

## BAB II

### TINJAUAN PUSTAKA

#### 2.1 Tinjauan Pustaka

Dalam sebuah penelitian tentunya tak bisa lepas dari pendukung yang dibutuhkan dalam sebuah penulisan yakni teori – teori yang berasal dari jurnal terdahulu yang berkaitan dengan klasifikasi menggunakan algoritma *Support Vector Machine* yang sudah dipublikasikan. Berikut merupakan referensi jurnal yang dipakai.

Penelitian yang dilakukan oleh Indri Monika Parapat, Muhammad Tanzil Furqon, dan Sutrisno dengan mengangkat permasalahan penyimpangan tumbuh kembang anak yang terlambat diketahui dapat berakibat jangka panjang dan sukar diperbaiki. Berdasarkan permasalahan tersebut, penelitian ini dilakukan menggunakan *Support Vector Machine* (SVM) untuk klasifikasi penyimpangan tumbuh kembang anak. Metode SVM terdiri dari proses *training* sebagai pembelajaran sistem dan testing untuk memperoleh hasil klasifikasi. Pengujian parameter yang dilakukan adalah pengujian lambda, *complexity*, dan iterasi maksimal. Data digunakan pada penelitian sebanyak 90 data yang terbagi menjadi 3 kelas. Kelas penelitian ini mewakili 3 jenis penyimpangan tumbuh kembang anak yaitu *Down Syndrome*, *Autisme*, dan *Attention Deficit Hyperactivity Disorder* (ADHD). Algoritma SVM merupakan metode klasifikasi linier, sehingga menggunakan *kernel* untuk mengatasi data yang bersifat non-linier. Hasil akhir dari penelitian ini menghasilkan rata-rata akurasi tertinggi sebesar 63,11%  $\lambda = 10$ ,  $C = 1$ , *itermax* = 200 dan juga menggunakan *kernel polynomial* [5].

Selanjutnya merupakan penelitian dari Tryse Rezza Biantong, Muhammad Tanzil Furqon, dan Arief Andy Soebroto dengan mengangkat permasalahan implementasi metode *Support Vector Machine* untuk klasifikasi jenis penyakit Malaria yang mempunyai 4 varian yaitu Malaria Tertiana, Malaria Tropika, Malaria Kuartana, dan Malaria Ovale Data yang digunakan sebanyak 200 data yang diambil dari Dinas Kesehatan Kabupaten Nabire, Papua. Pada pengujian ini digunakan *K-*

*fold Cross Validation* dengan nilai  $K\text{-fold}=10$ . Hasil akurasi terbaik yang dihasilkan oleh sistem ini adalah 72% dengan nilai parameter  $\lambda=0.1$ ,  $\sigma=1$ ,  $\gamma=0.001$ ,  $C=0.1$ ,  $\epsilon=1.10^{-5}$ ,  $itermax = 50$  pada data rasio 80% data *training* : 20% data testing [6].

Selanjutnya merupakan penelitian dari Daisy Kurniawaty, Imam Cholissodin, dan Putra Pandu Adikara dengan mengangkat permasalahan klasifikasi gangguan jiwa Skizofrenia menggunakan algoritma *Support Vector Machine*. Dalam penelitian ini menggunakan 75 data yang terbagi menjadi dua jenis gangguan jiwa skizofrenia yaitu paranoid dan *simplex*. Metode yang digunakan adalah algoritme SVM yang mana termasuk dalam kategori metode klasifikasi yang baik, yang mana memberikan suatu pendekatan statistik dalam pengenalan pola, dan merupakan metode yang bersifat linier, tetapi SVM menyediakan *kernel trick*, yang dapat menyelesaikan masalah terkait klasifikasi non-linier. Hasil pengujian akhir menghasilkan rata-rata akurasi optimal yaitu 100% dengan menggunakan rasio perbandingan 90%:10%, nilai gamma = 0,00001, lambda = 3, nilai C = 0,01, *kernel polynomial of degree*, dan iterasi maksimal adalah 1000 [7].

