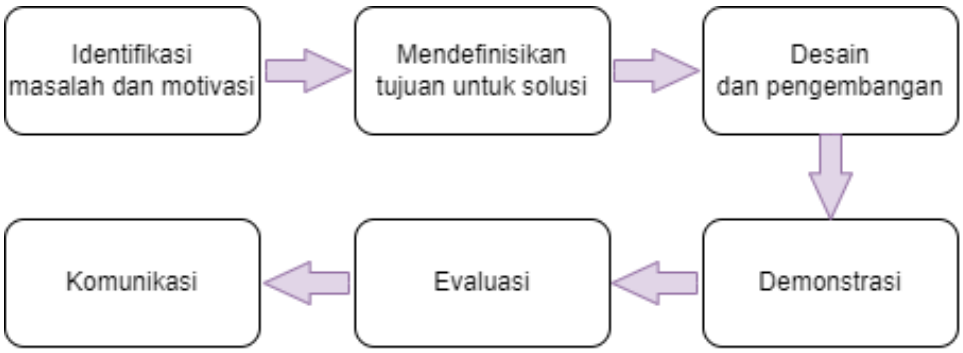
# BAB III
# METODE PENELITIAN

## 3.1 *Design Science Research Methode* (DSRM)

Metode penelitian yang digunakan adalah *Design Science Research Methode* (DSRM). DSRM adalah paradigma pemecahan masalah yang bertujuan untuk meningkatkan pengetahuan manusia melalui penciptaan benda atau alat inovatif. Secara sederhana, DSRM bertujuan untuk meningkatkan basis pengetahuan teknologi dan ilmu pengetahuan melalui penciptaan benda atau alat inovatif yang memecahkan masalah dan meningkatkan lingkungan di mana mereka diimplementasikan (vom Brocke dkk., 2020). DSRM bertujuan untuk memperluas batas kemampuan manusia dan organisasi dengan menciptakan benda atau alat baru dan inovatif, termasuk konstruk, model, metode, dan instansiasi (Hevner dkk., 2004).

Di bidang Sistem Informasi (SI), *Design Knowledge* (DK) mencakup pengetahuan tentang berbagai aspek seperti struktur dan konstruksi sistem basis data, pemodelan proses bisnis, penyesuaian SI dengan strategi organisasi, pengiriman analitik data untuk pengambilan keputusan (Becker dkk., 2015). DSRM tidak terbatas pada bidang SI tetapi juga merupakan paradigma penelitian sentral di domain lain termasuk teknik, arsitektur, bisnis, ekonomi, dan disiplin terkait teknologi informasi lainnya (vom Brocke dkk., 2020).

Model proses metodologi penelitian DSRM untuk mengembangkan sistem klasifikasi teks pada pemilihan legislatif 2024 dari media sosial *Twitter* menggunakan metode komparasi *Word2vec* dan TF-IDF dengan klasifikasi *Support Vector Machine* untuk melakukan klasifikasi teks pada *tweet* yang terkait dengan pemilihan legislatif 2024 ditunjukkan pada Gambar 3.1




Gambar 3.1 Tahapan *Design Science Research Methode* (DSRM)

Proses DSRM ini mencakup enam langkah: identifikasi masalah dan motivasi, definisi tujuan untuk solusi, desain dan pengembangan, demonstrasi, evaluasi, dan komunikasi (vom Brocke dkk., 2020).

Pada tahap awal penelitian yaitu identifikasi masalah dan motivasi. Aktivitas ini mendefinisikan masalah penelitian yang spesifik dan membenarkan nilai dari sebuah solusi. Membenarkan nilai dari sebuah solusi mencapai dua hal: itu memotivasi peneliti dan audiens penelitian untuk mengejar solusi tersebut dan membantu audiens untuk menghargai pemahaman peneliti tentang masalah tersebut. Sumber daya yang diperlukan untuk aktivitas ini termasuk pengetahuan tentang keadaan masalah dan pentingnya solusinya. Peneliti melakukan studi pustaka yang berhubungan dengan klasifikasi teks, pengumpulan data, metode klasifikasi teks, model algortima, model evaluasi dan model validasi sampai akhirnya peneliti dapat merumuskan permasalahan yang ada.

