

BAB III

METODOLOGI PENELITIAN

3.1 Subjek dan Objek Penelitian

3.1.1 Subjek



Gambar 3. 1 Unggahan *Instagram* Najwa Shihab

Berdasarkan gambar 3.1 penelitian ini akan menggunakan subjek penelitian pada akun *instagram* Najwa Shihab berdasarkan unggahannya yaitu “*vaksin siapa takut*”.

3.1.2 Objek

Objek penelitian ini adalah komentar yang berada pada akun *instagram* Najwa Shihab berdasarkan unggahannya yaitu “*vaksin siapa*

takut” untuk analisis sentimen menggunakan algoritma *Support Vector Machine*.

3.2 Alat dan Bahan

Untuk membantu dalam melakukan penelitian menggunakan algoritma *Support Vector Machine* dalam melakukan sentimen analisis terhadap komentar pada unggahan *instagram* Najwa Shihab membutuhkan alat dan bahan dalam penelitiannya, seperti :

3.2.1 Alat

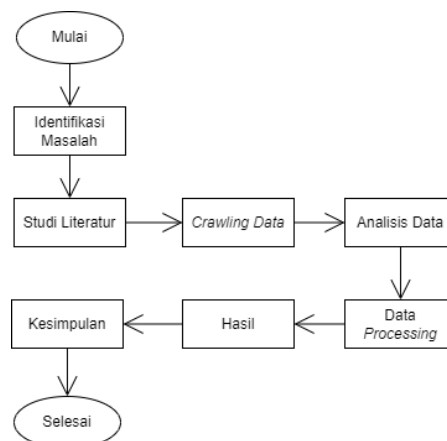
Alat yang digunakan dalam penelitian ini adalah perangkat keras yang akan digunakan yaitu Laptop 64-bit dengan kapasitas RAM 4 GB dengan HDD 1 T dengan kualifikasi CPU Intel Core i3. Sedangkan perangkat lunak yang akan membantu dalam pengolahan data dan melakukan analisis adalah *Windows 10*, *Microsoft Excel*, dan *Jupyter Notebook* dengan bahasa pemrograman *Python*.

3.2.2 Bahan

Penelitian ini membutuhkan bahan berupa data komentar mengenai unggahan *instagram* Najwa Shihab tentang “*vaksin siapa takut*” yang diunggah pada 13 Januari 2021.

3.3 Diagram Alur Penelitian

Alur penelitian akan digambarkan dalam diagram alur penelitian dibawah ini:



Gambar 3. 2 Diagram Alur Penelitian

Berdasarkan gambar diagram alur 3.2, diharapkan dapat memberikan gambaran mengenai tahapan penelitian yang akan dilakukan. Tahapan penelitian ini akan dimulai dengan identifikasi masalah berdasarkan topik yang dipilih, yang selanjutnya dilakukan studi literatur sebagai sumber referensi dalam penelitian, *crawling data* komentar, selanjutnya adalah proses analisis data terhadap data yang telah diambil, setelah data diamati dan dianalisis selanjutnya data akan dilakukan pemrosesan menggunakan algoritma SVM dan evaluasi menggunakan *confusion matrix*. Setelah pengolahan data sesuai dengan yang diharapkan, maka akan disampaikan hasil dari penelitian yang telah dilakukan dan disampaikan dalam kesimpulan.

3.3.1 Identifikasi Masalah

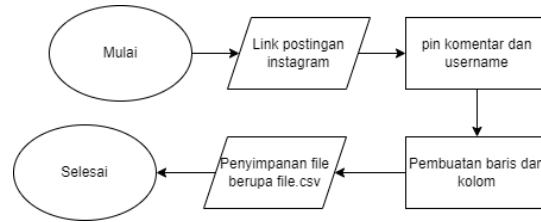
Identifikasi masalah menjadi dasar dalam penelitian ini, dimulai dengan penentuan topik yaitu vaksin *covid-19* yang diambil dan menjadi garis dasar penelitian ini. Sehingga penelitian ini akan melakukan analisis sentimen terhadap respon masyarakat mengenai vaksin *covid-19* berdasarkan unggahan dari akun Najwa Shihab tentang “vaksin siapa takut” menggunakan algoritma klasifikasi SVM. Identifikasi masalah akan digunakan untuk menjadi dasar dalam penelitian ini.

