



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 III

METODOLOGI PENELITIAN

3.1 Objek Penelitian

Objek yang menjadi sasaran penelitian ini adalah sentimen atau opini masyarakat mengenai pariwisata di kota Bali. Bali merupakan primadona salah satu daerah tujuan pariwisata di Indonesia yang mendatangkan wisatawan, hal ini dikarenakan pulau bali terkenal di mancanegara dan nusantara, sehingga bali banyak dikunjungi oleh wisatawan, baik wisatawan asing maupun lokal (Agribisnis et al., 2014). Pariwisata di kota Bali sendiri telah tumbuh sejak abad ke – 17 pada masa penjajahan belanda, tidak hanya wisatawan belanda teteapi wisatawan inggris dan lainnya juga berkunjung ke pulau Dewata ini sehingga bali sering disebut sebagai *the Island of Paradise, the Island of Gods* dan lain sebagainya (Malik et al., 2016). Alasan kota Bali menjadi destinasi wisata terkenal didunia karena kekayaan dan keanekaragaman budaya, keindahan paronama, alami dan infrastruktur dasar yang tersedia (Budiasa, 2011).

Penelitian ini akan menggunakan sosial media *Twitter* sebagai sumber data yang kemudian akan dianalisis untuk mendapatkan sentimen masyarakat terhadap kota Bali. Twitter adalah salah satu media sosial yang bertipe *microblogging* sebagai layanan interaksinya dan twitter telah menjadi salah satu sosial media yang paling terkenal di dunia dengan lebih dari 200 juta pengguna aktif dan lebih dari 10.6 miliar tweet yang telah dihasilkan (Nasution & Hayaty, 2019). Twitter didirikan pada maret 2006 oleh Jack Dorsey, pertumbuhan jumlah penggunanya

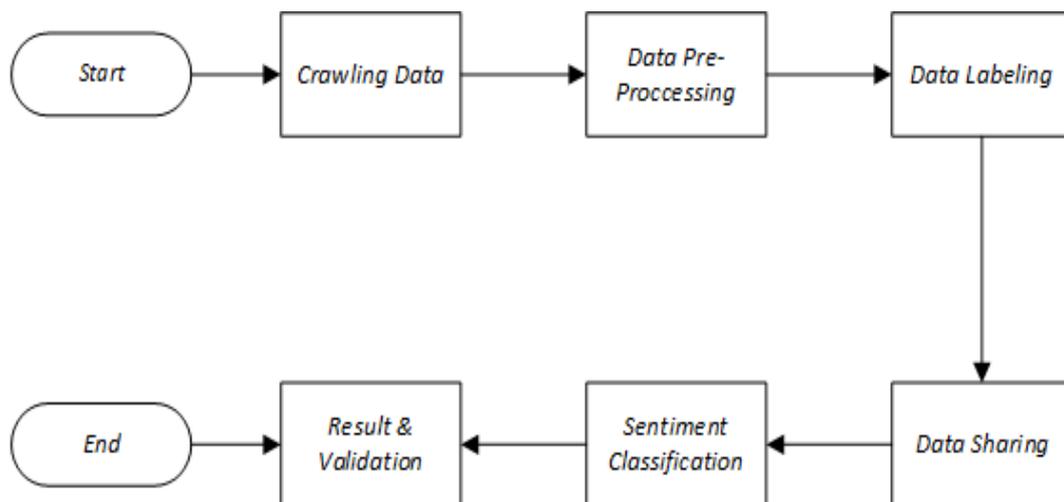
meningkat dengan pesat sehingga twitter memiliki lebih dari 300 juta pengguna dan berada di peringkat kedua sebagai jejaring sosial yang paling dikunjungi di dunia (Saputra, 2017). Twitter diciptakan untuk menjadi tempat penggunanya untuk saling berbagi pengalaman antar sesama tanpa adanya sekat penghalang, dengan adanya twitter penggunanya akan mudah untuk mengikuti tren, cerita, informasi, dan berita dari seluruh penjuru dunia (Muchammad Shiddieqy Hadna et al., 2016). Dalam penelitian ini *Twitter* dipilih untuk dijadikan sumber data karena twitter telah menyediakan API (Application Programming Interface) yang ditujukan kepada pengembang sistem untuk mempermudah pengambilan data dari twitter. Penelitian ini akan berfokus kepada hasil analisis sentimen masyarakat terhadap pariwisata di kota Bali, apakah sentimen masyarakat lebih mengarah ke hal yang positif, negatif ataupun netral. Terdapat 4 algoritma yang akan digunakan untuk membuat pemodelan analisis sentimen pada penelitian kali ini, yaitu algoritma *Naive Bayes*, *Support Vector Machine*, *K-Nearest Neighbors*, dan *Decision Tree*.

