

BAB II

KAJIAN PUSTAKA

2.1 Penelitian Sebelumnya

Dengan adanya penelitian sebelumnya akan membantu dan mempermudah penelitian berikutnya dalam pengolahan data. Penelitian sebelumnya menunjukkan bagaimana cara pengolahan data dan cara penggunaan algoritma yang digunakan dalam penelitian.

Penelitian yang dilakukan oleh Brian Laurensz dan Eko Sedyono dengan judul “*Analysis of Public Sentiment on Vaccination in Efforts to Overcome the Covid-19 Pandemic*” menggunakan algoritma *Support Vector Machine (SVM)* dan *Naïve Bayes* dengan tujuan untuk melakukan sentimen analisis dan perbandingan algoritma terhadap respon masyarakat di media sosial *twitter* dengan menggunakan kata kunci “vaksinmerahputih” dan “vaksinsinovac”. Pada penelitian ini juga melakukan sentimen terhadap respon masyarakat apakah bersifat positif atau netral pada masing-masing kata kunci dan algoritma [17].

Selanjutnya adalah penelitian dengan judul “*Analisis Sentimen Opini Publik pada Instagram mengenai Covid-19 dengan SVM*”. Penelitian ini dilakukan oleh Iin Kurniasari, Kusriani, dan Hanif Al Fatta dengan tujuan untuk melakukan penelitian mengenai opini publik dengan menggunakan algoritma *Support Vector Machine* pada *instagram* dengan akun AG243. Peneliti menggunakan lima ekstraksi fitur seperti *count vector*, *tf-idf word level*, *tf-idf ngram level*, dan *tf-idf char level*. Peneliti menggunakan teknik *scraping* data yang diambil menggunakan *Webharvy* yang disimpan dalam bentuk file *.csv*. Selanjutnya dilakukan klasifikasi dengan menggunakan algoritma SVM dengan membagi data menjadi data latih dan data uji [18].

Penelitian berikutnya adalah penelitian dengan tujuan untuk melakukan analisis sentimen terhadap aplikasi ruang guru dengan melakukan *crawling* data pada komentar di *google play store*. Jumlah data yang digunakan sebanyak 2000 data komentar dengan 60% data uji

dan 40% data latih. Penelitian ini menggunakan algoritma *Support Vector Machine* untuk membantu melakukan klasifikasi data. Pada pengujian data tidak hanya menggunakan data latih dan data uji, namun juga menggunakan sistem *K-Fold Cross Validation* dan Kernel Linear. Penelitian yang dilakukan oleh Faizal Fakhri Irfani, Mohamad Triyanto, Anggit Dwi Hartanto, dan Kusnawi dengan judul “*Analisis Sentimen Review Aplikasi Ruang Guru Menggunakan Algoritma Support Vector Machine*” memiliki hasil bahwasanya aplikasi ruang guru mendapatkan respon positif dari masyarakat untuk membantu sistem belajar mengajar [19].

Penelitian dengan judul “*Implementasi Support Vector Machine Untuk Analisis Sentimen Pengguna Twitter Terhadap Pelayanan Telkom dan Biznet*” yang dilakukan oleh Fadholi Fat Hartanto dan Bety Wulan Sari pada Tahun 2019, melakukan penelitian dengan tujuan untuk melakukan analisis terhadap pelayanan dua *provider* yaitu Telkom dan Biznet berdasarkan respon yang positif dan negatif. Penelitian ini menggunakan algoritma SVM dan menggunakan teknik *text mining* pada media *twitter*. Data yang digunakan pada penelitian ini menggunakan 500 data untuk kedua objek. Metode pengujian pada penelitian ini menggunakan *Confusion Matrix* dan *K-Fold Cross Validation* [20].

Penelitian selanjutnya yang dilakukan oleh Imam Santoso, Windu Gata, Atik Budi Paryanti dengan judul “*Penggunaan Feature Selection di Algoritma Support Vector Machine untuk Sentimen Analisis Komisi Pemilihan Umum*” memiliki tujuan untuk melakukan analisis pada media *twitter* dengan melakukan *crawling* data berdasarkan kata “KPU”. Penelitian ini menggunakan 319 data dengan sentimen positif sebanyak 142 dan negatif sebanyak 177. Penelitian ini juga bertujuan untuk membandingkan metode SVM dan SVM *feature selection* untuk melihat metode mana yang memiliki tingkat akurasi terbaik berdasarkan AUC (*Area Under Curve*). Pengujian pembobotan pada penelitian ini menggunakan *10-fold cross validation*. Hasil akhir dari penelitian ini

adalah metode SVM menggunakan teknik *feature selection* memiliki tingkat akurasi terbaik daripada metode SVM biasa [21].

Penelitian selanjutnya dilakukan oleh Oman Somantri dan Dyah Apriliani dengan penelitian yang berjudul “*Support Vector Machine Berbasis Feature Selection untuk Sentiment Analysis Kepuasan Pelanggan Terhadap Pelayanan Warung Dan Restoran Kuliner Kota Tegal*”. Penelitian ini memiliki tujuan memberikan rekomendasi terbaik terkait dengan pemilihan untuk melakukan kuliner di kota Tegal. Selain itu, tujuan penelitian ini adalah memiliki model terbaik dengan membandingkan algoritma SVM berbasis *feature selection* menggunakan *Information Gain* dan *Chi Square* berdasarkan tingkat akurasi yang diperoleh. Penelitian ini dilakukan berdasarkan klasifikasi komentar positif dan negatif pada salah satu tempat kuliner di kota Tegal [22].

