

**IMPLEMENTASI ALGORITMA *RANDOM FOREST* UNTUK
IDENTIFIKASI DINI SISWA YANG MENGALAMI
KENDALA AKADEMIK**

SKRIPSI

diajukan untuk memenuhi sebagian syarat untuk memperoleh gelar Sarjana
Pendidikan Teknik Elektro Konsentrasi Elektronika Industri



oleh

Iva Rachmawati

E.0451.1602314

**PROGRAM STUDI PENDIDIKAN TEKNIK ELEKTRO
DEPARTEMEN PENDIDIKAN TEKNIK ELEKTRO
FAKULTAS PENDIDIKAN TEKNOLOGI DAN KEJURUAN
UNIVERSITAS PENDIDIKAN INDONESIA**

2020

**IMPLEMENTASI ALGORITMA *RANDOM FOREST* UNTUK
IDENTIFIKASI DINI SISWA YANG MENGALAMI
KENDALA AKADEMIK**

Oleh
Iva Rachmawati

Sebuah skripsi yang diajukan untuk memenuhi salah satu syarat memperoleh gelar
Sarjana Pendidikan pada Departemen Pendidikan Teknik Elektro
Fakultas Pendidikan Teknologi dan Kejuruan

© Iva Rachmawati 2020
Universitas Pendidikan Indonesia
Juli 2020

Hak Cipta dilindungi undang-undang.
Skripsi ini tidak boleh diperbanyak seluruhnya atau sebagian,
dengan dicetak ulang, difoto kopi, atau cara lainnya tanpa ijin dari penulis.

LEMBAR PENGESAHAN

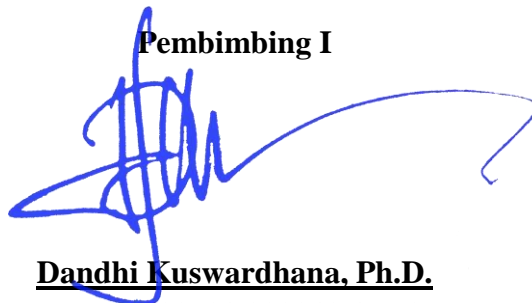
IVA RACHMAWATI

E.0451.1602314

Implementasi Algoritma *Random Forest* untuk Identifikasi Dini Siswa yang Mengalami Kendala Akademik

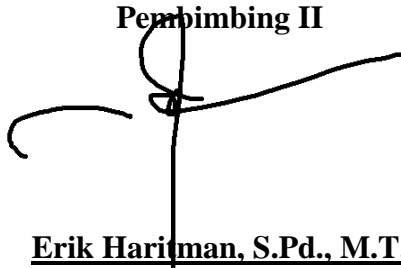
disetujui dan disahkan oleh:

Pembimbing I



Dandhi Kuswardhana, Ph.D.
NIP.19800623 200812 1 002

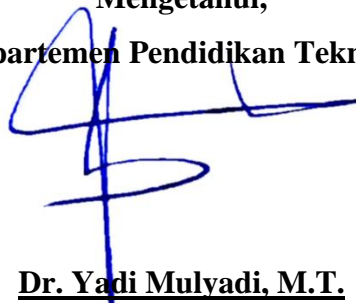
Pembimbing II



Erik Haritman, S.Pd., M.T.
NIP. 19760527 200112 1 002

Mengetahui,

Ketua Departemen Pendidikan Teknik Elektro



Dr. Yadi Mulyadi, M.T.
NIP. 19630727 199302 1 001

UCAPAN TERIMA KASIH

Puji syukur penulis panjatkan kehadirat Allah SWT, yang telah memberikan rahmat dan karunia-Nya sehingga penulis dapat melaksanakan dan menyelesaikan skripsi berjudul “Implementasi Algoritma *Random Forest* untuk Identifikasi Dini Siswa yang Mengalami Kendala Akademik”. Selama penyusunan skripsi, penulis mendapat bantuan dari berbagai pihak. Maka dari itu, penulis mengucapkan terima kasih kepada:

