

ANALISIS KINERJA DALAM MENDETEKSI STUDENT LOSES BERDASARKAN NILAI GAIN DENGAN *SPLITE FEATURE REDUCTION MODEL* PADA ALGORITMA C4,5

Mohammad Yusup^{*1}, Erwin S Panjaitan², Roni Yunis³

STMIK Mikroskil

Medan, Indonesia, 20212

¹luckyman513@gmail.com, ²erwin@mikroskil.ac.id, ³roni@mikroskil.ac.id

Abstrak— Perguruan tinggi diharapkan mampu menghasilkan sumber daya manusia yang berpengetahuan juga kreatif. faktor penilaian perguruan tinggi adalah persentase kemampuan mahasiswa dalam menyelesaikan studi. Saat ini, masalah *student loses* dan faktor-faktor yang mempengaruhinya telah menjadi topik yang menarik untuk diteliti. Institusi perlu mendeteksi kinerja mahasiswa yang berstatus "*student loses*" sehingga dapat diidentifikasi. Dengan metode klasifikasi algoritma C4.5 dalam mengambil keputusan yang dianggap memiliki kinerja yang baik, tetapi algoritma C4.5 memiliki beberapa kelemahan terhadap data yang tidak seimbang, yaitu dataset yang berbeda dua kelas yang berbeda antara kelas mayoritas dan kelas minoritas, berpengaruh pada algoritma C4.5, yang menghasilkan nilai hasil akurasi kurang optimal pada algoritma C4,5 untuk mengurangi pengaruh ketidakseimbangan kelas pada klasifikasi C4.5, maka perlu dilakukan pendekatan metode *Split Feature Reduction Model* hal ini dilakukan untuk memberi bobot pada setiap fitur yang direkomendasikan, sehingga ditemukan fitur yang merupakan klasifikasi yang kuat, berdasarkan perbandingan pada setiap fitur agar dapat direkomendasikan pada klasifikasi C4.5 dalam proses klasifikasi Hasil menunjukkan bahwa, kinerja akurasi klasifikasi C4.5 pada dataset mahasiswa dengan pendekatan SFRM sebelum proses Pengujian, 10 fold cross-validation, menunjukkan tingkat akurasi klasifikasi yang lebih baik yaitu akurasi 98% dalam penanganan ketidakseimbangan kelas.

Kata Kunci— *Student loses* klasifikasi C4.5, *Split Feature Reduction Model*, *Fold Cross-Validation*.

Abstract— College is expected to produce knowledgeable and creative human resources. Factors for college grades are students' ability to complete their studies. At present, students who have problems losing and the factors that influence it have become interesting topics to be fixed. Students with the status of "students loses" in order to support. With the classification method C4.5 algorithm in making decisions that are considered to have a good rating, but C4.5 algorithm analysis has a level of weakness against unbalanced data, namely different datasets two different classes between the corresponding class and a larger class, which examined in C4.5 algorithm, which results in less optimal test results on the C4.5 algorithm to reduce class imbalance in the C4.5 classification, it is necessary to use the *Split Feature Reduction Model* method this is done to give weight to each feature sought, can be found a feature which is a strong classification, based on ranking of each feature so that it can be recommended in the C4.5 classification in the classification process. shows better classification accuracy completed 98% in handling class imbalances.

Keywords— *Student loses* classification C4.5, *Split Feature Reduction Model*, *Fold Cross-Validation*.

I. PENDAHULUAN

Perguruan tinggi merupakan suatu instansi pengelola pendidikan akademik tingkat lanjut bagi mahasiswa. Lima lembaga perguruan tinggi diantaranya adalah universitas, institut, sekolah tinggi, akademi dan politeknik. Pusat Statistik Pendidikan Badan Penelitian dan Pengembangan Departemen Pendidikan Nasional Republik Indonesia memiliki data yang menyebutkan bahwa jumlah lembaga penyelenggara perguruan tinggi mengalami peningkatan setiap tahunnya. Pada tahun 2018 tercatat 3171 perguruan tinggi diselenggarakan di Indonesia.

