



ANALISIS POLA MENGGUNAKAN METODE C4.5 UNTUK PEMINATAN JURUSAN SISWA BERDASARKAN KURIKULUM (STUDI KASUS : SMAN 1 NATAR)

Vista Anestiviya¹, A. Ferico Octaviansyah Pasaribu²

Sistem Informasi, Fakultas Teknik dan Ilmu Komputer, Universitas Teknokrat Indonesia^{1,2}

vistaa31@gmail.com¹, fericopasaribu@teknokrat.ac.id²

Received: (8 Maret 2021) Accepted: (15 Maret 2021) Published: (29 Maret 2021)

Abstract

Specializing in majors is a decision made by new students to choose the “Science” and “IPS” majors at SMA Negeri 1 Natar according to their interests, achievements, zoning and transfer of duties of their parents. The selection of specializations is made on the basis of the interest of the prospective students. Accuracy in determining specialization can determine student learning success. On the other hand, a very good opportunity for students will be lost due to inaccuracy in determining specializations. The current curriculum used in SMA Negeri 1 Natar is the 2013 curriculum. Therefore, the school hopes that the right direction from the start will provide the end of a successful learning process so that the school needs basic variables to determine the success of the teaching and learning process (KBM) namely the value of Indonesian, English, Mathematics, Natural Sciences, and the interests of prospective students. The use of data mining with classification techniques through the C4.5 method is expected to provide an appropriate analysis of student majoring patterns so that the teaching and learning process can run smoothly and successfully. From the results of calculations through system and manual calculations, the results for data a have an accuracy rate of 100% and for data b of 80%, which means that C4.5 can be suggested to process student data in terms of helping to make the best decisions for selecting student majors.

Keywords: Data Mining, Classification, C4.5, Weka, Student majors data

Abstrak

Peminatan jurusan merupakan suatu keputusan yang dilakukan oleh peserta didik baru untuk memilih jurusan “IPA” dan “IPS” yang ada pada sekolah SMA Negeri 1 Natar sesuai dengan minat, Prestasi, Zonasi dan Perpindahan Tugas orangtua. Pemilihan peminatan dilakukan atas dasar minat calon peserta didik. Ketepatan dalam menentukan peminatan dapat menentukan keberhasilan belajar siswa. Sebaliknya, kesempatan yang sangat baik bagi siswa akan hilang karena kurang tepatan dalam penentuan peminatan. Kurikulum yang digunakan pada sekolah SMA Negeri 1 Natar saat ini adalah kurikulum 2013. Maka dari itu pihak sekolah berharap dengan adanya penjurusan yang tepat sejak awal akan memberikan akhir proses pembelajaran yang berhasil sehingga pihak sekolah membutuhkan variabel dasar penentuan keberhasilan proses kegiatan belajar mengajar (KBM) yaitu nilai Bahasa Indonesia, Bahasa Inggris, Matematika, IPA, serta minat dari calon siswa. Penggunaan data mining dengan teknik klasifikasi melalui metode C4.5 diharapkan dapat memberikan analisis pola peminatan jurusan siswa yang sesuai sehingga proses KBM dapat berjalan secara lancar dan berhasil. Dari hasil perhitungan melalui perhitungan sistem dan manual didapatkan hasil untuk data a memiliki tingkat akurasi sebesar 100% dan untuk data b sebesar 80%, yang mengartikan bahwa C4.5 dapat disarankan untuk mengolah data siswa dalam hal membantu memberikan keputusan terbaik pemilihan jurusan siswa.

Kata Kunci : Data Mining, Klasifikasi, C4.5, weka, Data penjurusan siswa

To cite this article:

Anestiviya, Pasaribu (2021). Analisis Pola Menggunakan Metode C4.5 Untuk Peminatan Jurusan Siswa Berdasarkan Kurikulum (Studi Kasus : SMAN 1 Natar). Jurnal Teknologi dan Sistem Informasi, Vol (2) No.1, 80 - 85

1. Pendahuluan

Peminatan jurusan merupakan suatu keputusan yang dilakukan oleh peserta didik baru untuk memilih jurusan "IPA" dan "IPS" yang ada pada siswa. Sebaliknya, kesempatan yang sangat baik bagi siswa akan hilang karena kurang tepatan dalam penentuan peminatan. Kurikulum yang digunakan pada sekolah SMA Negeri 1 Natar saat ini adalah kurikulum 2013. Ditahun sebelumnya pemilihan jurusan baru dilakukan pada saat kenaikan atau saat kenaikan kelas XI. Tentu saja hal ini bisa mengakibatkan sering terjadinya kesalahan dalam memilih jurusan oleh calon siswa SMA.