Tahapan selanjutnya mendefinisikan tujuan untuk solusi. Tujuan dari sebuah solusi dapat disimpulkan dari definisi masalah dan pengetahuan tentang apa yang mungkin dan layak dengan melakukan studi pustaka. Tujuan tersebut dapat bersifat kuantitatif, misalnya dalam hal di mana solusi yang diinginkan akan lebih baik daripada yang ada saat ini, atau bersifat kualitatif, misalnya deskripsi tentang bagaimana benda atau alat baru diharapkan dapat mendukung solusi untuk masalah yang sebelumnya belum teratasi. Tujuan tersebut harus disimpulkan secara rasional dari spesifikasi masalah (vom Brocke dkk., 2020). Penelitian mengenai klasifikasi teks tentunya sudah

banyak dilakukan. Namun Penelitian ini akan membatasi klasifikasi teks pada pemilihan legislatif 2024 di Indonesia dengan menggunakan data dari media sosial *Twitter*. Penelitian juga akan fokus pada komparasi metode klasifikasi teks dengan *Nature Language Processing* (NLP) yaitu *Word2vec* dan TF-IDF. Oleh karena itu tujuan dari solusi yang dihasilkan pada penelitian ini adalah menemukan metode yang terbaik berdasarkan tingkat akurasi, *presisi*, *recall* dan *f1-score* pada metode *Word2vec* dan TF-IDF dengan klasifikasi SVM dalam klasifikasi teks untuk pemilihan legislatif 2024 dari media sosial *Twitter*.

Tahapan Desain dan pengembangan. tahap ini terdiri dari 2 tahap, yaitu tahap perancangan sistem dan tahap pengembangan sistem. Tahap perancangan sistem mencakup pengumpulan data, mengklasifikasi teks, perancangan algoritma, dan perancangan pengujian sistem. Tahap pengembangan sistem meliputi implementasi program, pengujian program, dan evaluasi hasil implementasi program. Tahap perancangan dan pengembangan meliputi: algoritma pengumpulan data, membangun mode klasifikasi dan evaluasi model klasifikasi.

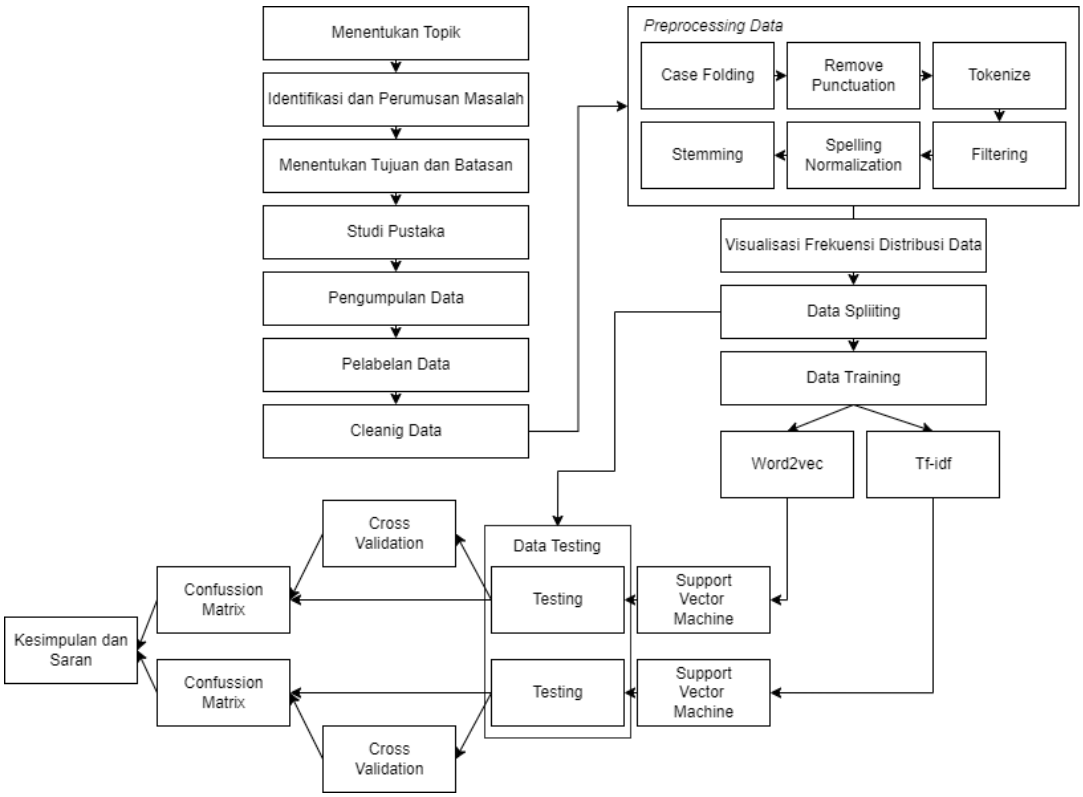
Tahap Demonstrasi. Pada tahap ini, rancangan yang dihasilkan dari tahap sebelumnya akan diimplementasikan. Penelitian ini akan mendemonstrasikan hasil rancangan apabila hasil tidak sesuai maka diperbaiki kembali ke tahap perancangan sampai mendapatkan hasil yang diinginkan. Penelitian mendemonstrasikan hasil rancangan dengan mengembangkan model klasifikasi menggunakan bahasa pemrograman *Python* dan beberapa tools yang dibutuhkan seperti *Google Collaboratory*, *Google Drive*, dan beberapa *library*.

