MODEL *EDUCATIONAL DATA MINING* BERBASIS *GRADIENT BOOSTED TREES* UNTUK PREDIKSI PERFORMA AKADEMIK MAHASISWA

Muhammad Arifin
30000320520026

PROGRAM STUDI DOKTOR SISTEM INFORMASI
SEKOLAH PASCASARJANA
UNIVERSITAS DIPONEGORO
SEMARANG
2024

HALAMAN PENGESAHAN

MODEL *EDUCATIONAL DATA MINING* BERBASIS *GRADIENT BOOSTED TREES* UNTUK PREDIKSI PERFORMA AKADEMIK MAHASISWA

Oleh

Muhammad Arifin

NIM 30000320520026

Telah diuji dan dinyatakan lulus ujian pada tanggal 20 bulan Maret tahun 2024 oleh Tim Penguji Program Studi Doktor Sistem Informasi Sekolah Pascasarjana Universitas Diponegoro

Promotor	**Co-Promotor**
Prof. Dr. Widowati, S.Si., M.Si.	Farikhin, S.Si., M.Si., Ph.D
NIP. 196902141994032002	NIP. 197312202000121001
Tanggal: 20 Maret 2024	Tanggal: 20 Maret 2024

Mengetahui,

Dekan Sekolah Pascasarjana Universitas Diponegoro	Ketua Program Studi Doktor Sistem Informasi Sekolah Pascasarjana Universitas Diponegoro
Dr. R.B Sularto, SH., M.Hum.	Prof. Dr. Rahmat Gernowo, M.Si.
NIP. 196701011991031005	NIP. 196511231994031003

HALAMAN PERSETUJUAN

MODEL *EDUCATIONAL DATA MINING* BERBASIS *GRADIENT BOOSTED TREES* UNTUK PREDIKSI PERFORMA AKADEMIK MAHASISWA

Oleh

Muhammad Arifin

NIM 30000320520026

Telah disetujui oleh:

Pimpinan Sidang

Dr. R.B Sularto, SH., M.Hum. .....

Sekretaris Sidang

Prof. Dr. Rahmat Gernowo, M.Si. .....

Anggota Tim Penguji

Dr. Maman Somantri, S.T., M.T. .....

Prof. Dr. Adian Fatchur Rochim, ST, MT. .....

Prof. Dr. Ridwan Sanjaya, S.E., S.Kom., MS.IEC .....

Prof. Dr. Widowati, S.Si., M.Si. .....

Farikhin, M.Si., Ph.D .....

PERNYATAAN PERSETUJUAN PUBLIKASI DISERTASI UNTUK KEPENTINGAN AKADEMIS

Sebagai sivitas akademik Universitas Diponegoro, saya yang bertanda tangan di bawah ini:

Nama	: Muhammad Arifin
NIM	: 30000320520026
Program Studi	: Doktor Sistem Informasi
Program	: Sekolah Pascasarjana
Jenis Karya	: Disertasi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Diponegoro Hak Bebas Royalti Non Eksklusif atas karya ilmiah saya yang berjudul:

Model *Educational Data Mining* Berbasis *Gradient Boosted Trees* Untuk Prediksi Performa Akademik Mahasiswa

beserta perangkat yang ada. Dengan Hak bebas Royalti Non Eksklusif ini Program Studi Doktor Sistem Informasi Sekolah Pascasarjana Universitas Diponegoro berhak menyimpan, mengalih media/formatkan, mengelola dalam bentuk pangkalan data (*database*) merawat, dan mempublikasikan Disertasi saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik hak cipta.

Dibuat di: Semarang
Pada tanggal: Maret 2024
Yang menyatakan

Muhammad Arifin
NIM. 30000320520026

HALAMAN PERNYATAAN ORISINALITAS

Saya yang bertanda tangan di bawah ini:

Nama : Muhammad Arifin
NIM : 30000320520026
Program Studi : Doktor Sistem Informasi
Sekolah Pascasarjana Universitas Diponegoro

Dengan ini saya menyatakan dengan sebenar-benarnya bahwa:

1. Disertasi dengan judul "**Model *Educational Data Mining* Berbasis *Gradient Boosted Trees* Untuk Prediksi Performa Akademik Mahasiswa"** adalah karya ilmiah asli dan belum pernah diajukan untuk mendapatkan gelar akademik (doktor) diperguruan tinggi manapun.
2. Disertasi ini adalah murni ide, rumusan dan hasil penelitian saya serta dilakukan tanpa bantuan orang lain, kecuali Tim Promotor dan Tim Penguji.
3. Disertasi ini tidak terdapat karya atau pendapat yang telah ditulis atau dipublikasikan orang lain, kecuali secara tertulis dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan judul aslinya serta dicantumkan dalam daftar pustaka.
4. Pernyataan ini dibuat dengan sesungguhnya dan apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyatan ini, saya bersedia menerima sanksi akademik berupa pencabutan gelar yang tela saya peroleh dan sanksi lain sesuai dengan norma yang berlaku di Universitas Diponegoro.

