

PREDIKSI KELULUSAN MAHASISWA TEPAT WAKTU MENGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE

Rooy Thaniket¹⁾,Kusrini²⁾,Emha Taufik Luthf³⁾

Magister Teknik Informatika universitas Amikom Yogyakarta

rooythaniket@gmail.com,

kusrini@amikom.ac.id,

emhataufiqluthfi@amikom.ac.id

ABSTRAK- Kelulusan adalah keinginan setiap mahasiswa untuk dapat menyelesaikan studi Untuk Mencapai kelulusan Mahasiswa harus menyelesaikan tahapan seperti menempuh 8 semester 144 sks dan itu sebagai aturan dalam perguruan tinggi. Pada penelitian kali ini peneliti menggunakan data mahasiswa lulusan dari tahun 2015-2019 yang peneliti ambil disalah satu universitas metode yang digunakan peneliti adalah .data mining. Data mining adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam database. Algoritma yang peneliti gunakan dalam memprediksi kelulusan adalah algoritma *Support Vector Machine* (SVM) karna mampum mempredik kelulusan yang baik. Dalam prediksi kelulusan menghasilkan rata-rata akurasi sebesar Accuracy 95,00%

Kata kunci : kelulusan Data ming Algoritma *Support Vector Machine*

Pendahuluan

Lulus adalah keinginan setiap mahasiswa, lulus berarti sudah terselesaikanya beberapa syarat ketentuan didalam sebuah universitas yaitu aturan akademik. Bagi universitas sendiri kelulusan ini untuk menunjang universitas dalam data pelaporan pendidikan DIKTI untuk setiap semester dan membantu naiknya nilai akreditasi setiap program studi.[1] Program Strata 1 Reguler adalah program pendidikan akademik setelah pendidikan yang memiliki beban studi sekurang-kurangnya 144 SKS yang dijadwalkan untuk 8 semester dan paling lama 14 semester[2]data kelulusan mahasiswa adalah salah satu yang data penting untuk bisa mengambil salah satu keputusan dalam database. Data bisa dia analisis menggunakan data mining setiap kumpulan data satu gudangset memberikan pengetahuan dalam menggali informasi dalam satu perguruan tinggi.

Dalam hal ini ada Salah satu teknik untuk mengatasi data agar bermanfaat serta menjadi informasi adalah data mining. Data mining adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam database. Data mining adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan dan machine learning untuk mengekstraksi dan mengidentifikasi informasi dan yang bermanfaat dan pengetahuan yang terakit dari berbagai database (Kusrini & Emha 2009)

Ada beberapa penelitian terkait dengan kelulusan tepat waktu yaitu menurut, [3] penelitian ini dilakukan Teknik klasifikasi yang digunakan adalah pohon keputusan dengan penerapan algoritma C4.5. Input yang digunakan adalah atribut data siswa termasuk asal daerah, jenis sekolah, cara masuk universitas, pengalaman pesantren, akumulasi nilai rata-rata (IPK), dan nilai rata-rata setiap semester, dari semester 1 hingga semester 5. Data siswa tersebut adalah data sampel pelatihan yang digunakan dalam menyusun pohon keputusan. Berdasarkan pengujian yang menggunakan data siswa lulus dari tahun 2005 hingga 2008, akurasi kompatibilitas dalam sistem ini mencapai 82,79%, sehingga dapat digunakan untuk memprediksi. namun algoritma algoritma C4.5 termasuk dalam kategori excellent classification.

Menurut penelitian [4] mengatakan bahwa Data mahasiswa yang digunakan sebagai atribut input adalah jenis sekolah asal, asal daerah, pekerjaan orang tua, dan kelas. Atribut output yang digunakan untuk mengklasifikasikan data mahasiswa adalah status, yang terdiri dari “lulus tepat waktu” dan “lulus tidak tepat waktu”. Hasil analisis menunjukkan bahwa, kasus pada penelitian ini algoritma C4.5 dapat memprediksi dengan nilai akurasi hanya sebesar 82%.

