

BAB II

TINJAUAN PUSTAKA DAN DASAR TEORI

2.1 Tinjauan Pustaka

Penulis menggunakan beberapa referensi sebagai bahan perbandingan dengan penelitian yang akan penulis lakukan.

Rizky Maulana (2016) melakukan penelitian “Analisis Sentimen Pengguna Twitter menggunakan *Support Vector Machine* berbasis *Cloud Computing*”. Penelitian tersebut menggunakan metode *support vector machine* dengan objek tokoh publik. Hasil penelitian didapatkan akurasi sebesar 79,5% dan ada beberapa saran dari penulis untuk melakukan deteksi pada tweet spam sehingga mengurangi duplikat *tweet*.

Septian Narsa Putra (2019) melakukan penelitian “Klasifikasi Topik Berita dan Analisis Sentimen pada *Tweets* Divisi Humas Polri dengan Metode *Naive Bayes Classifier*”. Penelitian tersebut menggunakan metode *naive bayes classifier* dengan objek divisi humas polri. Hasil penelitian didapatkan akurasi sebesar 86% dan klasifikasi topik sebesar 96%. Ada beberapa saran dari penulis untuk menambahkan *stop word* pada kalimat yang disingkat sehingga dapat meminimalkan kesalahan dalam proses klasifikasi, menggunakan metode baru untuk mencari akurasi seperti *confusion matrix*, serta aplikasi mampu menghilangkan *recor* kosong.

Penelitian yang lain “Analisis Sentimen Pengguna Gopay Menggunakan Metode *Lexicon Based* dan *Support Vector Machine*” oleh Mahendrajaya, R., Buntoro, G. A. dan Setyawan, M. B. (2019). Pada penelitian tersebut

menggunakan metode *suport vector machine* dengan objek gopay. Pengujian dilakukan menggunakan dua kernel yaitu *linier kernel* dan *polynomial kernel*. Hasil penelitian menunjukkan penggunaan *linier kernel* memiliki akurasi yang lebih besar yaitu 89.17% dibandingkan *polynomial kernel* dengan akurasi 84.38%.

Penelitian dengan judul “Metode *Lexicon Based* dan *Suport Vector Machine* untuk Menganalisis Sentimen pada Media Sosial Sebagai Rekomendasi Oleh-Oleh Favorit” oleh Hidayatul, H, Debby, G. P. Wilis, K (2018). Penelitian tersebut membandingkan hasil akurasi performa pengujian sistem yang dihasilkan dari masing-masing metode sehingga dapat diketahui performa mana yang lebih unggul dari kedua metode tersebut. Hasil penelitian menghasilkan akurasi terbesar adalah menggunakan metode *lexicon based* 87.78%, hasil presisi terbesar adalah menggunakan *lexicon based* sebesar 94.23%, sedangkan hasil *recall* terbesar adalah menggunakan metode *suport vector machine* sebesar 100%.

| Nama | Objek | Metode | Hasil |
|---|--------------------|---|---|
| Rizky Maulana (2016) | Tokoh Publik | <i>Suport Vector Machine</i> | Hasil penelitian didapatkan akurasi sebesar 79,5%. |
| Septian Narsa Putra (2019) | Devisi Humas Polri | <i>Naive Bayes</i> | Hasil penelitian didapatkan akurasi sebesar 86% dan klasifikasi topik sebesar 96%. |
| Mahendrajaya, R., Buntoro, G. A. dan Setyawan, M. B. (2019) | Pengguna Gopay | <i>Metode Lexicon Based dan Suport Vector Machine</i> | Hasil dari klasifikasi dengan metode SVM menggunakan <i>linier kernel</i> menghasilkan 89,17% dan 84,38% untuk <i>polynomial kernel</i> . |

| | | | |
|---|----------------------|---|--|
| Hidayatul, H. Debby, G. P. Dan Wilis, K (2018). | Oleh-oleh Favorit | <i>Metode Lexicon Based dan Suport Vector Machine</i> | Hasil penelitian menghasilkan akurasi terbesar adalah menggunakan metode <i>lexicon based</i> 87.78%, hasil presisi terbesar adalah menggunakan <i>lexicon based</i> sebesar 94.23%, sedangkan hasil <i>recall</i> terbesar adalah menggunakan metode <i>suport vector machine</i> sebesar 100%. |
|---|----------------------|---|--|

Tabel 2.1 Perbandingan Penelitian Sebelumnya

2.2 Dasar Teori

2.2.1 BPJS Kesehatan

BPJS Kesehatan adalah badan publik yang menyelenggarakan program jaminan kesehatan untuk masyarakat Indonesia yang dikelola oleh pemerintah. BPJS Kesehatan resmi beroperasi pada 1 Januari 2014, berawal pada tahun 2004 saat pemerintah mengeluarkan UU Nomor 40 Tahun 2004 tentang Sistem Jaminan Sosial Nasional (SJSN) dan kemudian pada tahun 2011 pemerintah menetapkan UU Nomor 24 Tahun 2011 tentang Badan Penyelenggara Jaminan Sosial (BPJS) serta menunjuk PT Askes (Persero) sebagai penyelenggara program jaminan sosial di bidang kesehatan, sehingga PT Askes (Persero) pun berubah menjadi BPJS Kesehatan.