Semarang, Maret 2024

Muhammad Arifin

DAFTAR RIWAYAT HIDUP

Dr. Ir. Muhammad Arifin, A.Md, S.Kom, M.Kom, MCE, lahir pada tanggal 21 April 1983, di Jepara, Jawa Tengah, Indonesia. Penulis merupakan putra pertama dari Bapak Ladri dan Ibu Sukirah. Penulis menyelesaikan pendidikan dasar di SD Negeri Bucu IV Jepara pada tahun 1995. Pendidikan menengah ditempuh penulis pada MTs. Matholiul Falah, Sumanding, Jepara pada tahun 1995-1997 dan SMK Bhakti Praja Jepara pada tahun 1997-2000. Penulis menyelesaikan program Diploma jurusan Teknik Elektronika di Universitas Muria Kudus pada tahun 2005. Penulis melanjutkan pendidikan Stratas I pada Prodi Sistem Informasi Universitas Muria Kudus 2008-2011. Penulis menempuh pendidikan Akta IV di Universitas Tunas Pembangunan Surakarta tahun 2012. Pendidikan Strata Dua ditempuh penulis di Universitas Dian Nuswantoro, Semarang, Prodi Teknik Informatika pada tahun 2011-2013. Pada bulan Januari tahun 2021 Penulis terdaftar sebagai mahasiswa doktoral pada program Doktor Sistem Informasi, Sekolah Pascasarjana, Universitas Diponegoro Semarang untuk periode 2021 sampai 2024, dengan bidang penelitian *Educational Data Mining* (EDM) dan pembelajaran mesin. Disertasi yang disusun berjudul Model *Educational Data Mining* Berbasis *Gradient Boosted Trees* Untuk Prediksi Performa Akademik Mahasiswa**,** berhasil diselesaikan dalam waktu 3 tahun, 1 bulan, 28 hari. Penulis bekerja sebagai dosen pada program studi Sistem Informasi, Fakultas Teknik, Universitas Muria Kudus sejak tahun 2014 sampai sekarang, dan berkonsentrasi untuk bidang penelitian sistem informasi dan data mining untuk pendidikan.

KATA PENGANTAR

Puji dan Syukur penulis panjatkan kehadirat Allah SWT, atas limpahan Karunia, Taufik, Hidayah dan Ridho-Nya sehingga penulis bisa menyelesaikan laporan Disertasi ini yang merupakan salah satu persyaratan akademik guna memperoleh gelar Doktor dalam Program Studi Doktor Sistem Informasi, Sekolah Pascasarjana, Universitas Diponegoro.

Judul yang diangkat dalam disertasi ini adalah **Model *Educational Data Mining* Berbasis *Gradient Boosted Trees* Untuk Prediksi Performa Akademik Mahasiswa**. Dari penelitian ini menghasilkan dua paper yang terbit di seminar internasional dan dua paper pada jurnal internasional, serta satu aplikasi, dua HKI dan satu buku ber-ISBN.

Penulis menyadari bahwa dalam proses penyelesaian Disertasi ini telah melibatkan berbagai pihak, baik secara langsung maupun tidak langsung, yang telah memberikan kontribusi dalam penyelesaian penyusunan laporan disertasi ini. Untuk itu dalam kesempatan ini penulis mengucapkan terima kasih dan penghargaan yang setinggi-tingginya kepada yang terhormat:

1. Bapak Dr. R.B. Sularto, S.H., M.Hum selaku Dekan Sekolah Pascasarjana Universitas Diponegoro.
2. Bapak Prof. Dr. Drs. Rahmat Gernowo, M.Si selaku Ketua Program Studi Doktor Sistem Informasi Sekolah Pascasarjana Universitas Diponegoro.
3. Ibu Prof. Dr. Widowati, S.Si., M.Si, selaku promotor.
4. Bapak Farikhin, M.Si, Ph.D, selaku ko promotor.
5. Bapak Dr. Maman Somantri, S.T., M.T., selaku penguji I.
6. Bapak Prof. Dr. Adian Fatchur Rochim, S.T., M.T. selaku penguji II.
7. Bapak Prof. Dr. Ridwan Sanjaya, S.E., S.Kom., MS.IEC. selaku penguji eksternal.
8. Semua pihak yang tidak dapat penulis sebutkan satu persatu, yang telah membantu selama proses penelitian sampai selesainya penyusunan laporan disertasi ini.
9. Orang Tua, Bapak Ladri, Ibu Sukirah dan Ibu mertua Sumini, Istri tercinta Sugiyani, Anak pertama Manggar Aflah, Anak kedua Bima Maimuun Argan dan Anak ketiga Khansa Rasydan Argan.

