

Universitas Pendidikan Indonesia

**PENGEMBANGAN MODEL KLASIFIKASI PENALIZE SUPPORT VECTOR
MACHINE DALAM MEMPREDIKSI KELULUSAN SELEKSI NASIONAL
BERDASARKAN PRESTASI (SNBP)**

Bagus Syamsu Rahmatullah
2003164

Overview

01 Latar Belakang

02 Rumusan Masalah

03 Tujuan Penelitian

04 Batasan Penelitian

05 Penelitian Terdahulu

Overview

06 Kajian Pustaka

07 Desain Penelitian

08 Alat & Bahan Penelitian

04 Metric Penelitian

Latar Belakang

- Seleksi Nasional Berdasarkan Prestasi Negeri atau SNBP adalah salah satu jalur untuk masuk kedalam perguruan tinggi negeri yang dikelola oleh Seleksi Nasional Penerimaan Mahasiswa Baru (SNPMB) dan di bentuk oleh Kementrian Pendidikan dan Kebudayaan (Website Resmi SNPMB, 2024).
- Pada umumnya sekolah menengah atas menggunakan peringkat keseluruhan siswa pada sekolah untuk mengetahui siswa mampu bersaing untuk dapat masuk Perguruan Tinggi Negeri (PTN) melalui SNBP. Oleh karena itu, dengan adanya analisis data dengan menggunakan regresi, klasifikasi atau clustering dapat membantu dalam mengetahui presentase kemungkinan lulusnya SNBP (Oloruntoba & Akinode, December 2017)

Latar Belakang

- Penggunaan *machine learning* menjadi solusi dalam menganalisis data, dapat membantu sekolah dalam memberikan keputusan dengan lebih akurat dalam memilih PTN mana yang kemungkinannya lebih tinggi. Tapi pemilihan metode juga harus disesuaikan dengan data yang akan digunakan pada saat analisis data menggunakan ML. Pemilihan metode yang akan digunakan dalam menganalisis sebuah data harus dipertimbangkan beberapa faktor diantaranya adalah, kelas atau label yang sudah diketahui klasifikasinya, kemiripan pada sebuah data yang ada dan hubungan keberlanjutan antara data sekarang dan data sebelumnya.
- Support Vector Machine (SVM) adalah metode ML yang dapat digunakan sebagai klasifikasi dan regresi sekaligus dan SVM menggunakan hyperplane untuk menganalisa data yang akan mengategorikan data menjadi dua keanggotaan kelas yang baru (Nisbet, Miner, & Yale, 2018). Penalize SVM adalah sebuah variasi dari algoritma SVM yang menambahkan penalti pada fungsi objektifnya

Rumusan Masalah

Bagaimana implementasi model klasifikasi SVM untuk memprediksi kelulusan SNBP berdasarkan nilai akademik siswa Sekolah Menengah Atas?

Bagaimana cara menentukan *hyperplane* dan *penalize* terbaik untuk meningkatkan akurasi prediksi kelulusan yang tinggi?

01

02

Bagaimana pengaruh pengaturan kelas yang tidak seimbang dalam memprediksi tingkat kelulusan SNBP pada algoritma SVM?

03

04

Bagaimana tingkat akurasi dari model dalam memprediksi kelulusan SNBP?

Tujuan Penelitian

Mengimplementasikan model klasifikasi dan regresi SVM untuk memprediksi kelulusan SNBP pada Sekolah Menengah Atas.

01

02

Melakukan evaluasi pengaruh dari pengaturan kelas yang tidak seimbang dalam memprediksi tingkat kelulusan SNBP menggunakan SVM.

Melakukan analisis terhadap pemilihan *hyperplane* dan *penalize* untuk menemukan cara peningkatan akurasi prediksi kelulusan yang tinggi.

03

04

Melakukan analisis terhadap model Penalize-SVM terhadap tingkat akurasi dari model dan memprediksi kelulusan SNBP.

Batasan Penelitian

- 01 Penelitian ini dilakukan di Sekolah Menengah Atas Negeri Bandung
- 02 Dataset yang digunakan untuk penelitian nilai akademik siswa dari tahun 2015-2023 dari semester 1 – semester 5.
- 03 Nilai yang digunakan hanya nilai jurusan Ilmu Pengetahuan Alam (IPA) pada Sekolah Menengah Atas Negeri Bandung
- 04 Penelitian ini menggunakan algoritma SVM classification, Support Vector Regression (SVR) dan melakukan penalize pada kedua metode tersebut

Penelitian Terdahulu

Student Academic Performance Prediction Using Support Vector Machine

Pada penelitian ini dilakukan pengujian dengan metode regresi yang diujikan ke dalam enam metode yang berbeda, pada pengujiannya SVR dipilih sebagai metode terbaik dengan nilai Mean Squared Error (MSE) paling rendah dan standar deviasi tertinggi (Oloruntoba & Akinode, December 2017).