Menurut [5] penerapan sistem cerdas untuk dapat mengklasifikasikan data prediksi kelulusan mahasiswa berdasarkan parameternya. Algoritme *Support Vector Machine* (SVM) mengklasifikasikan data menjadi 2 kelas menggunakan kernel Gaussian RBF dengan kombinasi nilai parameter $\lambda = 0,5$, konstanta $\gamma = 0,01$, dan ϵ (epsilon) = 0,001 itermax = 100, c = 1 dengan menggunakan data latih sebanyak 170 dataset. Penelitian ini menghasilkan rata-rata akurasi sebesar 80,55 %.

Menurut [6] bahwa metode yang digunakan adalah metode support vector hanya parameter kernel dalam satu parameter C yang memberikan pinalti pada titik data yang di klasifikasikan secara acak. data uji tahun 2011 sebanyak 310 set data dengan mahasiswa dengan mahasiswa non drop out 283 orang (KS) dan drop out 27 (KS) orang menghasilkan prediksi mahasiswa drop out 20 (HU) orang dan 290 orang (HU) non drop out. Diperoleh jumlah prediksi yang benar (sama) sebanyak 304 set data dan jumlah prediksi yang salah (tidak sama) sebanyak 6 set data. Untuk pengukuran tingkat akurasi diperoleh akurasi sebesar 98,06% dan nilai error sebesar 0.0193.

Dari semua keterangan penelitian tersebut maka peneliti akan mengadakan penelitian dengan judul prediksi kelulusan mahasiswa. Dalam penelitian tersebut peneliti berusaha agar mendapatkan nilai akurasi terbaik menggunakan menggunakan algoritma *Support Vector Machine*. Karena *Support Vector Machine* SVM merupakan algoritma yang bagus dalam memprediksi tingkat kelulusan mahasiswa

Kajian pustaka

Kelulusan

Kelulusan mahasiswa adalah hal yang penting diperhatikan, karena persentase jumlah kelulusan mempengaruhi penilaian pemerintah serta mempengaruhi status akreditasi program studi tingkat kelulusan dalam suatu perguruan tinggi itu sangat penting bagi setiap mahasiswa ada berepa faktor yang mempengaruhi kelulusan diantaranya biaya perkuliahan IPS semester dari IPS 1 sampai IPS 7, IPK, dan SKS persemester penghasilan orang tua. Adapun indikator yang yang sering digunakan untuk mentukan seorang mahasiswa bisa lulus adalah IPS banyak perguruan tinggi menggunakan hal ini dan variable yang akan digunakan adalah jenis kelamin IPS 1 – IPS 7 IPK dan SKS, SKS persemester dan penghasilan rata-rata orang tua.

Klasifikas data mining

Klasifikasi adalah teknik yang paling penting yang dipakai dalam data mining.[7] Pengkelompokan data mining adalah metode pengklasifikasi yang mempelajari data latih untuk diklasifikasin Adapun beberapa algoritma klasifikasi, antara lain *Bayesian Classification*, *K-Nearest Neighbor*, *Decision Tree Induction*, *Case-Based Reasoning*, *Genetic Algorithms*, dan *Support Vector Machine*. [8]

Support Vector Machine

Support Vector Machine (SVM) adalah suatu teknik yang relatif baru untuk melakukan prediksi, baik dalam kasus klasifikasi maupun regresi, yang terbaik saat ini. karena SVM baru mulai oprasikan pada tahun 1995. SVM berada dalam suatu kelas dengan ANN yang dapat menyelesaikan masalah dan Keduanya termasuk dalam kelas Supervised Learning [5] SVM paling disukai orang dalam bidang Data Mining dan Machine Learning, performancinya yang faktanya dalam memprediksi kelas suatu data baru akurat.