Tahapan Evaluasi mengukur sejauh mana sistem mendukung solusi terhadap masalah. Tahapan ini melibatkan membandingkan tujuan solusi dengan hasil yang diamati dari sistem. Penelitian ini akan mengevaluasi model dengan melakukan validasi dan pengukuran kinerja model. Model yang telah dihasilkan divalidasi menggunakan *K-fold cross validation*, yaitu membagi jumlah data ke dalam k kelompok dimana secara bergilir 1/k bagian dijadikan sebagai data uji dan sisanya sebagai data latih sebanyak k kali. Evaluasi kinerja model dilakukan dengan menggunakan metric evaluasi seperti *accuracy, precision, recall,* dan *f1-score*

Tahap akhir yaitu Komunikasi. Di sini, laporan dari hasil penelitian yang berupa kesimpulan untuk publikasi artikel ilmiah ini. Isi kesimpulan berupa hasil analisis evaluasi model yang telah dibuat dan telah diujikan.

## 3.2 Tahap Penelitian

Tahapan rancangan penelitian komparasi metode *Word2vec* dan TF-IDF klasifikasi teks pada pemilihan legislatif 2024 dari media sosial *Twitter* dengan klasifikasi SVM. yang menunjukkan tahapan penelitian dari awal sampai akhir ditunjukkan Gambar 3.2




Gambar 3.2 Rancangan Tahapan Penelitian

Tahapan rancangan penelitian komparasi metode *Word2vec* dan TF-IDF untuk klasifikasi teks pada pemilihan legislatif 2024 dari media sosial *twitter* dengan klasifikasi *support vector machine* yang menunjukkan tahapan penelitian dari awal sampai akhir ditunjukkan oleh Gambar 3.2

1) Pengumpulan Data

Tahap ini melibatkan pengumpulan data yang diperlukan untuk menjawab masalah penelitian. Tahap pengumpulan data akan melibatkan penggunaan *library snscrape Twitter* untuk mengakses data tweet terkait pemilihan legislatif 2024 dari media sosial *Twitter*.

2) Pelabelan Data

Pelabelan data akan melibatkan memberikan label atau kategori teks pada setiap data *tweet* yang telah dikumpulkan. Proses pelabelan data dapat dilakukan secara manual dengan melibatkan manusia dalam mengkategorikan teks *tweet*. Data yang telah melewati proses pembersihan akan diberi label sebagai kategori negatif atau positif (Waskito, 2019). Seluruh data yang diperoleh dari proses pengumpulan data akan diberi label secara manual sesuai dengan teks positif atau negatif yang terkandung dalamnya.

3) *Cleaning* Data

*Cleaning* data atau pembersihan data adalah proses untuk menghilangkan elemen yang tidak diinginkan, seperti duplikat dan nilai yang hilang (*missing values*). Tujuannya memastikan bahwa data yang digunakan dalam analisis atau pemodelan adalah berkualitas, konsisten, sesuai dengan kebutuhan dan mengurangi *buzzer* atau akun yang mendapat imbalan dari mempopulerkan topik tertentu dengan cara mengirim *tweet* secara berulang (Felicia dan Loisa, 2018).

