**BAB II**

**TINJAUAN PUSTAKA**

Bab ini menjelaskan tentang definisi dan teori-teori yang digunakan sebagai landasan penelitian yang berasal dari hasil publikasi dan penelitian dan/atau buku yang relevan.

(Perlu ditambah dan diisi untuk teori pada bagian berikut :

1. **Analisis Sentimen**

…

1. **Twitter**

…

1. **SVM**

Support Vector Machine(SVM) adalah salah satu algoritma pembelajaran mesin yang sering digunakan untuk melakukan klasifikasi. Algoritma ini menggunakan kernel untuk memetakan data ke ruang dimensi yang lebih tinggi untuk mengklasifikasikan data yang tidak dapat dipidahkan secara linear. [6] [7]. Penggunaan kernel pada algoritma SVM membutuhkan sekumpulan parameter yang disebut hyperparameter dan dapat diubah sesuai dengan kebutuhan. Parameter tersebut antara lain c dan gamma. Parameter c berfungsi sebagai parameter pengatur, sedangkan parameter gamma adalah koefesien yang digunakan pada RPF, Poli, atau kernel Sigmoid.

Penentuan nilai hyperparameter agar mendapatkan hasil yang optimal dilakukan dengan menggunakan metide GridSerachCV. GridSearchCv

adalah algoritma yang cukup sederhana untuk pencarian nilai hyperparameter optimal. Algoritma ini adalah algoritma pencarian lengkap melalui subset tertentu dari ruang hyperparameter dari algoritma pembelajaran. [6] [8]

Pada Data train yang didapat dengan menggunkan k-fold cross validation akan digunakan model SVM dengan nilai hyperparameter yang didapat untuk melatih model. Lalu setelahnya model yang sudah di-*train* akan disimpan dan digunakan untuk melakukan klasifikasi data *test*. Klasifikasi data *test* ini akan dihitung dengan menggunakan persamaan :

 (4)

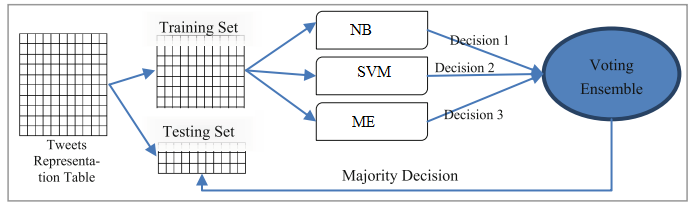
dimana; x adalah vektor fitur input , w adalah vektor bobot, dan b adalah bias.

Setelah melakukan klsifikasi, tingkat akurasi dari model akan diukur. Tingkat akurasi dihitung dnegan membandingkan hasil prediksi dengan hasil aktual. Nilai akurasi dihihtung pada setiap fold dan kemudian dijumlahkan untuk mendapatkan nilai akurasi rata-rata.

.

1. **Ensemble**

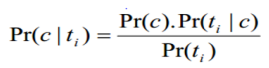
Ensemble model dengan metode majority voting adalah algoritma yang menggunakan satu atau lebih algoritma diklasifikasikan, dan hasil luarannya diberikan berdasarkan pemilihan predictor (nilai mode dari semua algoritma) [9]. Algoritma yang digunakan dalam Ensemble model ini aadalah Naïve Bayes, SVM, dan Maximum Entropy



***Ensemble Model : Majority Votting***

* 1. **Naïve Bayes**

Salah satu algoritma klasifikasi probabilistic yang paling mudah selama menggunakan algoritma naïve bayes yang kredibel [10]. Anggapan ini berarti bahwa setiap kata adalah bebas dan pada dasarnya tidak berhubungan. Pada algoritma ini, C adalah kumpulan label, dan ti adalah dokumen baru yang akan diklasifikasi. Secara matematis probabilitasnya adalah :

 (5)

Dengan menggunakan perhitungan diatas, dilakukan pelatihan model dengan data train dan setelahnya model yang sudah terbentuk akan digunakan untuk klasifikasi data train.

* 1. **Maximum Entropy**

Dalam klasifikasi menggunakan algoritma Maximum Entropy, tidak ada asumsi yang digunakan dalam hubungan antar fitur. Metode ini hanya bertujuan untuk memakksimalkan entorpu pada system dengan memprediksi kondisi distribusi label pada masing masing kelas. Distribusi ini kemudian didefinisikan sebagai Maximum Entropy yang tidak membuat asumsi apapun pada fitur-fiturnya [11]

Secara matematis ditulis dalam persamaan berikut:

 (6)

Hasil dari klasifikasi majority votting, akan diukur tingkat akurasinya dengan membandingkan hasil klasifikasi dengan data sebenarnya. Nilai akurasi dihihtung pada setiap fold dan kemudian dijumlahkan untuk mendapatkan nilai akurasi rata-rata.

.

Penelitian yang relevan :

Sentiment Analysis yang dilakukan menggunakan algoritma SVM dan Algoritma Ensamble untuk melihat perbandingannya. Hal ini karena dalam jurnal Abdullah Alsaeedi, penggunaan Algoritma Naïve Bayes dan SVM dalam melakukan sentiment analisis mencapai akurasi 80% dan untuk algoritma ensamble dan hybrid mencapai akurasi 85%. [6]. Perbandingan Naïve Bayes dan SVM sendiri, dalam penelitian R Vihdya,dkk dan Marcio,dkk mengatakan bahwa Algoritma SVM lebih unggul dibanding algoritma Naïve Bayes dan RF [7] [8]. Sejalan dengan pemikiran R Vidhya, Husein, dkk dan Abdul Mohimen,dkk juga mengatakan bahwa algoritma SVM memiliki nilai akurasi yang lebih tinggi dibanding Naïve Bayes dalam melakukan sentiment analisis. [9] [10].

Algoritma pendekatan ensemble adalah metodologi pembelajaran yang menggabbungkan beberapa model untuk membangun model tunggal yang lebih besar dan lebih kuat [11]. Pada penelitian ini akan menggunakan pendekatan klasifikasi ensemble dengan metode majority voting yakni satu atau lebih algoritma diklasifikasikan, dan hasil luarannya diberikan berdasarkan pemilihan predictor (nilai mode dari semua algoritma) [12]. Dengan menggunakan pendekatan ensemble ini diharapkan hasil yang didapat akan lebih tinggi. Disini untuk melakukan klasifikasi ensemble akan digunakan algoritma klasifikasi SVM, NB, dan Maximum Entropy. Pemilihan ketiga algoritma ini karena ketiganya merupakan algoritma klasifikasi yang cukup baik dalam melakukan sentiment analysis [6].