Cara naif membuat classifier nonlinear dari linear classifier adalah memetakan data dari ruang input X ke ruang fitur F menggunakan fungsi nonlinear $\phi: x \rightarrow f$. Dalam ruang F , fungsi adalah [9]

$$f(x) = w^T \phi(x) + b$$

Menggunakan notasi model linier $f(x, w)$ diberikan oleh

$$w = \sum_{i=1}^n \alpha_i x_i$$

$$f(x, w) = \sum_{i=1}^n \alpha_i x_i \varphi(x) + b$$

$$f(x) = \sum_{i=1}^n \alpha_i x_i^T x + b$$

f mengambil bentuk seperti berikut :

$$f(x) = \sum_{i=1}^n \alpha_i \varphi(x_i)^T \varphi(x) + b, \quad 0 \leq \alpha_i \leq C$$

$$K(x, x') = \varphi(x)^T \varphi(x')$$

Ada banyak fungsi kernel di SVM, jadi bagaimana memilih fungsi kernel yang baik Namun, untuk tujuan umum, ada beberapa fungsi kernel yang populer yaitu, Linear kernel: *Polynomial kernel*, RBF kernel, sigmoid kernel, tetapi pada penelitian kali ini peneliti menggunakan tipe kernel RBF. RBF Kernel sendiri menggunakan rumus sebagai berikut:

$$K(x_i, x_j) = \exp\left(-\gamma \|x_i - x_j\|^2\right)$$

dimana x' adalah „inti“ yang dipilih dari data pelatihan

Di sini C , γ dan r , dan d adalah parameter kernel. Itu sudah terkenal bahwa kinerja generalisasi SVM (akurasi estimasi) tergantung pada pengaturan meta-parameter, parameter C , γ dan r dan parameter kernel yang baik. Pilihan C , γ , dan mengendalikan kompleksitas model prediksi (regresi). Masalah pemilihan parameter optimal lebih rumit karena kompleksitas model SVM (dan karenanya kinerja generalisasi) tergantung pada ketiga parameter. Fungsi kernel digunakan untuk mengubah dimensi ruang input untuk melakukan klasifikasi.

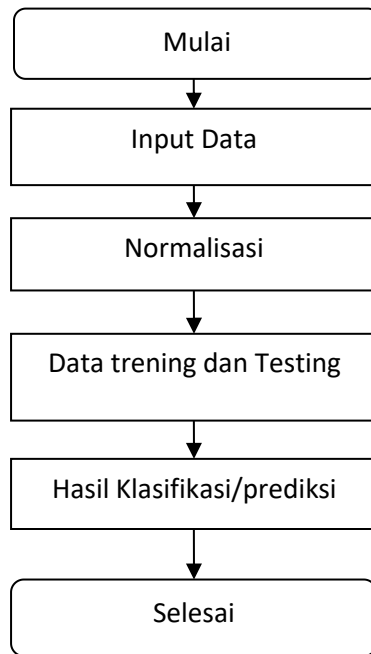
Metode Penelitian

Dalam penelitian ini peneliti menggunakan penelitian eksperimen dengan data yang diambil dari Universitas satya wiyata mandala berupa data lulus mahasiswa pada program studi Teknik informatika dengan jumlah data sampel 100 data sampel yang akan di bagi dua yaitu data testing dan latih. Gambar dibawah ini merupakan gambar alur penelitian agar terstruktur mulai dari penginputan data setelah itu

data tersebut dinormalisasi menggunakan max-min normalization. setelah data tersebut dinormalisasi akan training set dan di testing set sehingga mendapat klasifikasi

Alur Penelitian

Alur penelitian dapat dilihat pada gambar dibawah ini Alur Penelitian dimulai dari Input data dan berakhir, hasil jika prediksi berhasil dilakukan.



