Klasifikasi Nilai Ujian Siswa Menggunakan Decision Tree

**Ade Kukuh Setiawan , Sandy Yudha Avyan, Krishna Eka Bayu , Muhammad Shilni Sidqan**

Fakultas Teknologi Informasi/Universitas Yarsi

Jl. Letjen Suprapto/Cempaka Putih

Email : [adekukuh12@gmail.com](mailto:adekukuh12@gmail.com), [Alvyansandy.yudha@gmail.com](mailto:Alvyansandy.yudha@gmail.com), [krishnafarhan@gmail.com](mailto:krishnafarhan@gmail.com), [sidqanshilni@gmail.com](mailto:sidqanshilni@gmail.com).

***Abstract***

*Informasi di dalam dunia pendidikan merupakan bagian penting dan sangat berharga. Informasi yang akurat dan tepat akan membantu seseorang dalam mengambil keputusan dan menentukan langkah-langkah yang harus dilakukan untuk mempertahankan dan mengembangkan usahanya. Semua itu diperlukan sistem basis data yang lengkap, akurat dan dapat ditampilkan secara mudah dan cepat setiap kali diperlukan. Penelitian ini mengusulkan metode klasifikasi nilai ujian Siswa menggunakan Klasifikasi Decision Tree berdasarkan hasil data Decision Tree. Hasil dari penelitian ini menunjukan bahwa klasifikasi decision tree mendapat nilai akurasi sebesar 60%.*

***Keywords****: Nilai Siswa, Decision Tree, Data Mining , Classification ,Machine Learning ,*

**1. Kata Pengantar**

Semua lembaga pendidikan tinggi berusaha untuk memenangkan siswa yang termotivasi untuk belajar dan yang memiliki rekam jejak dalam keberhasilan studi. Ini biasanya anggapan untuk kesuksesan masa depan. Para siswa dengan profil dan pendidikan yang berbeda juga memiliki berbagai tingkat keberhasilan yang diukur sesuai dengan nilai rata-rata di sekolah atau ujian negara. Ini, bersama dengan tingkat pendidikan orangtua, kursus persiapan ujian dan etnis mungkin mempengaruhi keberhasilan mereka pada fase awal studi mereka. Memprediksi

Keberhasilan siswa pada tahap awal studi mereka membantu sekolah dalam mengarahkan lebih banyak kegiatan untuk siswa yang kurang berprestasi untuk meningkatkan keberhasilan mereka.

Menurut Simeunović dan Preradović [25], menganalisis kesuksesan akademik adalah penting bagi lembaga pendidikan tinggi, mengingat bahwa perencanaan strategis

program studi menyiratkan memperluas atau mengurangi ruang lingkup atau kedalaman

kurikulum serta memodifikasi proses pedagogis dan pendidikan, tergantung pada prestasi siswa.

Dalam karya ini, tiga pendekatan penambangan data diusulkan untuk memprediksi kinerja siswa. Prediksi dilakukan dengan menggunakan **data extraction, data exploration, dan data preprocessing.** Prediksi dilakukan dengan catatan akademik bersama dengan jenis kelamin, etnis, tingkat pendidikan orang tua, dan informasi skor akademik.

**2. Pendahuluan**

**A. Motivasi**

Kita mengetahui seberapa pentingnya mengetahui kelulusan kita dalam suatu pembelajaran di sekolah atau di universitas, Apalagi informasi tersebut akurat dan sangat tepat untuk membantu Terkadang ada begitu banyak rintangan dalam hidup sehingga hal-hal yang diinginkan sulit didapat dan dipertahankan. Maka itu kita termotivasi untuk membuat aplikasi ini.

**B. Tujuan**

Tujuan kita mengambil nilai klasifikasi ujian siswa untuk memprediksi seberapa banyak siswa yang lulus dan tidak lulus dengan mengetahuinya melalui algoritma Decision Tree. Manfaat utamanya mempermudah klasifikasi data pada nilai siswa. Menyaring data dari beberapa siswa yang berprestasi ataupun tidak. Menghasilkan hasil dari klasifikasi proses data yang di ada pada tabel indeks nilai siswa.

**C. Tinjauan Pustaka**

1. Bahasa pemrograman Python

Python adalah bahasa pemrograman interpretatif multiguna dengan filosofi perancangan yang berfokus pada tingkat keterbacaan kode. Python diklaim sebagai bahasa yang menggabungkan kapabilitas, kemampuan, dengan sintaksis kode yang sangat jelas, dan dilengkapi dengan fungsionalitas pustaka standar yang besar serta komprehensif. Python juga didukung oleh komunitas yang besar.  
  