Landasan hukum BPJS Kesehatan:

1. Undang-Undang Dasar 1945.
2. Undang-Undang Nomor 40 Tahun 2004 tentang Sistem Jaminan Sosial Nasional.
3. Undang-Undang Nomor 24 Tahun 2011 tentang Badan Penyelenggara Jaminan Sosial.

Visi BPJS Kesehatan:

Terwujudnya jaminan kesehatan yang berkualitas tanpa diskriminasi.

Misi BPJS Kesehatan:

1. Memberikan layanan terbaik kepada peserta dan masyarakat.

2. Memperluas kepesertaan program jaminan kesehatan mencakup seluruh penduduk Indonesia.
3. Bersama menjaga kesinambungan finansial program jaminan kesehatan.

2.2.2 Analisis Sentimen

Analisis sentimen atau *opinion mining* merupakan proses memahami, mengekstrak dan mengolah data tekstual secara otomatis untuk mendapatkan informasi sentimen yang terkandung dalam suatu kalimat opini. Analisis sentimen dilakukan untuk melihat pendapat atau kecenderungan opini terhadap sebuah masalah atau objek oleh seseorang, apakah cenderung berpandangan atau beropini negatif atau positif (B. Liu. 2010)(Maulana dan Redjeki 2016). Menurut (Pravina, Cholissodin, dan Adikara 2019) Analisis sentimen biasanya melakukan penelitian dalam bentuk analisis, salah satunya mengenai opini dan emosi banyak orang terhadap suatu entitas sebagai contoh permasalahan, topik atau layanan. Dengan penjelasan lain, analisis sentimen merupakan proses ekstraksi emosi atau opini dari sebuah teks atau bacaan. Dari beberapa definisi analisis sentimen dapat disimpulkan bahwa analisis sentimen adalah proses penilaian dari pandangan seseorang terhadap suatu isu yang sedang terjadi, bisa berupa pandangan positif, pandangan negatif atau pandangan netral.

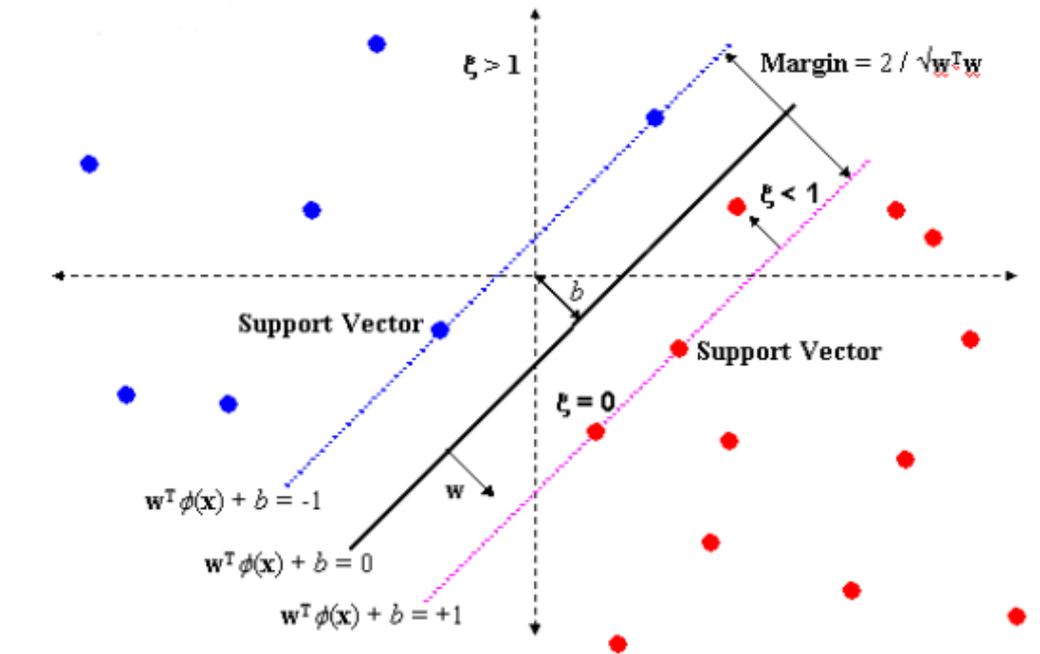
2.2.3 Twitter

Twitter adalah sebuah situs jejaring sosial yang sedang berkembang pesat saat ini karena pengguna dapat berinteraksi dengan pengguna lainnya dari

komputer ataupun perangkat *mobile* mereka dari manapun dan kapanpun. Setelah diluncurkan pada Juli 2006, jumlah pengguna Twitter meningkat sangat pesat. Pada September 2010, diperkirakan jumlah pengguna Twitter yang terdaftar sekitar 160 juta pengguna (Chiang, 2011). Jadi Twitter adalah media sosial yang banyak digunakan untuk berinteraksi antara pengguna dengan pengguna yang lain serta sebagai media menyampaikan opini publik terhadap suatu isu yang sedang terjadi.

