yang bersifat subjektif (seperti berpendapat) atau objektif (seperti faktual).

Ulasan-ulasan bersifat subjektif berdasarkan pendapat seseorang, perasaan pribadi, kepercayaan, dan menilai suatu entitas atau acara (*events*). Ulasan objektif berdasarkan fakta, pembuktian, observasi yang terukur. Ulasan-ulasan konsumen dan *post* pada sosial media secara tanpa sadar bahwa mengungkapkan sesuatu seperti kegembiraan, frustrasi, kekecewaan, menyenangkan dan perasaan-perasaan lainnya.

Analisa sentimen saat ini memiliki peran penting dalam area penelitian di berbagai bidang pada era digitalisasi baik yang berkaitan dengan produk ataupun jasa yang di tawarkan kepada konsumen sehingga diperoleh *feedback* berupa ulasan terhadap produk atau jasa yang telah digunakan konsumen melalui media sosial, website dan lain-lain.

Dua metode pendekatan yang umum dalam analisis sentimen adalah metode *lexicon-based* dan pendekatan *machine learning*, dimana pendekatan *lexicon-based* membutuhkan ketersediaan atau sebuah konstruksi apriori yang cocok dengan kamus (*lexicon*), sedangkan metode *machine learning* secara otomatis mengklasifikasi teks tetapi biasanya membutuhkan set pelatihan data (*data training sets*) yang berasal dari pemrograman manusia (Ge, Vazquez, & Ulrike, 2018; Paltoglou & Thelwall, 2017).

Pada metode *supervised*, pengklasifikasi menggunakan unigrams sebagai fitur. Dalam metode *lexicon-based*, *unigram* yang ditemukan dalam *lexicon* diberi skor *polarity*, skor *polarity* keseluruhan teks kemudian dihitung sebagai jumlah dari polaritas *unigram* (Kolchyna, Souza, Treleaven, & Aste, 2015).

Menurut (Alaei et al., 2017), analisa sentimen merepresentasikan masalah klasifikasi polaritas. Tersusun dari beberapa *classes* yang berbeda, klasifikasi polaritas sentimen (*sentiment polarity classification*). Pada klasifikasi binary (*binary classification*), kita dapat mengasumsikan bahwa seorang konsumen memberikan ulasan dimaksud subjektif. Sedangkan klasifikasi binary (*binary classification*) dapat diasumsikan kalimat–kalimat yang mengandung unsur terutama positif atau negatif, dan kemudian akan menjelaskan apakah ulasan-ulasan yang diberikan bersifat positif atau negatif. Definisi dari kedua sisi sentimen bersifat positif dan negatif bergantung dari aplikasi dan domain.

Menurut (Fuchs, Lexhagen, Hopken, & Schmunk, 2014), Perbandingan spesifik analisa sentimen dengan *text mining* secara umum terletak pada

ketertarikan suatu objek pada teks. *Text mining* berhubungan dengan mengidentifikasi pernyataan yang bersifat subjektif dan mengandung opini atau pendapat seseorang dan sentimen. Sedangkan analisa sentimen lebih melakukan penelitian secara mendalam terhadap ketertarikan masa kini. Pendekatan untuk mengklasifikasi dari analisa sentimen dibagi menjadi empat kategori: *machine learning*, berdasarkan kamus (*dictionary-based*), *statistical* dan *semantic approaches* (Fuchs, Lexhagen, Hopken, & Schmunk, 2014).

2.2. Text Mining

Text Mining bagian dari *data mining*, yang membedakan pada sumber datanya. *Text mining* merupakan pengolahan kumpulan-kumpulan dokumen dalam jumlah besar yang terakumulasi dari waktu ke waktu dengan menggunakan beberapa teknik analisis. Tujuan dengan adanya pengolahan teks adalah untuk menciptakan suatu informasi menjadi nilai yang memiliki arti dengan cara mengetahui dan mengekstrak informasi yang berguna dari sumber data dengan mengidentifikasi dan mengeksplorasi pola (Nurhadi, 2015).