3.3.2 Studi Literatur

Studi literatur dalam tahapan penelitian ini dimulai dengan mencari berbagai referensi berdasarkan jurnal-jurnal yang terkait dengan isu dan metode yang digunakan. Selain memperoleh referensi dalam penggunaan metode, studi literatur juga membantu untuk pengolahan data, langkah-langkah pada tahap pemrosesan data, dan penggunaan algoritma-algoritma yang digunakan dalam penelitian.

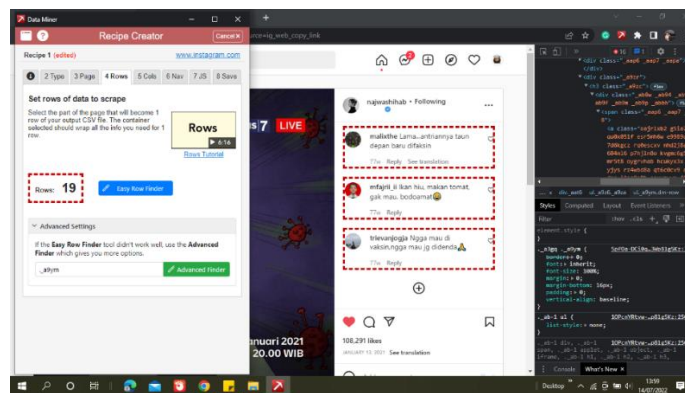
3.3.3 *Crawling Data*

Proses pengambilan data atau *crawling data* dalam penelitian ini menggunakan *tools data miner*. Berikut diagram alur dalam proses *crawling data* :



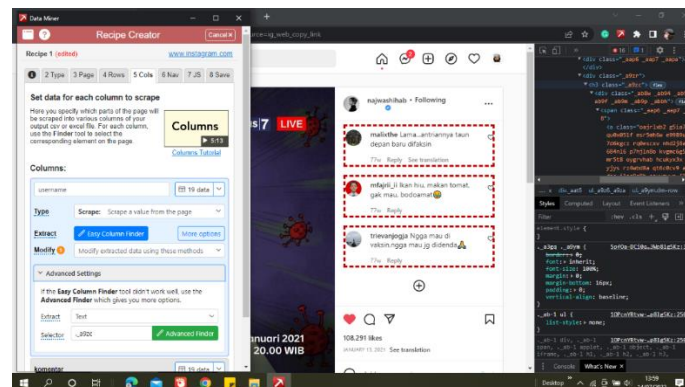
Gambar 3. 3 Proses *Crawling* Data

Berdasarkan gambar 3.3, menunjukkan proses pengambilan data dalam penelitian ini menggunakan *tools data miner*. Pengambilan data ini akan disimpan dalam bentuk file *.csv* yang berisi baris dan kolom berupa *username* dan komentar untuk menyimpan data yang diambil.



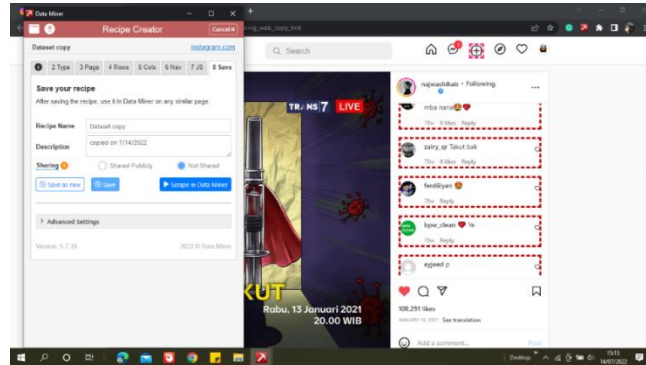
Gambar 3. 4 Pin Komentar dan *Username*

Gambar 3.4 merupakan langkah yang dilakukan setelah melakukan pengaksesan pada akun instagram *Najwa Shihab* dan unggahannya mengenai vaksin siapa takut. Langkah awalnya adalah melakukan *inspect* terhadap halaman *web* mengenai postingan “vaksin siapa takut” untuk melakukan pin pada komentar dan *username* yang akan diambil datanya. Data awal yang bisa diambil adalah sebanyak 19 data atau baris.