Penelitian selanjutnya adalah penelitian yang dilakukan oleh Fiktor Imanuel Tanesab, Irwan Sembiring, dan Hindrayanto Dwi Purnomo. Penelitian ini memiliki judul “*Sentiment Analysis Model Based On Youtube Comment Using Support Vector Machine*” dan tujuan penelitian ini adalah melakukan klasifikasi berdasarkan opini yang bersifat positif, negatif, dan netral berdasarkan komentar pada youtube terhadap Ahok sebagai gubernur. Penelitian ini menggunakan 1000 data dan menggunakan metode *Lexicon Based* dan *Confusion Matrix* dalam melakukan pengujian untuk memperoleh hasil akurasi dari metode SVM yang dipilih. Penelitian ini memperoleh hasil akurasi sebesar 84% [23].

Penelitian yang dilakukan oleh Ashima Garg dan Neelan Duhon dengan judul “*Sarcasm Detection On Twitter Data Using Support Vector Machine*” merupakan penelitian yang dilakukan dengan maksud dan tujuan untuk melakukan sentimen terhadap opini yang bersifat sarkas pada media *twitter*. Penelitian ini digunakan untuk melakukan klasifikasi berdasarkan komentar positif atau negatif berdasarkan kata kunci yang berbeda. Penelitian ini juga menggunakan fitur interjeksi dengan melakukan kombinasi dengan *unigram* untuk mendapatkan hasil

klasifikasi yang baik. Hasil akurasi dari penelitian ini memperoleh hasil yang baik dengan persentase sebanyak 91% [24].

Penelitian selanjutnya adalah penelitian dengan judul “*Perbandingan Metode Lexicon-based dan SVM untuk Analisis Sentimen berbasis Ontologi pada Kampanye Pilpres Indonesia Tahun 2019 di Twitter*” yang dilakukan oleh Ahmad Choirun Najib, Akhmad Irsyad, Ghiffari Assamar Qandi, Nur Aini Rakhmawati. Penelitian ini memiliki tujuan untuk analisis sentimen dengan pendekatan ontologi. Penggunaan ontologi dalam penelitian ini adalah untuk membantu dalam penyaringan data yang digunakan. Penelitian ini menggunakan metode *Lexicon-based* dan SVM dengan data yang digunakan sebanyak 700.000 *tweet*. Hasil penelitian ini memperoleh hasil *lexicon-based* sebanyak 39% dan SVM sebanyak 83%. Penelitian memperoleh hasil dengan SVM merupakan metode dengan performa yang lebih baik [25].

Penelitian dengan judul “*Analisis Sentimen Opini Terhadap Vaksin Covid-19 pada Media Sosial Twitter Menggunakan Support Vector Machine dan Naïve Bayes*” yang dilakukan oleh Frizka Fitriana, Ema Utami, Hanif Al Fatta, memiliki tujuan untuk melakukan perbandingan performa algoritma antara algoritma SVM dengan *Naïve Bayes*, berdasarkan komentar masyarakat mengenai kegiatan vaksinasi Covid-19. Penelitian ini menggunakan 1000 dataset dengan kata kunci pada *twitter vaksin covid-19*. Penelitian ini juga menggunakan perbandingan tingkat akurasi, presisi, recall, dan menggunakan perbandingan waktu selama melakukan uji latih. Selain itu, penelitian ini juga melakukan klasifikasi berdasarkan komentar positif, negatif, dan netral. Hasil penelitian ini memperoleh hasil bahawasanya algoritma SVM memiliki hasil yang lebih baik [1].

Tabel 2. 1 Penelitian Sebelumnya

No.	Judul	<i>Comparing</i>	<i>Contrasting</i>	<i>Criticize</i>	<i>Synthesize</i>	<i>Summarize</i>
1.	“ <i>Analysis of Public Sentiment on Vaccination in Efforts to Overcome the Covid-19 Pandemic</i> ” (Brian Laurensz, Eko Sedyono, 2021)	Melakukan sentimen analisis terhadap tindakan vaksinasi covid-19 terhadap respon masyarakat menggunakan algoritma SVM. Penelitian ini bertujuan untuk melakukan klasifikasi terhadap komentar yang bersifat positif atau negatif mengenai kegiatan vaksinasi. Selain itu penelitian ini juga melakukan perbandingan terhadap dua algoritma yang digunakan.	Menggunakan algoritma SVM dan <i>Naïve Bayes</i> untuk penelitiannya, serta menggunakan kata kunci “vaksinmerahputih” dan “vaksinsinovac” pada media <i>twitter</i> .	Perlu adanya tindakan untuk melakukan tahapan <i>pre-processing</i> , penambahan data dari media sosial yang lain, serta percobaan untuk menggunakan algoritma yang lain.	Penelitian ini menggunakan 854 data dan dibagi menjadi data latih dan data uji masing-masing 253 dan 592 data. Pada tahap pelabelan menggunakan sistem pelabelan manual untuk mengetahui sentimen positif dan negatif. Selanjutnya adalah tahapan pembobotan, penelitian ini menggunakan <i>tf-idf</i> untuk melakukan pembobotan. Hasil dari proses pembobotan akan digunakan untuk melakukan klasifikasi menggunakan algoritma SVM dan <i>Naive Bayes</i> . Setelah melakukan proses klasifikasi, penelitian melakukan pengujian dengan menggunakan <i>k-fold cross validation</i> .	Hasil akhir dari penelitian ini memperoleh nilai untuk sentiment positif algoritma <i>Naïve Bayes</i> dengan kata kunci “vaksinsinovac” adalah 66% dan negatif 34%. Untuk “vaksinmerahputih” memiliki sentimen positif dan negatif sebesar 89% dan 11%. Untuk algoritma SVM nilai dari sentimen positif dan negatif “vaksinsinovac” adalah 96% dan 4%. Sedangkan untuk “vaksinmerahputih” sebesar 98% dan 2%. Hasil akurasi terbaik adalah menggunakan algoritma <i>Naïve Bayes</i> dengan hasil 85,59% dan SVM 84,41%.