Pada *text mining*, sumber data yang digunakan terdiri dari kumpulan atau koleksi dokumen terstruktur, dan tidak terstruktur perlu adanya pengelompokan untuk diketahui informasi yang seragam, Rosanensi et al., (2018) dan Sinaga, (2017). Dengan penerapan dan mengembangkan *text mining* akan muncul teknik-teknik baru dari bidang lain yang dapat memberikan solusi baru, seperti *machine learning*, *data mining*, *natural language processing (NLP)*, *linguistic*, *statistik* dan *matematika*, dan *visualization* (Nurhadi, 2015).

Tingkatan dalam *text mining* secara umum terdapat pada *pre-processing* teks dan seleksi fitur (Rosanensi et al., 2018). Menurut (Nurhadi, 2015), proses pengolahan data terdiri dari serangkaian tahapan seperti tahap pengumpulan data, tahap *pre-processing* yang di dalamnya terdapat proses kategorisasi (*categorization*), tahap ekstraksi fitur (*feature/term extraction*), setelah itu dokumen dilakukan pemberian *keyword*, *labeled* dan *time-stamped*, dan tahap terakhir proses *core mine operations* dan *presentation*.

2.3. *Lexicon-based Approach*

Pendekatan yang sering digunakan dalam melakukan analisa sentimen adalah pendekatan kamus atau leksikal (*Lexicon-based*). Metode ini disebut *lexical based approach*. *Lexical-based approach* adalah metodologi untuk melakukan analisa sentimen menggunakan kamus sebagai sumber bahasa atau yang berhubung dengan kamus (leksikal) (Rosanensi et al., 2018).

Pendekatan ini menggunakan sekumpulan kata-kata opini bersifat positif dan negatif untuk mengukur suatu kalimat itu digolongkan ke dalam kalimat positif, negatif atau netral. Pendekatan ini sangat populer dan membutuhkan fungsi skoring untuk menghitung nilai dari setiap kalimat berdasarkan kemunculan kata-kata positif atau negatif pada kamus (Puspitarani, 2015).

2.4. Labuan Bajo

Labuan Bajo ialah kota kecil yang terdapat di Indonesia, dengan nuansa penampakan yang tenang dan indah, serta tergolong memiliki keindahan alami yang luar biasa ke tujuh di dunia (“About Labuan Bajo,” 2019). Labuan Bajo terletak di ujung Flores bagian Barat dan untuk menjelajahi penampakan Flores yang eksotik

sangat di rekomendasikan mulai dari *Varanus Komodoensis*, Taman Nasional Komodo dengan keindahan kehidupan bawah laut, hingga tradisi memburu Paus.

Tempat wisata Labuan Bajo mempunyai banyak tempat wisata yang sangat layak dikunjungi yang memberikan pengalaman liburan yang luar biasa, dan sudah banyak wisatawan dari lokal ataupun mancanegara yang mengunggah foto dan *video* hasil perjalanan liburan mereka ke Labuan Bajo pada media sosial *Instagram* yang di tambahkan *Hashtags* #labuanbajo tercatat sekitar 720.000 *posting*.

Terdapat media berita yaitu *Metrotv news* meliput informasi mengenai wisatawan yang berkunjung ke Labuan Bajo meningkat pada tahun 2018 sebesar 163.807 orang, sedangkan di tahun 2019 sebesar 184.206 dimana kunjungan ini di dominasi oleh kunjungan wisatawan mancanegara. Peningkatan kunjungan yang terjadi juga berdampak pada pendapatan asli daerah dengan angka tahun 2018 mencapai 34 miliar dan di tahun 2019 mencapai 60 miliar. Kunjungan wisatawan mancanegara terbanyak berasal dari Jerman, Inggris, Perancis dan Spanyol.

