BAB II
TINJAUAN PUSTAKA

2.1 Kajian Pustaka Peneltian

Dalam penelitian ini, kajian pustaka dilakukan dengan mencari informasi dari sumber jurnal penelitian yang telah dilakukan. Jurnal yang digunakan adalah jurnal yang relevan sebagai acuan penelitian. Data yang diperoleh pada penelitian sebelumnya dapat digunakan sebagai pembanding penelitian yang akan dilakukan.

Kurniawan dan Maharani (2020) meneliti tentang Analisis Sentimen Twitter Bahasa Indonesia dengan Word2Vec model skip-gram dan CBOW. Penelitian tersebut menggunakan word2vec untuk mengekstraksi fitur dan support vector machine (SVM) sebagai pemodelan klasifikasi sentimen. Penelitian tersebut bertujuan menganalisis faktor-faktor yang mempengaruhi akurasi model klasifikasi menggunakan gabungan skip-gram dan SVM. Data yang digunakan dari tweet dengan tema seputar bencana. Percobaan menggunakan variasi 6 model yaitu CBOW dengan dimensi 100, CBOW dengan dimensi 200, CBOW dengan dimensi 300, skip-gram dengan dimensi 100, skip-gram dengan dimensi 200, dan skip-gram dengan dimensi 300. Hasil penelitian menyatakan bahwa model skip-gram dengan parameter 100 dimensi memberikan hasil klasifikasi paling baik, ditunjukan dengan nilai tingkat presisi sebesar 64,4%, recall sebesar 58%, dan f-score sebesar 61,1%. Hasil klasifikasi paling baik didapatkan ketika menggunakan data latih sebanyak 1010 data. Pada pengujian juga diketahui terdapat 577 kata dalam komentar yang tidak dikenali oleh model klasifikasi [6].

Kurnia, Tangkuman, dan Girsang (2020) meneliti tentang Classification of User Comment Using Word2vec and SVM Classifier. Penelitian tersebut menggunakan Word2vec untuk untuk membuat vektorisasi dari data yang telah dikumpulkan dan klasifikasi menggunakan algoritma SVM. Metodologi yang digunakan web crawler untuk mengumpulkan data *(play.google.com)*, *text prepocessing* yang digunakan

pada penelitian tersebut adalah case folding, tokenization, character cleansing, stemming, and stop word removal [12].

Rofiqoh, Perdana, dan Fauzi (2017) meneliti tentang Analisis Sentimen Tingkat Kepuasan Pengguna Penyedia Layanan Telekomunikasi Seluler Indonesia Pada Twitter Dengan Metode Support Vector Machine dan Lexicon Based Features. Penelitian tersebut menggunakan metode support vector machine dan Lexicon based sebagai pembaharuan fitur, opini masyarakat mengenai penyedia layanan telekomunikasi seluler. Lexicon Based digunakan sebagai perhitungan frekuensi kemunculan kata sentimen positif atau negatif, menghasilkan nilai accuracy sebesar 79%, precision sebesar 65%, recall sebesar 97%, dan f-measure sebesar 78% [8].

Acosta, dkk (2019) meneliti tentang *Sentiment Analysis of Twitter Messages Using Word2Vec*. Algoritma yang yang digunakan meliputi *Naïve Bayes, Logistical Regression* dan *Support Vector Machine* untuk mengklasifikasikan lebih dari 4000 *tweet* setelah melatih *word2vec* dengan lebih dari 10000 *tweet.* Akurasi tertinggi diraih oleh *Support Vector Classifier* dan *Skip Gram* sebagai *word2vec training model* dengan angka 72%[13].

Niasita, Adikara dan Adinugroho (2019) meneliti tentang Analisis Sentimen Pembangunan Infrastruktur di Indonesia dengan *Automated Lexicon Word2Vec* dan *Naive-Bayes*. Penelitian tersebut bertujuan untuk mengetahui berbagai komentar mengenai proyek pembangunan infrastruktur di Indonesia, pada sosial media Twitter dan melihat Jumlah perbandingan pendapat pro dan kontra dari masyarakat, dengan menggunakan kamus leksikon untuk mengetahui apakah data tersebut bernilai positif atau negatif, dan dibuat secara otomatis menggunakan Word2vec pada data komentar, selanjutnya menentukan kelas sentimen menggunakan metode *Naive-Bayes*. Dengan 150 data latih kemudian dilakukan data uji menggunakan metode Naive Bayes dan Lexicon Based. Mendapat nilai precision 0,36, recall 0,818, f-measure 0,5 dan akurasi 64%[14].

