



IMPLEMENTASI DATA MINING UNTUK MENGETAHUI MINAT SISWA TERHADAP PERGURUAN TINGGI

Hendri Noviyanto^a, Arma Fauzi^b

^a Sistem Komputer, hendrinoviyantoo@gmail.com, Universitas Surakarta

^b Sistem Komputer, armafauziunsa@gmail.com, Universitas Surakarta

ABSTRAK

College is a place for further learning after completing high school education or vocational high school. The number of universities, especially private ones, makes competition to attract prospective students. This makes every university compete in making innovations ranging from learning techniques, facilities, laboratories and much more. Therefore, a prediction of the interest of prospective students is carried out by involving several factors that encourage a prospective student to choose a university with the intention that the university is more appropriate in making decisions for the development process. In this study, the dataset was taken through interview techniques to the research subjects as many as 100 people. The data processing uses the WEKA application, and the validation uses the Cross Validation technique with $K = 10$. The attributes used in this study were 7 attributes plus 1 class attribute. The proposed algorithm is Sequential Minimal Optimization which is successful in classifying data with an accuracy of 0.9. The algorithm used for the comparison process gets a better score of 0.92 using Decision Tree.

Keywords: data mining, interest, college, prediction, smo.

Abstrak

Perguruan Tinggi merupakan sebuah tempat belajar lebih lanjut setelah menyelesaikan pendidikan sekolah menengah atas atau sekolah menengah kejuruan. Banyaknya Perguruan Tinggi, khususnya swasta menjadikan persaingan untuk menarik minat calon mahasiswa. Hal ini membuat setiap Perguruan Tinggi berlomba dalam membuat sebuah inovasi mulai dari teknik pembelajaran, fasilitas, laboratorium dan masih banyak lagi. Oleh karena itu, dilakukan sebuah prediksi minat calon mahasiswa dengan melibatkan beberapa faktor pendorong seorang calon mahasiswa untuk memilih sebuah Perguruan Tinggi dengan maksud supaya Perguruan Tinggi lebih tepat dalam mengambil keputusan untuk proses pengembangan. Pada penelitian ini, dataset diambil melalui teknik wawancara kepada subject penelitian sebanyak 100 orang. Proses pengolahan data menggunakan aplikasi WEKA, dan validasi menggunakan teknik *Cross Validation* dengan $K=10$. Atribut yang digunakan pada penelitian ini sebanyak 7 buah atribut ditambah 1 atribut kelas. Algoritma yang diusulkan adalah *Sequential Minimal Optimization* yang berhasil mengklasifikasikan data dengan perolehan akurasi sebesar 0,9. Algoritma yang digunakan untuk proses komparasi mendapatkan nilai lebih baik sebesar 0,92 menggunakan *Decision Tree*.

Kata Kunci: data mining, minat, perguruan tinggi, prediksi, smo.

1. PENDAHULUAN

Pendidikan tinggi merupakan sebuah kebutuhan pokok untuk bisa mendapatkan kehidupan yang layak. Berbekal pendidikan seseorang akan memiliki pandangan yang luas dan kemampuan yang lebih sesuai dengan yang ditekuni [1]. Sebuah Perguruan Tinggi bisa dikatakan berkembang, salah satunya adalah jumlah mahasiswa yang terdaftar pada setiap tahunnya semakin meningkat. Perguruan Tinggi swasta pada khususnya, mahasiswa menjadi penopang utama dalam keberlangsungan dan perkembangan perguruan tinggi tersebut [2]. Penurunan jumlah penerimaan mahasiswa pada sebuah Perguruan Tinggi swasta membutuhkan sebuah perhatian khusus dan mencari penyebabnya. Pada saat ini, Program Studi Sistem Komputer Universitas Surakarta mengalami penurunan minat. Maka dibutuhkan sebuah usaha untuk mendapatkan calon mahasiswa baru.