Ditambah pemerintah di era Presiden Jokowi khususnya melakukan rapat pengembangan destinasi pariwisata Labuan Bajo, Kabupaten Manggarai Barat, Nusa Tenggara Timur, pada tanggal 20 Januari 2020 yang disampaikan melalui akun *Youtube* Sekretariat Presiden melalui link https://youtu.be/y1_AvyKP-70. Pengembangan destinasi wisata Labuan Bajo yang bersegmen super premium akan dimulai awal tahun 2020. Presiden Jokowi menekankan beberapa hal dalam rapat, seperti penataan kawasan yang mencakup 5 zona yang harus di tata di Labuan Bajo yaitu Bukit Pramuka, Kampung Air, Pelabuhan Peti Kemas, dermaga penumpang, kawasan marina dan di zona Kampung Ujung.

Kedua berkaitan dengan infrastruktur, seperti landasan pacu dan terminal bandara. Ketiga berkaitan penyiapan SDM, keempat mengenai sampah baik sampah darat maupun sampah laut, kelima berkaitan dengan air baku, keenam berkaitan keamanan para wisatawan dan terakhir mengenai promosi.

2.5. Instagram

Instagram merupakan aplikasi untuk membagikan foto dan *video* yang memiliki fitur-fitur pada *Instagram* seperti pengikut, mengunggah foto, kamera, efek foto, judul foto, *Arroba*, label foto, perlombaan, publikasi kegiatan, publikasi organisasi, *Geotagging*, jejaring sosial, tanda suka, *Popular*, peraturan *Instagram*, penandaan foto dengan bendera, dan banyak fitur tambahan lainnya. Pada tahun 2019, pengguna aplikasi Instagram di Indonesia mencapai 62 juta pengguna aktif (“Indonesia Digital 2019 : Media Sosial - Websindo,” 2019).

Alasan penelitian menggunakan media *Instagram* sebagai analisa sentimen terhadap tempat wisata Labuan Bajo, karena banyak tanggapan, komentar dan *caption* teks untuk diolah mengenai Labuan Bajo dari foto yang di unggah ke Instagram akan memudahkan untuk memperoleh sumber data untuk dilakukan analisis sentimen.

2.6. Hashtag

(Zhang & Murphy, 2016) menjelaskan *hashtags* sebagai kata-kata yang tidak memiliki spasi, merupakan ringkasan atau singkatan, kalimat yang diikuti dengan simbol #. *Hashtag* dapat di klasifikasi untuk menginformasikan dan *hashtag* untuk mengomentari (seperti, opini dan penilaian). Penggunaan yang relevan, *hashtags* yang terarah pada *posts* dan *stories* adalah langkah terbaik untuk menemukan

pendatang baru (*new audience*) pada Instagram. Hal ini dapat diartikan ke dalam untuk membangun hubungan, untuk menambah pengikut, dan menambah konsumen pada bidang bisnis. Seperti Twitter dan *platform* lainnya, *Instagram hashtag* berfungsi untuk mengelola dan mengategorikan konten foto, gambar dan *video*.

Jika memiliki sebuah akun Instagram publik dan menambahkan sebuah *hashtag* pada *post*, dimana *post* tersebut dapat muncul pada halaman *hashtag* yang sesuai. Dengan adanya penggunaan *hashtag* untuk mencari sebuah konten yang belum pernah dikunjungi, pengetikan *hashtag* yang tepat dapat oleh pengguna dapat merujuk pada konten tersebut, walau mereka belum saling berteman pada sosial medianya.

Contoh pada *travel blogger* ingin mengunggah sebuah gambar berpose foto di depan komodo dan kemudian ditambahkan *hashtag* #labuanbajotrip, #labuanbajo, #komodoisland, #wonderfulindonesia, #ilovelabuanbajo dan lain-lain ketika akan di unggah ke *Instagram*. Dengan menggunakan beberapa *hashtags* seperti pada Gambar 2.1. *Hashtags Instagram*, gambar tersebut di katalog sehingga pengguna *Instagram* dengan mudah untuk menemukannya.