Mahendrajaya, Buntoro dan Setyawan (2019) meneliti tentang Analisis Sentimen Pengguna Gopay Menggunakan Metode Lexicon Based dan Support Vector Machine. Penelitian tersebut bertujuan untuk mengetahui kepuasan pengguna terhadap layanan Go-Pay dan tingkat akurasi 2 kernel menggunakan metode SVM untuk klasifikasi komentar pengguna Go-pay. Data yang digunakan berupa opini tentang ulasan Go-Pay dari media sosial Twitter yang berjumlah 1210. Hasil akurasi tertinggi didapatkan dengan menggunakan kernel linear sebesar 89.17% dengan rincian melakukan klasifikasi sebanyak 1109 data, sedangkan kernel polynomial 84.38% atau 1021 data. Penelitian tersebut tidak ada ektraksi fitur sementara penelitian ini menggunakan ekstraksi fitur word2vec[15].

2.2 Tabel Kajian Pustaka

No.	Peneliti	Judul	Tujuan	Metodologi	Kesimpulan	Perbedaan dengan Penelitian
1.	Farhan Wahyu Kurniawan, Dr. Warih Maharani, S.T., M.T. (2020)	Analisis Sentimen Twitter Bahasa Indonesia dengan *Word2Vec* .	Menganalisis faktor-faktor yang mempengaruhi hasil penerapan metode word2vec seperti jenis arsitektur model dan ukuran dimensi pada analisis sentimen Twitter berbahasa Indonesia.	Menggunakan word2vec model CBOW dan skip-gram untuk mengekstraksi fitur dan support vector machine (SVM) untuk pemodelan klasifikasi sentimen. Menerapkan nilai parameter dimensi vektor, yaitu 100, 200, 300 pada masing-masing CBOW dan skip-gram	Model terbaik adalah model skip-gram dengan parameter dimensi sebesar 100. Nilai presisi model sebesar 64,4%, recall sebesar 58%, dan f-score sebesar 61,1%. Hasil klasifikasi paling baik didapatkan ketika menggunakan data latih sebanyak 1010 data.	Penelitian tersebut menggunakan dataset tweet bencana sedangkan penelitian ini menggunakan dataset tweet sicepat. Penelitian ini hanya menggunakan model skip-gram dengan beberapa perubahan parameter window dan size (dimensi).

2.	Sari dkk (2020)	Klasifikasi Teks Multilabel pada Artikel Berita Menggunakan Long ShortTerm Memory dengan Word2Vec	Bertujuan untuk mengategorikan teks ke dalam satu atau lebih kategori dan mengelompokkan empat label dari artikel berita.	Menggunakan Word2vec sebagai ektsraksi fitur dan LSTM sebagai fitur masukan, selanjutnya pada tahap pra-pemrosesan dengan one hote encoding.	Metode klasifikasi teks menggunakan LSTM dengan fitur Word2Vec mendapatkan nilai akurasi sebesar 95,38% dari delapan model, rata-rata nilai presisi, recall, dan F1-score sebasar 95%.	Penelitian tersebut menggunakan penerapan arsitektur Long Short-Term Memory (LSTM) sementara penelitian ini menggunakan arsitektur skipgram.
3.	Rofiqoh, Perdana, dan Fauzi (2017)	Analisis Sentimen Tingkat Kepuasan Pengguna Penyedia Layanan Telekomunikasi Seluler Indonesia Pada Twitter Dengan Metode Support Vector Machine dan Lexicon Based Features	Mengetahui tentang opini masyarakat mengenai penyedia layanan telekomunikasi Seluler, menggunakan metode support vector machine dan Lexicon based sebagai pembaharuan fitur	Pada tahapan praproses, pembobotan kata berupa tf-idf, pembobotan, Lexicon Based digunakan sebagai perhitungan frekuensi kemunculan kata sentimen positif atau negatif. Data yang digunakan pada penelitian ini sebanyak 300 data yang dibagi	Metode Support Vector Machine dan Lexicon Based Features menghasilkan nilai accuracy sebesar 79%, precision sebesar 65%, recall sebesar 97%, dan f-measure sebesar 78%.	Data yang digunakan pada penelitian tersebut, sebanyak 300 data dan menggunakan TF-IDF sebagai pembobotan kata. Sedangkan penelitian ini menggunakan 2667 data

