

BAB II

LANDASAN TEORI

2.1. Raja Ampat



Gambar 2.1. Raja Ampat

Sumber: (Maulia Indriana Ghani, 2020)

Raja Ampat disebut dengan nama tersebut berawal dari sejarah bahwa wilayah tersebut dahulu dikuasai oleh 4 raja yang tinggal di empat pulau terpisah, seluruh raja tersebut mengabdikan pada Kesultanan Tidore (Pusat Data dan Analisa Tempo, 2019) . Raja Ampat adalah sebuah kabupaten di barat laut Pulau Papua,

tepat di Provinsi Papua Barat. Raja Ampat yang memiliki luas 4,5 juta hektare, merupakan kabupaten yang sebagian besar berupa perairan, Pulau Waigeo, Pulau Bantata, Pulau Salawati, dan Pulau Misol, selain itu terdapat 600 lebih pulau kecil lain di sekitarnya seperti yang terlihat pada gambar 2.1 (Halim, 2014).

Raja Ampat merupakan salah satu objek wisata yang daya tarik utamanya adalah wisata menyelam. Terdapat 1320 spesies ikan karang, lebih dari setengah jenis *soft coral* dan tiga perempat jenis *hard coral* yang ditawarkan ketika melakukan wisata di Raja Ampat. Tempat penyelaman tersebut tersebar dari kawasan dari Selat Dampier (antara pulau Bantata dan pulau-pulau kecil Kri dan Mansuar), sedikit di bagian utara hingga Teluk Kabui (antara Pulau Gam dan Waigeo), lalu pada bagian barat yaitu ke Pulau Fam, Pulau Fainemo, Pulau Gag, kearah utara yaitu Pulau Kawe, dan ujung bagian utara yaitu Pulau Wayag, lalu batas di arah selatan Pulau Kofiau dan Pulau Misool (Halim, 2014).

Pada kenyataannya, kawasan wisata Raja Ampat memiliki potensi yang alangkah besar dan keanekaragaman hayati pesisir juga laut kepulauan sebagai aset yang menjanjikan pada bidang pengembangan pariwisata. Serta memberikan kesempatan bagi pariwisata pulau kecil dan bahari berbasis ekosistem (Parinusa, Ashar, & Kaluge, 2019).

2.2. Media Sosial

Manusia merupakan makhluk sosial, dengan majunya teknologi manusia sekarang dapat melakukan komunikasi jarak jauh menggunakan media sosial. Namun, tidak semua media dapat dikatakan sebagai media sosial, beberapa hanya

mendukung komunikasi antar manusia tanpa menjadi media bagi manusia untuk bersosialisasi (Fuchs, 2017).

Media sosial adalah istilah umum yang terdiri dari situs jejaring sosial dan *platform* bertukar pesan yang paling sering digunakan (Wartberg, Kriston, & Thomasius, 2020).

Media sosial menggambarkan sebuah teknologi atau aplikasi yang digunakan manusia dalam mengembangkan dan menjaga situs jejaring sosial mereka dalam *posting* berbentuk *multimedia* (teks, gambar, suara, *video*, dsb.), lokasi, *gaming*, dan masih banyak lagi (Albarran, 2013)

Dari beberapa pengertian dan definisi yang telah dijelaskan di atas maka dapat disimpulkan bahwa media sosial adalah media yang dibuat oleh manusia untuk melakukan komunikasi jarak jauh, terdiri dari situs jejaring sosial yang menyediakan komunikasi berbentuk *multimedia* agar dapat disebar luaskan ke khalayak publik. Terdapat berbagai *platform* media sosial, diantaranya yaitu *microblogging* (*Twitter*), *Instagram*, *Youtube*, *Facebook*, dan masih banyak lagi..

2.3. Big Data

Saat ini, *volume* data yang besar dihasilkan setiap hari pada tingkat yang tidak diterima dari sumber *heterogen* (misalnya kesehatan, pemerintahan, jejaring sosial, pemasaran, keuangan). Ini karena banyak tren teknologi, termasuk *Internet Of Things*, menjamurnya *Cloud Computing* (Botta, de Donato, Persico, & Pescapé, 2016) serta penyebaran perangkat-perangkat pintar. Di belakang layar, sistem yang

kuat dan aplikasi terdistribusi mendukung sistem koneksi ganda misalnya Sistem *smart grid*, sistem perawatan kesehatan, sistem *Walmart*, sistem pemerintah, dan sebagainya (Oussous, Benjelloun, Ait Lahcen, & Belfkih, 2018).