Menurut (Zhang & Murphy, 2016), *hashtags* diklasifikasi ke dalam topik (#labuanbajo), sentimen (#beautiful), dan topik-sentimen (#ilovelabuanbajo) dan kebanyakan *hashtags* ditemukan untuk menandai topik dan keterangan kata kunci. *Hashtag* juga dapat dilakukan sentimen. *Hashtag* yang mengandung emosional menunjukkan sebagai sebuah data, pengetahuan atau objek baru yang dapat diteliti.

Selain itu, hashtag positif dan negatif juga dapat diterapkan pada komentar atau ulasan untuk menunjukkan kepuasan seseorang yang disertakan kata sentimen.



Gambar 2.1. Hashtags Instagram

2.7. Python

Python merupakan bahasa pemrograman tingkat tinggi yang dibuat oleh Guido van Rossum. Bahasa ini sangat populer dengan kepraktisan untuk membaca *code* dan garis *code* (Jessica & Oetama, 2020). Menggunakan *white spaces* untuk membatasi blok. Python menyediakan *library* dengan standar yang besar untuk digunakan berbagai aplikasi seperti pemrosesan bahasa alami (*Natural Language Processing* disingkat NLP), pembelajaran mesin (*machine learning*), analisis data, dan lainnya.

Instagram-scraper adalah package dari *script* yang digunakan untuk mengambil data *Instagram* seperti foto, *caption*, *comment* dan lain-lain, dengan

menggunakan pemrograman *Python* agar konten bisa diambil. Instagram-scraper untuk saat ini belum ada batasan untuk menambang data tersebut dan tidak memerlukan REST API seperti Twitter untuk mengakses dan membaca data twitter.

2.8. Confusion Matrix

Confusion Matrix digunakan untuk menyimpulkan kinerja klasifikasi dari suatu pengklasifikasi terhadap suatu data uji (Wiratama & Rusli, 2020). *Confusion Matrix* merupakan matriks dua dimensi, satu dimensi di *indeks* oleh kelas sebenarnya dari suatu objek dan dimensi lainnya di *indeks* oleh kelas yang ditentukan oleh pengklasifikasi.

Tabel 2.1. Confusion Matrix

	Prediksi	Positif	Negatif
Aktual			
Positif		TP (<i>True Positive</i>)	FN (<i>False Negative</i>)
Negatif		FP (<i>False Positive</i>)	TN (<i>True Negative</i>)

Pada pengukuran kinerja menggunakan *confusion matrix*, terdapat 4 istilah sebagai representasi hasil proses klasifikasi. Keempat istilah tersebut adalah TP, TN, FP, dan FN. Nilai *false negative* (FN) adalah jumlah data negatif namun terklasifikasi salah oleh sistem, nilai *true negative* (TN) adalah jumlah data negatif yang terklasifikasi dengan benar oleh sistem, Nilai *false positive* (FP) adalah jumlah data positif namun terklasifikasi salah oleh sistem, dan Nilai *true positive* (TP) adalah jumlah data positif yang terklasifikasi dengan benar oleh sistem.

Keempat nilai tersebut digunakan untuk perhitungan performa model. Nilai akurasi (*accuracy*) digunakan untuk mengukur ketepatan prediksi yang dihasilkan. Nilai *accuracy* sendiri tidak cukup untuk mengukur performa model, maka diperlukan metrik pengukuran lainnya untuk membantu pengukuran performa

model (Wiratama & Rusli, 2020). Nilai *precision*, *recall*, dan *F1 Score* digunakan untuk mendukung performa model.