2.	Analisis Sentimen Opini Publik Pada Instagram Mengenai Covid-19 dengan SVM (Iin Kurniasari, Kusrini, Hanif Al Fatta, 2021)	Melakukan sentimen analisis pada komentar masyarakat pada <i>instagram</i> mengenai Covid-19 menggunakan metode SVM.	Menggunakan penambahan data dari <i>facebook</i> dan <i>instagram</i> . Jumlah data <i>facebook</i> adalah 1027 dan data <i>instagram</i> adalah 1000 data. Data ini diambil dari akun <i>instagram</i> AG234.	Memiliki hasil akurasi yang rendah untuk algoritma SVM dan perlu adanya penggunaan algoritma yang lain seperti HRRN dan LTSM. Perlakuan pelabelan pada penelitian ini tidak dijelaskan secara rinci serta banyak tahap yang kurang dijelaskan.	Pengambilan data pada penelitian ini menggunakan <i>WebHarvy</i> dengan memasukan URL dari <i>instagram</i> lalu menyimpannya dalam bentuk <i>.csv</i> . Setelah melakukan pengambilan data dilakukan <i>pre-processing</i> , yang mana hasil dari <i>pre-processing</i> ini akan dilanjutkan ke tahap klasifikasi menggunakan algoritma SVM. Pembobotan pada penelitian ini menggunakan <i>count vector</i> , <i>tf-idf word level</i> , <i>tf-idf ngram level</i> , <i>tf-idf char level</i> . Pengujian pada data latih dilakukan selama lima kali, dengan rasio data latih sebanyak 60%, 70%, 80%, dan 90%.	Memperoleh tingkat akurasi yang rendah dengan nilai 59%. Pada pendekatan <i>count vector</i> , <i>tf-idf word level</i> , <i>tf-idf ngram level</i> , dan <i>tf-idf char level</i> memperoleh masing-masin nilai sebanyak 59%, 57%, 59% dan 57%.
----	--	--	--	--	---	--

3.	<p>Analisis Sentimen Review Aplikasi Ruangguru Menggunakan Algoritma Support Vector Machine (Faizal F I, Moh. Triyanto, Anggit D H, Kusnawi, 2020)</p>	<p>Melakukan sentiment terhadap <i>review</i> aplikasi Ruangguru dan melakukan klasifikasi komentar positif dan negatif, serta melihat nilai akurasi.</p>	<p>Pengujian penelitian ini menggunakan tiga metode, metode kernel, metode pembagian data uji dan data latih, dan <i>k-fold cross validation</i>.</p>	<p>Melakukan perbandingan algoritma untuk melihat performa dari masing-masing algoritma dengan melihat nilai akurasi yang diperoleh.</p>	<p>Dataset yang digunakan sebanyak 2000 data yang diambil menggunakan <i>crawling</i> data menggunakan bahasa pemrograman <i>python</i> dari tanggal 5-9 November 2019, yang kemudian dilakukan <i>text mining</i> untuk memperbaiki struktur kalimat. Hasil dari <i>pre-processing</i> akan dilakukan untuk melakukan pembobotan menggunakan <i>tf-idf</i>. Hasil dari pembobotan ini akan digunakan untuk melakukan klasifikasi menggunakan algoritma SVM. Setelah itu adalah tahap pengujian dengan menggunakan metode kernel, pembagian data latih dan data uji, serta <i>k-fold cross validation</i>.</p>	<p>Hasil pada tahap pengujian menggunakan kernel memperoleh nilai kernel linear dengan akurasi tertinggi sebesar 0,897. Pada pengujian data latih dan data uji, nilai terbaik terletak pada pengujian data latih dan data uji 60%-40% dengan nilai akurasi sebesar 0,900. Pada pengujian k-fold cross validation yang menggunakan 2-10 fold dan menggunakan data 60-40, memperoleh nilai terbaik pada nilai k-fold 6,9,10 dengan akurasi 0.903. Rata-rata pengujian pada penelitian ini memperoleh nilai akurasi sebesar 90%.</p>
----	--	---	---	--	--	---