Dalam proses pemilihan sebuah perguruan tinggi, [2] menjelaskan bahwa setiap calon mahasiswa akan mencari informasi terkait dengan biaya kuliah, lokasi perguruan tinggi, fasilitas, perpustakaan, laboratorium, kegiatan mahasiswa, akreditasi dan program studi yang terdapat di sebuah universitas tujuan. Faktor lain yang mempengaruhi calon mahasiswa memilih perguruan tinggi adalah faktor pengaruh orang tua, teman, kerabat, beasiswa, dan peluang kerja [3]. Faktor-faktor yang telah disebutkan diatas, diharapkan

pihak Perguruan Tinggi dapat memprediksi segala bentuk kemungkinan untuk membuat keputusan dalam merekrut calon mahasiswa sebanyak mungkin. Informasi dan faktor yang dapat mempengaruhi minat calon mahasiswa untuk mendaftar di sebuah Perguruan Tinggi dapat diolah menggunakan teknik data mining.

Penelitian terkait dengan prediksi minat calon mahasiswa dalam memilih Perguruan Tinggi telah banyak dilakukan, seperti penelitian [3] menganalisis faktor pembentukan keputusan pemilihan Perguruan Tinggi. Penelitian [2] menggunakan algoritma C4.5 berusaha menentukan minat calon mahasiswa memilih Perguruan Tinggi. Penelitian ini menggunakan 25 data sampel dan 5 kategori penilaian. Data testing yang digunakan sebanyak 120 dataset. Hasil akhir adalah tingkat akurasi yang sempurna sebesar 100%. Penelitian yang dilakukan oleh [4] menerapkan algoritma C4.5 untuk membantu proses pemilihan jurusan pada sebuah Universitas. Pada penelitian ini dijelaskan bahwa, pemilihan jurusan ataupun Universitas yang dipilih dipengaruhi oleh orang tua, kerabat dan teman-teman sehingga belum tentu jurusan yang dipilih sesuai dengan minat. Akurasi yang didapatkan pada penelitian ini sebesar 95%. Penelitian yang dilakukan oleh [5] menerapkan algoritma K-Nearest Neighbour untuk melakukan prediksi penerimaan calon mahasiswa fakultas kedokteran. Nilai akurasi yang didapatkan pada penelitian ini sebesar 76.1%. Penelitian yang dilakukan oleh [6] menerapkan algoritma C4.5 untuk melakukan prediksi minat calon mahasiswa pada sebuah Perguruan Tinggi. Pada penelitian ini, variabel yang digunakan meliputi jenis kelamin, minat, jurusan asal sekolah dan hobi. Penelitian dari [7] membandingkan algoritma klasifikasi untuk menelusuri minat calon mahasiswa baru. Pada penelitian ini algoritma yang dibandingkan meliputi Decision Tree (J48), K-Nearest Neighbour (KNN), Naïve Bayes (NB). Hasil perbandingan melihat nilai *Area Under Curve* (AUC) dari J48 dengan nilai 0,965, KNN diperoleh 0,9343, dan NB sebesar 0,9461. Penelitian yang dilakukan oleh [8] melakukan prediksi heregistrasi untuk calon mahasiswa baru. Penelitian ini menggunakan algoritma NB dengan nilai akurasi sebesar 92,67 dengan AUC sebesar 0,841. Variabel yang digunakan pada penelitian ini cukup banyak, yaitu nama, tahun pendaftaran, program kelas, jenis kelamin, usia, prodi, kota asal, pekerjaan ayah, pekerjaan ibu, penghasilan orang tua, jurusan sekolah asal, nilai UN, informasi pendaftaran. Penelitian [9] melakukan prediksi calon mahasiswa yang akan melakukan daftar ulang. Variabel yang digunakan pada penelitian ini meliputi asal kota, gelombang, program studi, penghasilan orang tua, nilai UN dan status registrasi. Penelitian ini menggunakan algoritma naïve bayes untuk melakukan prediksi dengan hasil nilai akurasi sebesar 92,88%. Penelitian [10] mengklasifikasikan dan merekomendasikan jurusan yang akan diambil pada Perguruan Tinggi menggunakan algoritma Naïve Bayes – WP. Nilai akurasi yang didapatkan sebesar 82% menggunakan nilai akademik siswa siswi SMA.