				menjadi dua jenis data dengan rasio yaitu 70% untuk data latih 30% untuk data uji		
4.	Kurnia, Tangkuman, dan Girsang (2020)	Classification of User Comment Using Word2vec and SVM Classifier	Membangun model klasifikasi komentar menggunakan Word2vec, dan Pengklasifikasi SVM yang dapat mengklasifikasikan komentar berdasarkan peringkat skala dari 1-5.	Menggunakan web crawler untuk mengumpulkan data (play.google.com), text prepocessing yang digunakan pada penelitian tersebut adalah case folding, tokenization, character cleansing, stemming, and stop word removal, sementara word2vec untuk membuat vektorisasi dari data	Pada penelitian tersebut menggunakan 4 kali percobaan, dengan percobaan yang menggunakan data training dengan 3 class dan menggunakan stopword removal mencapai skor tertinggi untuk F1 dengan skor	Penelitian tersebut mengklasifikasikan teks komentar menjadi 5 kelas yang berbeda, 5 kelas didasarkan pada skala peringkat dari 1-5.

				yang telah dikumpulkan dan klasifikasi menggunakan algoritma svm	0.795	
5.	Acosta, dkk (2019)	Sentiment Analysis of Twitter Messages Using Word2Vec	Mengetahui tweet dapat diklasifikasikan, sebagai menampilkan positif, negatif, atau netral sentimen dari postingan Twitter yang terkait dengan perusahaan penerbangan U.S	penelitian ini menggunakan dataset yang diambil dari website kaggle yang berisi tweet tentang pengalaman pengguna terkait U.S Airlines. di tahap preprocessing, data dibagi menjadi 2 dengan rasio 70% untuk data training dan 30% untuk data testing. lalu data dibersihkan dengan metode lower case, stowpords, dan	Algoritma yang yang dipakai meliputi *Naïve Bayes, Logistical Regression* dan *Support Vector Machine* untuk mencoba mengklasifikasikan lebih dari 4000 *tweet* setelah melatih *word2vec* dengan lebih dari 10000 *tweet.* Akurasi tertinggi diraih	Penelitian tersebut menggunakan perbandingan dari beberapa algoritma. Sedangkan pada penelitian ini hanya berfokus pada penerapan *word2vec* dan *lexicon based* dengan SVM sebagai klasifikasinya.

				tokenizing. lalu data yang sudah dibersihkan dijadikan vektor dengan word2vec yang lalu diklasifikasikan dengan beberapa metode sebagai pembandingnya.	oleh *Support Vector Classifier* dan *Skip Gram* sebagai *word2vec training model* dengan angka 72%.	
6.	Niasita, Adikara dan Adinugroho (2019)	Analisis Sentimen Pembangunan Infrastruktur di Indonesia dengan *Automated Lexicon Word2Vec* dan *Naive-Bayes*	Untuk mengetahui berbagai komentar mengenai proyek pembangunan infrastruktur di Indonesia, pada sosial media Twitter dan melihat Jumlah perbandingan pendapat pro dan kontra dari masyarakat	Menggunakan kamus leksikon untuk mengetahui apakah data tersebut bernilai positif atau negatif, dan dibuat secara otomatis menggunakan Word2vec pada data komentar, selanjutnya menentukan kelas sentimen menggunakan metode *Naive-Bayes*.	Untuk membuat automated lexicon menggunakan algoritma word2vec, dengan 150 data latih kemudian dilakukan data uji menggunakan metode Naive Bayes Lexicon Based. Mendapat nilai precision 0,36, recall 0,818, f-measure	Penelitian ini menggunakan dataset tentang pembangunan infrastuktur di Indonesia, menggunakan metode Naive Bayes sebagai klasifikasinya, sedangkan peneliti svm sebagai algoritma klasifikasinya