1. Orang tua dan nenek penulis, Bapak Dodang, Ibu Siti Romlah, dan Ma Uka, yang senantiasa memberikan dukungan, semangat, do'a dan materi selama pengerjaan skripsi.
2. Bapak Dr. H. Yadi Mulyadi, M.T., selaku Ketua Departemen Pendidikan Teknik Elektro.
3. Bapak Tasma Sucita, S.T.,M.T., selaku Ketua Prodi S1 Pendidikan Teknik Elektro.
4. Bapak Dr. H. Bambang Trisno, MSIE., selaku Dosen Wali.
5. Bapak Dandhi Kuswardhana, Ph.D., selaku Dosen Pembimbing I yang selalu memberikan bimbingan dan motivasi dalam menyelesaikan skripsi.
6. Bapak Erik Haritman, S.Pd.,M.T., selaku Dosen Pembimbing II yang selalu memberikan bimbingan dan motivasi dalam menyelesaikan skripsi.
7. Seluruh dosen dan staf Departemen Pendidikan Teknik Elektro.
8. Sahabat SMK penulis, Annissa Nur Fitriani, R. Farida Nur Ardila, Syauqi Asyhabira, dan Latifah Nur Permatasari yang selalu ada, menerima, mendengarkan, memberi dukungan, dan percaya kepada penulis.
9. Triska ‘bubun’ Junita, yang selalu kebersamai penulis dari awal dalam setiap langkah yang ditempuh semasa perkuliahan sampai saat ini.
10. Akhwat PTE-A 2016, Slamet Tita, Eka Putri, Ghina Nadifah, Ai Asiah, Khaesa Ananda, dan Imas Devina, yang selalu ada setiap saat untuk penulis.
11. Rekan-rekan seperjuangan PTE-A 2016 dan ELIND 2016, terutama Hendra Fauzi, M. Rashif Farhan, Azhar Adhiyatso, dan Teh Olsa yang banyak membantu, memberi motivasi, dan kebersamai penulis dalam pengerjaan skripsi ini.

12. dan pihak lain yang tidak dapat penulis sebutkan satu persatu.

Semoga Allah SWT senantiasa membalas seluruh kebaikan.

Bandung, Juli 2020

Iva Rachmawati

IMPLEMENTASI ALGORITMA *RANDOM FOREST* UNTUK IDENTIFIKASI DINI SISWA YANG MENGALAMI KENDALA AKADEMIK

oleh

Iva Rachmawati

E.0451.1602314

ABSTRAK

Penelitian ini bertujuan untuk mengidentifikasi dini siswa yang mengalami kendala akademik dengan mengimplementasikan Algoritma *Random Forest*. Algoritma klasifikasi ini digunakan agar hasil identifikasi lebih dini dan akurat daripada pendekatan secara tradisional. Desain penelitian yang digunakan adalah metode kuantitatif dan studi literatur, dengan data sekunder yang diperoleh dari *UCI Machine Learning Repository*. Tahapan untuk mengimplementasikan Algoritma *Random Forest* yaitu pra-pemrosesan data, kemudian proses *training* dan *testing* model. Atribut yang berpengaruh secara signifikan terhadap hasil identifikasi siswa adalah G2 dan G1 (penilaian 2 dan 1), atribut lainnya adalah *absences* (jumlah ketidakhadiran) dan *failures* (kegagalan pada kelas sebelumnya). Implementasi Algoritma *Random Forest* untuk identifikasi dini siswa yang mengalami kendala akademik ini memiliki akurasi 88,89% dan *AUC Score* 0,971763 yang dikategorikan sangat baik. Berdasarkan hasil implementasi, diperoleh identifikasi 36 siswa yang mengalami kendala akademik dan 63 siswa yang tidak mengalami kendala akademik, dari data sebenarnya yaitu 33 siswa yang mengalami kendala akademik dan 66 siswa yang tidak mengalami kendala akademik.