Nilai *precision* untuk mengukur perbandingan prediksi positif yang benar (*true positive*) terhadap total prediksi positif. Nilai *recall* mengukur perbandingan prediksi positif yang benar (*true positive*) terhadap semua prediksi di kelas aktual. Sedangkan *F1 Score* merupakan rata-rata berbobot dari nilai *precision*, dan *recall*. Perhitungan nilai *accuracy*, *precision*, *recall* dan *F1 Score*, seperti di bawah ini.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} * 100\%$$

Rumus 2.1. Rumus Accuracy

$$Precision = \frac{TP}{(TP + FP)} * 100\%$$

Rumus 2.2. Rumus Precision

$$Recall = \frac{TP}{(TP + FN)} * 100\%$$

Rumus 2.3. Rumus Recall

$$F1 = \frac{(2 \times Precision \times Recall)}{(Precision + Recall)} * 100\%$$

Rumus 2.4. Rumus F1 Score atau F-Measure

Dimana:

- TP adalah *True Positive*, yaitu jumlah data positif yang terklasifikasi dengan benar oleh sistem.
- TN adalah *True Negative*, yaitu jumlah data negatif yang terklasifikasi dengan benar oleh sistem.

- FN adalah *False Negative*, yaitu jumlah data negatif namun terklasifikasi salah oleh sistem.
- FP adalah *False Positive*, yaitu jumlah data positif namun terklasifikasi salah oleh sistem
- *Precision* adalah tingkat ketepatan antara informasi yang diminta oleh pengguna dengan jawaban yang diberikan oleh sistem.
- *Recall* adalah tingkat keberhasilan sistem dalam menemukan kembali sebuah informasi.
- *Accuracy* didefinisikan sebagai tingkat kedekatan antara nilai prediksi dengan nilai aktual.

2.9. *Natural Language Toolkit (NLTK)*

Natural Language Toolkit dibuat pada tahun 2001 sebagai bagian dari mata kuliah komputasi linguistik pada *Department of Computer and Information Science* di *University of Pennsylvania*. NLTK memiliki sejumlah modul penting, dapat dilihat pada Tabel 2.2. Modul NLTK.

Penelitian ini akan menggunakan modul `nltk.corpus` untuk melakukan penghapusan *stopword* bahasa Inggris dan diterapkan untuk analisa sentimen metode *lexicon-based*.

Tabel 2.2. Modul NLTK

<i>Language Processing Task</i>	<i>NLTK Modules</i>	<i>Functionality</i>
<i>Accessing corpora</i>	<i>nltk.corpus</i>	<i>Standardized interfaces to corpora and lexicons</i>
<i>String processing</i>	<i>nltk.tokenize,</i> <i>nltk.stem</i>	<i>Tokenizer, sentence tokenizers, stemmers</i>
<i>Collocation discovery</i>	<i>nltk.collocations</i>	<i>t-test, chi-squared, point-wise mutual information</i>

<i>Part-of-speech-tagging</i>	<i>nlk.tagging</i>	<i>n-gram, backoff, Brill, HMM, TnT</i>
<i>Classification</i>	<i>nlk.classify, nlk.cluster</i>	<i>Decision tree, maximum entropy, naïve Bayes, EM, K-means, SVM</i>
<i>Chunking</i>	<i>nlk.chunk</i>	<i>Regular expressions, n-gram, named entity</i>
<i>Parsing</i>	<i>nlk.parse</i>	<i>Chart, feature-based, unification, probabilistic, dependency</i>
<i>Semantic interpretation</i>	<i>nlk.sem, nlk.inference</i>	<i>Lamda calculus, first-order logic, model-checking</i>
<i>Evaluation metrics</i>	<i>nlk.metric</i>	<i>Frequency distributions, smoothed probability distributions</i>
<i>Application</i>	<i>nlk.app, nlk.chat</i>	<i>Graphical concordance, parsers, WordNet browser, chat-bots</i>
<i>Linguistic fieldwork</i>	<i>nlk.toolbox</i>	<i>Manipulate data in SIL Tool-box format</i>

Sumber: (Tunggawan & Soelistio, 2017)

2.10. Excel

Excel merupakan alat bantu untuk pengolahan data atau memanipulasi data dalam bentuk teks dan angka dari Microsoft Office. Microsoft Excel sebuah aplikasi yang dapat mengelola data secara otomatis melalui operasi perhitungan dasar seperti penjumlahan, pengurangan, pembagian dan perkalian, penggunaan rumus dan fungsi yang hasilnya dapat ditampilkan dalam bentuk angka, tabel atau grafik, dan Excel memiliki lembar kerja yang disebut *spreadsheet* yang terdiri dari baris dan kolom untuk menempatkan data (“Introduction to Excel Starter - Excel,” 2010).