4) *Preprocessing Data*

*Preprocessing* data adalah tahap dimana data *tweet* yang telah dikumpulkan dari media sosial *Twitter* akan diolah dan disiapkan sebelum digunakan dalam klasifikasi teks. Tahap *preprocessing* bisa dilihat pada tabel

Tabel 3.1

Tahapan dan Penjelasan *Preprocessing* Teks

| No | Text Preprocessing | |
|---|---|---|
| | Tahap | Penjelasan |

| | | |
|---|---|---|
| 1 | *Spelling Normalization* | Normalisasi ejaan bertujuan untuk menangani variasi ejaan yang mungkin terjadi dalam data tweet. Ini dapat mencakup standarisasi ejaan, pemotongan (*trimming*) kata, atau penggantian kata-kata yang salah eja dengan bentuk yang benar. |
| 2 | *Case Folding* | Mengubah semua karakter dalam teks menjadi huruf kecil atau huruf besar. |
| 3 | *Remove Punctuation* | digunakan untuk menghapus atau mengganti tanda baca atau simbol-simbol yang digunakan dalam teks, seperti titik, koma, tanda tanya, tanda seru, tanda kurung, dan sebagainya dalam teks |
| 4 | *Tokenizing* | Proses memecah teks menjadi unit-unit yang lebih kecil yang disebut "token". Tokenisasi dilakukan dengan memisahkan kata-kata berdasarkan spasi atau tanda baca. |
| 5 | *Filtering* | penghilangan kata-kata yang tidak relevan atau tidak bermanfaat dalam klasifikasi teks. Kata-kata seperti *stopwords* (kata-kata umum seperti "dan", "atau", dll.), tautan URL, karakter khusus, atau tanda baca yang tidak memberikan kontribusi signifikan terhadap teks dapat dihapus. |
| 6 | *Stemming* | proses mengubah kata-kata menjadi bentuk dasarnya atau kata dasar. Hal ini dilakukan untuk menghapus awalan (*prefix*) dan akhiran (*suffix*) dari kata-kata yang berbeda tetapi memiliki akar kata yang sama dapat diperlakukan sebagai satu entitas dalam klasifikasi teks. |

5) *Data Spliting*

Pembagian Data yang telah selesai di preprocessing akan membagi data menjadi dua subset, yaitu data latih (*data train*) dan data uji (*data test*). Data pelatihan digunakan untuk melatih model klasifikasi SVM dengan metode *Word2vec* dan TF-IDF. Data pengujian digunakan untuk menguji performa model yang telah dilatih dan mengevaluasi hasil klasifikasi teks.

6) Pelatihan dan pengujian dengan Klasifikasi *Support Vector Machine*

Melatih model klasifikasi *Support Vector Machine* menggunakan fitur-fitur yang diperoleh dari *Word2vec* atau TF-IDF. Dalam tahap pelatihan, akan ada dua skenario yang digunakan. Pertama, menggunakan metode *Word2vec* dengan *Support*

*Vector Machine*, dan kedua, menggunakan metode TF-IDF dengan *Support Vector Machine*. Setelah dilakukan pelatihan, model yang telah terlatih akan diuji menggunakan data uji pada setiap skenario.

7) Kesimpulan dan Saran

Tahap kesimpulan melibatkan evaluasi performa model algoritma *Support Vector Machine* saat menggunakan *dataset* yang sama, yaitu TF-IDF dan *Word2vec*. Pada tahap ini, akan digunakan matriks kebingungan (*confusion matrix*) yang merupakan tabel untuk menghitung dan mengevaluasi kinerja model klasifikasi dengan memperhitungkan jumlah objek penelitian yang diprediksi dengan benar dan salah. Secara singkat, *confusion matrix* dan *K-fold cross validation* memberikan rincian tentang hasil evaluasi dan validasi kinerja klasifikasi yang terjadi. Tahap saran merupakan tahap di mana peneliti memberikan saran untuk penelitian selanjutnya atau pengembangan lebih lanjut terkait topik yang telah diteliti.

## 3.3 Objek dan Subjek Penelitian

Objek penelitian ini adalah *tweet* yang berisikan persepsi pengguna media sosial *twitter* terhadap pemilihan legislatif 2024. Sedangkan subjek pada penelitian ini merupakan pengguna media sosial *Twitter* yang *tweet* terkait pemilihan legislatif 2024 dari rentang waktu April - Juni 2023.

## 3.4 Populasi dan Sampel Penelitian

Dalam konteks klasifikasi teks untuk pemilihan legislatif 2024 dari media sosial *Twitter*, populasi penelitian adalah semua *tweet* yang dibuat oleh pengguna *Twitter* terkait pemilihan legislatif 2024. Namun, karena jumlah *tweet* yang ada dalam populasi tersebut sangat besar, maka seringkali digunakan sampel sebagai representasi dari populasi tersebut.