4.	Implementasi Support Vector Machine Untuk Analisis Sentimen Pengguna Twitter Terhadap Pelayanan Telkom dan Biznet (Fadholi Fat Haranto, Betty W, 2019)	Melakukan analisis sentimen menggunakan algoritma SVM dan melakukan klasifikasi terhadap komentar yang bersifat positif atau negatif pada provider Telkom dan Biznet untuk meningkatkan layanan kepada para pengguna pada media twitter.	Metode pengujian dilakukan dengan menggunakan <i>Confusion Matrix</i> dan <i>k-fold cross validation</i>	Penggunaan dataset yang kurang serta dapat ditambahkan metode pengujian agar dapat memperoleh nilai akurasi yang lebih baik untuk masing-masing provider.	Penelitian ini melakukan <i>crawling</i> data pada twitter dengan 500 data. Masing-masing objek memiliki 250 data yang akan digunakan untuk dijadikan sebagai data latih dan data uji. Pada objek Telkom memiliki 103 sentimen yang bersifat positif dan 147 tweet yang bersifat negatif. Sedangkan Biznet memiliki 88 tweet positif dan 162 tweet negatif. Setelah melakukan <i>crawling</i> data adalah tahapan <i>text mining</i> atau tahapan <i>preprocessing</i> . Penelitian ini menggunakan <i>tf-idf</i> untuk menghitung pembobotan . Pada tahapan klasifikasi akan menggunakan algoritma SVM.	Metode pengujian yang dilakukan dengan menggunakan <i>10-fold cross validation</i> dan <i>Confusion matrix</i> memiliki fungsi yang berbeda. Fungsi dari <i>10-fold cross validation</i> adalah untuk melakukan pembagian untuk metode pengujian data latih dan data uji, sedangkan <i>confusion matrix</i> digunakan untuk memperoleh nilai dari <i>recall</i> dan <i>F1-score</i> . Hasil dari penelitian ini adalah untuk Telkom memperoleh nilai sentimen positif sebanyak 41% dan negatif sebanyak 58,8%. Sedangkan Biznet memperoleh nilai sebesar 35,2% untuk sentimen positif dan negatif sebesar 64,8%.
----	--	--	--	---	--	--

5.	Penggunaan <i>Feature Selection</i> di Algoritma <i>Support Vector Machine</i> untuk Sentimen Analisis Komisi Pemilihan Umum (Imam Santoso, Windu Gata, Atik Budi Paryanti, 2019)	Melakukan sentiment terhadap KPU mengenai kegiatan pemilu dengan mengambil <i>tweet</i> dan balasan <i>retweet</i> pada media sosial <i>twitter</i> . Algoritma pada penelitian menggunakan SVM dan menggunakan optimasi <i>Feature Selection</i> dan melakukan perbandingan penggunaan algoritma tanpa <i>feature selection</i> dan menggunakan <i>feature selection</i> dilihat dari nilai akurasi dan AUC.	Penelitian ini menggunakan tahapan evaluasi 10 <i>fold cross validation</i> sebagai metode evaluasinya.	Penelitian selanjutnya diharapkan untuk menggunakan jumlah data yang lebih banyak dan menggunakan pembobotan yang lain dari metode pembobotan <i>feature selection</i> .	Penelitian ini menggunakan metode CRIS-DM untuk melakukan tahapan <i>text mining</i> , dan pembobotan menggunakan metode <i>feature selection weight by correlation</i> untuk membantu dalam menaikkan tingkat akurasi yang penggunaannya adalah sebelum melakukan tahapan pemodelan.	Penggunaan <i>feature selection</i> pada penelitian ini meningkatkan nilai akurasi dan AUC sebesar 81.18% dan 0.943, dibandingkan tanpa <i>feature selection</i> hanya memperoleh nilai sebesar 66.49% saja.
----	---	---	---	--	---	--

6.	<p><i>Support Vector Machine</i> Berbasis <i>Feature Selection</i> Untuk <i>Sentiment Analysis</i> Kepuasan Pelanggan Terhadap Pelayanan Warung dan Restoran Kuliner Kota Tegal (Oman Somantri, Dyah Apriliani, 2018)</p>	<p>Melakukan sentimen analisis menggunakan algoritma SVM dan pemodelan menggunakan SVM dalam klasifikasinya.</p>	<p>Penelitian ini bertujuan untuk melakukan optimalisasi model dengan penerapan <i>feature selection</i> dan penggunaan algoritma IG dan Chi Square.</p>	<p>Perlu adanya penelitian selanjutnya untuk memperoleh nilai akurasi yang terbaik dengan menemukan sebuah model yang lebih baik.</p>	<p>Penelitian ini dilakukan dengan melakukan <i>crawling</i> data pada web www.tripadvisor.co.id terhadap warung dan restoran di kota Tegal, dengan membagi ke kelas “Bagus” dan “Rata-rata”. Pembobotan pada penelitian ini menggunakan metode TF-IDF dan melakukan kombinasi data latih dan data uji sebesar 90% dan 10%. Untuk memperoleh hasil yang baik, pengujian validasi dari model yang dihasilkan dari SVM dilakukan dengan uji <i>k-fold cross validation</i> untuk mendapatkan nilai akurasi yang diperoleh dengan menggunakan SVM <i>feature selection</i> dan tanpa <i>feature selection</i>. Tahapan evaluasi ini akan diuji dengan SVM yang dilakukan kombinasi</p>	<p>Hasil dari penelitian ini memperoleh tingkat akurasi terbaik terjadi pada kombinasi algoritma SVM + IG dengan nilai akurasi 72,45% daripada kombinasi SVM+CS.</p>
----	---	--	--	---	--	--