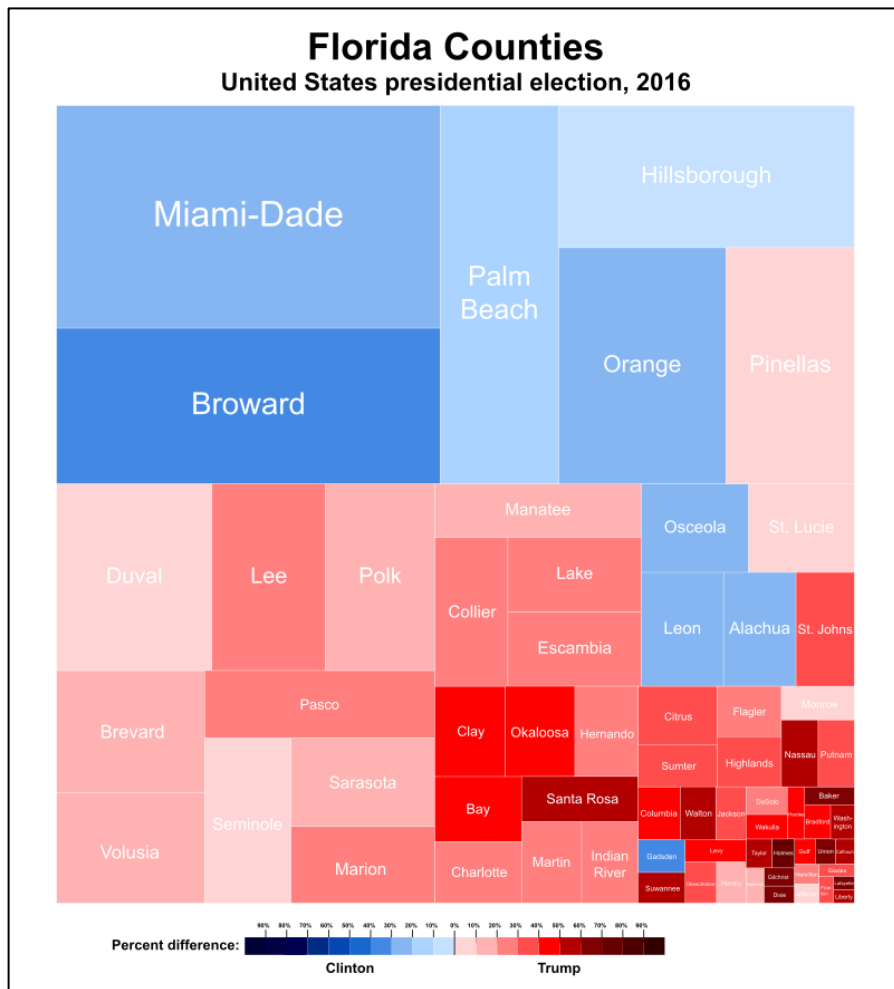
2.11. Google Sheet

Google *Sheet* merupakan salah satu komponen dalam aplikasi Google *cloud*. Google *sheet tools* dari Google secara komprehensif menawarkan sekumpulan fitur dan fungsi standar *spreadsheet* seperti aplikasi Microsoft Excel, serta terdapat dukungan terbaru seperti fungsi *Query* dan *Regular Expression* (“Google Sheets: Getting Started with Google Sheets,” 2020).