Selanjutnya merupakan penelitian dari Edgar Maulana Thoriq, Dian Eka Ratnawati, dan Bayu Rahayudi dengan mengangkat permasalahan Analisis Sentimen Opini Publik pada Media Sosial Twitter terhadap Vaksin Covid-19 menggunakan Algoritma *Support Vector Machine* dan *Term Frequency-Inverse Document Frequency*. Penelitian ini menggunakan data sebanyak 450 tweet, kemudian pengujian dilakukan dengan menggunakan metode cross validation dengan jumlah fold = 10. Performa terbaik dari algoritma klasifikasi yang diperoleh adalah *accuracy* sebesar 86%, *precision* sebesar 88%, *recall* sebesar 82%, dan *f-measure* sebesar 85%. Nilai dari performa tersebut diperoleh dengan nilai C sebesar 1 dan nilai iterasi maksimum sebesar 300 [8].

Selanjutnya merupakan penelitian dari Ratna Ayu Wijayanti, Muhammad Tanzil Furqon, dan Sigit Adinugroho dengan mengangkat permasalahan Penerapan Algoritme *Support Vector Machine* Terhadap Klasifikasi Tingkat Pasien Gagal Ginjal Di Indonesia menurut data dari penetri (Persatuan Netrologi Indonesia) diperkirakan mencapai 70 ribu penderita penyakit gagal ginjal kronik. Untuk

membantu mengetahui status fungsi ginjal seseorang dibuat suatu sistem yang dapat melakukan proses klasifikasi tingkat risiko pasien gagal ginjal menggunakan algoritme *Support Vector Machine* (SVM) dan strategi *one-against-all*. Alur dari penelitian yaitu menggunakan analisis korelasi untuk melihat hubungan antar fitur, melakukan normalisasi agar nilai data berada pada interval yang sama, perhitungan kernel RBF, melakukan proses *training* dengan *sequential training*, kemudian menggunakan *one-against-all* untuk proses klasifikasi. Pengujian akhir dari penelitian ini menghasilkan nilai rata-rata akurasi sebesar 83,998% dan akurasi tertinggi sebesar 98,33% dengan menggunakan rasio perbandingan data 80%:20%, nilai parameter  $\lambda$  (*lambda*) = 1,  $\gamma$  (*gamma*) = 0,0001,  $\sigma$  kernel RBF = 2, C (*Complexity*)=0,0001 dan jumlah iterasi SVM=100. Berdasarkan hasil tersebut dapat disimpulkan bahwa algoritme SVM dan strategi *one-against-all* dapat digunakan untuk klasifikasi tingkat risiko pasien gagal ginjal [9].

Selanjutnya merupakan penelitian dari Hilda Apriyani dan Kurniati dengan mengangkat permasalahan Perbandingan Metode *Naive Bayes* Dan *Support Vector Machine* Dalam Klasifikasi Penyakit Diabetes Melitus dengan adanya permasalahan tersebut maka penulis melakukan analisis perbandingan dari dua algoritma yaitu algoritma *Naive Bayes* dan algoritma *Support Vector Machine* untuk klasifikasi penyakit diabetes dengan menggunakan alat bantu WEKA dengan tools options *Cross Validation* dan *Confussion Matrix* dengan hasil akurasi tertinggi yaitu algoritma *Support Vector Machine* dengan *kernel polynomial* yang hasilnya 96.2704% dan tingkat error sebanyak 3.7296% dapat disimpulkan algoritma yang akurat dalam klasifikasi penyakit diabetes yaitu algoritma *Support Vector Machine* dengan *kernel polynomial* [10].