Penelitian Terdahulu

Student Academic Performance Prediction Using Support Vector Machine

S/N	Algoritma	MSE	Standar Deviasi
1	Linear Regression	-0.042915	0.021999
2	LASSO	-0.157838	0.059085
3	Elastic Net	-0.157838	0.059085
4	KNN	-0.119799	0.038263
5	Desicion Tree	-0.083867	0.051384
6	SVR	-0.170131	0.059279

Penelitian Terdahulu

Educational Data Mining Techniques Approach to Predict Student's Performance

Penelitian kedua yaitu terkait penggunaan metode linear regresi dan multi-class SVM classification dengan menggunakan data yang tersedia di Universitas Islam Indonesia's information system (UNISYS) (Khasanah & Harwati, 2019). Berbeda dengan penelitian yang pertama pada penelitian disini penggunaan SVM menunjukan akurasi yang lebih rendah dibanding dengan penggunaan linear regresi untuk memprediksi performa pendidikan, dilihat dari nilai Root Mean Square Error pada SVM nilai yang diberikan lebih besar dibandingkan dengan linear regresi.

Penelitian Terdahulu

Educational Data Mining Techniques Approach to Predict Student's Performance

Data	LR	SVM
No Cluster	0.084	2.182
4 Cluster	0.076	0.271

Kajian Pustaka

01 Data Mining

Data Mining (DM) adalah proses pengerukan, pengolahan dan pengumpulan informasi penting dari suatu data yang besar. Untuk mengidentifikasi, mengklasifikasikan informasi pada pengolahan data dalam data mining selalu menggunakan konsep statistika, matematika hingga memanfaatkan teknologi kecerdasan buatan dan machine learning.

Educational Data Mining (EDM) adalah sebuah pengembangan bidang DM dan pendidikan, dimana proses pengolahan informasi menggunakan konsep statistika, matematika hingga memanfaatkan teknologi kecerdasan buatan dan machine learning untuk mencari informasi penting didalam bidang pendidikan (Moharekar, Pol, & Pantakar, 2019).

Kajian Pustaka

02 Machine Learning

Machine Learning (ML) adalah cabang dari artificial intelligence (AI) dan ilmu komputer yang berfokus pada penggunaan data dan algoritma untuk meniru cara manusia belajar. ML akan meningkatkan kecerdasan komputer dengan meningkatkan kinerjanya pada tugas melalui pengalaman, pengalaman didapatkan dengan memproses data yang sudah tersedia untuk dipelajari dari keunikan data tersebut (Ottervanger, Mitra, et al, 2021).

03 Algoritma Support Vector Machine

Support Vector Machine (SVM) adalah metode ML supervised learning yang dapat digunakan sebagai classification dan regression. SVM menggunakan hyperplane untuk menganalisa data dan akan mengategorikan data menjadi dua keanggotaan kelas yang baru (Nisbet, Miner, & Yale, 2018).