Python mendukung multi paradigma pemrograman, utamanya; namun tidak dibatasi; pada pemrograman berorientasi objek, pemrograman imperatif, dan pemrograman fungsional. Salah satu fitur yang tersedia pada python adalah sebagai bahasa pemrograman dinamis yang dilengkapi dengan manajemen memori otomatis. Seperti halnya pada bahasa pemrograman dinamis lainnya, python umumnya digunakan sebagai bahasa skrip meski pada praktiknya penggunaan bahasa ini lebih luas mencakup konteks pemanfaatan yang umumnya tidak dilakukan dengan menggunakan bahasa skrip. Python dapat digunakan untuk berbagai keperluan pengembangan perangkat lunak dan dapat berjalan di berbagai platform sistem operasi.

1. Neural Network

Neural Network merupakan kategori ilmu Soft Computing. Neural Network sebenarnya mengadopsi dari kemampuan otak manusia yang mampu memberikan stimulasi/rangsangan, melakukan proses, dan memberikan output. Output diperoleh dari variasi stimulasi dan proses yang terjadi di dalam otak manusia. Kemampuan manusia dalam memproses informasi merupakan hasil kompleksitas proses di dalam otak. Misalnya, yang terjadi pada anak-anak, mereka mampu belajar untuk melakukan pengenalan meskipun mereka tidak mengetahui algoritma apa yang digunakan. Kekuatan komputasi yang luar biasa dari otak manusia ini merupakan sebuah keunggulan di dalam kajian ilmu pengetahuan.

Fungsi dari Neural Network diantaranya adalah:

* Pengklasifikasian pola
* Memetakan pola yang didapat dari input ke dalam pola baru pada output
* Penyimpan pola yang akan dipanggil kembali
* Memetakan pola-pola yang sejenis
* Pengoptimasi permasalahan
* Prediksi

1. Data Mining

Data mining didefinisikan sebagai sebuah proses untuk menemukan hubungan, pola dan trend baru yang bermakna dengan menyaring data yang sangat besar, yang tersimpan dalam penyimpanan, menggunakan teknik pengenalan pola seperti teknik Statistik dan Matematika (Kamagi dan Seng, 2014). Pengolahan data mining terdiri dari beberapa metode pengolahan (Kamagi dan Seng, 2014), yaitu:

1) Model prediksi merupakan pengolahan data mining dengan melakukan prediksi/peramalan.   
 Tujuan metode ini untuk membangun model prediksi suatu nilai yang mempunyai ciri-ciri   
 tertentu.   
  
2) Association (Asosiasi) merupakan teknik dalam data mining yang mempelajari hubungan  
 antar data.   
  
3) Clustering (Pengelompok) merupakan teknik untuk mengelompokkandata ke dalam suatu   
 kelompok tertentu.   
  
4) Classification merupakan teknik mengklasifikasikan data.

D. **Metode Penelitian**

Proses persiapan data berada di bawah pengumpulan / pengumpulan data, dan pembersihan data melalui transformasi data. Proses pengumpulan data melibatkan pengumpulan data mentah tentang siswa.

Data yang digunakan untuk analisis dimasukkan ke dalam file spreadsheet

Excel. Jumlah matrikulasi aktual siswa adalah tidak digunakan: Format data ditunjukkan pada tabel model untuk memprediksi keberhasilan siswa dirancang menggunakan **data extraction, data exploration, dan data preprocessing** . Penelitian dilakukan di antara siswa dari nilai Math Score, Reading Score dan Writing Score .

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1. Data Understanding   Data understanding digunakan untuk memahami dataset yang digunakan   |  |  | | --- | --- | | Nama Atribut | Tipe Data | | Gender | Object | | race/ethnicity | Object | | parental level of education | Object | | Lunch | Object | | test preparation course | Object | | math score | int64 | | reading score | int64 | | writing score | int64 |  1. Missing Value   Pada dataset ini tidak ditemukan adanya missing value.   1. Membuat anotation   Anotation dibuat untuk menentukan tingkat kelulusan siswa di dalam dataset   1. Membuat correlation matriks     Dari Correlation Matriks diatas, terdapat 3 atribut yang saling berhubungan yaitu math score, reading score dan writing score.   1. One hot encoding   One Hot encoding digunakan untuk mengubah nilai teks menjadi binaries   1. Decision Tree   Decission tree merupakan salah satu algoritma yang digunakan untuk klasifikasi.   1. Confussion Matriks   Confussion matriks diperlukan untuk mendapatkan hasil evaluasi dari klasifikasi Decision Tree |  |