2.12. Visualisasi Data

Visualisasi data terdapat 3 kriteria agar menghasilkan sebuah informasi yang berharga, yaitu (1) proses didasarkan pada data kualitatif dan kuantitatif, (2) menghasilkan keluaran yang berasal dari data awal, dan keluaran dapat dibaca oleh penggunaan dan mendukung proses eksplorasi, pemeriksaan dan komunikasi data (Azzam, Evergreen, Germuth, & Kistler, 2013). Visualisasi data bertujuan untuk mengkomunikasikan dan menyajikan data secara jelas, efektif dan interaktif dalam bentuk grafis, *chart*, peta agar tampilan menarik namun tetap informatif (Azzam et al., 2013; Prasetya & Susilowati, 2016).

Visualisasi dapat ditampilkan dalam beberapa chart, salah satunya *Treemap*. *Treemap* adalah representasi dari pohon data, setiap *node* dapat tidak memiliki atau memiliki satu atau lebih anak *node* dan satu *node parent*. Setiap *node* di tunjukkan dengan bentuk persegi yang berbeda warna sesuai dengan nilai atau *value*, seperti *tree view*, warna dapat dipetakan ke dalam nilai atau kategori, tetapi akan lebih baik untuk membaca total area daripada dengan *tree view*.



Gambar 2.2. Treemap

Sumber: (“TreeMap,” 2016)

2.13. Google Translate

Google *Translate* merupakan mesin penerjemah bahasa yang tersedia secara online atau diakses melalui browser internet. Google translate sebuah mesin penerjemahan bahasa yang tersedia terjemahan ke dalam beberapa bahasa, dan dapat menerjemahkan kalimat teks lebih dari 90 bahasa.

Google *Translate* banyak digunakan pada kalangan pelajar, yang biasanya digunakan untuk memudahkan pengguna untuk memahami informasi seperti teks

yang ditulis bahasa Inggris. Para pelajar tidak sepenuhnya mengandalkan Google *Translate*, tetapi diperlukan pengeditan atau penyesuaian selama proses menggunakannya agar tidak memberikan kesalahpahaman dari teks atau informasi yang dibaca (Gestanti, Nimasari, & Mufanti, 2019).

2.14. *Rapid Miner*

Rapid miner merupakan alat untuk pengelolaan data yang tersedia open source. Rapid miner merupakan software yang sering digunakan oleh data science dan dikembangkan oleh Ralf Klinkenberg, Ingo Mierswa dan Simon Fischer (Pynam, Spanadna, & Srikanth, 2018). Rapid Miner memberikan banyak fungsi seperti digunakan untuk *data preparation, machine learning, deep learning, text mining, predictive analytical tools, business analytic* dan *analysis sentiment tools*.

2.15. Penelitian Terdahulu

Dalam penelitian yang akan dilakukan, digunakan beberapa jurnal penelitian sebelumnya sebagai media referensi dan pembelajaran untuk mendukung informasi-informasi pembuatan skripsi ini. Berikut ini adalah jurnal-jurnal yang dipakai selama penelitian:

Tabel 2.3. Jurnal Terdahulu

No.	Judul	Nama Penulis (Tahun)	Publikasi Jurnal
1	Metode <i>Lexicon Based</i> Dan <i>Support Vector Machine</i> Untuk Menganalisis Sentimen Pada Media Sosial Sebagai Rekomendasi Oleh-Oleh Favorit	(Himawan et al., 2018)	Seminar Nasional Informatika 2018
Ringkasan			
Hasil Pengujian sistem dengan metode <i>Lexicon-based</i> menghasilkan tingkat akurasi, presisi, dan <i>recall</i> , yaitu 87.78%, 94.23%, dan 91.87%.			

Hasil pengujian sistem dengan metode SVM menghasilkan tingkat akurasi, presisi dan *recall*, yaitu 86.19%, 86.19%, dan 100%.

Hasil pengujian pada sistem yang dibangun memperlihatkan bahwa metode *lexicon based* memberikan akurasi dan presisi yang lebih baik daripada metode *support vector machine* dengan akurasi sebesar 87.78% dan presisi sebesar 94.23%. Sedangkan untuk hasil *recall*, metode *support vector machine* lebih baik daripada metode *lexicon based* dengan *recall* sebesar 100%

Metode *lexicon based* sangat bergantung pada jumlah data kamus sentimen yang digunakan, semakin banyak jumlah data kamus yang digunakan maka semakin besar nilai keakurasiannya, dan sebaliknya semakin sedikit jumlah data kamus yang digunakan maka semakin kecil nilai keakurasiannya yang dihasilkan.