Dari beberapa metode yang digunakan oleh peneliti terdahulu, algoritma yang populer seperti *naïve bayes* dan *decision tree* sangat banyak digunakan, dengan menghitung nilai akurasi ataupun nilai AUC para peneliti dapat mengetahui kemampuan algoritma tersebut ketika diimplementasikan. Pada penelitian ini, peneliti menggunakan algoritma *Sequential Minimal Optimization* (SMO) untuk mengklasifikasikan minat calon mahasiswa yang mendaftar di Perguruan Tinggi. Kemudian kemampuan algoritma tersebut akan akan dibandingkan dengan algoritma populer lainnya seperti NB, KNN, dan *Decision Tree* (J48).

2. TINJAUAN PUSTAKA

2.1. Minat

Minat merupakan suatu kecenderungan dan keairahan yang tinggi atau keinginan yang besar terhadap sesuatu. Minat dapat timbul karena terdapat daya tarik dari luar dan datang dari hati sanubari [2]. Dari minat ini kita dapat mengetahui keinginan calon mahasiswa yang menjadi faktor pendorong untuk memilih sebuah Perguruan Tinggi.

2.2. *Sequential Minimal Optimization* (SMO)

Sequential Minimal Optimization (SMO) merupakan optimasi dari algoritma *Support Vector Machine* (SVM). SMO bekerja secara sederhana dan memecahkan masalah pemrograman kuadratik (QP) ke dalam submasalah pemrograman kuadratik. SMO memperbaiki kekurangan SVM dengan menyelesaikan masalah QP dalam algoritma SVM tanpa menggunakan ruang matriks tambahan dan tanpa mengulang nilai numerik yang sama untuk setiap submasalah. Cara kerja SMO tidak seperti metode pendahulunya, SMO akan memecahkan masalah-masalah kecil yang mungkin diselesaikan di setiap langkah. SMO melibatkan dua pengganda Lagrange untuk optimasi pada masalah QP. SMO merupakan metode dekomposisi yang bekerja berdasarkan prinsip himpunan kerja dua elemen dan bekerja secara analitis, hal ini menghasilkan jumlah iterasi yang sangat banyak, namun karena setiap iterasi cukup kecil maka total waktu yang dibutuhkan juga lebih singkat [11].

Nilai yang dimasukkan ke dalam working set adalah variabel yang paling melanggar ketentuan KTT; berikut ini adalah persamaan kondisi SMO (1) dan KTT (2).

$$\max_{\alpha} \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n y_i y_j K(x_i, x_j) \alpha_i \alpha_j, \quad (1)$$

Harus memenuhi,

$$\begin{aligned} K &= (\vec{x}_i, \vec{x}_j) = \Phi(\vec{x}_i) \cdot \Phi(\vec{x}_j) \\ \sum_{i=1}^n y_i \alpha_i &= 0 \\ \alpha_i = 0 &\Rightarrow y_i f(\vec{x}_i) \geq 1, \\ 0 < \alpha_i < C &\Rightarrow y_i f(\vec{x}_i) = 1, \\ \alpha_i = C &\Rightarrow y_i f(\vec{x}_i) \leq 1, \end{aligned} \quad (2)$$

3. METODOLOGI PENELITIAN

3.1. Dataset

Dataset pada penelitian ini memiliki 7 variabel dengan mengadopsi dari penelitian [2] dan hasil diskusi dengan beberapa ketua program studi yaitu lokasi Perguruan Tinggi, biaya kuliah, fasilitas Perguruan Tinggi, laboratorium, perpustakaan, kegiatan mahasiswa, akreditasi. Dataset dikumpulkan dari hasil interview di beberapa tempat dengan subject siswa siswi sekolah menengah atas dan sekolah menengah kejuruan dengan jumlah sebanyak 100 data.

3.2. Analisa Data

Tahap pertama pada penelitian ini adalah memprediksi dataset yang telah didapatkan menggunakan algoritma *Sequential Minimal Optimization* (SMO). Setelah mendapatkan hasil dari tingkat akurasi, presisi, dan recall, kemudian dilakukan proses prediksi menggunakan algoritma Naïve Bayes, KNN, dan J48 sebagai pembandingan hasil kinerja dari algoritma yang diusulkan.