					dengan algoritma IG dan CS	
7.	Sarcasm Detection On Twitter Data Using Support Vector Machine (Ashma Garg, Neelam Dhuha, 2020)	Melakukan sentiment analisis berdasarkan kalimat sarkas pada <i>twitter</i> menggunakan SVM. Penelitian ini melakukan pelabelan berdasarkan kalimat sarkas dan non-sarkas dengan data <i>tweet</i> sebanyak 3892 dengan hasil kalimat sarkas sebanyak 1945 dan non sarkas sebanyak 1947 .	Melakukan perbandingan nilai akurasi pada algoritma SVM dengan kernel <i>polynomial</i> dan menggunakan <i>injeksi</i> dan <i>unigram feature</i> .	Penelitian ini akan melakukan perbandingan algoritma dalam deteksi kalimat sarkas pada <i>twitter</i> menggunakan algoritma SVM kernel <i>polynomial</i> dan injeksi <i>unigram selection</i> dan memanfaatkan metode <i>feature selection</i> . Dimana <i>unigram feature</i> digunakan untuk mengumpulkan setiap kata yang unik dan injeksi digunakan untuk mempresentasikan perasaan manusia yang diekspresikan berdasarkan kata seperti “waw” dan “nah”. Pemodelan akan dilakukan pada tahapan klasifikasi	Penelitian selanjutnya harap dapat dilakukan dengan skema yang sama menggunakan big data.	Hasil akurasi pada penelitian ini yang paling banyak adalah menggunakan algoritma SVM dengan memperoleh nilai sebanyak 90.94% dengan metode K-Cross.

				dengan SVM dengan melihat nilai <i>hyperplane</i> dan nilai maksimum dari marginnya. Pembobotan pada penelitian ini menggunakan pembobotan TF-IDF dan metode evaluasinya menggunakan <i>k-fold cross validation</i> .		
8.	Sentiment Analysis Model Based On Youtube Comment Using <i>Support Vector Machine</i> (Imanuel, Fiktor T, Sembiring Irwan, Purnomo Hindrayanto D, 2017)	Melakukan sentimen analisis dan membagi kategori menjadi sentimen positif, negatif, dan netral dengan kata kunci Ahok menggunakan algoritma SVM sebagai model klasifikasinya dan melihat tingkat akurasinya dengan menghitung nilai akurasi, presisi,	Melakukan kombinasi dengan metode <i>lexicon based</i> dan <i>confusion matrix</i> dalam evaluasi pengujiannya.	Untuk penelitian selanjutnya dapat dilakukan dengan menambah data untuk meningkatkan tingkat akurasi yang lebih tinggi.	Tahapan penelitian ini dimulai dengan melakukan pengambilan data secara acak sebanyak 1000 data pada tahun 2015-2016 dengan kata kunci Ahok di <i>Youtube</i> . Selanjutnya dilakukan klasifikasi berdasarkan komentar yang bersifat positif, negatif, dan netral, dan dilakukan analisis menggunakan algoritma SVM. Selanjutnya dilakukan pengujian menggunakan metode	Hasil daripada penelitian ini adalah menghasilkan nilai dari <i>true positive</i> sebanyak 91.1% dengan nilai akurasi 84%, presisi 91%, dan recall 80%.

		dan recall.			<i>lexicon based</i> dan <i>confusion matrix</i> .	
9.	Perbandingan Metode Lexicon Based dan SVM untuk Analisis Sentimen Berbasis Ontologi pada Kampanye Pilpres Indonesia Tahun 2019 di Twitter (Ahmad Choirn Najib, Akhmad Irsyad, Ghiffari Assmar Qandi, Nur Aini Rakhmawati, 2019)	Penelitian ini menggunakan 700.000 data <i>tweet</i> dengan melakukan pelabelan manual dan menggunakan kombinasi algoritma <i>lexicon based</i> dan SVM. atribut yang digunakan pada <i>finansial</i> dan <i>kesejahteraan</i> menandakan sentimen positif dan <i>lapangan kerja</i> bersifat negatif.	Penelitian ini menggunakan pendekatan Ontologi untuk melakukan penyaringan sebuah data. Dan menggunakan metode kombinasi <i>lexicon based</i> dan SVM.	Sentimen analisis pada penelitian ini memiliki kendala dalam evaluasi menggunakan <i>lexicon based</i> karena memiliki bahasa yang kurang baku dan memiliki kata yang bersifat positif lebih dominan dalam <i>tweet</i> yang bersifat negatif dan sebaliknya. Pada penelitian selanjutnya akan dilakukan proses ekstraksi pada ontologi secara semi-otomatis untuk membantu penyempurnaan ontologi yang dibuat.	Penelitian ini menggunakan Ontologi sebagai pendekatannya untuk membantu. Setelah melewati tahapan <i>processing</i> , penelitian ini juga melakukan pelabelan secara manual untuk sentiment yang bersifat positif, negatif, atau netral. Dengan hasil sentimen positif sebanyak 600 <i>tweet</i> , 600 negatif, dan 600 netral. Berdasarkan metode <i>lexicon based</i> dengan melakukan batas <i>threshold</i> batas atas dan bawah. Jika nilai positif maka batasnya adalah atas, sedangkan jika negatif adalah batasnya bawah. Pada perhitungan menggunakan SVM dengan melakukan pembobotan term menggunakan TF-IDF Vector, dan	Hasil dari penelitian ini adalah pada metode <i>lexicon based</i> memiliki sentiment yang bersifat netral, sedangkan pada SVM memiliki sentiment positif pada atribut finansial dan kesejahteraan serta netral dalam atribut lapangan kerja. Di dalam penelitian ini memiliki performa paling baik adalah metode SVM dengan tingkat akurasinya adalah 83% dan metode <i>lexicon based</i> memiliki tingkat akurasi sebesar 39%.