					0,5 dan akurasi 64%.	
7.	Mahendrajaya, Buntoro dan Setyawan (2019)	Analisis Sentimen Pengguna Gopay Menggunakan Metode Lexicon Based dan Support Vector Machine	Dapat mengetahui kepuasan pengguna terhadap layanan Go-Pay dan tingkat akurasi 2 kernel menggunakan metode SVM untuk klasifikasi komentar pengguna Go-pay	Penelitian ini menggunakan prepocessing dan labelling kelas positif dan negatif dengan metode lexicon Based. Sedangkan untuk klasifikasinya menggunakan metode Support Vector Machine (SVM). Data yang digunakan berupa opini tentang ulasan Go-Pay dari media sosial Twitter	Hasil penelitian menyebutkan bahwa hasil akurasi tertinggi didapatkan dengan menggunakan kernel linear sebesar 89.17% dengan rincian melakukan klasifikasi sebanyak 1109 data, sedangkan kernel polynomial	Penelitian tersebut tidak ada ektraksi fitur sementara penelitian ini menggunakan ekstraksi fitur word2vec.

				yang berjumlah 1210	84.38% atau 1021 data.	

2.3 Dasar Teori

2. 2.1 Jasa Pengiriman Barang

Jasa pengiriman barang adalah layanan kepada pengguna jasa untuk melakukan pengiriman baik itu dokumen maupun barang untuk dikirim ke alamat yang dituju, proses pengiriman barang saat ini dibutuhkan dalam menunjang kebutuhan masyarakat Indonesia mulai dari pelaku usaha dibidang penjualan online, ibu rumah tangga bahkan perusahaan[2].

2.2.2 Sicepat

SiCepat Express merupakan perusahaan yang bergerak dalam bidang ekspedisi yang sudah berdiri sejak tahun 2004. Visi dari perushaan ini adalah menjadi penyedia jasa pengiriman barang modern pertama di Indonesia dengan mengedepankan solusi parktis bagi e-commerce yang hingga 2019 kini telah memiliki lebih dari 200 kantor cabang yang tersebar di seluruh Indonesia. SiCepat sebagai perusahaan yang terus mengalami perkembangan memiliki komitmen untuk menjadi partner terpercaya untuk para konsumennya, khususnya para pelaku usaha online. Untuk dapat terus menjadi rekan kerja yang baik bagi anggota atau mitra bisnisnya, SiCepat menyediakan berbagai layanan yang beragam bagi para konsumen[16].

Berikut merupakan layanan yang diberikan oleh SiCepat Express:

1. H3LO

   H3LO kepanjangan dari heboh 3 kilo. Layanan dengan ongkos kirim barang yang lebih irit dan terjangkau. Cukup bayar 2kg untuk kiriman barang 3,3kg dan hanya berlaku di Social Commerce, drop paket di gerai atau PUDO.

2. Siuntung

   Layanan siuntung adalah pengiriman cepat yang murah dengan harga regular, tanpa biaya tambahan. Khusus Wilayah Jabodetabek dan Bandung.

3. BEST

   BEST merupakan singkatan dari besok sampai tujuan, dengan estimasi 1 hari sampai di kota-kota besar Indonesia. Berlaku diseluruh Kota Besar Indonesia.

4. GOKIL

   GOKIL merupakan singkatan dari cargo gokil, layanan pengiriman barang dalam jumlah banyak/ukuran yang lebih besar. Dengan pengiriman Barang minimum charge 10kg dengan harga yang lebih ekonomis estimasi 1- 3 hari. Berlaku diseluruh Kota Besar Indonesia.

5. COD

   COD atau yang disebut Cash on Delivery merupakan transaksi belanja dengan pembayaran cash pada saat barang diterima, setiap belanja online di market place. Layanan ini hanya ada di wilayah Jabodetabek, Bandung, Yogyakarta, Solo, Semarang, Surabaya.