Kata Kunci: *Random Forest*, Identifikasi, Kendala Akademik

IMPLEMENTATION OF RANDOM FOREST ALGORITHM FOR STUDENT AT RISK EARLY IDENTIFICATION

by

Iva Rachmawati

E.0451.1602314

ABSTRACT

This research aims to early identify student at risk by implementing Random Forest Algorithm. This classification algorithm is used for early and more accurate results than traditional method. Research design used is quantitative method and literature study, with secondary data from *UCI Machine Learning Repository*. Steps to implement Random Forest algorithm are pre-processing data, then training and testing the model. The attributes that significantly influences identification results are G2 and G1 (grade in periode 2 and 1), other attributes are absences and failures. Implementation of Random Forest algorithm for student at risk early identification has 88,89% of accuracy and 0,971763 of AUC Score which is categorized as an excellent classification. Based on the implementation results, 36 student at risk and 63 student not at risk has been identified from the actual data of 33 student at risk and 66 student not at risk.

Keywords: Random Forest, identification, student at risk.

DAFTAR ISI

LEMBAR PENGESAHAN	i
UCAPAN TERIMA KASIH.....	ii
ABSTRAK.....	iv
ABSTRACT.....	v
DAFTAR ISI.....	vi
DAFTAR TABEL.....	viii
DAFTAR GAMBAR	ix
DAFTAR LAMPIRAN.....	x
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah	2
1.3 Tujuan Penelitian.....	3
1.4 Batasan Masalah	3
1.5 Manfaat Penelitian.....	3
1.6 Sistematika Penulisan	3
BAB II KAJIAN PUSTAKA	5
2.1 Kendala Akademik	5
2.2 Identifikasi Dini Siswa Terkendala Akademik.....	7
2.3 <i>Machine Learning</i>	7
2.3.1 <i>Algoritma Random Forest</i>	8
2.4 <i>Machine Learning</i> dalam Pendidikan.....	9
BAB III METODE PENELITIAN	11
3.1 Desain Penelitian	11
3.2 Objek Penelitian	11
3.3 Data Penelitian.....	11
3.4 Prosedur Penelitian	15
3.5 Analisis Data	16

BAB IV TEMUAN DAN PEMBAHASAN	18
4.1 Tahapan Implementasi Algoritma Random Forest.....	18
4.1.1 Pra-pemrosesan Data	18
4.1.2 Proses <i>Training</i> dan <i>Testing</i>	19
4.2 <i>Feature Importance</i>	22
4.3 Evaluasi Model Algoritma <i>Random Forest</i>	24
4.3.1 <i>Confusion Matrix</i>	25
4.3.2 Kurva ROC dan <i>AUC Score</i>	28
4.4 Hasil Identifikasi Siswa yang Mengalami Kendala Akademik	30
 BAB V SIMPULAN, IMPLIKASI, DAN REKOMENDASI	 34
5.1 Simpulan.....	34
5.2 Implikasi	34
5.3 Rekomendasi	35
 DAFTAR PUSTAKA	 36
 Lampiran 1 Dataset Penelitian	 40
Lampiran 2 Model Algoritma <i>Random Forest</i>	54

DAFTAR TABEL

Tabel 3.1 Atribut Data Penelitian.....	12
Tabel 4.1 Hasil Training dan Testing.....	21
Tabel 4.2 Hasil Training dan Testing menggunakan GridSearchCV	22
Tabel 4.3 Persentase Feature Importance.....	23
Tabel 4.4 Pengkategorian AUC Score	30
Tabel 4.5 Hasil Identifikasi dan Hasil Sebenarnya Siswa KA dan TKA.....	30

DAFTAR GAMBAR

Gambar 3.1 Diagram Alir Prosedur Penelitian	15
Gambar 4.1 Perbandingan Jumlah Siswa KA dan TKA	18
Gambar 4.2 Sebagian Atribut Fitur	18
Gambar 4.3 Atribut Fitur Pasca-pemrosesan	19
Gambar 4.4 Grafik Pengaruh n_{est} terhadap F1-Score	21
Gambar 4.5 Feature Importance	23
Gambar 4.6 Confusion Matrix	25
Gambar 4.7 Kurva ROC	29

DAFTAR LAMPIRAN

Lampiran 1 Dataset Penelitian	40
Lampiran 2 Model Algoritma <i>Random Forest</i>	54