Gambar 3. 5 Pembuatan Kolom

Gambar 3.5 merupakan langkah selanjutnya yaitu pembuatan kolom *username* dan kolom komentar. Pembuatan kolom ini disesuaikan dengan *selector* atau nomor pada elemen *inspect*. Pembuatan kolom ini untuk menyimpan data berupa baris dari komentar dan *username*.



Gambar 3. 6 Penyimpanan *File*

Gambar 3.6 merupakan langkah akhir yaitu penyimpanan file komentar yang telah diambil menjadi file.csv.

Tabel 3. 1 Dataset Komentar

username	komentar
mardiyahisalaris	Siapa takuuuut....ikan hiu ikan asin kita siap di vaksin...
mijahi22	silahkan suntik vaksin bagi yg mau, kl aku sih ogah dikadali pino
period.pooh_spammm	♥www.fansup.vip♥ #getrealfollowers 🤞 @nofans.vip
ykyohan	SAYA TAKUT
saptaajimato	Mf,..pertama sy ngeri dengan jarum suntik alias(takut klw disuntik). Juga blum Yakin dengan Vaksin tersebut...terus bisa kebalkah dengan "Vaksin itu" blum lagi efek dengan tubuh seseorang satu sama "Vaksin ini"...dan seberapa jauhkah "Vaksintersebut " satu bln,setahun, dan seterusnya..sy belum "Yakin dengan Vaksin tsb. Gak ah...🙏🙏
bagas.kxx	Jadi kesimpulan nya yg 90% yg komen tidak mau di vaksin 🤪
urriasmanda5	gak Update ah !
venommist12	Aku siap ² aja sih kalau aku knp ² itu kehendak Tuhan :))

laksamana.chengho1	Undang dokter @dr_lois7 @najwashihab . Anda juga ingin pandemi ini segera berakhir kan? Biar seluruh Indonesia melek penjelasan ilmiah nya dan Indonesia segera keluar pandemi
--------------------	--

Tabel 3.1 merupakan dataset hasil *crawling* komentar dari akun instagram *Najwa Shihab* mengenai unggahannya “vaksin siapa takut”.

3.3.4 Analisis Data

Tahap analisis data ini digunakan untuk mengetahui dan memahami isi dari data yang telah diambil. Selain untuk memahami isi dari data komentar yang telah diambil, data ini juga melakukan pengkajian ulang terhadap kalimat-kalimat komentar.

Tabel 3. 2 Dataset Komentar

suju9301	Ciptaning aja kkg mau dokter padahal,,Ayooo Lantang Diajeng Nana,,ktnya klo mau berhasil hrs berani...sekarang saatnya Nana diuji...,wokey 😊
zulapip	Bismillah colossal titan
boziq24	Jadi Titan gak sih mbak?
ajvnx_o	Bismillah, titan kolosal... 🙏
monique.sinaga.shaikh	Org2 muda kapan dapat nya? Urutan nya aja gak ada? 1. Tenaga medis, 2. Orang2 Public, 3 Lansia, Anak2 Muda mungkin 5 tahun lagi?? @najwashihab
wisnuabdulmalik_	Tmn w d vaksin bsokanya brubah jadi TITAN..... TINAN KEBAL DRI CORONA
ardi_227	Walking Dead real life is coming 😊 😊
ahmadraya_5575	Ada bisnis baru nih
mln_fery	Bismillah founding titan
dora_chuby	Minum jamu sambil makan tahu ..ya udah kamu aja dulu.. 😊 😊

Pada tabel 3.2 diatas, dataset komentar masih berisi banyak kalimat yang tidak beraturan, seperti penggunaan huruf besar yang tidak sesuai tempatnya, kata-kata yang tidak baku, dan masih terdapat karakter gambar atau biasa disebut dengan emoji atau *emoticon*.