2.2.4 Support Vector Machine

Support Vector Machine (SVM) adalah suatu teknik yang relatif baru untuk melakukan prediksi, baik dalam kasus klasifikasi maupun regresi. Menurut (Buntoro, 2017) *Support Vector Machine (SVM)* adalah seperangkat metode pembelajaran terbimbing yang menganalisis data dan mengenali pola, digunakan untuk klasifikasi dan analisis regresi. Algoritma SVM asli diciptakan oleh Vladimir Vapnik dan turunan standar saat ini (margin lunak) diusulkan oleh Corinna Cortes dan Vapnik Vladimir. Jadi *Support Vector Machine* adalah metode klasifikasi yang termasuk kedalam kategori *supervised learning* yang melakukan pembelajaran terhadap data *training* untuk mendapatkan model terbaik. Klasifikasi *support vector machine* dilakukan dengan mencari *hyperplane* terbaik dengan memaksimalkan margin agar meminimalisir *error*.



Sumber : Anandan, Varma, Joy 2014

Gambar 2.1 Suport Vector Machine

2.2.5 Algoritma Nazief dan Andriani

Algoritma Nazief & Adriani dikembangkan pertama kali oleh Bobby Nazief dan Mirna Adriani. Algoritma ini berdasarkan pada aturan morfologi bahasa Indonesia yang mengelompokkan imbuhan, yaitu imbuhan yang diperbolehkan atau imbuhan yang tidak diperbolehkan. Pengelompokan ini termasuk imbuhan di depan (awalan), imbuhan dibelakang (akhiran), imbuhan di tengah (sisipan) dan kombinasi imbuhan awalan dan akhiran (konfiks). (Wirayasa, Wirawan, dan Pradnyana 2019).

Menurut (Wirayasa, Wirawan, dan Pradnyana 2019) Algoritma Nazief & Adriani yang dibuat oleh Bobby Nazief dan Mirna Adriani ini memiliki tahap-tahap sebagai berikut :

1. Pertama cari kata yang akan di- *stemming* dalam kamus kata dasar. Jika ditemukan maka diasumsikan kata adalah *root word*. Maka algoritma berhenti.
2. *Inflection Suffixes* (“-lah”, “-kah”, “-ku”, “-mu”, atau “-nya”) dibuang. Jika berupa *particles* (“-lah”, “-kah”, “-tah” atau “-pun”) maka langkah ini diulangi lagi untuk menghapus *Possesive Pronouns* (“-ku”, “-mu”, atau “-nya”), jika ada.
3. Hapus *Derivation Suffixes* (“-i”, “-an” atau “-kan”). Jika kata ditemukan di kamus, maka algoritma berhenti. Jika tidak maka ke langkah 3a.
 - a. Jika “-an” telah dihapus dan huruf terakhir dari kata tersebut adalah “- k”, maka “-k” juga ikut dihapus. Jika kata tersebut ditemukan dalam kamus maka algoritma berhenti. Jika tidak ditemukan maka lakukan langkah 3b.
 - b. Akhiran yang dihapus (“-i”, “-an” atau “-kan”) dikembalikan, lanjut ke langkah 4.
4. Hapus *Derivation Prefix*. Jika pada langkah 3 ada sufiks yang dihapus maka pergi ke langkah 4a, jika tidak pergi ke langkah 4b
 - a. Periksa tabel kombinasi awalan- akhiran yang tidak diijinkan. Jika ditemukan maka algoritma berhenti, jika tidak pergi ke langkah 4b.
 - b. *For i = 1 to 3*, tentukan tipe awalan kemudian hapus awalan. Jika *root word* belum juga ditemukan lakukan langkah 5, jika sudah

maka algoritma berhenti. Catatan: jika awalan kedua sama dengan awalan pertama algoritma berhenti.

5. Melakukan *Recoding*.
6. Jika semua langkah telah selesai tetapi tidak juga berhasil maka kata awal diasumsikan sebagai *root word*. Proses selesai.

2.2.6 Term Frequency Inverse Document Frequency (Tfidf)

Metode TFIDF merupakan metode untuk menghitung bobot setiap kata yang paling umum digunakan pada *information retrieval*. Metode ini juga terkenal efisien, mudah dan memiliki hasil yang akurat. Metode *Term Frequency - Inverse Document Frequency* (TFIDF) adalah cara pemberian bobot hubungan suatu kata (*term*) terhadap dokumen. TFIDF ini adalah sebuah ukuran statistik yang digunakan untuk mengevaluasi seberapa penting sebuah dokumen atau dalam kelompok kata. Untuk dokumen tunggal tiap kalimat dianggap sebagai dokumen. Frekuensi kemunculan kata di dalam dokumen yang diberikan menunjukkan seberapa penting kata itu di dalam dokumen tersebut. Frekuensi dokumen yang mengandung kata tersebut menunjukkan seberapa umum kalimat tersebut. Bobot kata semakin besar jika sering muncul dalam suatu dokumen dan semakin kecil jika muncul dalam banyak dokumen. (Wirayasa, Wirawan, dan Pradnyana 2019).