Pemintaan jurusan siswa yang dilakukan pada SMA Negeri 1 Natar sesuai minat dan dilakukan sejak peserta didik mendaftar ke sekolah yang memiliki tujuan untuk memberikan kesempatan kepada peserta didik mengembangkan kompetensi sikap, kompetensi keterampilan peserta didik sesuai dengan minat, bakat, dan kemampuan akademik dalam sekelompok mata pelajaran keilmuan. Dalam hal penentuan peminatan siswa terlebih lagi proses dalam penentuan peminatan ini masih bersifat manual sehingga memerlukan waktu yang lama dan hasilnya pun belum tentu akurat. Proses penentuan jurusan di SMA Negeri 1 Natar itu merupakan hal yang sangat penting, karena nantinya siswa dibekali pembelajaran sesuai dengan jurusan yang telah dipilih sebagai bekal memasuki dunia kerja. Maka dari itu pihak sekolah perlu mengetahui pola minat siswa yang sesuai jurusan yang tersedia untuk menghindari kesalahan dalam pemilihan jurusan. Pola tersebut juga nantinya memudahkan pihak sekolah untuk menentukan jurusan mana yang tepat untuk calon siswa tersebut. Seiring dengan perkembangan teknologi, penggalian pola dapat dilakukan dengan teknik klasifikasi data yang didapat dari proses *data mining*.

Dari hasil wawancara awal yang dilakukan terhadap Bapak Sutrisno, S.Pd yang menjabat sebagai Wakil Kepala Bidang Kurikulum SMAN 1 Natar. Diketahui bahwa saat ini SMAN 1 Natar memiliki kesulitan untuk proses peminatan jurusan karena masih menggunakan sistem zonasi dimana sistem tersebut sistem yang menerima calon siswa berdasarkan jarak terdekat dari sekolah, dimana jumlah pendaftar mencapai jumlah 900 sedangkan yang diterima hanya berjumlah 400. Dimana kapasitas kelas untuk jurusan "IPA" hanya berjumlah 8 kelas sedangkan "IPS" berjumlah 4 kelas, masing-masing kelas berjumlah 36 siswa. Dengan penerapan metode C4.5 untuk pemilihan minat calon siswa SMA ini diharapkan bisa membantu bagian kurikulum dalam mengelompokkan minat calon siswa SMA. Berdasarkan pemaparan singkat tentang kelebihan yang sudah peneliti jelaskan diatas, maka peneliti akan membuat sebuah analisis yang diharapkan dapat

menampilkan hasil data mining dengan Algoritma *Decision Tree C4.5* dalam proses peminatan jurusan siswa berdasarkan kurikulum pada SMAN 1 Natar. Untuk menentukan jurusan yang akan diambil oleh siswa sesuai dengan latar belakang, minat dan kemampuannya sendiri *Decision Tree C4.5* diharapkan dapat menyumbang saran agar dapat dilakukan secara objektif dan terstruktur dengan adanya penentuan serta berbagai pertimbangan dan analisis berdasarkan kriteria yang sesuai dengan alternatif yang ada [1].

Dengan demikian peluang untuk sukses dalam studi di masa depan semakin besar. Penggunaan data mining dapat menjadi salah satu cara untuk mengolah data penerimaan siswa sebagai bahan prediksi bagi siswa yang akan diterima dengan tambahan label jurusan yang sesuai, guna menemukan pola pengetahuan sebagai bahan prediksi pada tahun 2020, dimana pengambilan keputusan dapat memberikan kebijakan terbaik dalam penerimaan siswa dengan melihat keterkaitan dari setiap pola yang akan dihasilkan guna mendapatkan hasil siswa yang akan berhasil dalam proses kegiatan belajar mengajar (KBM). Sehingga penggunaan data mining dirasa cocok untuk mengolah data penerimaan siswa untuk memperoleh informasi dan pola pengetahuan yang berkualitas. Beberapa teknik dalam data mining antara lain prediksi, klasifikasi, *clustering*, dan asosiasi.