Tidak seperti data tradisional, istilah *Big Data* mengacu pada set data yang tumbuh besar yang mencakup format yang lebih beragam: data terstruktur, tidak terstruktur, dan semi-terstruktur. *Big data* memiliki sifat kompleks yang membutuhkan teknologi canggih dan algoritma canggih. Jadi alat *Business Intelligence* statis tradisional tidak lagi efisien dalam kasus pengaplikasian *Big data*.

Sebagian besar ilmuwan dan pakar data mendefinisikan *Big data* dengan tiga karakteristik utama yang disebut 3V. berikut adalah penjelasannya (Furht & Villanustre, 2016):

1. *Volume*: data *digital* berskala besar dihasilkan secara terus menerus dari jutaan perangkat dan aplikasi (*smartphones*, kode produk, *social network*, *log*, dan sebagainya). Diperkirakan sekitar 2,5 *exabytes* dihasilkan setiap hari pada 2012. Jumlah ini dua kali lipat setiap 40 bulan. Pada 2013, total data *digital* yang dibuat, direplikasi, dan dikonsumsi diperkirakan oleh *International Data Corporation* (sebuah perusahaan yang menerbitkan laporan penelitian) sebesar 4,4 *Zettabytes* (ZB). Ini dua kali lipat setiap 2 tahun. Pada 2015, data *digital* tumbuh hingga 8 ZB. Menurut laporan IDC, *volume* data akan mencapai 40

ZettaBytes pada tahun 2020 dan meningkat 400 kali lipat sekarang (Oussous et al., 2018)

2. *Velocity* (Kecepatan): Data dihasilkan dengan cara cepat dan harus diproses dengan cepat untuk mengekstrak informasi yang berguna dan wawasan yang relevan (Oussous et al., 2018). Salah satu contohnya ialah, *Youtube* yang memiliki kecepatan perpindahan *Big data* yang cepat
3. *Variety* (Variasi): *Big data* dihasilkan dari berbagai sumber yang didistribusikan dan dalam berbagai format data (contohnya video, dokumen, komentar, log). Kumpulan data besar terdiri dari data terstruktur dan tidak terstruktur, publik atau pribadi, lokal atau jauh, dibagi atau rahasia, lengkap atau tidak lengkap, dan sebagainya (Oussous et al., 2018).

Tapi ditemukan semakin banyak V dan karakteristik lainnya pada *Big data*. Antara lain, *Vision* (tujuan), *Verification* (data yang diproses sesuai dengan beberapa spesifikasi), *Validation* (Terpenuhinya tujuan), *Value* (informasi terkait dapat diekstraksi untuk banyak sektor), *Complexity* (sulit untuk mengatur dan menganalisis *Big data* karena hubungan data berkembang) dan *Immutability* (*Big data* dapat dikumpulkan dan disimpan permanen jika dikelola dengan baik) (Oussous et al., 2018)

2.4. *Data Mining*

Data Mining mencari sejumlah besar data untuk menentukan pola yang jika tidak dicari akan terlewat. Penerbit kartu kredit merupakan salah satu contoh dalam *data mining*, mencari jutaan transaksi kartu kredit yang disimpan dalam *database* mereka untuk menemukan tanda-tanda penipuan. Aturan *data mining* sederhana, seperti "X atau lebih banyak pembelian, dalam waktu Y, di tempat Z" dapat digunakan untuk menemukan kartu kredit yang telah dicuri dan digunakan tidak sebagaimana seharusnya (Conrad et al., 2016)

Data Mining didefinisikan secara sederhana yaitu menemukan pola yang berguna dalam data. *Data mining* juga disebut sebagai penemuan pengetahuan, pembelajaran mesin, dan analitik prediktif (Kotu & Deshpande, 2015).