6. SiCepat Syariah

   SiCepat Syariah merupakan diskon ongkir yang diterima Sahabat SiCepat, dapat didonasikan sebesar 2,5% untuk sekolah literasi dan pesantren di wilayah Cilacap dan Lombok melalui kerjasama dengan Dompet Dhuafa. Berlaku untuk transaksi sahabat SiCepat non market place.

7. SiCepat GO

   SiCepat GO layanan yang melayani pengiriman internasional dengan harga hemat, pengiriman aman dan cepat sampai. Layanan ini juga memberikan layanan penjemputan gratis dengan berat per paket maksimal 70 kilogram. Layanan ini hanya ada di wilayah Jabodetabek, Bandung, Yogyakarta, Solo, Semarang, Surabaya.

8. SiCepat Klik

   SiCepat Klik merupakan aktifitas SiCepat dengan teknologi WhatsApp Bot. Dapat mengakses informasi SiCepat Ekspres seperti cek b

iaya pengiriman, tracking pengiriman, produk PPOB, Dapat di akses 24 jam dan informasi lainnya. Dengan cara ketik “SICEPAT” dan kirim ke WhatsApp 0813-1920-0030 atau dengan masuk ke halaman web bit.ly/SiCepatklik[17].

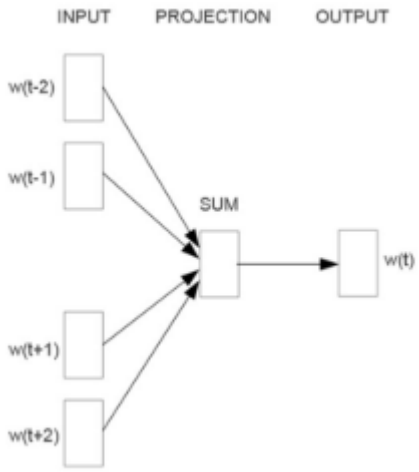
2.2.3 Sentimen Analisis

Sentimen yaitu pendapat seseorang terhadap suatu topik, produk, atau layanan tertentu. Analisis sentimen digunakan untuk menganalisis opini terhadap suatu topik tertentu apakah bersifat positif atau negatif. Melalui media sosial Twitter digunakan oleh masyarakat Indonesia untuk menuliskan opini mereka dalam bentuk cuitan (tweet) [6]. Dengan sentimen analisis dapat diketahui opini publik terkait suatu hal baik berupa produk, jasa, aplikasi, film, musik, transportasi dan lain sebagainya melalui review pengguna pada berbagai situs.

2.2.4 Word2Vec

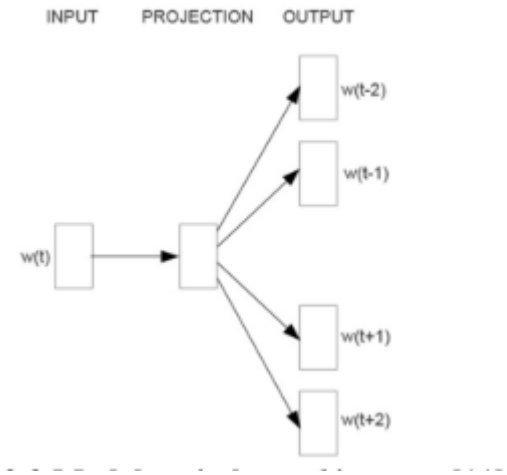
Word2vec adalah algoritma yang mempelajari posisi kedekatan semantic antar kata dari sebuah teks masukan, metode word to vec terbagi dua algoritma utama word embedding didalamnya, yaitu: continuous bag of word (CBOW) dan skip gram. Algoritma CBOW biasanya untuk melihat panjang tertentu di dokumen masukan pada sebuah kata. Sementara algoritma skip gram yaitu dapat memprediksi konteks kata dengan melihat kedekatan sebuah kata dari kata lain, letaknya sebelum atau sesudah kata tersebut. Secara garis besar arsitektur word to vector merupakan jaringan syaraf tiruan yang tidak memiliki banyak hidden layer, seperti node dalam tiap layer dan banyaknya layer[18].