Oleh karena itu, peneliti memilih teknik yang sesuai untuk penyelesaian masalah yang terjadi dengan menggunakan teknik prediksi di jadikan sebagai solusi karena dengan teknik ini proses pengestimasi nilai prediksi dilakukan berdasarkan pola-pola di dalam sekumpulan data. Prediksi menggunakan beberapa variabel untuk memprediksi nilai-nilai variabel masa mendatang yang diperlukan, yang belum diketahui saat ini. Tujuan dari teknik ini berkaitan dengan pembuatan sebuah model yang dapat melakukan pemetaan dari setiap himpunan variabel ke setiap targetnya untuk memodelkan prediksi suatu nilai yang mempunyai ciri-ciri tertentu. Kemudian menggunakan model tersebut untuk memberikan nilai target pada himpunan baru yang didapat. Beberapa metode yang dapat digunakan dalam melakukan analisis dengan teknik prediksi yaitu metode *K-nearest Neighbors*, *Naive Bayes*, *Decision Tree*, dan *Support Vector Machine* [2].

Dari algoritma yang ada maka peneliti memutuskan untuk melakukan analisis algoritma *Decision Tree* dalam hal ini yaitu metode yang digunakan adalah C4.5 karena dalam penelitian yang telah dilakukan sebelumnya dengan menggunakan komparasi algoritma klasifikasi *data mining C4.5* dan *Naive Bayes* kemudian dilakukan optimasi berbasis *Particle Swarm Optimization* menghasilkan hasil pengujian nilai akurasi algoritma C4.5 sebesar 85.40% dan nilai akurasi algoritma *Naive Bayes* sebesar 85,09%. Dari kedua algoritma tersebut kemudian

dilakukan kombinasi dengan optimasi *Particle Swarm Optimization*, dengan hasil algoritma C4.5+PSO memiliki nilai tertinggi berdasarkan nilai accuracy sebesar 87.61%, AUC sebesar 0.860 dan precision sebesar 88.96% (Rifai and Aulianita, 2018). Pada penelitian ini penulis mengharapkan untuk mendapatkan hasil prediksi terbaik untuk pemilihan siswa yang akan diterima dengan label jurusan yang sesuai dengan minat siswa.

2. Tinjauan Pustaka

2.1. Data Mining

Pada tahun 1990-an istilah Data mining mulai populer dikomunitas pengguna basisdata, akan tetapi sebenarnya perkembangan awal data mining berawal pada tahun 1763 ketika Thomas Bayes mempublikasikan Teorema Bayes [3]. Data Mining didefinisikan sebagai proses pengekstraksian informasi dari sekumpulan data yang sangat besar melalui penggunaan dapat didefinisikan sebagai algoritma dan teknik penarikan dalam bidang statistik, pembelajaran mesin dan sistem manajemen basis data [4]. Sebagai suatu rangkaian proses, data Mining dapat dibagi menjadi beberapa tahap proses. Tahap-tahap tersebut bersifat interaktif, pemakai terlibat langsung atau dengan perantaraan knowledge base [5].

2.2. Decision Tree

Decision Tree atau pohon keputusan adalah sebuah metode klasifikasi yang menggunakan struktur pohon, dimana setiap node mewakili atribut dan cabangnya mewakili nilai atribut, sedangkan daun adalah digunakan untuk mewakili kelas [6]. Pohon yang digunakan sebagai prosedur penalaran untuk menda-patkan jawaban dari masalah yang dimasukkan. Pohon yang dibentuk tidak selalu berupa pohon biner. Jika semua fitur dalam set menggunakan 2 macam nilai kategorikal maka bentuk pohon yang didapatkan berupa pohon biner. Jika dalam fitur berisi lebih dari 2 macam nilai kategorikal atau menggunakan tipe numerik maka bentuk pohon yang didapatkan biasanya tidak berupa pohon biner. Setiap simpul internal merepresentasikan variabel, dan simpul daun merepresentasikan suatu kelas [7].