					perbandingan data uji dan data latih sebesar 60%/40%, menggunakan random sate 0, dan menggunakan kernel RBF.	
10.	Analisis Sentimen Opini Terhadap Vaksin Covid-19 pada Media Sosial Twitter Menggunakan <i>Support Vector Machine</i> dan <i>Naïve Bayes</i> (Frizka F, Emma U, Hanif A F, 2021)	Penelitian ini melakukan sentimen analisis terhadap vaksin <i>covid-19</i> pada media <i>twitter</i> dengan melakukan perbandingan algoritma antara <i>Naïve Bayes</i> dan SVM.	Melakukan perbandingan performa algoritma yang digunakan, yaitu SVM dan <i>Naïve Bayes</i> .	Penelitian selanjutnya diharapkan dengan melakukan optimalisasi menggunakan metode <i>particle swarm optimization</i> .	Dalam tahapan pengambilan data, dilakukan menggunakan API <i>twitter</i> dengan menginputkan kata kunci “vaksin <i>covid-19</i> ”. Penelitian ini menggunakan bahasa PHP dan JavaScript dan database MySQL. Tahapan evaluasi pada penelitian ini menggunakan metode <i>k-fold cross validation</i> dengan dilakukan perulangan sebanyak lima kali dengan melihat performa akurasi, presisi, recall, dan waktu pelatihan pada data latih.	Performa dengan menggunakan algoritma SVM memiliki tingkat akurasi yang lebih baik daripada <i>Naïve Bayes</i> . Dimana memiliki tingkat akurasi, presisi, dan recall adalah 90,47%, 90,23%, 90,78%. Dan untuk algoritma <i>Naïve Bayes</i> memiliki nilai 88,64%, 87,32%, 88,13%. Namun dalam kecepatan waktu performa algoritma <i>Naive Bayes</i> memiliki 8,1 detik daripada SVM dengan lama waktu 11 detik.

2.2 Dasar Teori

Penelitian ini menggunakan beberapa landasan sebuah teori untuk membantu dan mempermudah tahapan penelitian yang akan dilakukan. Berikut ini beberapa dasar teori yang akan digunakan dalam penelitian ini, seperti :

2.2.1 Najwa Shihab

Najwa Shihab merupakan salah satu aktris sekaligus jurnalistik Indonesia yang lahir pada Tanggal 16 September 1977 di Makassar, Sulawesi Selatan. Najwa Shihab merupakan salah satu Wanita Indonesia yang terkenal dengan citra diri dan motivasinya terhadap khalayak muda yang selalu jadi andalan. Najwa Shihab merupakan tokoh publik yang memiliki ciri khas dengan selalu bersikap kritis terhadap kebijakan publik atau kebijakan-kebijakan pemerintah dengan melakukan analisa melalui fakta dan opini berdasarkan dengan topik yang diangkat [10].

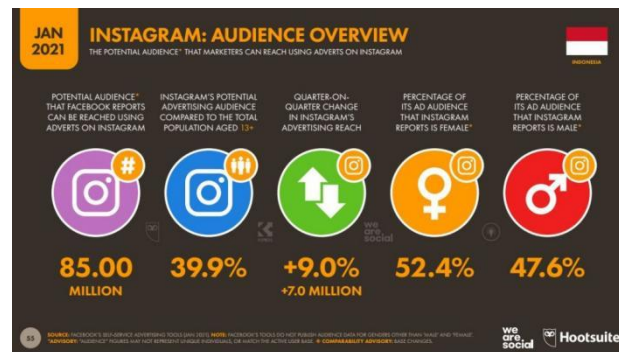
2.2.2 Vaksin Covid-19

Pada akhir tahun 2019, dunia digemparkan dengan adanya penyakit yang dimulai pada kota Wuhan, China. Penyakit ini disebabkan oleh virus jenis SARS-Cov-2 yang menyerang pada sistem pernafasan. Virus ini dikenal juga dengan sebutan Covid-19 yang ditetapkan oleh WHO sebagai salah satu pandemi yang menyerang seluruh dunia. Penyebaran virus ini berkembang dengan cepat, hingga tercatat 175 juta kasus di seluruh dunia.

Akibat adanya *covid-19*, menyebabkan berbagai pemerintah di seluruh dunia bahkan Indonesia melakukan berbagai cara untuk mengurangi dampak penularan virus ini. Mulai dari penggunaan masker, mencuci tangan, dan menjaga jarak hingga diberlakukannya pembatasan sosial.