Kajian Pustaka

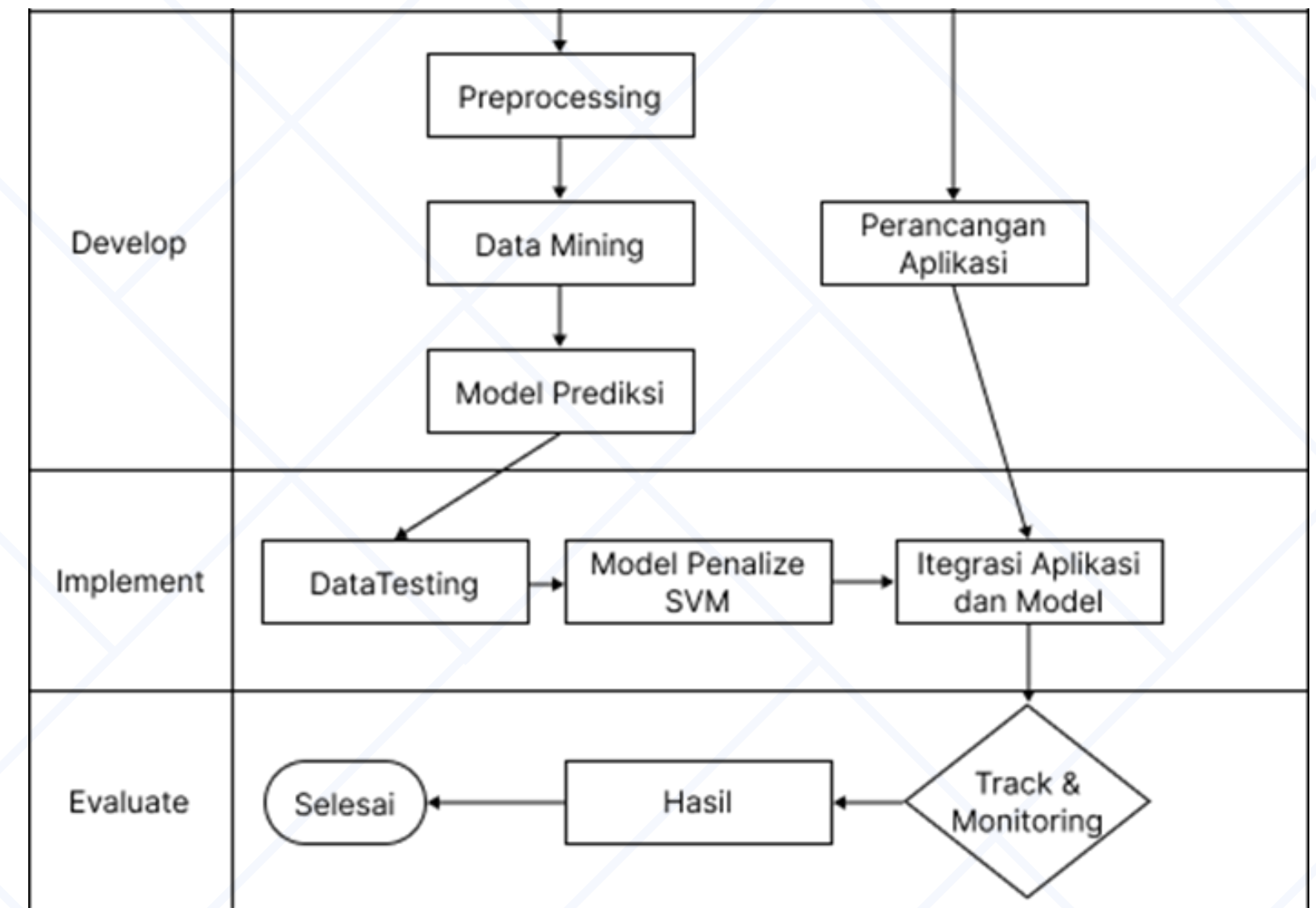
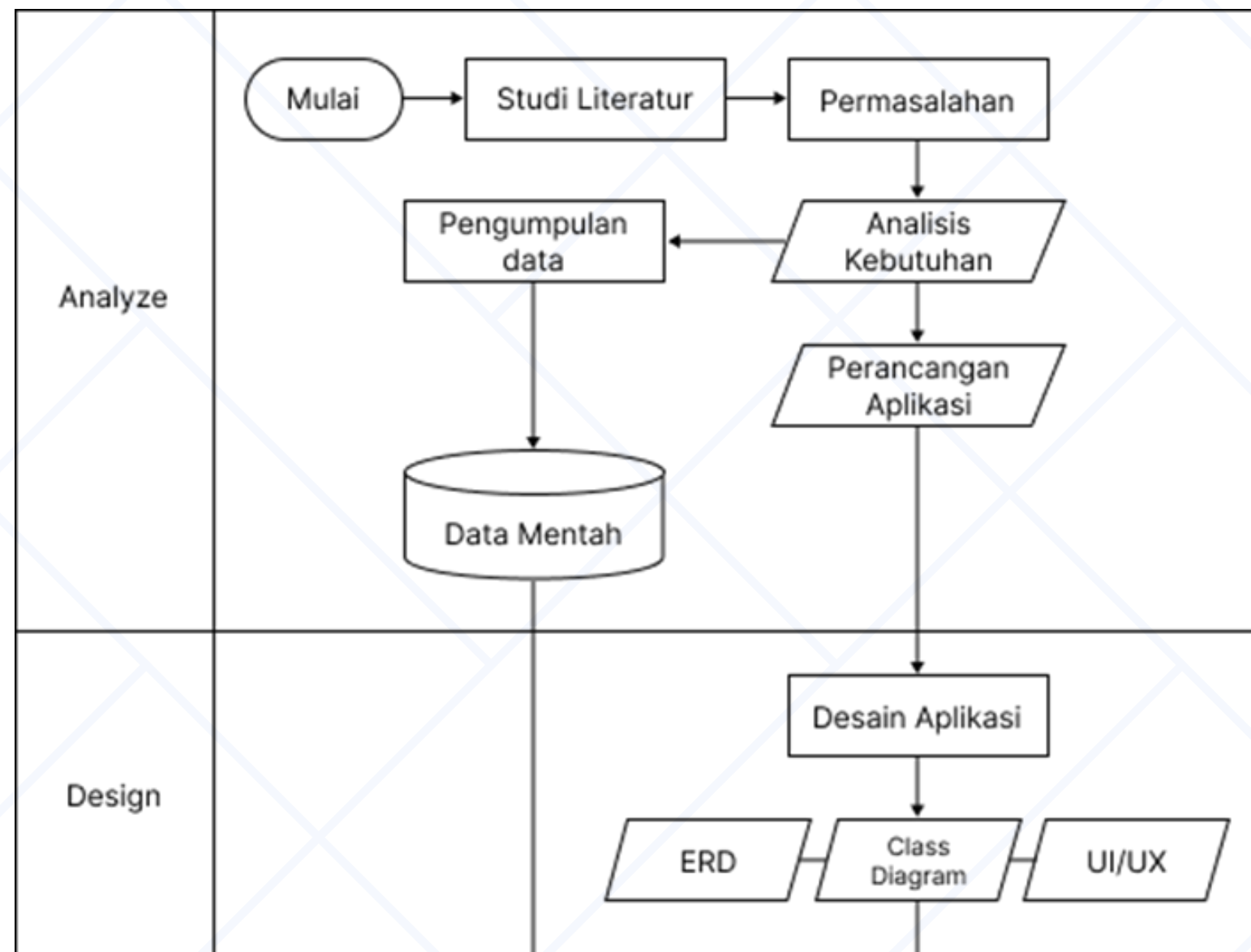
04 Penalize Support Vector Machine