Kefleksibelan membuat metode ini atraktif, khususnya karena memberikan keuntungan berupa visualisasi saran (saran bentuk Decision Tree) yang membuat procedure prediksinya dapat diamati (Gorunescu, 2011). Beberapa peneliti telah mempresentasikan metode pohon keputusan banyak digunakan untuk menyelesaikan kasus penentuan keputusan seperti di bidang kedokteran (diagnosis penyakit pasien), ilmu komputer (struktur data), psikologi (teori pengambilan keputusan), dan sebagainya [8].

2.3. Algoritma C4.5

Algoritma C4.5 adalah rangkaian algoritma untuk masalah klasifikasi dalam pembelajaran mesin dan penggalian data [9]. Algoritma C4.5 merupakan pengembangan dari algoritma ID3. Algoritma C4.5 dan

ID3 diciptakan oleh seorang peneliti dibidang kecerdasan buatan bernama J. Rose Quinlan pada akhir tahun 1970-an. Algoritma C4.5 membuat pohon keputusan dari atas ke bawah, dimana atribut paling atas merupakan akar, dan paling bawah dinamakan daun (Rahayu, 2014). Rumus algoritma C4.5 berupa pohon keputusan untuk mengetahui nilai entropi adalah sebagai berikut:

$$(-) = \sum - * \log 2$$

Dengan :

- S = Tetapkan Kasus
- J = Fitur
- N = Jumlah Partisi S
- Pi = Proposi Si Ke S

$$(S. A) = (S) - \sum = 1 || * (S)$$

Dengan :

- S = Tetapkan Kasus
- J = Atribut
- N = Jumlah Partisi Atribut A
- [Si] = Jumlah Kasus Di Partisi Ke - I
- |S| = Jumlah Kasus

Secara umum algoritma C4.5 untuk membangun pohon keputusan adalah:

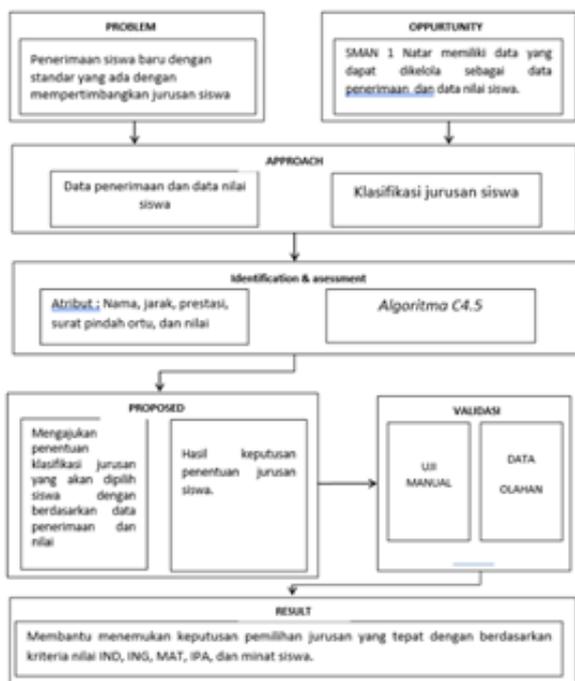
1. Pemilihan atribut sebagai akar
2. membuat cabang untuk setiap nilai
3. Pembentukan untuk kasus cabang
4. Ulangi proses masing-masing cabang sampai kasus berlanjut sehingga cabang memiliki kelas yang sama (Prasetyo, 2014).

2.4. K-Fold Cross Validation

Cross Validation adalah teknik validasi dengan membagi data secara acak kedalam k bagian dan masing-masing bagian akan dilakukan proses klasifikasi [10]. Teknik ini dapat memperkirakan keakurasian prediksi dari sebuah model ketika dilakukan penerapan dalam proses prediksi. Salah satu teknik dari validasi silang adalah *k-fold cross validation*, yaitu memecah data menjadi dua bagian dengan ukuran yang sama. Metode ini digunakan pada proses pengujian untuk mengukur performa dari metode klasifikasi yang digunakan [11].

3. Metode Penelitian

Metode penelitian pada penelitian ini yaitu metode penelitian kuantitatif. Penelitian kuantitatif merupakan penelitian yang datanya berbentuk angka dan dianalisis dengan menggunakan teknik statistic. Kerangka penenilitan pada dasarnya adalah kerangka hupromosin antara konsep konsep yang ingin diamati atau diukur melalui penelitian yang akan dilakukan [12] seperti pada gambar 1.