Salah satunya adalah mulai dikembangkannya vaksin *covid-19* oleh pemerintah. Beberapa vaksin yang diedarkan di Indonesia dimulai sejak tahun 2021 hingga Maret 2022 [26].

2.2.3 Instagram



Gambar 2. 1. Pengguna Instagram Per Januari 2021

Pada Gambar 2.1, berdasarkan data dari *Hootsuite "we are social"* per Januari 2021, *instagram* merupakan salah satu sosial media dengan pengguna paling aktif nomor tiga setelah *youtube* dan *whatsapp*. Dengan penggunanya yang mencapai 85 juta jiwa dengan rentang usia pengguna paling banyak adalah 25-34 tahun [8].

Instagram juga merupakan salah satu media yang digunakan sebagai penyebaran informasi. *Instagram* memanfaatkan fitur foto dan video untuk membantu dalam penyebarannya. Dengan adanya penggunaan fitur yang relatif mudah membuat para kalangan muda terus menggunakannya. *Instagram* juga membantu dalam memberikan edukasi secara tidak langsung bagi penggunanya, sehingga *instagram* dapat dikatakan menjadi salah satu fenomena baru dalam sistem edukasi yang terus berkembang [4].

2.2.4 Analisis Sentimen

Analisis sentimen merupakan sebuah langkah yang dilakukan untuk mengkategorikan sebuah opini atau pendapat masyarakat tentang perasaan yang sedang dirasakan berdasarkan kalimat-kalimat atau teks untuk menghasilkan sebuah nilai sentimen yang positif, negatif, atau netral [27]. Dalam bidang *machine learning* tahapan pelatihan pada analisis sentimen yang dikenal dengan kesulitannya. Kesulitan ini berdasarkan dengan data berupa teks yang didasarkan dengan sebuah perasaan manusia yang memiliki pendapat berbeda setiap orangnya [28].

Tujuan dilakukannya analisis sentimen adalah melihat dan menentukan sebuah tingkah-laku atau tindakan seseorang berdasarkan sebuah berita dan topik yang dituangkan dalam sebuah kalimat atau teks [29].

2.2.5 Preprocessing

Preprocessing merupakan sebuah tahapan atau langkah yang dilakukan pada pemrograman bahasa alami. Cara ini digunakan untuk membantu dalam membuat kalimat ringkasan terhadap kalimat yang tidak terstruktur [30]. Proses ini dilakukan untuk membantu membersihkan data dan menghilangkan *noise* untuk membantu dalam pengolahan data agar lebih terstruktur. Berikut adalah tahapan *preprocessing* :

1. *Case Folding*, langkah ini digunakan untuk menyamakan setiap kata atau mengubah kata dari huruf besar ke huruf kecil.
2. *Tokenizing*, merupakan tahapan yang dilakukan untuk memisahkan kata dasar dengan spasi.
3. *Cleansing*, pada tahap ini dilakukan penghapusan karakter tanda baca dalam kalimat [31].
6. *Stemming*, pada proses ini maka akan menghilangkan kata imbuhan pada sebuah kalimat [32].

2.2.6 Pembobotan TF-IDF

Term Frequency Inverse Document Frequency (TF-IDF) merupakan pembobotan yang dilakukan untuk memberikan nilai pada masing-masing kata dengan memberikan bobot. Metode ini digunakan untuk memberikan hubungan antar kata terhadap suatu kalimat. Metode ini digunakan untuk menghitung nilai frekuensi kemunculan setiap kata dan seberapa sering kata tersebut muncul (TF), seberapa umum sebuah kata dalam kalimat atau dokumen yang digunakan (IDF), dan menentukan bobot nilai pada hubungan kata dengan dokumen (Wt).

1. Menentukan nilai *Trem Frequency* (TF)

$$Tf = (t, d)$$

(2. 1)

2. Menentukan IDF

$$IDF_t = \log \frac{n}{df_t} + 1$$

(2. 2)

3. Menentukan nilai TF-IDF

$$W = TF_t \times IDF_t$$

(2. 3)

Keterangan :

TF_t = Bobot setiap kata

(t, d) = kemunculan kata dalam dokumen

n = banyaknya dokumen

df = banyaknya kata dalam dokumen

W_t = Nilai TF-IDF

[33]

2.2.7 Klasifikasi SVM

Support Vector Machine (SVM) merupakan bagian dari *supervised learning* yang dikenalkan oleh Vapnik untuk membantu dalam melakukan sebuah klasifikasi terhadap beberapa objek di berbagai bidang. Seperti dalam melakukan sebuah klasifikasi pada citra tanda tangan, klasifikasi teks, klasifikasi gambar, dan pengenalan objek lainnya [34].