Gambar 1 Alur penelitian

Python

Python merupakan bahasa pemrograman interpretatif multiguna dengan filosofi perancangan yang berfokus pada tingkat keterbacaan kode. Bahasa pemrograman *Python* dipilih karena banyak digunakan untuk membuat berbagai macam program, seperti : program CLI, Program GUI, Aplikasi Mobile, Web, IoT. Game, dan lain-lain. *Python* dipilih karena bahasa pemrograman tersebut dirasa tepat dalam melakukan penelitian untuk memprediksi kelulusan mahasiswa.

Input Data

Untuk penginputan data peneliti mengambil data dari salah satu perguruan tinggi swasta di papua yang yaitu di universitas satya wiyata mandala pada program study Teknik informatika dengan data yang diambil adalah data kelulusan mahasiswa. Dan peneliti mengolah data tersebut. Atribut yang digunakan

adalah Jenis kelamin, Penghasilan orang tua, Index Prestasi Semester (IPS) 1 sampai ndex Prestasi Semester(IPS) 7, Index Prestasi komulatif (IPK) dan Satuan Kredit Semester (SKS) 1 sampai Satuan Kredit Semester (SKS) semester 7 setelah itu Label keterangan Lulus dan Tidak Lulus [11] jumlah data keseluruhan adalah 100 data Sampel dan dibagi dua data yaitu data latih 80% dan 20% data uji.

Normalisasi data

Normalisasi dipakai agar supaya nilai antar fitur yang berada pada interval yang sama. Min-Max Normalization adalah metode normalisasi yang sering digunakan untuk mengatasi permasalahan nilai antar fitur yang memiliki jarak terlampau jauh[12].

$$X^* = \left(\frac{X - \min(x)}{\max(x) - \min(x)} \right)$$

dimana

X^* adalah hasil dari normalisasi

x adalah data yang belum dinormalisasi

Min (X) adalah nilai minimum

Max (X) adalah nilai maksimum

Data Trenning dan data testing

Data yang gunangkan adalah data mahasiswa lulus dengan jumlah keseluruhan data yaitu 100 data dengan data latih 80% dan 20% data latih. Secara umum dalam proses klasifikasi memiliki dua proses yaitu :

1. Proses training : pada proses training digunakan training set yang telah diketahui label-labelnya untuk membangun model atau fungsi.
2. Proses testing : untuk mengetahui keakuratan model atau fungsi yang akan dibangun pada proses training, maka digunakan data yang disebut dengan testing set untuk memprediksi label-labelnya.

Prediksi

Metode yang digunakan untuk memprediksi kelulusan mahasiswa adalah algoritma yang saat ini populer yaitu *Support Vector Machine* (SVM). Variabel input yang digunakan dalam penelitian ini yaitu data mahasiswa dan data Index Prestasi Semester (IPS) 1 sampai ndex Prestasi Semester(IPS) 7, Index Prestasi komulatif (IPK) dan Satuan Kredit Semester (SKS) 1 sampai Satuan Kredit Semester (SKS) semester 7 ditambah dengan penghasilan orang tua . Jumlah kategori yang akan di prediksi memiliki dua kategori output adalah mahasiswa mahasiswa lulus dan tidak lulus. Tahapan algoritma *Support Vector Machine* sebagai berikut : Melakukan proses transformasi data sesuai dengan format *Support Vector Machine*. Menentukan fungsi kernel yang akan digunakan Penelitian ini menggunakan tipe kernel yang

digunakan adalah *Radial basis function* (RBF). Menentukan nilai-nilai parameter kernel dan parameter cost (C) untuk melakukan optimasi. Memilih parameter terbaik untuk optimasi data training dan untuk prediksi data testing dan Menghitung ketepatan prediksi.