Pembahasan

Pada jurnal ini mengenai analisa sentimen dengan menggunakan metode salah satunya *lexicon-based* dengan data dari Instagram, hasil yang diperoleh untuk menentukan rekomendasi oleh-oleh favorit melalui penggunaan *Hashtag*.

Untuk menghitung akurasi dengan metode SVM menggunakan *k-fold cross validation*, sedangkan *lexicon-based* untuk menguji performa akurasi menggunakan metode *confusion matrix*.

No.	Judul	Nama Penulis (Tahun)	Publikasi Jurnal
2	Pemanfaatan Analisis Sentimen Untuk Peningkatan Popularitas Tujuan Wisata	(Sinaga, 2017)	Jurnal Penelitian Pos dan Informatika, Vol.07 No 02 Desember 2017

Ringkasan

Metode klasifikasi yang digunakan menggunakan Naïve Bayes.

Data yang digunakan destinasi wisata Bali berasal dari Facebook, Instagram, dan forum, dari hasil pemeringkatan 5 tertinggi yaitu, pantai Kuta, pantai Sanur, Pantai Lovina, Tanjung Bena, dan Pantai Legian dan Seminyak.

Pengujian validitas akurasi algoritma mendapatkan hasil prosentase yang digolongkan menjadi 2 pengujian, yaitu 100 frase memperoleh akurasi sebesar 65.65%, dan 5000 frase memperoleh 82.67%, sehingga kesimpulan semakin banyak frase yang dimiliki sebagai core dari algoritma, maka semakin akurat analisa sentimen yang disajikan.

Pembahasan

Jurnal ini mengenai analisa sentimen untuk pemeringkatan destinasi popularitas destinasi wisata dari berbagai sumber *online*, salah satunya yang sesuai dengan penelitian yang akan dilakukan berasal dari media sosial Instagram dan teknik yang teknik yang dapat dipakai pada tahap *text pre-processing* dan metode pemeringkatan.

No.	Judul	Nama Penulis (Tahun)	Publikasi Jurnal
3	Sentiment Analysis : Extracting Decision- Relevant Knowledge from UGC	(Fuchs et al., 2014)	Information and Communication Technologies in Tourism 2014
Ringkasan			
<p>2 Metode pendekatan yang digunakan, yaitu <i>dictionary</i> dan <i>machine learning</i>. Inti penambangan dari analisa sentimen dibagi 3 kategori <i>recognition</i>, yaitu <i>recognition of property</i>, <i>recognition of sentimen</i>, <i>recognition of subjectivity</i>.</p> <p><i>dictionary-based</i> yang memiliki hasil pencapaian klasifikasi yang baik yaitu 71.28%. Selain itu untuk membantu pengambilan keputusan yang baik berdasarkan pengetahuan terutama dalam bagian bidang bisnis destinasi wisatawan, dan data warehouse.</p> <p>Untuk mengklasifikasi dengan metode SVM untuk menentukan suatu kalimat dalam sifat positif, dan negatif, hasil akurasi yang di peroleh mencapai 76.80% dengan tambahan menggunakan bigrams dalam pre-processing data.</p> <p>Metode data mining memiliki pendekatan secara otomatis mengekstrak dan menganalisa teks secara bebas seperti tanggapan konsumen dari ulasan online atau platform travel dan khususnya pendekatan klasifikasi.</p>			
Pembahasan			
<p>Pendekatan <i>dictionary-based</i> approach tidak memerlukan sebuah model pelatihan (<i>data training</i>), tetapi berdasarkan kumpulan-kumpulan kata yang terkandung pada suatu kalimat yang lebih dominan. Kumpulan kata yang dimaksud kata-kata positif, dan kata negatif.</p>			