3.2 Teknik Pengumpulan Data

Pengumpulan data dilakukan dengan menggunakan library *rtweet* menggunakan aplikasi R Studio dan menghasilkan data yang dapat disimpan kedalam format JSON atau csv. Pengumpulan data dilakukan dengan menggunakan *Twitter Streaming API*. Twitter sendiri telah memberikan API ini untuk memudahkan para pengembang sistem untuk mengambil data dari aplikasi *twitter*. Total *tweet* yang dikumpulkan berjumlah 6542 data dan disimpan kedalam format csv. Potongan source code untuk crawling data dapat dilihat pada lampiran

3.3 Metodologi Penelitian

Metode penelitian yang digunakan dalam penelitian ini diadaptasi dari (Kusumawati et al., 2019) dengan sejumlah penyesuaian. Penelitian ini terdiri dari *crawling data*, *data-pre-processing*, *data labeling*, *data sharing*, *sentiment classification*, dan *result & validation*. Alur tahapan penelitian dapat dilihat pada gambar 3.1 dibawah ini.



Gambar 3. 1 Kerangka Pikir

Pengumpulan data dilakukan dengan menggunakan library *rtweet* pada aplikasi R Studio dan menghasilkan data berupa tweet dari akun twitter yang melakukan *tweet* mengenai "Bali" dalam bentuk file *csv*. Tahap data *pre-processing* dilakukan untuk menghilangkan *URL*, *punctuation*, simbol, *special character*, serta penghapusan *stopwords*. Data yang telah melalui tahap *pre-processing* akan diberikan label apakah kalimat yang ada pada *tweet* tersebut memiliki sentimen positif, negatif atau netral. Setelah diberikan label data tersebut dibagi menjadi ke dua bagian, yaitu *training* data dan *testing* data.

Data training akan digunakan untuk mengklasifikasikan sentimen dengan menggunakan algoritma *SVM*. Sedangkan data testing digunakan untuk melakukan validasi terhadap data training, hasil klasifikasi data training akan di bandingkan dengan data testing sehingga akan menghasilkan akurasi algoritma *SVM* dalam melakukan analisis sentimen

3.3.1 *Crawling Data*

Pengumpulan data yang berupa tweet dikumpulkan dengan menggunakan *API* yang telah disediakan oleh twitter, *API* ini mengizinkan penggunaanya untuk melakukan pengambilan data dari aplikasi *twitter* maksimal 7 hari sebelumnya. Untuk dapat melakukan pengambilan data tweet digunakan aplikasi R Studio. Versi R yang digunakan merupakan versi 3.6.2 dan *library* yang digunakan untuk pengambilan data adalah *library rtweet*. Pada penelitian ini, *tweet* yang diambil adalah yang mengandung *keyword* “Bali” untuk mendapatkan opini atau pengalaman pengguna twitter mengenai pariwisata di kota bali. *Tweet* yang sudah dikumpulkan ditulis dalam bahasa Inggris dan disimpan dalam bentuk csv.

3.3.2 Data Pre-Processing

Pada tahapan ini dan selanjutnya masih menggunakan R Studio sebagai *tools*. Tahapan ini akan membersihkan *tweet* yang dikumpulkan dan melakukan beberapa perubahan.