Tabel 2.1 Penelitian Terdahulu

No.	Judul	Tahun	Peneliti	Hasil
1.	Penerapan Metode Support Vector Machine (SVM) Pada Klasifikasi Penyimpangan Tumbuh Kembang Anak	2018	Indri Monika Parapat, Muhammad Tanzil Furqon, dan Sutrisno	Hasil akhir dari penelitian ini menghasilkan rata-rata akurasi tertinggi sebesar 63,11% $\lambda = 10$ , $C = 1$ , $itermax = 200$ dan juga menggunakan <i>kernel polynomial</i>
2.	Implementasi Metode Support Vector Machine Untuk Klasifikasi Jenis Penyakit Malaria	2019	Tryse Rezza Biantong, Muhammad Tanzil Furqon, dan Arief Andy Soebroto	Pada pengujian ini digunakan <i>K-fold Cross Validation</i> dengan nilai <i>K-fold=10</i> . Hasil akurasi terbaik yang dihasilkan oleh sistem ini adalah 72% dengan nilai parameter $\lambda=0.1$ , $\sigma=1$ , $\gamma=0.001$ , $C=0.1$ , $\epsilon=1.10^{-5}$ , $itermax = 50$ pada data rasio 80% data <i>training</i> : 20% data testing
3.	Klasifikasi Gangguan Jiwa Skizofrenia Menggunakan Algoritma Support Vector Machine	2018	Daisy Kurniawaty, Imam Cholissodin, dan Putra Pandu Adikara	Hasil pengujian akhir menghasilkan rata-rata akurasi optimal yaitu 100% dengan menggunakan rasio perbandingan 90%:10%, nilai $\gamma = 0,00001$ , $\lambda = 3$ , nilai C

				= 0,01, <i>kernel polynomial of degree</i> , dan iterasi maksimal adalah 1000
4.	Analisis Sentimen Opini Publik pada Media Sosial Twitter terhadap Vaksin Covid=19 menggunakan Algoritma Support Vector Machine dan Term Frequency-Inverse Document Frequency	2021	Edgar Maulana Thoriq, Dian Eka Ratnawati, dan Bayu Rahayudi	Dengan menggunakan metode <i>Cross Validation</i> dengan jumlah <i>fold</i> = 10. Performa terbaik dari algoritma klasifikasi yang diperoleh adalah <i>accuracy</i> sebesar 86%, <i>precision</i> sebesar 88%, <i>recall</i> sebesar 82%, dan f-measure sebesar 85%. Nilai dari performa tersebut diperoleh dengan nilai C sebesar 1 dan nilai iterasi maksimum sebesar 300
5.	Penerapan Algoritma Support Vector Machine Terhadap Klasifikasi Tingkat Risiko Pasien Gagal Ginjal	2018	Ratna Ayu Wijayanti, Muhammad Tanzil Furqon, dan Sigit Adinugroho	Pengujian akhir dari penelitian ini menghasilkan nilai rata-rata akurasi sebesar 83,998% dan akurasi tertinggi sebesar 98,33% dengan menggunakan rasio perbandingan data 80%:20%, nilai parameter $\lambda$ (lambda) = 1, $\gamma$ (gamma) = 0,0001, $\sigma$ <i>kernel RBF</i> = 2, $C$ ( <i>Complexity</i> )=0,0001 dan jumlah iterasi SVM=100. Berdasarkan hasil tersebut

				dapat disimpulkan bahwa algoritme SVM dan strategi <i>oneagaints-all</i> dapat digunakan untuk klasifikasi tingkat risiko pasien gagal ginjal
6.	Perbandingan Metode Naive Bayes dan Support Vector Machine Dalam Klasifikasi Penyakit Diabetes Melitus	2020	Hilda Apriyani dan Kurniati	Klasifikasi penyakit diabetes dengan menggunakan alat bantu WEKA dengan tools options <i>Cross Validation</i> dan <i>Confussion Matrix</i> dengan hasil akurasi tertinggi yaitu algoritma Support Vector Machine dengan kernel polynomial yang hasilnya 96.2704% dan tingkat error sebanyak 3.7296% dapat disimpulkan algoritma yang akurat dalam klasifikasi penyakit diabetes yaitu algoritma Support Vector Machine dengan kernel polynomial