Data Mining adalah Penggunaan algoritma *machine learning* untuk menemukan pola hubungan yang samar antara elemen data dalam kumpulan data yang besar, berisik, dan berantakan, yang dapat mengarah pada tindakan untuk meningkatkan manfaat dalam beberapa bentuk (diagnosis, laba, deteksi, dan sebagainya) (Nisbet, Yale, & Miner, 2017).

Dengan berbagai definisi yang sudah dijelaskan diatas dapat disimpulkan bahwa *data mining* ialah suatu proses untuk menemukan pola atau hubungan pada data yang memiliki makna atau berguna menggunakan algoritma *machine learning*.

2.5. Visualisasi Data

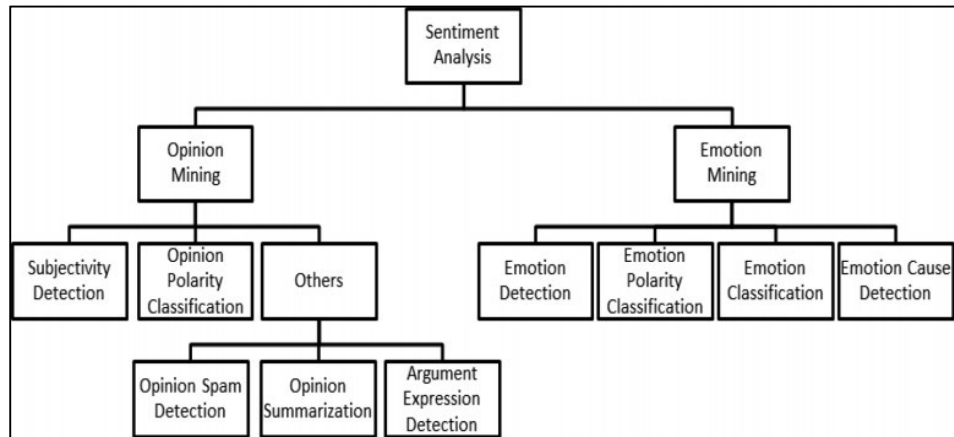
Visualisasi data mengacu pada proses transformasi data, yang dihasilkan dengan cara pengukuran berbagai proses yang terjadi secara fisik atau dibuat oleh aplikasi komputer seperti simulasi, untuk dijadikan gambar. Tujuan dari transformasi ini terbagi 2, untuk membantu pengguna memahami data dengan lebih baik dan lebih mudah dan untuk membiarkan mereka menemukan fakta yang tidak diketahui tentang fenomena yang mendasari dari mana data berasal. Untuk dapat melakukan ini, beberapa kelas metode visualisasi telah dikembangkan, masing-masing berfokus pada tipe data, pengguna, atau domain aplikasi tertentu (Frey, 2018).

2.6. Analisa Sentimen

Analisis sentimen merupakan proses deteksi, analisis, dan evaluasi pemikiran manusia terhadap berbagai masalah, layanan, dan kejadian lainnya. Lengkapnya, analisis sentimen ditujukan untuk melihat opini, sentimen dan emosi berdasarkan observasi pada tindakan orang-orang yang bisa didapatkan dari tulisan, mimik wajah, cara bicara, musik, pergerakan dan hal lainnya (Yadollahi, Shahraki, & Zaiane, 2017).

Analisis sentimen terbagi menjadi 2, yaitu *opinion mining* merupakan sentimen dengan ekspresi pada opini-opini dan *emotion mining* yang mengacu pada emosi dalam pengucapan atau artikulasi. *Opinion mining* lebih memperhatikan ekspresi pada opini dalam suatu kalimat, seperti ekspresi positif, negatif, atau netral, sedangkan *emotion mining* lebih mengarah pada emosi

(seperti senang, sedih) yang ditunjukkan dalam suatu teks (Yadollahi et al., 2017).

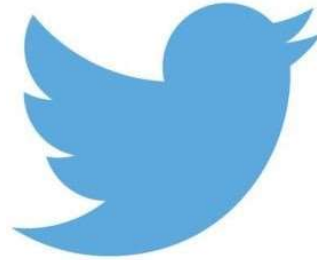


Gambar 2.2. Hierarki tugas analisis sentiment

Sumber: (Yadollahi et al., 2017)

Dari struktur yang terlihat pada gambar 2.2, penelitian ini menggunakan jenis analisis sentimen *opinion polarity classification*, salah satu bagian dari *opinion mining*, dimana teks yang terkumpul akan dikelompokkan ke dalam 3 kategori, yaitu positif, negatif dan netral.