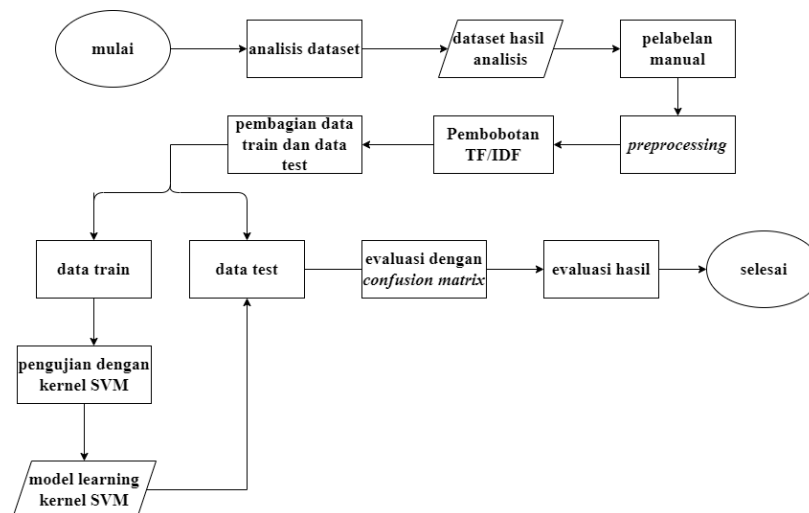
Tabel 3. 3 Data Yang Tidak Sesuai

muhammad.z.husein	Bismillah: Silakan kak antiseptic gell nya merk ONESCRA B 500ML, harga 80rb. Yg berminat boleh DM silakan.
k0h_wan	#prayforkasel #savemeratuss #kalimantanmerdeka #borneomerdeka , LIHAT KAMI
abdhadisjy_	hallo ka @najwashihab mohon bantuannya UP info kasel yg sedang dilanda banjir yg sangat dalam siapa tau dngn banyaknya informasi bnyak bantuan masuk 🙏 silahkan cek ig wargabanua
penyairsalon_	Hallo semuaa, Jika berkenan yuk follow akun baru @donpabloquotes 🙏

Pada tabel 3.3 merupakan dataset komentar pada akun *instagram* *Najwa Shihab* mengenai unggahannya yaitu “vaksin siapa takut” terdapat komentar yang tidak sesuai dengan topik dan pembahasan dari unggahannya tersebut. Data yang tidak sesuai ini akan dilakukan pembersihan pada langkah selanjutnya.

3.3.5 Data Processing

Pemrosesan atau pengolahan data komentar *instagram* dalam penelitian ini akan ditunjukkan pada Gambar 3.7 :



Gambar 3. 7 Diagram Alur Data Processing

Gambar 3.7 menunjukkan alur dari tahapan *processing* data atau pemrosesan data komentar. Setelah dilakukan analisis pada data yang telah diambil atau pada tahap sebelumnya, kemudian data akan disimpan menjadi dataset setelah analisis. Data hasil dari analisis kemudian akan dilakukan tahap pelabelan secara manual, hasil dari pelabelan ini akan digunakan pada *preprocessing* untuk membuat data lebih terstruktur dan mempermudah klasifikasi. Selanjutnya data hasil dari *preprocessing* akan dilakukan pembobotan menggunakan pembobotan *term* TF-IDF, pembobotan pada tahap TF-IDF digunakan untuk melihat frekuensi kemunculan suatu kata dalam sebuah dokumen dan melihat kemunculan suatu *term* dalam dokumen yang lain. Setelah dilakukan pembobotan, dataset ini akan dibagi menjadi data *train* dan data *test*. Data *train* ini akan dilakukan pelatihan dengan algoritma SVM untuk menghasilkan model *learning* SVM. Model *learning* ini yang akan digunakan untuk pengujian menggunakan kernel *linear*, *sigmoid*, *polynomial*, dan *rbf* kernel, hasil dari pemodelan ini akan diuji dengan data *test* untuk melihat performa dari algoritma *support vector machine* atau SVM. Hasil dari pengujian ini kemudian akan dilakukan evaluasi algoritma menggunakan *confusion matrix* untuk melihat performa algoritma berdasarkan nilai dari akurasi, *presisi*, *recall*, dan *f1-score*.

3.3.5.1 Pelabelan Manual

Proses pelabelan ini digunakan untuk memberikan label atau tanda pada kalimat komentar yang telah diambil dan dilakukan analisis sebelumnya. Pelabelan pada komentar ini dilakukan secara manual dan secara subjektif oleh satu orang untuk mengelompokkan komentar yang bersifat positif, negatif, dan netral. Berdasarkan komentar yang telah diambil pada unggahan akun *instagram* Najwa Shihab mengenai unggahannya “vaksin siapa takut”, komentar-komentar ini akan dilabeli 1 (positif) jika komentar tersebut bersifat mendukung adanya vaksinasi covid-19, bernilai -1 (negatif) jika komentar tersebut menolak adanya vaksinasi covid dan memberikan pernyataan yang kurang mendukung, dan “0” (netral) untuk komentar yang netral atau tidak mengandung unsur keduanya namun masih sesuai dengan topik pembahasan. Proses pelabelan ini memiliki keterangan kriteria yang dimaksudkan sebagai

keyword kenapa kalimat tersebut digolongkan sebagai kalimat positif, negatif atau netral.