Hasil Pembahasan

penulis akan coba melakukan implementasi kedalam bentuk pemrograman dengan menggunakan bahasa pemrograman *Python*. Dataset disimpan kedalam *Google Drive* dan Penulis menggunakan mount drive untuk memberi kode agar dapat mengelola dataset yang sudah disimpan dalam *Google Drive*.

```
[1] from google.colab import drive
    drive.mount('/content/drive')

Mounted at /content/drive
```

Gambar 2 Drive Mount untuk Google Drive

Selanjutnya, penulis melakukan import library yang dibutuhkan. Ditahap ini penulis melakukan import library seperti sklearn, pandas, pydotplus dan Ipyhton.

```
# Packages for analysis
import pandas as pd
import numpy as np
from sklearn import svm

# Packages for visuals
import matplotlib.pyplot as plt
import seaborn as sns; sns.set(font_scale=1.2)

# Allows charts to appear in the notebook
%matplotlib inline

# Pickle package
import pickle

[ ] # Read in muffin and cupcake ingredient data
data_mhs = pd.read_csv('/content/drive/My Drive/TESIS/dataset/data_training.csv')
data_mhs = data_mhs.drop(['No', 'Nama MHS', 'IPK', 'NIM'], axis=1)
data_mhs
```

Gambar 3 Import Library

Selanjutnya peneliti melakukan import dataset. Dataset ini adalah data yang tersimpan diberi nama “Data training svm .csv” dengan atribut yang sudah ditentukan seperti jenis kelamin penghasilan orangtua, Index Prestasi Semester (IPS) 1 sampai ndex Prestasi Semester(IPS) 7, Index Prestasi kumulatif (IPK)

dan Satuan Kredit Semester (SKS) 1 sampai Satuan Kredit Semester (SKS) semester dan Label seperti terlihat pada gambar 4.

	Jenis Kelamin	Penghasilan Orang tua	IPS_I	IPS_II	IPS_III	IPS_IV	IPS_V	IPS_VI	IPS_VII	SKS S1	SKS 2	sks 3	SKS 4	SKS 5	SKS 6	SKS 7	Label
0	2	1500000.0	3.05	3.07	3.14	3.14	3.16	3.14	2.92	23.0	20	20	21	20	20	13	1
1	2	1500000.0	3.10	3.14	3.19	3.32	3.42	3.52	3.15	23.0	20	20	21	20	20	13	1
2	1	1500000.0	3.03	2.91	2.98	3.13	1.20	1.20	2.92	23.0	20	20	21	20	6	13	1
3	1	500000.0	3.10	3.12	3.13	3.17	3.17	3.12	3.38	23.0	20	20	21	0	17	13	2
4	2	1500000.0	3.10	3.21	3.10	3.14	3.12	2.98	3.38	23.0	20	20	21	20	3	13	1
...
95	2	1000000.0	3.26	3.05	2.00	3.06	0.00	2.00	3.00	23.0	20	20	15	0	15	13	2
96	1	1000000.0	3.31	3.14	0.00	2.00	2.00	2.94	3.00	23.0	20	0	15	14	15	13	2
97	1	1000000.0	3.13	2.98	2.94	3.02	3.05	3.26	2.85	23.0	20	20	21	20	20	13	1
98	1	2500000.0	3.25	3.23	3.21	3.26	3.26	3.23	3.23	23.0	20	20	21	20	20	13	1
99	2	1000000.0	3.33	3.05	3.08	3.10	3.24	3.14	3.00	23.0	20	20	21	20	20	13	1

100 rows x 17 columns

Gambar 4 Pembagian data

Penelitian ini menggunakan SVM dengan memanfaatkan library *sklearn* pada *Python* dengan menggunakan parameter tipe kernel *Radial basis function* (RBF). Parameter nilai *gamma* otomatis, dan nilai *Cost* (C) = 2 dengan menggunakan perintah seperti terlihat pada gambar 5.

```

from sklearn import svm

model = svm.SVC(kernel='rbf', gamma='auto', C=2)
model.fit(X_train, y_train)
predictions = model.predict(X_test)
    
```

Gambar 5 Parameter SVM

setelah itu mencari nilai *confusion matrix* hasil pencarian nilai *confusion matrix* bisa dilihat pada gambar dibawah ini.