CBOW adalah model arsitektur didasarkan pada neural network di mana non-linear hidden layer yang dihilangkan dan projection layer dibagikan pada semua kata, CBOW dapat memprediksi suatu kata dari kata-kata lain, yang terdapat dalam satu kalimat. Berbeda dengan model bag-of-words, digunakan untuk representasi yang terdistribusi secara kontinyu[6].



Gambar 2.1 Model arsitektur *continous bags-of-words* (CBOW) [6]

Model skip-gram adalah representasi kata yang berfungsi untuk memprediksi kata-kata terdekat pada suatu kalimat atau dokumen. Cara kerja Skip-gram dengan proses prediksi, melihat kata yang ada di sebelum dan sesudah current word[6].



Gambar 2.2 Model arsitektur Skip-gram[6]

Model skip-gram menggunakan nilai rata-rata log probability, berikut adalah persamaanya :

$$\frac{1}{T}\sum_{t=1}^{T}\sum_{-c\leq j\leq c, j\neq 0}\log p\left(w_{t+j}\middle|w_t\right)$$

(2.1)

w_t : kata center

w_{t+j} : kata setelah kata center

c : ukuran training context

Nilai vektor pada dokumen merupakan nilai vektor dari cuitan (tweet). Mencari nilai vektor dokumen dengan menghitung nilai rata-rata vektor dari semua kata dalam satu cuitan (tweet)

2.2.5 Support Machine Vector

SVM yaitu metode machine learning (supervised learning) yang digunakan untuk klasifikasi, karena dapat memprediksi kelas berdasarkan pola dari hasil proses training yang ditemukan oleh Vladimir Vapnik. Klasifikasi dilakukan dengan garis pembatas (hyperlane) untuk memisahkan antara kelas opini positif dan opini negatif. Menurut intuitif suatu garis pembatas yang baik adalah yang memiliki jarak terbesar ke titik data pelatihan terdekat dari setiap kelas, semakin besar margin tersebut, maka akan rendah error generalisasi dari pemilah [19]. Persamaan Support Vector Machine:

$$f(x) = \sum_{i=1}^{m} a_i y_i K(x, x_i) + b$$

(2.2)

Keterangan :

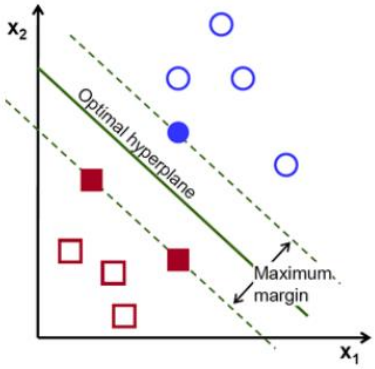
w : parameter hyperplane yang dicari (garis yang tegak lurus antara garis hyperplane dan titik support vector)

x : titik data masukan Support Vector Machine

a_i : nilai bobot setiap titik data

$K\ x, x_i$: fungsi kernel

b : parameter hyperplane yang dicari (nilai bias)



Gambar 2.3 Ilustrasi Metode Support Vector Machine[8]

Konsep klasifikasi menggunakan *Support Vector Machine* yaitu mencari *hyperplane* terbaik, sebagai pemisah dua kelas data. *Support Vector Machine* dapat mencari dataset yang berdimensi tinggi dengan cara kernel trik. *Support Vector Machine* menggunakan beberapa titik data terpilih untuk berkontribusi (support vector) model tersebut akan digunakan dalam proses klasifikasi[8].

2.2.6 Labelling

Labelling merupakan kata label. Label yaitu himpunan karakter yang digunakan untuk mengidentifikasi suatu variabel terdapat pada data atau berkas[20]. Proses melakukan pelabelan pada setiap tweets tersebut. dilakukan dengan cara memilih orientasi data pada tweet apakah positif, negatif dan netral. Pelabelan data tweet dilakukan secara manual. Pelabelan dilakukan untuk menghitung akurasi dari hasil sentimen yang didapat secara otomatis.