Semarang, Maret 2024

Penulis

DAFTAR ISI

DAFTAR GAMBAR

DAFTAR TABEL

DAFTAR LAMPIRAN

1. Artikel-artikel luaran penelitian
2. Data buku EDM ber-ISBN 9786234171730
3. HKI Buku
4. HKI Aplikasi Sistem Informasi Prediksi Performa Akademik Mahasiswa
5. Panduan Aplikasi Sistem Informasi Prediksi Performa Akademik Mahasiswa

DAFTAR ARTI LAMBANG DAN SINGKATAN

1. n_estimators adalah parameter yang digunakan untuk menentukan jumlah pohon yang akan digunakan dalam algoritma.
2. Subsample adalah parameter yang digunakan untuk menentukan jumlah sampel yang diambil dari data untuk diproses.
3. Learning rate adalah parameter yang digunakan untuk mengontrol kecepatan pembelajaran algoritma.
4. Max_depth adalah parameter yang digunakan untuk menentukan kedalaman maksimum suatu pohon yang akan dibangun dalam algoritma.

ABSTRAK

Pendidikan memberikan dampak yang sangat penting terhadap pertumbuhan ekonomi suatu bangsa karena pendidikan berperan besar dalam menentukan kualitas tenaga kerja. Penggunaan teknologi informasi pada dunia pendidikan menghasilkan sejumlah data besar yang berkaitan dengan mahasiswa dalam bentuk elektronik. Sangat penting bagi pemangku kepentingan untuk secara efektif mengubah kumpulan data ini menjadi informasi yang membantu pengajar, administrator, dan pembuat kebijakan untuk menganalisisnya guna meningkatkan kualitas pengambilan keputusan. *(Educational Data Mining)* EDM adalah disiplin ilmu yang berkembang, berkaitan dengan perluasan metode *Data Mining* klasik dan mengembangkan metode baru untuk menemukan data yang berasal dari sistem pendidikan. Prediksi performa akademik mahasiswa bertujuan untuk memperkirakan nilai yang tidak diketahui dari variabel yang menggambarkan mahasiswa. Nilai-nilai yang biasanya diprediksi adalah performa, skor, atau nilai. Memprediksi performa akademik mahasiswa dapat membantu mahasiswa dan pengajar dalam melacak kemajuan mahasiswa. Penelitian dalam memprediksi performa akademik mahasiswa berusaha untuk menemukan fitur yang memiliki pengaruh terhadap performa akademik mahasiswa. Data performa akademik mahasiswa sebagian besar menggunakan dua jenis kumpulan data yaitu data dari database perguruan tinggi dan platform pembelajaran online. Penelitian ini bertujuan untuk membangun sistem informasi dengan menggunakan model regresi berbasis algoritma *boosting* untuk memprediksi performa akademik dan menggabungkan fitur akademik dan non-akademik, selanjutnya menemukan apakah fitur-fitur tersebut berpengaruh terhadap performa akademik mahasiswa. Hasil penelitian menunjukkan bahwa model yang diusulkan memiliki tingkat kesalahan paling kecil dibandingkan dengan model yang dibangun menggunakan algoritma *generalized linear model, deep learning, decision tree, random forest,* dan *support vector machine*, selain itu setelah dilakukan optimasi tingkat kesalahan dari model semakin kecil. Gabungan fitur akademik dan non-akademik masing-masing memiliki pengaruh terhadap prediksi sehingga fitur-fitur ini dapat digunakan untuk memprediksi performa akademik mahasiswa. Sistem informasi prediksi ini dapat memudahkan pengguna untuk memprediksi performa akademik mahasiswa.

Kata kunci: Model Regresi, Algoritma *Boosting*, GBT, EDM, Prediksi Performa Akademik Mahasiswa, LMS, akademik, non-akademik

ABSTRACT

Education holds a critically significant impact on a nation's economic growth as it plays a major role in determining the quality of the workforce. The integration of information technology in the field of education generates substantial electronic data related to students. Effectively transforming this data into meaningful information is crucial for stakeholders, aiding educators, administrators, and policymakers in analysis for enhanced decision-making quality. Educational Data Mining (EDM) is an evolving scientific discipline, involved in extending classical Data Mining methods and developing new methods to unearth data originating from the education system. Student academic performance prediction aims to estimate unknown values of variables describing students. The values typically predicted include performance, scores, or grades. Predicting student academic performance can assist students and teachers in monitoring student progress. Research in predicting student academic performance seeks to identify influential features on student performance. Student academic performance data largely leverages two types of datasets: data from university databases and online learning platforms. This research aims to build an information system using a regression model based on a boosting algorithm to predict academic performance, combining academic and non-academic features. Subsequently, the research aims to determine the impact of these features on student academic performance. The research results indicate that the proposed model has the smallest error rate compared to models built using generalized linear model algorithms, deep learning, decision trees, random forests, and support vector machines. Furthermore, after optimization, the model's error rate decreases. The combination of academic and non-academic features each has an impact on predictions, indicating their utility in predicting student academic performance. This predictive information system facilitates users in predicting student academic performance.

Keywords: Regression models, Boosting algorithm, GBT, EDM, Predicting student academic performance, LMS, Academic, Non-academic