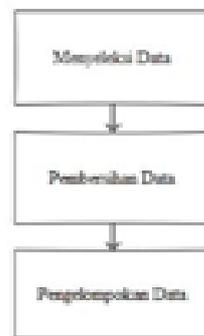
Gambar 1. Kerangka Penelitian

1. *Problem* (Permasalahan)
Permasalahan yang terjadi yaitu dalam penerimaan siswa baru pihak sekolah selain mempertimbangkan jalur zonasi, prestasi, surat tugas orangtua (bagi yang siswa pindahan) juga melihat klasifikasi jurusan yang tepat bagi siswa yang akan diterima.
2. *Opportunity* (Peluang)
Adanya data penerimaan siswa dan data nilai siswa yang dapat digunakan untuk menentukan pola penerimaan siswa dengan keputusan jurusan pada tahun berikutnya dengan menggunakan algoritma C4.5.
3. *Approach* (Pendekatan)
Tahapan pendekatan dalam penelitian ini merupakan bagaimana cara peneliti melakukan pendekatan melalui masalah yang ada sehingga menemukan solusi dalam penelitian ini melalui penggunaan teknik data mining dengan menggunakan metode C4.5.
4. *Identification dan Assesment* (Identifikasi dan Penugasan)
Dalam tahapan identifikasi dan penugasan penelitian adalah hal yang berkaitan dengan atribut yang digunakan di dalam penelitian ini, sehingga dengan adanya penelitian ini dapat menghasilkan informasi yang sesuai dengan tujuan yang diharapkan yaitu menghasilkan informasi penentuan siswa yang akan diterima dengan tambahan sudah mendapatkan klasifikasi jurusan yang tepat.

4. Hasil dan Pembahasan

4.1. Preprocessing Data

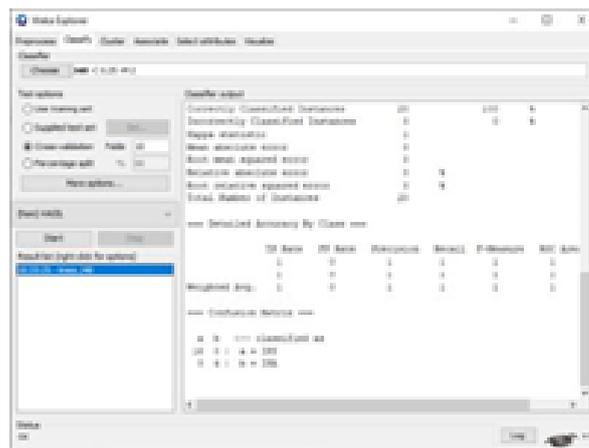
Tahap praproses umumnya ditujukan untuk membersihkan data sehingga pada proses pengolahan berikutnya lebih terstruktur [13]. Pada tahapan ini data set akan diproses sebelum dilakukan analisis dan proses *training data* [14], berikut ini adalah tahapan *preprocessing* :



Gambar 2. Preprocessing Data

4.2. Implementasi C4.5 pada Weka

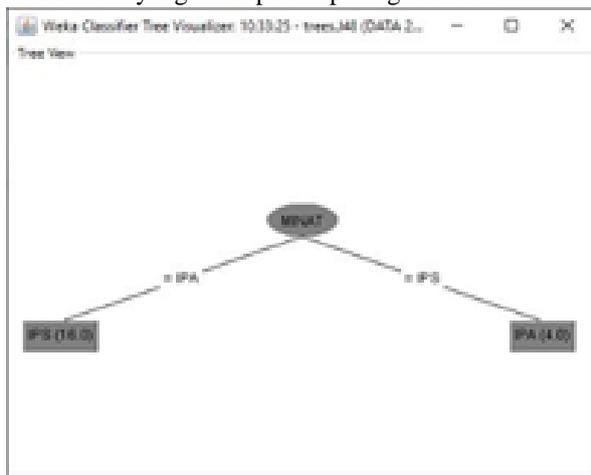
Algoritma C4.5 diimplementasikan pada weka dengan nama J48 yang merupakan variasi dari C4.5 yang telah mengalami proses optimasi. Berikut hasil hasil penggunaan algoritma C4.5 untuk klasifikasi data siswa.