Sampel penelitian adalah *subset* dari populasi yang dipilih untuk dianalisis secara lebih mendalam. Sampel dapat dipilih secara acak atau dengan menggunakan metode tertentu, seperti pengambilan sampel stratifikasi berdasarkan wilayah geografis atau topik tertentu terkait pemilihan legislatif. Sampel *tweet* diambil menggunakan

*Twitter* dengan *library snscrape* dengan menggunakan kata kunci, seperti: pemilihan legislatif 2024 dan calon legislatif 2024 dan jumlah tweet sebanyak 10.202 data dari rentang waktu April sampai Juni 2023, karena pada rentang waktu tersebut adanya pencalonan anggota DPR, DPRD provinsi dan DPRD kabupaten/kota (InfoPublik, 2023), adanya parpol dan caleg gencarkan sosialisasi (Rahayu, 2023), pengumuman bakal calon legislatif oleh masing masing partai (Abduh, 2023).

## 3.5 Instrumen Penelitian

Instrumen penelitian dalam penelitian ini menggunakan perangkat keras atau *hardware* dan perangkat lunak atau *software*, sebagai berikut:

1) Perangkat Keras

Berikut adalah perangkat keras yang digunakan selama penelitian yaitu laptop acer dengan spesifikasi sebagai berikut:

1. *Processor intel core* i3 *gen* 7
2. RAM 4 GB
3. Hardisk 250 GB

2) Perangkat Lunak

Adapun perangkat Lunak yang digunakan selama penelitian sebagai berikut:

1. *Google colaboratory*
2. *Google Drive*
3. *Google chrome*
4. *Python* 3.6.9

## 3.6 Metode Pengumpulan Data

Pada penelitian ini, metode pengumpulan data yang digunakan adalah metode *web scraping*. pengambilan data menggunakan *Twitter* dengan *library snscrape*. Berikut adalah langkah-langkah umum dalam pengumpulan data menggunakan Twitter dengan *library snscrape*:

1) Menggunakan *Snscrape* untuk *Scrapping* data: *Snscrape* adalah *library Python* yang populer untuk mengumpulkan data dari *Twitter* secara luas dan fleksibel

tanpa perlu menggunakan *API Twitter*. Anda dapat *install Snscrape* menggunakan *pip* dan *import library* tersebut ke dalam kode *Python* Anda.

2) Mengambil data *tweet*: Setelah *install snscrape* berhasil, Anda dapat menggunakan beberapa baris kode untuk mengambil data *tweet* berdasarkan kata kunci atau parameter lain yang sesuai dengan kebutuhan Anda. Misalnya, Anda dapat menggunakan metode *search* untuk mengambil tweet berdasarkan kata kunci tertentu.

3) Mengolah dan menyimpan data: Setelah Anda mendapatkan data *tweet*, Anda dapat melakukan langkah-langkah *preprocessing* yang diperlukan, seperti membersihkan teks *tweet*, menghapus karakter khusus, atau melakukan tokenisasi. Selanjutnya, peneliti menyimpan data tweet dalam format yang sesuai, seperti file *csv*, untuk analisis selanjutnya.

## 3.7 Jenis dan Sumber Data

Jenis data yang digunakan dalam penelitian ini merupakan data primer. Sumber data utama adalah media sosial *Twitter*. Dalam hal ini, pengguna *Twitter* yang aktif menyampaikan pendapat, dukungan, kritik, atau pendapat mereka melalui *tweet-tweet* yang mereka buat. Data tersebut diambil dengan menggunakan pengambilan data melalui *library snscrape* dari media sosial *twitter* dengan menggunakan kata kunci atau parameter pencarian terkait pemilihan legislatif 2024.

## 3.8 Metode Analisis Data

Dalam penelitian ini, digunakan *platform Google Colaboratory* dan *Twitter Developer* untuk mengakses data. Penelitian bersifat deskriptif adalah menggambarkan karakteristik fenomena yang sedang diteliti secara mendalam, luas, dan terperinci. penelitian melakukan pendekatan secara kuantitatif dengan menghasilkan angka, tabel, atau diagram yang menggambarkan hasil kinerja model *Word2vec* dan TF-IDF dengan klasifikasi SVM. Terdapat beberapa analisis data yang diterapkan, antara lain:

a. Analisis deskriptif, digunakan untuk memberikan gambaran tentang *tweet* dari pengguna *Twitter* yang ada di media sosial tersebut.

b. Metode *machine learning*, khususnya *Support Vector Machine* (SVM), digunakan untuk melakukan klasifikasi ulasan ke dalam kategori positif dan negatif.
c. *Wordcloud* digunakan untuk visualisasi kata-kata yang paling sering muncul dalam ulasan.
d. *Confusion Matrix* dan *K-fold Cross Validation* digunakan untuk melihat kinerja pada model.