2.7. *Twitter*



Gambar 2.3. Logo *Twitter*

Sumber: (“*Twitter* bird logo,” 2016)

Twitter adalah suatu *platform* elektronik dimana penggunanya bisa membagikan pemikiran mereka, berita-berita, dan ide-ide dengan jumlah karakter di bawah 280 karakter. Para pengguna saling “mengikuti” satu sama lain untuk tetap memperbaharui status masing-masing atau untuk berkomunikasi dengan individu ataupun kelompok lain (Rigolin, 2018). Logo *Twitter* pada gambar 2.2 yaitu burung yang melambangkan cuitan (*tweets*), fitur utama dari *Twitter*. *Twitter* mulai dikembangkan pada tahun 2006, dengan *co-founder* Jack Dorsey, Noah Glass, Biz Stone, dan Evan Williams. Ide awalnya adalah sebagai sebuah *platform* komunikasi bagi sekelompok teman saling berkomunikasi. Sejak saat itu, jejaring sosial ini meningkatkan popularitasnya. Pada kuartal ketiga tahun 2017, *Twitter* memiliki lebih dari 330 juta pengguna aktif per bulan (Rigolin, 2018).

2.7.1. *Twitter API*

Twitter API merupakan fasilitas dari *developer Twitter* yang memungkinkan untuk mengakses informasi yang ada pada *Twitter*. Ketika

ingin mengambil data (*crawling*) pada *Twitter*, pertama kita harus memiliki akun *Twitter*, lalu kita masuk ke situs web *Twitter developers* (<https://developer.Twitter.com/en.html>) untuk mendaftarkan aplikasi agar nantinya mendapatkan ID Aplikasi *API* dan token akses (Dewi, Meiliana, & Chandra, 2019), lebih jelas, gambar 2.3 menjelaskan tahapan mendaftar pada *Twitter API*.



Gambar 2.4. Tahapan mendapatkan akses *API Twitter*

Sumber: (Dewi et al., 2019)

2.8. *RStudio*

RStudio adalah *Integrated Development Environment* (IDE) yang memiliki fitur-fitur modern seperti penyorotan sintaksis dan *Code refactoring* ke R. Kelebihan *RStudio* adalah sudah terdapat semua fitur yang diperlukan pada satu tempat (Campbell, 2019). Dapat dikatakan bahwa *RStudio* telah menyediakan hampir semua fitur yang dibutuhkan sebuah IDE dengan cara yang baru, membuatnya lebih mudah dan lebih produktif untuk menggunakan R. Lebih lanjut, terdapat beberapa kelebihan dari *RStudio* (Verzani, 2011):

- Komponen utama IDE semuanya diintegrasikan dengan baik ke dalam tata letak empat panel yang mencakup konsol untuk session R interaktif, editor kode sumber tab untuk mengatur file proyek.
- *Source code Editor* mudah digunakan, kaya fitur, memiliki fitur navigasi kode yang sangat baik, dan terintegrasi dengan baik ke konsol bawaan.
- Konsol dan *Source code editor* terkait erat dengan sistem bantuan internal R melalui penyelesaian tab dan komponen penampil halaman bantuan.
- Fitur proyek memudahkan untuk mengatur alur kerja yang berbeda. Menyiapkan proyek yang berbeda adalah mudah, dan beralih di antara mereka bahkan lebih mudah.
- IDE tersedia untuk tiga sistem operasi utama dan dapat dijalankan melalui *browser* web untuk akses remot.

2.9. *Hashtag*

Hashtag atau tagar merupakan suatu bentuk pengaturan atau penataan pernyataan yang sebelumnya terbatas hanya pada kalangan profesional yang sangat khusus yang menjadi ciri penggunaan media sehari-hari. Sejak maraknya penggunaan jejaring sosial selama sepuluh tahun terakhir telah menjadi kepentingan untuk mengatur kata kunci (Bernard, 2019).