Tabel 3. 4 Label Positif

username	komentar	label	kriteria
mardiyahisalaris	Siapa takuuuut...ikan hiu ikan asin kita siap di vaksin...	1	siap di vaksin
venommist12	Aku siap ² aja sih kalau aku knp ² itu kehendak Tuhan :))	1	siap
vinsbbcr435	Siap gue tungguin jadwalnya semoga cepet dpt jadwal divaksin nya walaupun ada tegang juga 😊😊	1	siap cepet dapat jadwal vaksin
mhdarafly	Ayoo vaksin. Dah capek Kuliah Di HUM 🤔	1	ayo vaksin
bratawala	Gue juga ga takut di vaksin masa kalah sama bapak presiden .. mesti berani dong	1	ga takut

Tabel 3.4 merupakan tabel untuk pemberian label bernilai 1 (positif) pada data dengan memperhatikan kriteria komentar seperti siap di vaksin, siap, semoga cepat dapat vaksin, ayo vaksin, dan tidak takut vaksin.

Tabel 3. 5 Label Negatif

username	komentar	Label	Kriteria
mijahi22	silahkan suntik vaksin bagi yg mau, kl aku sih ogah dikadali pino	-1	ogah dikadali pino
bagas.kxx	Jadi kesimpulan nya yg 90% yg komen tidak mau di vaksin 😞	-1	tidak mau di vaksin
urriasmanda5	gak Update ah !	-1	gak update
saddamyrr	Ikan hiu makan ketan, aku mau asal war hammer titan	-1	war hamer titan

Tabel 3.5 merupakan pemberian label untuk komentar negatif, dengan memperhatikan kriteria komentar seperti ogah dikadali, tidak mau di vaksin, gak update, dan war hamer titan.

Tabel 3. 6 Label Netral

username	komentar	label	kriteria
rizkth_13	Jjr dri hti sy yg sedlm dlmnya bahwasanya saya takut jarum suntik sekian terimauang 😊	0	takut jarum suntik
quludut_	Klo sya gk mau gimna? Trus dipaksa? Apakah termasuk melanggar HAM?	0	pertanyaan
ivander_dc	Aku lebih takut jarum suntiknya drpd dampaknya 🤢. Gw takut suntik!	0	takut suntik
pricedeal.id	@wicaksonodanutri Nah bener sekali mas, vaksin ini efeknya jangka panjang. Dulu sebelum adanya vaksin orang hidup sehat dan alami. Klu emang mau sehat dan jauh dari penyakit ikuti aja Sunnah Rasulullah	0	pernyataan pendapat

Tabel 3.6 merupakan labeling untuk kelas netral dengan memperhatikan kriteria komentar seperti belum siap, pertanyaan, dan takut jarum suntik.

3.3.5.2 Preprocessing

Preprocessing merupakan tahapan yang digunakan untuk membantu dalam pembersihan dataset. Tahap ini digunakan untuk membuat data menjadi lebih tertata, bersih, dan mempermudah untuk diolah. Dalam tahapan ini biasanya akan menyelesaikan masalah mengenai *noise* data atau membersihkan dan memperbaiki data komentar yang tidak terstruktur. Tahapan *preprocessing* pada penelitian ini dijelaskan pada gambar dibawah ini.

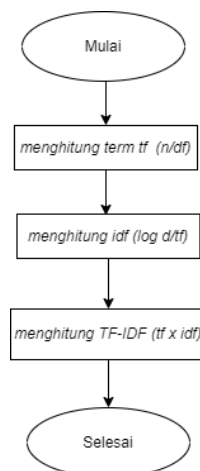


Gambar 3. 8 Alur *Preprocessing*

Gambar 3.8 menunjukkan diagram alur pada tahapan *preprocessing* yang dimulai dengan *case folding* dimana pada proses ini akan dilakukan penyamarataan huruf besar menjadi huruf kecil semua misalnya “A-Z” menjadi “a-z”, selanjutnya adalah *cleaning* yaitu untuk menghapus karakter-karakter yang tidak digunakan seperti tanda baca, angka, dan emoji yang ada pada komentar. Setelah melakukan proses *cleaning* akan dilakukan tahapan *tokenizing* yang mana akan dilakukan untuk memisahkan setiap kata pada kalimat dengan koma. Proses selanjutnya adalah *stemming*, yang mana proses ini akan digunakan untuk menghilangkan kata imbuhan pada sebuah kata atau kalimat.