1. Mengubah semua kata menjadi huruf kecil, semua kata diubah menjadi huruf kecil. Hal ini dilakukan karena *library* yang ada pada aplikasi R akan membaca kata per kata dalam huruf kecil
2. Menghapus *punctuation*, semua tanda baca yang ada pada *tweet* akan dihapus
3. Menghapus username, apabila dalam suatu *tweet* menyebut atau *mention user* lain, maka username tersebut akan dihapus
4. Menghapus *hashtag*, menghapus semua simbol “#” pada data *tweet* yang telah dikumpulkan
5. Menghapus URL, *tweet* yang diambil dari *twitter* mengandung banyak URL baik itu http atau https ataupun *link* gambar sehingga perlu dihapus untuk mempermudah proses labeling
6. Menghapus RT atau *Retweet*, semua kata RT atau retweet akan dihapus karena dinilai tidak memiliki makna
7. Menghapus *emoticon*, menghapus *emoticon* atau emoji yang ada pada *tweet*
8. Menghapus pengulangan data, *tweet* yang diambil teridentifikasi memiliki beberapa pengulangan data yang perlu dihapus untuk mempermudah proses analisa sentimen

3.3.3 Data Labeling

Setelah melakukan penarikan dan pembersihan data, maka langkah selanjutnya adalah memberikan *label* kepada setiap tweet yang didapat. Data yang diberi *label* adalah data yang mengandung opini user terhadap pariwisata di kota Bali. Data tweet yang mengandung iklan dan tidak relevan dengan pariwisata di kota Bali akan dihapus. Data labeling digunakan untuk memberikan label positif atau negatif dari setiap tweet yang ada. Pada penelitian ini, proses labeling secara manual akan dilakukan oleh narasumber. Narasumber yang dipilih merupakan orang yang memiliki pemahaman bahasa Inggris yang baik (Lulusan Sastra Inggris). Narasumber dapat memberikan label positif, negatif atau netral terhadap tweet yang berkaitan dengan pariwisata di kota Bali. Label yang diberikan oleh narasumber berdasarkan interpretasinya terhadap isi dari tweet.

3.3.4 Data Sharing

Pada tahapan ini data yang telah diberikan label akan dibagi kedalam menjadi 2 bagian, yaitu data *training* dan data *testing*. Algoritma yang akan digunakan merupakan algoritma *Supervised Learning*, yaitu metode belajar dengan adanya latihan dan pelatih. Dalam pendekatan ini, untuk menemukan fungsi keputusan, fungsi pemisah atau fungsi regresi, diperlukan data training sebagai contoh data yang mempunyai output atau label selama proses training (Ridwan et al., 2013). Berdasarkan penelitian yang dilakukan (Fitriyaningsih & Basani, 2019) menyebutkan bahwa prediksi terbaik didapat dengan menggunakan rasio 70 : 30

3.3.4 Sentimen Classification

Proses *sentiment classification* akan membandingkan 4 algoritma klasifikasi sentimen, yaitu algoritma *Naive Bayes*, *Support Vector Machine*, *K-Nearest Neighbors*, dan *Decision Tree* dengan menggunakan pariwisata di kota bali sebagai objek penelitian. Hal ini dilakukan untuk mengetahui algoritma yang tepat digunakan untuk melakukan analisa sentimen masyarakat terhadap kota Bali. Model akan dilatih dengan data yang telah dikumpulkan dan telah ditentukan sentimennya. Selanjutnya data testing yang sentimennya belum ditentukan akan digunakan untuk memperlihatkan apakah model yang dibuat sudah dapat merepresentasikan data yang telah dilatih.

3.3.5 Hasil dan Validasi

Setelah melakukan klasifikasi sentimen dan telah didapatkan hasil prediksi analisis sentimen dengan menggunakan algoritma ke empat algoritma, yaitu *Naive Bayes*, *Support Vector Machine*, *K-Nearest Neighbors*, dan *Decision Tree* maka selanjutnya hasil tersebut akan divalidasi untuk mengetahui seberapa besar akurasi yang dihasilkan masing masing algoritma dalam melakukan klasifikasi sentimen pariwisata di kota Bali. Nilai *Area Under Curve* yang didapatkan akan dijadikan dasar perbandingan keempat algoritma tersebut untuk menentukan untuk algoritma yang tepat digunakan dalam melakukan klasifikasi sentimen di industri pariwisata Indonesia.