DAFTAR PUSTAKA

- Ahadi, A., Lister, R., Haapala, H., & Vihavainen, A. (2015). Exploring machine learning methods to automatically identify students in need of assistance. *ICER 2015 - Proceedings of the 2015 ACM Conference on International Computing Education Research*, 121–130. <https://doi.org/10.1145/2787622.2787717>
- Ahmad, A. (2017). *Mengenal Artificial Intelligence, Machine Learning, Neural Network, dan Deep Learning*. June.
- Ariyani, L. (2016). Kajian Penerapan Model C45 , Support Vector Machine (SVM), dan Neural Network Dalam Prediksi Kenaikan Kelas. *Faktor Exacta*, 9(1), 72–86.
- Aulck, L., Velagapudi, N., Blumenstock, J., & West, J. (2016). *Predicting Student Dropout in Higher Education*. 16–20. <http://arxiv.org/abs/1606.06364>
- Babbie, E. (2010). *The Practice of Social Research* (12th ed.). Cengage Learning.
- Beaulac, C., & Rosenthal, J. S. (2019). Predicting University Students' Academic Success and Major Using Random Forests. *Research in Higher Education*, 60(7), 1048–1064. <https://doi.org/10.1007/s11162-019-09546-y>
- Breiman, L. (2011). Random forests. *Machine Learning*, 45, 5–32. <https://doi.org/10.1201/9780367816377-11>
- Brownlee, J. (2020). *How to Fix k-Fold Cross-Validation for Imbalanced Classification*. <https://machinelearningmastery.com/cross-validation-for-imbalanced-classification/>
- Cortez, P., & Silva, A. (2008). Using data mining to predict secondary school student performance. *15th European Concurrent Engineering Conference 2008, ECEC 2008 - 5th Future Business Technology Conference, FUBUTEC 2008, 2003(2000)*, 5–12.
- Davis, J., & Goadrich, M. (2016). *The Relationship Between Precision-Recall and ROC Curves*. 233–240. <http://arxiv.org/abs/1609.07195>

- Doll, J. J., Eslami, Z., & Walters, L. (2013). Understanding why students drop out of high school, According to their own reports: Are they pushed or pulled, Or do they fall out? a comparative analysis of seven nationally representative studies. *SAGE Open*, 3(4). <https://doi.org/10.1177/2158244013503834>
- Galdi, P., & Tagliaferri, R. (2018). Data mining: Accuracy and error measures for classification and prediction. *Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics*, 1–3(January), 431–436. <https://doi.org/10.1016/B978-0-12-809633-8.20474-3>
- Gholamy, A., Kreinovich, V., & Kosheleva, O. (2018). *Why 70 / 30 or 80 / 20 Relation Between Training and Testing Sets : A Pedagogical Explanation*. 1–6.
- Jacobs, J., & Rudis, B. (2014). *Data Driven Security: Analysis, Visualization, and Dashboards*. John Wiley & Sons, Inc.
- Jordan, M. I., & Mitchell, T. M. (2015). *Machine learning: Trends, perspectives, and prospects*. 349(6245). <https://doi.org/10.1126/science.aaa8415>
- Kleinbaum, D. G., & Klein, M. (2010). *Logistic Regression: A Self Learning Text*. Springer. <https://doi.org/https://doi.org/10.1007/978-1-4419-1742-3>
- Kučak, D., Juričić, V., & Đambić, G. (2018). Machine learning in education - A survey of current research trends. *Annals of DAAAM and Proceedings of the International DAAAM Symposium*, 29(1), 0406–0410. <https://doi.org/10.2507/29th.daaam.proceedings.059>
- Lakkaraju, H., Aguiar, E., Shan, C., Miller, D., Bhanpuri, N., Ghani, R., & Addison, K. L. (2015). A machine learning framework to identify students at risk of adverse academic outcomes. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2015-Augus*, 1909–1918. <https://doi.org/10.1145/2783258.2788620>
- Mduma, N., Kalegele, K., & Machuve, D. (2019). A survey of machine learning approaches and techniques for student dropout prediction. *Data Science Journal*, 18(1), 1–10. <https://doi.org/10.5334/dsj-2019-014>