Gambar 3. Output dari klasifikasi J4.8 (a)

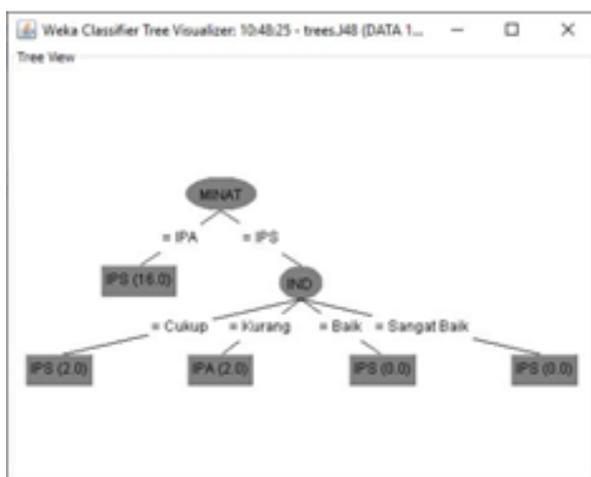
Hasil dari validasi dengan penggunaan algoritma C4.5 ditampilkan pada bagian Summary. Dimana dari hasil pengujian tersebut terdapat 0 data atau 0 % data terklasifikasi salah dan 20 data atau 100% terklasifikasi secara benar. Sedangkan pengujian dari Confusion Matrix yang diperuntukan untuk membagi informasi training serta informasi [15] terdapat 0 data dengan label IPS yang salah diklasifikasikan sebagai IPA dan 0 data dengan label IPA yang salah diklasifikasikan sebagai IPS. Untuk bentuk dari decision tree ditampilkan pada *Classifier model*. Gambar 4.5 menampilkan visualisasi decision tree

dengan percabangan pertama dari tree menggunakan atribut Minat. Berikut hasil visualisasi dari pembentukan decision tree yang ditampilkan pada gambar 4.



Gambar 4. Visualisasi Decision Tree (a)

Sedangkan untuk hasil visualisasi dari data b disajikan pada Gambar 5.



Gambar 6. Visualisasi Decision Tree (b)

Dari hasil pembahasan sebelumnya, maka pihak sekolah dapat mengetahui pola nilai dan minat untuk pemilihan jurusan bagi calon siswa pendaftar, sehingga pihak sekolah dapat memberikan arahan yang tepat bagi siswa dalam pemilihan jurusan. Dari hasil perhitungan melalui perhitungan sistem dan manual didapatkan hasil untuk data a memiliki tingkat akurasi sebesar 100% dan untuk data b sebesar 80%, yang mengartikan bahwa C4.5 dapat disarankan untuk mengolah data siswa dalam hal membantu memberikan keputusan terbaik pemilihan jurusan siswa.

5. Kesimpulan dan Saran

5.1. Kesimpulan

Kesimpulan dari penelitian ini adalah :

1. Dari hasil pembahasan sebelumnya, maka pihak sekolah dapat mengetahui pola nilai dan minat untuk pemilihan jurusan bagi calon siswa pendaftar, sehingga pihak sekolah dapat memberikan arahan yang tepat bagi siswa dalam pemilihan jurusan. Dari hasil perhitungan melalui perhitungan sistem dan manual didapatkan hasil untuk data a memiliki tingkat akurasi sebesar 100% dan untuk data b sebesar 80%, yang mengartikan bahwa C4.5 dapat disarankan untuk mengolah data siswa dalam hal membantu memberikan keputusan terbaik pemilihan jurusan siswa.
2. Berdasarkan hasil pembahasan sebelumnya, dapat disimpulkan bahwa pihak sekolah (wakil bidang kurikulum) dapat memberikan keputusan penerimaan siswa dengan tambahan label jurusan siswa sesuai dengan minat siswa.

5.2. Saran

Pada penelitian ini, peneliti hanya menggunakan satu metode prediksi, sehingga diharapkan pada penelitian selanjutnya dapat melakukan komparasi dengan metode prediksi yang lain untuk pengujian menggunakan data penerimaan siswa.