Terlebih semenjak *Twitter* dan *Instagram* memperkenalkan "tagar" pada tahun 2007 dan 2010. Dalam istilah linguistik, tagar berada diantara teks dan metateks. sebagai elemen yang mengikat antara komunikasi sehari-hari dan kode komputer, tagar telah menjadi bahasan populer saat ini (Bernard, 2019). Pada

penelitian kali ini penulis akan menggunakan *hashtag* untuk mengambil data lewat *API Twitter*.

2.10. Power BI

Pada penelitian kali ini, *tool* untuk melakukan visualisasi yaitu *Power BI*. *Power BI* dibagi menjadi 2, *desktop* dan *service*. *Power BI desktop* adalah aplikasi gratis berbasis *Windows* yang diinstall pada *desktop* lokal atau komputer. *desktop Power BI* adalah alat utama yang digunakan untuk mengambil, membersihkan, dan mengubah berbagai jenis data, menggabungkan data menjadi model, dan kemudian menganalisis dan memvisualisasikan data melalui pembuatan perhitungan, visualisasi, dan laporan.

Begitu laporan dibuat di *desktop Power BI*, laporan tersebut bisa dipublikasikan ke *Power BI service* yang merupakan *cloud-based* atau *platform* berbasis *online*, yang dimana dapat digunakan untuk *sharing* dan berkolaborasi dalam satu *report* atau proyek yang sama (Deckler, 2019).

2.11. Naïve Bayes

Naïve Bayes mengasumsikan bahwa pengaruh nilai atribut dari kelas yang diberikan tidak tergantung pada nilai atribut lainnya (García, Luengo, & Herrera, 2015). Metode ini telah banyak digunakan dalam berbagai penelitian terkait *text mining*, kelebihan *Naïve Bayes* yaitu algoritma sederhana tapi hasil akurasi tinggi, dapat menangani data dengan jumlah yang besar (Ria, Huaturuk, Rahmadani, & Ak, 2018), semua atribut pada setiap kategori tidak memiliki ketergantungan

satu sama lain (independen), dan hanya memerlukan sejumlah kecil data latih untuk menentukan parameter *mean* dan *varians* dari variabel yang diperlukan untuk klasifikasi (Devita, Herwanto, & Wibawa, 2018).

Naïve Bayes adalah penggolong probabilistik berdasarkan *Teorema Bayes*, dengan fitur-fitur yang independen dari satu sama lain. Setiap fitur dianggap berkontribusi pada probabilitas setiap contoh uji yang diberikan untuk menjadi kelas tertentu. Pertimbangkan n fitur untuk direpresentasikan sebagai vektor. Dengan menerapkan teorema *Bayes* dalam formula yang ada pada Rumus 2.1.

$$p(C_k|x) = \frac{p(C_k) p(x |C_k)}{p(x)}$$

Rumus 2.1. Rumus perhitungan *Naïve Bayes*

2.12. Penelitian Terdahulu

Tabel 2.1. Tabel Penelitian Terdahulu

Judul	Peneliti	Permasalahan	Metode	Hasil	Yang Diadopsi
Analisis Sentimen Pariwisata di Kota Malang Menggunakan Metode <i>Naive Bayes</i> dan Seleksi Fitur <i>Query Expansion Ranking</i> (2018) Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer Vol. 2	Shima Fanissa, M. Ali Fauzi Sigit, Adinugroho	Penelitian bertujuan untuk menganalisis ulasan dari masyarakat tentang pariwisata Kota Malang melalui analisis sentimen	Metode yang digunakan adalah <i>Naive Bayes</i> dengan seleksi fitur <i>Query Expansion Ranking</i> Metode <i>Multi-nominal Naive Bayes</i> dapat diterapkan pada proses analisis sentimen pariwisata Malang menggunakan data latih dari ulasan-ulasan pada <i>website TripAdvisor</i> . Data latih dilabeli kelas sentiment untuk proses klasifikasi. Tahap dalam analisis sentimen penelitian ini adalah <i>preprocessing</i> (tokenisasi, <i>filtering</i> , <i>Stemming</i>), seleksi fitur <i>Query Expansion Ranking</i> , dan <i>Naive Bayes</i> .	Metode <i>Multinomial Naive Bayes</i> dapat diterapkan pada proses analisis sentimen menggunakan data latih dari ulasan-ulasan pada <i>website TripAdvisor</i> . Analisis sentimen bertujuan untuk mengurangi dimensi fitur, dan metode <i>Query Expansion Ranking</i> bekerja dengan baik bersama metode <i>Multinomial Naive Bayes</i> .	Teknik penggunaan algoritma <i>naive bayes</i> , Konsep kerangka pikir dalam analisa sentimen
<i>Sentiment analyses of multilingual tweets on halal tourism</i> (2020)	Sulaiman Ainin,	Bagaimana sentimen masyarakat global terhadap industri	menggunakan data dari <i>Twitter</i> berupa <i>tweets</i> yang didapatkan dari <i>hashtag</i> berkaitan dengan	Terdapat 10 negara yang paling banyak <i>ditweets</i> mengenai <i>halal</i>	Teknik pengambilan