- Muijs, D. (2004). *Doing Quantitative Research in Education*. Sage Publications.
<https://doi.org/10.7748/ns.29.31.44.e8681>
- Nur, M. R. F., & Oktora, S. I. (2020). Analisis Kurva Roc Pada Model Logit Dalam Pemodelan Determinan Lansia Bekerja Di Kawasan Timur Indonesia. *Indonesian Journal of Statistics and Its Applications*, 4(1), 116–135.
<https://doi.org/10.29244/ijisa.v4i1.524>
- Olufunke, A. (2014). *Attendance Dilemma and its Effects on the Academic Performance of Secondary Schools ' Students in Osun State ,. 1(4)*, 13–20.
- Power, C., & Sophister, J. (2014). *the Student Economic Review* vol. Xxviii *Education Development: Importance, Challenges and Solutions*. 149–157.
https://www.tcd.ie/Economics/assets/pdf/SER/2014/Colin_Power.pdf
- Putri, A. (2016). *STUDI KOMPARASI IMPLEMENTASI SOFT TOTAL QUALITY MANAGEMENT PADA SEKOLAH MENENGAH ATAS DAN SEKOLAH MENENGAH KEJURUAN NEGERI DI KOTA BANDUNG Universitas Pendidikan Indonesia | repository.upi.edu | perpustakaan.upi.edu*.
- Sasaki, Y. (2015). *The truth of the F-measure The truth of the F-measure. January 2007*, 1–6.
- scikit-learn user guide*. (2020).
- Silalahi, U. (2015). Metode Penelitian Sosial Kuantitatif. *Journal of Visual Languages & Computing*, 11(3), 287–301.
- Smith, L. M. (2015). *ENGAGING AT-RISK VOCATIONAL EDUCATION STUDENTS WITH MOBILE DEVICES. October*.
- Smith, T. (2016). *Building a Student Intervention System*.
github.com/Tsmith5151/Student-intervention/
- Sugianto, C. A., & Pujiyanita, N. S. (2019). OPTIMALISASI ALGORITMA C4.5 MENGGUNAKAN ALGORITMA GENETIKA UNTUK PREDIKSI KELULUSAN SISWA SMKN 2 CIMAHI. *Infotekmesin*, 10(02), 46–51.
- Suryanto, A., Alfarobi, I., & Tutupoly, T. A. (2018). Komparasi Algoritma C4.5,

Naive Bayes Dan Random Forest Untuk Klasifikasi Data Kelulusan Mahasiswa Jakarta. *Mitra Dan Teknologi Pendidikan*, iv nomor 1, 2–14. <https://www.publikasiilmiah.com/jurnal-mitra-dan-teknologi-pendidikan-volume-iv-nomer-1-februari-2018/>

Suwardika, G. S., & Suniantara, I. K. P. (2019). Analisis Random Forest Pada Klasifikasi Cart Ketidaktepatan Waktu Kelulusan Mahasiswa Universitas Terbuka. *BAREKENG: Jurnal Ilmu Matematika Dan Terapan*, 13(3), 177–184. <https://doi.org/10.30598/barekengvol13iss3pp177-184ar910>

Taş, A., Selvitopu, A., Bora, V., & Demirkaya, Y. (2013). Reasons for dropout for vocational high school students. *Kuram ve Uygulamada Egitim Bilimleri*, 13(3), 1561–1565. <https://doi.org/10.12738/estp.2013.3.1398>

Tumini, & Damayanti, L. (2018). Aplikasi Data Mining Untuk Memprediksi Kelulusan Siswa Dengan Metode Naive Bayes Studi Kasus SMP Negeri 11 Kotabumu Utara. *Informatika SIMANTIK*, 3(2), 23–30.

Udacity. (2016). *Student Intervention*. Udacity. https://github.com/udacity/machine-learning/tree/master/projects/student_intervention

UNESCO Institute for Statistics. (2020). *No Title*. <http://data.uis.unesco.org/index.aspx?queryid=157>

Wahid, F., Ghazali, R., Fayaz, M., Shah, A. S., Tun, U., & Onn, H. (2017). *A Simple and Easy Approach for Home Appliances Energy Consumption Prediction in Residential Buildings Using Machine Learning Techniques*. 7(3), 108–119.

Zaenuri. (2019). PENERAPAN DATA MINING UNTUK MEMPREDIKSI KELULUSAN SISWA DENGAN METODE DECISION TREE. *Artikel Skripsi Universitas Nusantara PGRI Kediri*.