## 2.2 Landasan Teori

### 2.2.1 Machine Learning

Machine Learning adalah sebuah teknik yang mengembangkan performa sistem dengan mempelajari lewat metode komputasi. Dalam sistem komputer pengalaman ada dalam bentuk data, dan tugas utama dari Machine Learning ialah mengembangkan algoritma pembelajaran yang membangun sebuah model dari data. Dengan memberikan algoritma pembelajaran dengan data, kita mendapatkan model yang bisa membuat prediksi (contoh apel yang sudah matang) dalam sebuah observasi baru (contoh apel yang belum dipotong). Jika kita menganggap *computer science* sebagai subjek algoritma, maka *machine learning* merupakan sebuah subjek pembelajaran algoritma [11].

### 2.2.2 Klasifikasi

Klasifikasi adalah proses pengelompokan objek, entitas, atau konsep ke dalam kategori atau kelas yang memiliki karakteristik serupa. Tujuan utamanya adalah mengurangi kompleksitas dan memahami hubungan antara objek atau konsep yang dikelompokkan [12].

### 2.2.3 Support Vector Machine

*Support Vector Machine* atau SVM adalah salah satu algoritma pembelajaran mesin / *machine learning* yang populer dan efektif untuk klasifikasi dan regresi. Secara khusus, SVM umumnya digunakan untuk tugas klasifikasi data dengan dua kelas (data biner), meskipun dapat diperluas untuk tugas klasifikasi *multiclass* juga .

SVM sendiri pertama kali diperkenalkan oleh Vapnik pada tahun 1992 sebagai salah satu metode *machine learning* yang bekerja dengan prinsip *Structural Risk Minimization* atau SRM yang bertujuan menemukan *hyperlane* terbaik yang memisahkan dua buah *class* pada *input space*. Metode ini memakai hipotesis fungsi linier dalam sebuah ruang fitur yang berdimensi tinggi, dengan mengaplikasikan *learning bias* yang berasal dari teori pembelajaran statistik. Tingkat akurasi pada model yang akan dihasilkan oleh proses peralihan dengan SVM sangat bergantung pada fungsi kernel dan parameter yang digunakan [14].

#### **2.2.4 Python Programming Language**

*Python* adalah bahasa pemrograman tingkat tinggi yang populer dan serbaguna. Diciptakan oleh Guido Van Rossum dan pertama kali dirilis pada tahun 1991, *Python* dirancang dengan filosofi yang menekankan kejelasan, kesederhanaan, dan mudah dibaca. Kelebihan ini membuatnya menjadi bahasa pemrograman yang ideal untuk pemula, sambil tetap menjadi pilihan yang kuat untuk para pengembang yang lebih berpengalaman.

Penelitian ini menggunakan *Python* dengan bantuan *framework* atau modul yang ada untuk *Python*. *Framework* (modul/package) merupakan sebuah kerangka kerja yang digunakan untuk mempermudah para developer software dalam membuat dan mengembangkan aplikasi. *Framework* berisikan perintah dan fungsi dasar yang umum digunakan untuk membangun sebuah *software* aplikasi sehingga diharapkan aplikasi dapat dibangun dengan lebih cepat serta tersusun dan terstruktur dengan cukup rapi.

#### **2.2.5 Natural Language Processing**

NLP (*Natural Language Processing*) adalah cabang dari kecerdasan buatan (*Artificial Intelligence*) yang berfokus pada interaksi antara manusia dan komputer melalui bahasa manusia alami. Tujuan utama NLP adalah memungkinkan komputer untuk memahami, memproses, dan menghasilkan bahasa manusia seperti layaknya manusia.

### **2.2.6 Fake Job Posting**

Job posting palsu (*Fake Job Posting*) adalah tawaran pekerjaan atau lowongan pekerjaan yang dipublikasikan dengan maksud menipu atau menipu calon pencari kerja. Tujuan dari job posting palsu adalah untuk mencuri informasi pribadi, mengirimkan penipuan, atau bahkan meminta uang dari calon pencari kerja.