4. HASIL DAN PEMBAHASAN

Penelitian ini menggunakan software Weka [12] untuk membantu proses prediksi minat calon mahasiswa. Proses pengujian validasi data menggunakan teknik *10-fold Cross Validation*. Dataset dibagi menjadi 2 bagian yaitu data training sebesar 80% dan data testing sebesar 20%. Dataset berasal dari proses interview dari beberapa tempat dengan subject siswa siswi sekolah menengah atas dan sekolah menengah kejuruan dengan jumlah data yang akan di test sebesar 100 instances dan 7 attribute. Proses prediksi dilakukan menggunakan algoritma SMO, kemudian dilakukan pengujian dengan algoritma NB, KNN dan J48 yang kemudian hasilnya akan dikomparasikan. Hasil pengujian kemudian dimasukkan kedalam tabel supaya lebih mudah untuk mengevaluasinya.

Tabel 1. Hasil Prediksi

	Algoritma			
	SMO	NB	KNN	J48
Presisi	0,911	0,886	0,906	0,921
Recall	0,900	0,880	0,900	0,920

Pada Tabel 1. Hasil prediksi dapat dilihat proses pengklasifikasian dari setiap algoritma menghasilkan Presisi dan Recall yang saling berdekatan. Melihat tabel diatas, algoritma J48 memiliki nilai presisi dan recall tertinggi sebesar 0,921 untuk presisi dan 0,920 untuk recall. Kemudian untuk menghitung nilai akurasi pada setiap hasil algoritma kita dapat merujuk pada tabel *confusion matrix* yang ada dibawah ini.

Nilai akurasi dapat dihitung menggunakan rumus 3 dibawah ini.

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (3)$$

```

=== Confusion Matrix ===      === Confusion Matrix ===      === Confusion Matrix ===      === Confusion Matrix ===
a b <-- classified as      a b <-- classified as      a b <-- classified as      a b <-- classified as
42 9 | a = Tidak           42 9 | a = Tidak           43 8 | a = Tidak           46 5 | a = Tidak
1 48 | b = Dipilih         3 46 | b = Dipilih         2 47 | b = Dipilih         3 46 | b = Dipilih

```

Gambar 1. SMO

Gambar 2. NB

Gambar 3. KNN

Gambar 4. J48

Melihat Gambar 1, 2, 3 dan 4 dapat dihitung nilai akurasi pada setiap algoritma pengklasifikasi yang digunakan menggunakan persamaan (3). Nilai akurasi dapat dilihat pada tabel 2 dibawah ini.

	Algoritma			
	SMO	NB	KNN	J48
Akurasi	0,9	0,88	0,9	0,92

Melihat Tabel 2 dapat diketahui bahwa nilai akurasi tertinggi didapatkan ketika proses prediksi menggunakan algoritma decision tree atau J48 sebesar 0,92. Disusul dengan algoritma SMO dan KNN dengan nilai akurasi sebesar 0,9. Naïve Bayes mendapatkan nilai akurasi terendah sebesar 0,88. SMO merupakan algoritme yang diusulkan belum mampu mendapatkan nilai akurasi terbaik, hal ini dapat dipengaruhi oleh banyak faktor salah satunya adalah kondisi data input yang belum sesuai dengan algoritma SMO sendiri. Merujuk pada penelitian terdahulu, algoritma *Decision Tree* sangat banyak digunakan untuk proses temu kembali informasi dan ternyata algoritma ini memang *robust* terhadap jenis data input yang berbeda.

5. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian yang telah dilaksanakan maka dapat diambil sebuah kesimpulan bahwa algoritma *Decision Tree* merupakan algoritma yang cukup *Robust* terhadap data input dan mampu menghasilkan nilai akurasi yang cukup tinggi sebesar 0,92. Nilai akurasi pada algoritma yang diusulkan yaitu SMO hanya 0,9 disusul KNN 0,9 dan terakhir NB sebesar 0,88. Namun, nilai akurasi tidak terpaut begitu besar, sehingga algoritma SMO masih layak digunakan dan diimplementasikan untuk sistem prediksi minat calon mahasiswa dalam memilih sebuah Perguruan Tinggi.

