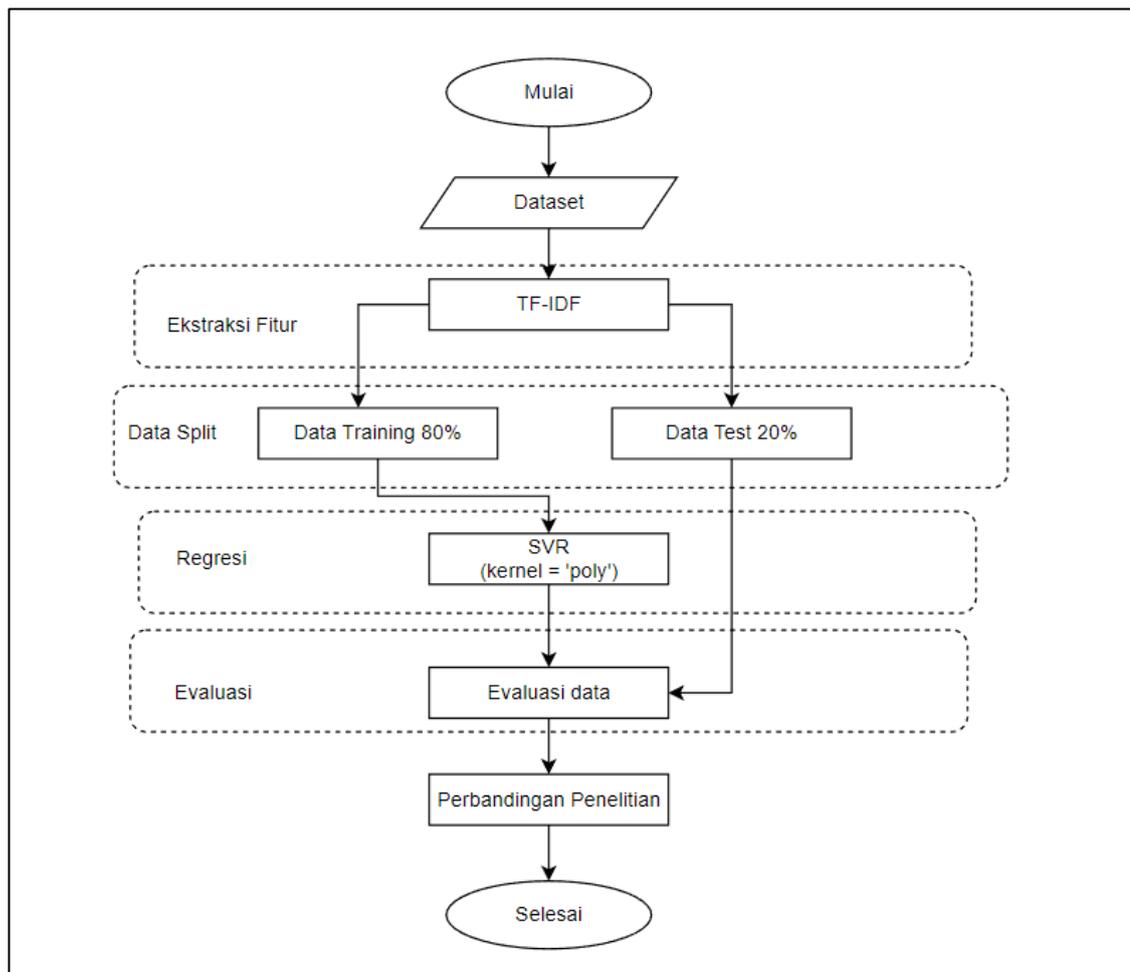


BAB 3

METODOLOGI PENELITIAN

3.1 Tahapan Penelitian

Penelitian ini akan melakukan pencarian hasil evaluasi RMSE dari penggunaan metode SVR dengan kernel *polynomial* lalu membandingkan hasilnya dengan penelitian yang dilakukan oleh Verdikha (2021) yang menggunakan kernel rbf, dan menghasilkan nilai evaluasi RMSE sebesar 2,166. Hasil dari penelitian Verdikha (2021) akan dibandingkan dengan hasil evaluasi RMSE yang didapatkan pada penelitian ini, tahapan penelitian dapat dilihat pada gambar 3.2.