3.3.5.3 Pembobotan TF-IDF

Untuk mempermudah pengolahan data, maka data yang digunakan sebaiknya berupa data numerik. Oleh karena itu, pembobotan TF-IDF digunakan untuk mengubah data setelah *preprocessing* menjadi data numerik. Pembobotan TF-IDF juga digunakan untuk melakukan perhitungan frekuensi kemunculan sebuah kata dalam dokumen dan keseluruhan dokumen, serta keterkaitan antar dokumen dan seluruh dokumen.



Gambar 3. 9 Proses Pembobotan TF-IDF

Gambar 3.9 menunjukkan alur pembobotan dengan TF-IDF. Data hasil proses *preprocessing* akan diolah pada pembobotan TF-IDF untuk menghitung kemunculan sebuah kata dalam dokumen dan keseluruhan dokumen. Pada proses ini akan menghitung frekuensi dalam sebuah dokumen (D) yang memiliki *term* tersebut, kemudian menghitung frekuensi dari kemunculan kata dalam sebuah dokumen itu sendiri dengan dokumen yang lain (TF). Selanjutnya adalah menghitung nilai dari IDF untuk menghitung kemunculan sebuah *term* dalam keseluruhan dokumen. Selanjutnya adalah melakukan perhitungan TF-IDF dengan rumus $W=TF \times IDF$. Nilai terbaik *term* adalah sering ditemukan dalam dokumen namun jarang ditemukan dalam dokumen yang lainnya.

3.3.5.4 Pembagian Data *Test* dan Data *Train*

Setelah dilakukan pembobotan pada proses TF-IDF, selanjutnya adalah melakukan pembagian untuk data *train* dan data *test*. Pembagian dataset ini digunakan untuk mencari nilai akurasi terbaik dari model dataset yang digunakan. Pembagian dataset ini menggunakan rasio 90:10, 80:20, 70:30, 60:40, dan 50:50. Perbandingan ini adalah seperti membagi data *train* menjadi 90% dan data *test* 10%.

3.3.5.5 Klasifikasi SVM *Multiclass*

Pada pembagian data yang telah dilakukan menjadi data *train* dan data *test*, langkah selanjutnya adalah melakukan klasifikasi menggunakan algoritma SVM untuk memperoleh *model learning* yang akan digunakan pada pengujian data *test* untuk melakukan klasifikasi komentar positif, negatif, dan netral.

Seperti yang diketahui, bahwasanya SVM melakukan klasifikasi biner secara linear yang dengan memperoleh satu model dan satu *hyperplane*. Namun, pada penelitian ini, klasifikasi SVM melakukan klasifikasi komentar untuk tiga kelas, yaitu kelas positif, negatif, dan netral. Dengan menggunakan pendekatan *one against one*, maka akan dilakukan fungsi dasar SVM yaitu dengan melakukan klasifikasi pada masing-masing kelas secara biner dengan persamaan linear. Langkah pertama yang dilakukan adalah menentukan

model yang akan di buat untuk memperoleh nilai mayoritas pada setiap model yang akan digunakan untuk pengujian pada data *test* untuk klasifikasi komentar. Persamaan yang digunakan adalah sebagai berikut:

$$\frac{k(k-1)}{2} = \frac{3(3-1)}{2} = 3, \text{ dengan } k \text{ adalah jumlah kelas}$$

Berdasarkan persamaan diatas, diperoleh model *learning SVM* adalah tiga. Masing-masing model ini akan menghasilkan tiga *hyperplane* yang digunakan untuk memisahkan antar kelas secara *linear* dengan persamaan $w_x^T + w_o$.