DAFTAR PUSTAKA

- [1] H. Noviyanto and B. Mukti, "IMPLEMENTASI ALGORITME NAÏVE BAYES UNTUK MENENTUKAN KELAYAKAN CALON PENERIMA BEASISWA," vol. 1, no. 2, pp. 7–12, 2022.
- [2] C. Nas, "Data Mining Prediksi Minat Calon Mahasiswa Memilih Perguruan Tinggi Menggunakan Algoritma C4.5," *J. Manaj. Inform.*, vol. 11, no. 2, pp. 131–145, 2021, doi: 10.34010/jamika.v11i2.5506.
- [3] M. Fakhri *et al.*, "Analisis Faktor Pembentuk Keputusan Pemilihan Perguruan Tinggi Swasta Universitas Telkom (Studi Pada Mahasiswa Prodi S1 Administrasi Bisnis Fakultas)," *Bisnis Entrep.*, vol. 11, no. 1, pp. 1–12, 2017, [Online]. Available: <https://www.neliti.com/publications/77086/analisis-faktor-pembentuk-keputusan-pemilihan-perguruan-tinggi-swasta-universita>.
- [4] D. R. Sari, A. P. Windarto, D. Hartama, and S. Solikhun, "Decision Support System for Thesis Graduation Recommendation Using AHP-TOPSIS Method," *J. Teknol. dan Sist. Komput.*, vol. 6, no. 1, pp. 1–6, 2018, doi: 10.14710/jtsiskom.6.1.2018.1-6.
- [5] A. Yobioktabera, "Penerapan Data Mining Untuk Memprediksi Penerimaan Calon Mahasiswa Baru Fakultas Kedokteran Menggunakan Algoritma K-NN," *JTET (Jurnal Tek. Elektro Ter.)*, pp. 16–19, 2021, [Online]. Available: <https://jurnal.polines.ac.id/index.php/jtet/article/view/2550>.

*Implementasi Data Mining Untuk Mengetahui Minat Siswa Terhadap Perguruan Tinggi
(Hendri Noviyanto)*

- [6] Z. Situmorang, S. Mandasari, Y. Franciska, K. Andriyani, and P. S. Ramadhan, "Algoritma C45 Dalam Memprediksi Minat Calon Mahasiswa," *J. Sci. Soc. Res.*, vol. 5, no. 1, p. 125, 2022, doi: 10.54314/jssr.v5i1.809.
- [7] B. Budiman, "Perbandingan Algoritma Klasifikasi Data Mining untuk Penelusuran Minat Calon Mahasiswa Baru," *Nuansa Inform.*, vol. 15, no. 2, pp. 37–52, 2021, doi: 10.25134/nuansa.v15i2.4162.
- [8] G. W. N. Wibowo and M. A. Manan, "Penerapan Algoritma Naive Bayes Untuk Prediksi Heregistrasi Calon Mahasiswa Baru," *JTINFO J. Tek. ...*, vol. 1, no. 1, pp. 1–10, 2022, [Online]. Available: <https://journal.unisnu.ac.id/JTINFO/article/view/126>.
- [9] A. Damayanti, "Algoritma Naive Bayes Untuk Prediksi Jumlah Pendaftar Ulang Pada Penerimaan Mahasiswa Baru," *J. Multi Media dan IT*, vol. 5, no. 1, pp. 2–6, 2019, doi: 10.46961/jommit.v3i2.338.
- [10] R. Fitriawanti, I. Cholissodin, and R. K. Dewi, "Klasifikasi dan Rekomendasi Jurusan Kuliah Bagi Pelajar SMA Menggunakan Algoritme Naive Bayes -WP," *J. Pengemb. Teknol. Inf. dan Ilmu Komput. Univ. Brawijaya*, vol. 2, no. 11, pp. 4914–4922, 2018, [Online]. Available: <http://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/3129>.
- [11] B. Santosa, "Tutorial Support Vector Machine 1 Ide Dasar Support Vector Machine," pp. 1–23, 2011.
- [12] E. Frank, M. A. Hall, and I. H. Witten, "The WEKA workbench," *Data Min.*, pp. 553–571, 2017, doi: 10.1016/b978-0-12-804291-5.00024-6.