Gambar 3.1 Tahapan penelitian

3.2 Dataset

Data yang digunakan pada penelitian ini diperoleh dari penelitian yang dilakukan oleh Thamrin (2021). Data ini diambil dari hasil jawaban essay pelajaran bahasa indonesia yang dilakukan di UMKT pada semester 2 tahun 2020, beberapa data penelitian yang akan digunakan dapat dilihat pada tabel 3.1

Tabel 3.1 Data penelitian

nilai	kelas	jawaban
8	A	faktor yg sebab sebab bahasa melayu angkat jadi bahasa indonesia bahasa melayu rupa lingua franca indonesia bahasa hubung
8	A	bahasa melayu angkat jadi bahasa satu indonesia 28 oktober 1928 peristiwa sebut sumpah pemuda sejak itu bahasa
8	A	empat faktor sebab bahasa melayu angkat jadi bahasa indonesia 1 bahasa melayu rupa lingua franca indonesia bahasa hubung bahasa
8	A	alas bahasa melayu pilih jadi bahasa indonesia 1 bahasa melayu rupa lingua
.....		
4	H	cakup jumlah bahasa saling mirip tutur wilayah nusantara di semenanjung melayu bahasa luas pakai bahasa jadi bahasa resmi
4	H	empat faktor bahasa melayu angkat jadi bahasa indonesia bahasa melayu rupa lingua franca indonesia bahasa hubung bahasa dagang
4	H	memang banyak guna bagi besar masyarakat indonesia mudah ajar bahkan ajar sekolah bumi putera waktu
4	H	bahasa melayu ini akhir jadi bahasa indonesia alas bahasa melayu jadi bahasa indonesia 1 bahasa melayu fungsi lingua franca

Tabel 3.1 menunjukkan bahwa data berisi beberapa atribut, pertama atribut nilai yang berisi nilai siswa dengan nilai dari 0 sampai 12, lalu terdapat atribut kelas yang berisi data kelas siswa dari kelas A sampai kelas H, terakhir terdapat atribut jawaban yang berisi semua jawaban siswa. Data ini sebelumnya sudah melalui beberapa tahapan pra proses diantaranya adalah :

- *Data Cleansing*, untuk menghapus *non-alphanumeric characters* seperti () ` ~ ! @ # \$ % ^ & * - + = | \ { } [] : ; " ' < > , . ? / _ . dan menghapus jawaban yang belum di isi,
- *Case folding*, untuk mengubah semua karakter menjadi huruf kecil,
- *White space removal*, untuk menghilangkan spasi yang berlebih,
- *Spell correction*, untuk mengoreksi kesalahan ejaan dalam sebuah penulisan,
- *Stop word*, digunakan untuk menghapus kata umum (*common words*) yang biasanya muncul dalam jumlah besar dan dianggap tidak memiliki makna, misalnya “dan”, “yang”, “di”,
- *Stemming*, digunakan untuk menghilangkan kata imbuhan pada teks, seperti imbuhan kata awal dan akhir, contohnya “menghabiskan” kata “meng-” merupakan imbuhan awal dan “-kan” adalah imbuhan akhir.

Tahapan - tahapan praproses ini dibutuhkan untuk mempersiapkan data mentah sebelum dilakukan proses lain. Praproses data dilakukan dengan cara mengeliminasi data yang tidak sesuai atau mengubah data menjadi bentuk yang lebih mudah diproses oleh sistem. Praproses juga dilakukan untuk mendapatkan hasil yang lebih akurat, pengurangan waktu perhitungan untuk large scale problem, dan membuat nilai data menjadi lebih kecil tanpa merubah informasi yang didalamnya (Nasution et al., 2019).

3.3 Ekstraksi Fitur TF-IDF

Melakukan ekstraksi fitur menggunakan TF-IDF untuk mengekstrak keywords (kata-kata inti) pada data jawaban mahasiswa, menghitung bobot setiap kata, dan mencari suatu kata yang sering muncul, jadi hasil dari ekstraksi

fitur ini akan digunakan untuk mengetahui seberapa penting suatu kata pada dokumen dengan melakukan pembobotan menggunakan TF-IDF.

3.4 Data Split

Membagi data menjadi data training dan data test dengan perbandingan 80:20 karena mengikuti rasio data split penelitian sebelumnya sehingga hasil dari evaluasi RMSE dapat dibandingkan dengan penelitian ini. Menurut Saifudin (2018) *Data training* digunakan untuk melatih algoritma klasifikasi, sedangkan *data test* digunakan untuk menguji algoritma/model yang telah dilatih.

3.5 Regresi SVR dan Evaluasi

Selanjutnya menggunakan *Support Vector Regression (SVR)* dengan parameter kernel *polynomial* akan mengatasi masalah regresi pada data yang bertujuan untuk mengecilkan persentase error atau selisih yang terdapat pada nilai prediksi dengan nilai sebenarnya, nilai prediksi disini adalah y_{pred} dan nilai sebenarnya adalah y_{test} .

Untuk mengatasi bilangan non-linear data akan di naikan ke dimensi yang lebih tinggi menggunakan kernel *polynomial* agar nilai mendekati garis batas, sehingga masih bisa di toleransi. Selanjutnya melakukan evaluasi data menggunakan *Root Mean Square Error (RMSE)* untuk mengetahui tingkat kesalahan penggunaan metode SVR dengan parameter kernel *polynomial*.

3.6 Perbandingan Kinerja

Hasil dari regresi SVR dan evaluasi yang sudah didapatkan pada penelitian ini akan dilakukan perbandingan dengan penelitian sebelumnya untuk mengetahui apakah hasil yang didapatkan pada penelitian ini lebih baik dibandingkan penelitian sebelumnya atau sebaliknya.