Tabel 3. 7 Tabel Klasifikasi *One Against One*

Model SVM	One Against One
SVM1	P vs N
SVM2	N vs Net
SVM3	P vs Net

Keterangan:

P = Positif

N = Negatif

Net = Netral

Langkah selanjutnya adalah menghitung probabilitas mayoritas kemunculan kelas untuk 3 model yang telah diperoleh. Jika pada model SVM1 yaitu P vs N dengan P = 1 dan N = -1, diperoleh nilai x adalah 1, lalu pada model kedua yaitu N vs Net dengan N = -1 dan Net = 0, dengan nilai x adalah -1, dan pada model SVM3 yaitu P = 1 dan Net = 0, diperoleh nilai x adalah 1, maka dapat diperoleh bahwa klasifikasi yang sering muncul adalah komentar 2P dengan nilai 1. Maka hasil ini akan digunakan pada data *test* untuk melakukan klasifikasi komentar pada data yang baru.

Tabel 3. 8 Probabilitas SVM *One Against One*

Model SVM	One Against One	Probablitas
SVM1	P vs N	1
SVM2	N vs Net	-1
SVM3	P vs Net	1

3.3.5.6 Pengujian dengan Kernel SVM

Hasil yang diperoleh pada *model learning* akan dilakukan juga pengujian menggunakan kernel SVM. Penelitian ini menggunakan fungsi kernel yang umum digunakan yaitu kernel *linear*, *sigmoid*, *polynomial*, dan *rbf*. Hasil pada pengujian kernel ini juga akan digunakan untuk melakukan pengujian data *test*. Berikut merupakan persamaan-persamaan pada kernel SVM:

Tabel 3. 9 Persamaan Kernel Pada SVM

Kernel	Persamaan
Linear	$K(X_i, X_j) = X_i^T X_j$
RBF	$K(X_i, X_j) = \exp(-\gamma X_i - X_j ^2), \gamma > 0$
Polynomial	$K(X_i, X_j) = (\gamma \cdot X_i^T X_j + r)^d, \gamma > 0$
Sigmoid	$K(X_i, X_j) = \tanh(\gamma \cdot X_i^T X_j + r)$

Salah satu contoh penggunaan fungsi kernel ini adalah menggunakan kernel RBF. Kernelisasi dilakukan pada data *train* untuk mendapatkan data dengan fitur dimensi tinggi berdasarkan fungsi kernel RBF $K(X_i, X_j) = \exp(-\gamma |X_i - X_j|^2), \gamma > 0$, sehingga akan diperoleh data set dengan dimensi $N \times N$, dimana N adalah banyaknya data.

Pengujian ini digunakan untuk melihat nilai akurasi dan performa dari algoritma SVM berdasarkan model yang telah dibuat. Pengujian ini akan menghasilkan nilai akurasi, presisi, *recall*, dan *f1-score* berdasarkan hasil dari masing-masing pengujian berdasarkan kernel *linear*, *sigmoid*, *polynomial*, dan *rbf* berdasarkan pembagian data *train* dan data *test*.

3.3.5.7 Evaluasi Confusion Matrix

Seperti yang diketahui, penggunaan *confusion matrix* pada penelitian ini adalah untuk melihat nilai akurasi, *presisi*, *recall*, *f1-score* dari performa algoritma dan model yang digunakan dengan perhitungan nilai aktual dan prediksi yang digunakan. *Confusion matrix* mempresentasikan nilai dari klasifikasi komentar berdasarkan dengan nilai *True Positif*, *True Negatif*,

False Positif, dan *False Negatif*. Evaluasi menggunakan *confusion matrix* ini akan menggunakan data *train* dan data *test* berdasarkan pengujian kernel yang memiliki tingkat akurasi tertinggi, seperti yang diketahui pengujian kernel pada penelitian ini menggunakan kernel *linear*, *sigmoid*, *polynomial*, dan *rbf*.

3.3.6 Evaluasi Hasil

Pada tahapan ini akan disampaikan sebuah kesimpulan dari hasil dan tahapan penelitian yang telah dilakukan dan dilakukan pemaparan hasil penelitian yang telah dilakukan berdasarkan komentar dari akun *instagram* Najwa Shihab mengenai unggahannya “vaksin siapa takut”.

3.3.7 Kesimpulan

Penarikan ringkasan berdasarkan topik, pengolahan data, hasil dari evaluasi metode yang digunakan, hasil dari penelitian yang diperoleh, serta hambatan dan kekurangan yang ada di dalam penelitian ini.