**3. Penelitian terkait sebelumnya**

Penelitian dilakukan di antara siswa contoh terdiri dari 1001 siswa. Variabel output dari model adalah Grade average dari nilai mereka sebelumnya yang dinyatakan sebagai dua kategori - kurang dari, sedangkan variabel inputnya adalah Gender, Race/ethnicity, Parental Level of Education dan Test Preparation Course and Lunch. Waktu yang didedikasikan untuk belajar, bahan ujian, Siswa mengambil ujian pendahuluan, Neural Network memberikan tingkat akurasi yang lebih rendah (60%), dan uji statistik menunjukkan perbedaan yang signifikan secara statistik. mengeksplorasi penggunaan metode penambangan data dalam pendidikan tinggi dan menciptakan model klasifikasi untuk decision tree. Analisis dilakukan pada sampel 1001 siswa.

melakukan penelitian yang bertujuan memprediksi akademik

kinerja siswa. Tujuan dari makalah ini adalah untuk mengklasifikasikan siswa

menjadi ... kelompok : (1) .... (2) ... (3) ....

Tujuan dari penelitian Shah [24] adalah untuk menyelidiki faktor-faktor yang mempengaruhi akademik kinerja siswa dengan membandingkan keakuratan berbagai pengklasifikasi. Siswa dikategorikan dalam lima kelompok berdasarkan kinerja seperti: “sangat siswa yang pintar ”- kemungkinan besar untuk berhasil; Nilai Siswa "Besar" - di atas rata-rata hasil dengan sedikit usaha dan yang mungkin berhasil dengan nilai bagus;

"Sedang" siswa - mereka yang mungkin berhasil; Siswa “Kurang” - mereka berinvestasi lebih banyak upaya untuk berhasil; dan siswa “gagal”.

**4. Hasil dan Analisis**

Dari Penelitian yang kami kerjakan, kami mendapatkan hasil akurasi sebesar 60% dari confussion matriks berikut :

|  |  |  |
| --- | --- | --- |
|  | **prediction/f** | **prediction/p** |
| **actual/f** | **22** | **42** |
| **actual/p** | **38** | **99** |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Precision** | **Recall** | **Fi score** | **Support** |
| F | 0.37 | 0.34 | 0.35 | 64 |
| P | 0.70 | 0.72 | 0.71 | 137 |
|  |  |  |  |  |
| Micro avg | 0.60 | 0.60 | 0.60 | 201 |
| Macro avg | 0.53 | 0.53 | 0.53 | 201 |
| Weighted avg | 0.60 | 0.60 | 0.60 | 201 |

**5. Kesimpulan**

1. Dari hasil klasikasi menggunakan decision tree didapatkan hasil akurasi sebesar 60%.
2. Dari data pengujian yang diberikan berupa data pelajar sebanyak 201 pelajar yang akan ditentukan, didapatkan 137 pelajar lulus dan 64 pelajar tidak lulus.

**6. Saran**

1. Penelitian ini harus ditambah dengan standar kelulusan siswa sesuai dengan asal negara dataset diambil, sehingga validasi keputusannya semakin valid.

**Referensi :**

1. Berry, M. J. A., & Linoff, G. (1997). Data Mining Techniques: For Marketing, Sales, and Customer Support. Wiley Computer Publishing.
2. Sherrod, P. H. (2006). DTREG: Classification and Regression Trees and Support Vector Machines for Predictive Modeling and Forecasting. Retrieved from http://www.dtreg.com/ DTREG.pdf
3. Bouckaert, R. R., Frank, E., Hall, M., Kirkby, R., Reutemann, P., Seewald, A. and Scuse, D. (2016). WEKA Manual for Version 3-8-0.
4. Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5-32.
5. Cheewaprakobkit, P. (2015). Predicting student academic achievement by using the decision tree and neural network techniques. Catalyst, 12(2), 34-43.
6. Delen, D. (2010). A comparative analysis of machine learning techniques for student retention management. Decision Support Systems, 49(4), 498-506.
7. Ibrahim, Z. and Rusli, D. (2007). Predicting students' academic performance: comparing artificial neural network, decision tree and linear regression. 21st Annual SAS Malaysia Forum, 5th September 2007, Kuala Lumpur.