Judul	Peneliti	Permasalahan	Metode	Hasil	Yang Diadopsi
<i>Journal Tourism Management Perspectives 34</i>	Ali Feizollah, Nor Badrul Anuar, Nor Aniza Abdullah	<i>halal tourism</i> , dengan melihat negara mana yang paling populer dan pro kontra dari fasilitas yang dimiliki pada wilayah tersebut	<i>halal tourism</i> , melakukan data <i>cleansing</i> dengan metode <i>MD5 Hash Algorithm</i> . Data bersih yang didapatkan sebanyak 85.259 <i>tweets</i> .	<i>tourism</i> . Sentimen yang dihasilkan mayoritas positif namun terdapat beberapa <i>tweets negative</i> yang menggaris bawahi tentang ketidaksesuaian pelayanan dan fasilitas yang harusnya ada.	data, teknik <i>data cleansing</i>
<i>A spatial and sentiment analysis of tourism related tweets in Romania (2019)</i> <i>Proceeding ISCONTOUR 2019 Tourism Research Perspective</i>	Lupu Cristina, Stoleriu Oana Mihaela	Objek wisata yang paling diminati di Romania dengan menggunakan analisis sentimen	Penelitian menggunakan <i>Sentistrength</i> dan disajikan dalam <i>software</i> ArcGIS 10.1	Hasil analisa menunjukkan bahwa secara garis besar tujuan wisata yang diminati berada kota-kota utama seperti Bucharest, Black sea seaside, dan juga daerah Transylvania.	Teknik pengambilan data
Visualisasi Data dan Prediksi Perkembangan Impor Barang di Indonesia Tahun 2010 – 2014 (2018)	Billy Riadi, Johan Setiawan	Perlunya mengetahui pertumbuhan impor, untuk mengetahui sektor mana saja yang menjadi penyumbang impor	Metode yang digunakan pada penelitian ini adalah <i>Visual Data Mining</i> , karena tahapan metode ini lengkap dan terstruktur, selain itu tahapan dalam VDM juga dijelaskan	hasil dari penelitian ini menunjukkan bahwa produk mineral mendominasi sektor impor di Indonesia seperti bahan tambang,	<i>Tools</i> yang digunakan pada penelitian ini yaitu <i>software Power BI</i>

Judul	Peneliti	Permasalahan	Metode	Hasil	Yang Diadopsi
Jurnal Ultima InfoSys		terbesar, dan membantu pemerintah untuk meningkatkan produksi di sektor tersebut agar mengurangi tingkat impor dan menghemat devisa negara	secara terperinci sehingga mudah dipahami.	yang mana ini berarti Indonesia masih bergantung terhadap negara lain untuk pemenuhan kebutuhan bahan tambang.	untuk melakukan visualisasi.

Dari seluruh penelitian diatas (Fanissa, Fauzi, & Adinugroho, 2018), (Ainin, Feizollah, Anuar, & Abdullah, 2020), (Lupu & Stoleriu, 2019), (Riadi & Setiawan, 2019) seluruh sumber membantu dalam penyelesaian penelitian ini dengan memberikan wawasan tentang menggunakan kerangka pikir penelitian, konsep analisa klasifikasi *naïve bayes* untuk *text mining*, teknik pengambilan data dan pembersihan data serta *Tools* visualisasi yang akan digunakan yaitu *Power BI*.