Daftar Pustaka

- [1] J. F. B. Logo, A. Wantoro Dan E. R. Susanto, "Model Berbasis Fuzzy Dengan Fis Tsukamoto Untuk Penentuan Besaran Gaji Karyawan Pada Perusahaan Swasta," *Jurnal Teknoinfo*, Vol. 14, No. 2, Pp. 125-130, 2020.
- [2] H. Sulistiani Dan Y. T. Utami, "Penerapan Algoritma Klasifikasi Sebagai Pendukung Keputusan Pemberian Beasiswa Mahasiswa," Dalam *Snti Vi*, Jakarta, 2018.
- [3] A. Wibisono, S. D. Rizkiono Dan A. Wantoro, "Filtering Spam Email Menggunakan Metode Naive Bayes," *Telefortech : Journal Of Telematics And Information Technology*, Vol. 1, No. 1, Pp. 9-17, 2020.
- [4] M. Bakri, "Penerapan Data Mining Untuk Clustering Kualitas Batubara Dalam Proses Pembakaran Di Pltu Sebalang Menggunakan Metode K-Means," *Jurnal Teknoinfo*, Vol. 11, No. 1, Pp. 1-4, 2017.
- [5] R. I. Borman Dan M. Wati, "Penerapan Data Mining Dalam Klasifikasi Data Anggota Kopdit Sejahtera Bandar Lampung Dengan Algoritma Naive Bayes," *Jurnal Ilmiah Fakultas Ilmu Komputer*, Vol. 9, No. 1, Pp. 25-34, 2020.

- [6] A. A. Aldino Dan H. Sulistiani, "Decision Tree C4.5 Algorithm For Tuition Aid Grant Program Classification (Case Study: Department Of Information System, Universitas Teknokrat Indonesia)," *Jurnal Ilmiah Edutic*, Vol. 7, No. 1, Pp. 40-50, 2020.
- [7] A. Nurkholis, M. Muhaqiqin Dan T. Susanto, "Analisis Kesesuaian Lahan Padi Gogo Berbasis Sifat Tanah Dan Cuaca Menggunakan Id3 Spasial," *Juita: Jurnal Informatika*, Vol. 8, No. 2, P. 235 – 244, 2020.
- [8] H. Sulistiani, K. Muludi Dan A. Syarif, "Implementation Of Dynamic Mutual Information And Support Vector Machine For Customer Loyalty Classification," Dalam *Journal Of Physics: Conference Series*, United Kingdom, 2018.
- [9] Kisworo, "Alternative Model Base As An Enabler For Success Of Business Intelligence-Based Companies Using C4.5," *Journal Of Theoretical And Applied Information Technology* , Vol. 95, No. 14, Pp. 3210-3216, 2017.
- [10] I. Ahmad, H. Sulistiyani Dan H. Saputra, "Using Fuzzy K-Nearest Neighbor For Predicting University Students Graduation In Teknokrat," *Indonesian Journal Of Artificial Intelligence And Data Mining (Ijaidm)* , Vol. 1, No. 1, Pp. 47-52, 2018.
- [11] D. Alita Dan A. Rahman, "Pendeteksian Sarkasme Pada Proses Analisis Sentimen Menggunakan Random Forest Classifier," *Jurnal Komputasi*, Vol. 8, No. 2, Pp. 50-58, 2020.
- [12] D. Alita, I. Sari, A. R. Isnain Dan Styawati, "Penerapan Naïve Bayes Classifier Untuk Pendukung Keputusan Penerima Beasiswa," *Jurnal Data Mining Dan Sistem Informasi*, Vol. 2, No. 1, Pp. 17-23, 2021.
- [13] M. A. Assuja Dan Saniati, "Analisis Sentimen Tweet menggunakan Backpropagation Neural Network," *Jurnal Teknoinfo*, Vol. 10, No. 2, Pp. 23-28, 2016.
- [14] D. Alita Dan Y. Fernando, "Multiclass Svm Algorithm For Sarcasm Text In Twitter," *Jurnal Teknik Informatika Dan Sistem Informasi*, Vol. 8, No. 1, Pp. 118-128, 2021.
- [15] D. Darwis, E. S. Pratiwi Dan A. F. O. Pasaribu, "Penerapan Algoritma Svm Untuk Analisis Sentimen Pada Data Twitter Komisi Pemberantasan Korupsi Republik Indonesia," *Jurnal Ilmiah Edutic*, Vol. 7, No. 1, Pp. 1-11, 2020.