Penalize SVM adalah sebuah variasi dari algoritma SVM yang menambahkan penalty pada fungsi objectifnya. Hal ini dilakukan untuk meningkatkan performa SVM pada beberapa kasus, seperti:

- Data yang tidak terpisahkan sempurna
- Overfitting

Desain Penelitian

Analyze Design Develop Implement Evaluate.



Desain Penelitian

Design

merancang desain penelitian perangkat lunak dan model ML yang akan dikembangkan dalam penelitian ini

Analyze

masalah yang ada pada penelitian diklarifikasi, tujuan dan sasaran ditetapkan, menetapkan batasan-batasan penelitian dan menetapkan pengetahuan yang akan digunakan pada penelitian ini

Develop

dataset yang telah didapatkan dari SMA Negeri dilanjutkan ke dalam preprocessing, pada tahap ini data akan melewati 3 proses, pertama pembersihan data, kedua pemilihan data dan ketiga pengubahan data

Desain Penelitian

Implement

data training yang telah didapat setelah melakukan preprocessing pada tahap develop menjadi sebuah model SVM yang terbaik berdasarkan tingkat accuracy score, precision, recall dan mean squared error

Develop

tahap dilanjutkan ke proses data mining, dalam penelitian ini akan dilakukan dua kali proses data mining. Pertama akan dilakukan dengan algoritma SVM tanpa melakukan penalize, dan kemudian SVM dengan penalize

Evaluate

setelah model penalize-SVM di integrasikan kedalam perangkat lunak dilakukan pengujian untuk memeriksa akurasi model dan kestabilan perangkat lunak ketika di integrasikan dengan model.

Alat & Bahan Penelitian

01 Perangkat Keras

- AMD Ryzen 5 5500U with Radeon Graphics 2.10 GHz
- AMD Radeon(TM) Graphics
- RAM 8 GB DDR4 3200MHz
- ROM 512 SSD

02 Studi Literatur

Jurnal, buku, buku elektronik, artikel ilmiah, dan lainnya akan digunakan. Selain itu terdapat alat dan teknologi yang digunakan akan membantu proses implementasi.

Metric Penelitian

Accuracy mengukur persentase prediksi yang benar dari seluruh prediksi. Nilai accuracy berkisar antara 0 dan 1, dengan 1 menunjukkan prediksi sempurna dan 0 menunjukkan prediksi tidak benar sama sekali.

Recall mengukur rasio dari data positif yang benar yang ditemukan dari seluruh data positif. Recall memberikan informasi tentang seberapa baik model machine learning menemukan semua data positif. Nilai recall berkisar antara 0 dan 1.

01

02

03

04

Precision mengukur rasio prediksi positif yang benar dari semua prediksi positif. Precision memberikan informasi tentang seberapa sering model machine learning membuat prediksi positif yang benar. Nilai precision berkisar antara 0 dan 1

Mean squared error mengukur rata-rata kuadrat dari perbedaan antara prediksi dan target. MSE digunakan untuk memperkirakan ketidakakuratan model machine learning dalam memprediksi target.



Universitas Pendidikan Indonesia

Thank You

Bagus Syamsu Rahmtullah

2003164