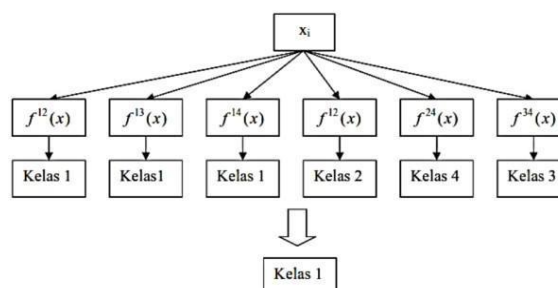
Metode SVM merupakan sebuah metode yang digunakan untuk melakukan sebuah klasifikasi dalam mencari sebuah nilai maksimum berdasarkan *margin* dari sebuah *hyperplane*. Margin ini dapat dikatakan sebagai jarak antara satu sisi *hyperplane* dengan *hyperplane* yang lain. Jika jarak sebuah *margin* semakin lebar, maka dapat dikatakan bahwa nilai akurasi akan semakin besar pula nilainya [35]. Selain itu, SVM juga dapat melakukan generalisasi yang tinggi terhadap data latih yang sedikit,

ruang dimensi yang tinggi dan ruang input untuk data yang bersifat *linear*. Sedangkan untuk data yang bersifat *non-linear*, SVM memanfaatkan fitur kernel untuk melakukan transformasi data ke ruang dimensi yang tinggi. Adanya kernel ini membantu SVM untuk melakukan klasifikasi model dengan data yang terpilih. Terdapat empat kernel yang biasa digunakan dalam penelitian, seperti kernel *linear*, *polynomial*, *RBF*, dan *Sigmoid* [36].

Algoritma SVM tidak hanya melakukan klasifikasi dua kelas namun dapat melakukan klasifikasi untuk beberapa kelas atau biasa disebut dengan *multiclass*. Salah satunya adalah jika ingin melakukan klasifikasi pada *text mining* dengan tiga sentimen yaitu positif, negatif, dan netral. Pada SVM terdapat beberapa pendekatan yang bisa dilakukan untuk membantu dalam melakukan klasifikasi terhadap tiga kelas tersebut atau lebih, seperti :

1. *One Against One*

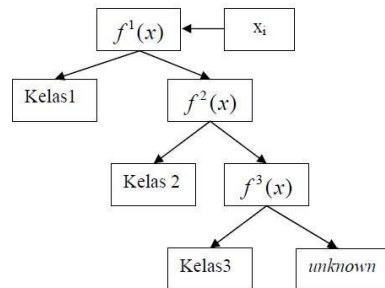
Pendekatan berdasarkan *One Against One* dikenal juga dengan *pairwise coupling* adalah klasifikasi dengan membuat golongan terhadap masing-masing pasangan kelas. N merupakan jumlah kelas yang digunakan untuk mengetahui jumlah *classifier*. Fungsi dari *classifier* ini akan dilakukan pelatihan untuk menjadi pembeda antar kelas. Jika hasil dari klasifikasi masih memiliki pola yang tidak beraturan atau tidak diketahui akan dilakukan *vote* maksimum pada *classifier* masing-masing kelas memperoleh nilai satu untuk dilakukan akumulasi untuk memilih kelas [37].



Gambar 2. 2 Gambaran *One Against One*

2. One Against All

Pendekatan selanjutnya adalah pendekatan *One Against All* yang memiliki satu *classifier* pada tiap kelas untuk membedakan *sample* kelas yang satu berdasarkan keseluruhan *sample* kelas yang ada [37].



Gambar 2. 3 Gambaran *One Against All* [37]

2.2.8 Confusion Matrix

Dalam klasifikasi analisis sentimen, dalam pengukurannya akan menghitung nilai dari akurasi, *presisi*, *recall*, dan *f1-score* untuk menentukan nilai dengan tingkat akurasi tertinggi [38].

1. Akurasi, bertujuan untuk melihat klasifikasi yang bernilai benar.

Rumus dari akurasi adalah :

$$\text{Akurasi} = \frac{(TP + TN)(TP + FP + TN + FN)}{(TP + FP + TN + FN)} \quad (2.4)$$

2. *Presisi*, digunakan untuk menentukan sebuah nilai dari kelas yang ditentukan, seperti menentukan nilai positif dalam kelas positif yang telah diklasifikasi. Rumusnya *presisi* adalah :

$$\text{Presisi} = \frac{TP}{(TP + FP)} \quad (2.5)$$

3. *Recall*, menentukan nilai keseluruhan pada dokumen berdasarkan dengan teks. Seperti melakukan perbandingan hasil positif dengan seluruh data yang bersifat positif. Rumus *recall* adalah :

$$\text{Recall} = \frac{TP}{(TP + FN)} \quad (2.6)$$

4. *F1-Score*, merupakan rata-rata dari *presisi* dan *recall* yang dapat dihitung berdasarkan dengan bobot perbandingannya.

$$\text{Recall} = \frac{2(\text{Presisi} \times \text{Recall})}{(\text{Recall} + \text{Presisi})} \quad (2.9)$$

Keterangan :

TP = *true positif*

TN = *true negatif*

FP = *false positif*

FN = *false negatif*

2.2.9 Data Latih dan Data Uji

Dalam tahapan klasifikasi memiliki beberapa langkah seperti pembuatan, implementasi, dan hasil dari sebuah model. Pada tahap pembuatan model digunakan data latih yang sudah memiliki jenis dan kelasnya. Hasil dari data latih ini akan digunakan untuk menentukan sebuah kelas pada data baru dengan melihat akurasi dari penerapan model ke data baru. Proses ini memiliki dua tahapan yaitu proses *training* dan *testing*. Proses *training* digunakan untuk membuat sebuah model untuk menguji dengan data yang baru, sedangkan proses *testing* adalah untuk menentukan nilai akurasi berdasarkan dengan hasil pengujian menggunakan data latih [39].