Gambar. 6 Confusion Matrix

Dari hasil pengujian menggunakan tools *Python* untuk prediksi kelulusan mahasiswa pada universitas satya wiyata mandala pada program study Teknik informatika dengan dataset 100 sampel dengan dibagi 80 data latih dan 20 data uji maka terdapat *accuracy* yang lebih baik yaitu 95.00%.

Kesimpulan

Dari hasil ujicoba diatas penulis menyimpulkan bahwa

1. Algoritma *Support Vector Machine* dapat memprediksi tingkat kelulusan Mahasiswa pada unviversitas satya wiyata mandala
2. Akurasi yang dapat dimiliki oleh algoritma *Support Vector Machine* dengan tiper kernel RBF untuk memprediksi kelulusa mahasiswa dengan menggunakan phyton sebesar 95,00%

5.1 Saran

Dari data diatas penelitian ini masih kurang memuaskan untuk penelitian selanjutnya diharapkan menggunakan data yang besar supaya bisa dicoba dengan algoritma-algoritma lain agar mendapat akurasi lebih baik.

Refrensi

- [1] L. Ilmu and P. Indonesia, “Pedoman akreditasi majalah ilmiah,” 2011.
- [2] D. Salmu, S. and A. Solichin, “Prediksi Tingkat Kelulusan Mahasiswa Tepat Waktu Menggunakan Naïve Bayes : Studi Kasus UIN Syarif Hidayatullah Jakarta Prediction of Timeliness Graduation of Students Using Naïve Bayes : A Case Study at Islamic State University Syarif Hidayatullah Jakarta,” *Pros. Semin. Nas. Multidisiplin Ilmu*, no. April, pp. 701–709, 2017.
- [3] SUSI MASHLAHAH, *Prediksi Kelulusan Mahasiswa Menggunakan Metode Decision Tree Dengan Penerapan Algoritma C4.5*. 2013.
- [4] I. P. Astuti, “Prediksi Ketepatan Waktu Kelulusan Dengan Algoritma Data Mining C4.5,” *Fountain Informatics J.*, vol. 2, no. 2, p. 5, 2017.
- [5] A. Pratama, R. C. Wihandika, and D. E. Ratnawati, “Implementasi Algoritme Support Vector Machine (SVM) untuk Prediksi Ketepatan Waktu Kelulusan Mahasiswa,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. March, pp. 1704–1708, 2018.
- [6] S. Nurhayati, Kusri, and E. T. Luthfi, “Prediksi Mahasiswa Drop Out Menggunakan Metode Support Vector,” *Ilm. Sisfotenika*, vol. 5, no. x, pp. 82–93, 2015.
- [7] S. Džeroski, *Data Mining*. 2008.
- [8] . رساله ی دکتری دانشگاه شهید بهشتی, No Title, رسوله ی .
- [9] A. H. Haghiabi, H. M. Azamathulla, and A. Parsaie, “Prediction of head loss on cascade weir using ANN and SVM,” *ISH J. Hydraul. Eng.*, vol. 5010, no. October, pp. 1–9, 2016.
- [10] J. Sumpena and N. K. H, “Analisis prediksi kelulusan siswa pkbm paket c dengan metoda algoritma naïve beyes,” vol. 13, no. 2, pp. 127–133, 2019.
- [11] W. W. Ariestya, Y. E. Praptiningsih, and W. Supriatin, “Decision Tree Learning Untuk Penentuan Jalur Kelulusan Mahasiswa,” *J. Ilm. FIFO*, vol. 8, no. 1, p. 97, 2016.
- [12] R. A. Wijayanti, M. T. Furqon, and S. Adinugroho, “Penerapan Algoritme Support Vector Machine Terhadap Klasifikasi Tingkat Risiko Pasien Gagal Ginjal,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput. Univ. Brawijaya*, vol. 2, no. 10, pp. 3500–3507, 2018.