Perguruan tinggi diharapkan mampu menghasilkan intelektual muda yang berkualitas dan menghasilkan sumber daya manusia yang berpengetahuan juga kreatif, semakin meningkat jumlah sumber daya manusia yang berkualitas yang menjadi *outcome* bagi Institusi perguruan tinggi. Salah satu faktor penentu penilaian kualitas perguruan tinggi adalah persentase kemampuan mahasiswa untuk menyelesaikan studi tepat waktu. Berdasarkan matriks penilaian instrument akreditasi program studi (BAN-PT) [1]. bahwa unsur penilaian akreditasi Perguruan Tinggi, data yang didapat dari Pusat Statistik Pendidikan Badan Penelitian dan Pengembangan Departemen Pendidikan

Nasional Republik Indonesia [2] pada tahun akademik 2014/2015 sampai dengan 2017/2018 menunjukkan bahwa perguruan tinggi swasta di Indonesia menerima rata-rata sebanyak 966.587 mahasiswa baru dan meluluskan rata-rata 691.124 mahasiswa setiap tahunnya atau hanya mencapai 72% saja dari data tersebut diketahui bahwa 28% mahasiswa tidak diketahui keberadaannya. status tersebut bisa jadi karena mahasiswa menempuh studi tidak tepat waktu, memiliki status tidak aktif, atau bahkan *student loses*. Universitas Pembangunan Pancabudi merupakan salah satu perguruan tinggi swasta di Sumatera Utara yang memiliki 19.881 mahasiswa. Pada tahun ajaran 2014/2015 sampai dengan data 2018/2019 meluluskan 17%, terdapat 21 % dengan status tidak aktif. Tingginya persentase mahasiswa yang memiliki status tersebut menyebabkan tingginya persentase mahasiswa lulus tidak tepat waktu. Hal ini menjadi sangat penting bagi Institusi mengingat persentase mahasiswa lulus tepat waktu adalah merupakan unsur penting penilaian akreditasi yang ditetapkan oleh Badan Akreditasi Nasional. Dan Institusi perlu melakukan tindakan untuk mengetahui faktor-faktor penyebab mahasiswa memiliki status *student loses*. Institusi perlu melakukan prediksi perilaku terhadap mahasiswa dalam mencegah kegagalan akademik secara dini untuk mencegah *student loses*. Database institusi perguruan tinggi menyimpan data akademik, administrasi biodata yang terkait dengan mahasiswa dan Data tersebut apabila digali dengan tepat maka dapat diketahui pola atau *knowledge* untuk mengambil keputusan [3]. Serangkaian proses mendapatkan *knowledge* atau pola dari data base disebut dengan data mining (Witten, Ian 2016). Data mining dapat membantu mengatasi masalah dengan menganalisis data yang telah ada dalam *database*. Diantaranya, Penelitian yang dilakukan oleh Nagy, Marcell [4] , melakukan perbandingan untuk memprediksi mahasiswa *student losses* dengan jumlah 15.825 data akademik mahasiswa dari Universitas Teknologi dan Ekonomi Budapest terdaftar antara 2010 sampai dengan 2017 dengan beberapa algoritma diuji dengan 10-fold cross-validation dan penelitian Kelly J. de O. Santos [5] di Federal University of Sergipe (UFS) Brazil. Mahasiswa gagal pada semester 4. Prodi *System Information* dan *Computer Science* dengan menggunakan 23,690 data set mahasiswa.

II. LITERATURE REVIEW

DATA MINING

Penambangan data menghasilkan pola yang dapat ditransformasi dalam bentuk atau format matematika. Meskipun dimungkinkan untuk mengartikan makna dari permukaan ruang vektor, pola yang diperoleh dari objek yang lebih kompleks mungkin tidak mudah ditafsirkan bahkan para ahli itu sendiri. Penambangan, data semakin rumit, dan juga pola yang diperoleh jauh lebih kompleks [6] Penambangan Data diartikan

sebagai proses untuk menemukan hubungan, pola, dan tren baru dengan menyaring sejumlah data, disimpan dalam penyimpanan, menggunakan teknik pengenalan model seperti teknik Matematika dan Statistik. [7] Penambangan data adalah proses kunci ketika suatu metode diterapkan untuk menemukan (*knowledge*) pengetahuan yang berharga dan tersembunyi dari database.

Algoritma C4,5

Algoritma C4.5 adalah salah satu algoritma klasifikasi yang berfungsi untuk membuat pohon keputusan. [8]. Algoritma C4.5 merupakan algoritma klasifikasi *desicion tree* termasuk paling banyak digunakan karena memiliki keunggulan dibandingkan algoritma lainnya, Kelebihan algoritma C4.5 mudah diimplementasikan, dengan akurasi yang dapat diterima, dan efisien dalam penanganan atribut tipe diskrit serta dapat menangani atribut tipe numerik [9].

Splice Feature Reduction Model

Teknik reduksi data yaitu memperoleh data volume kecil dari volume besar tetapi masih mempertahankan integritas data asli, Tujuan pengurangan fitur adalah untuk menghilangkan fitur yang tidak relevan [10]. Pemilihan fitur berdasarkan Gain Ratio menggunakan metode ranker. Ranker bekerja dengan menyortir fitur dari nilai gain tertinggi ke nilai gain terendah.

Cross Validation

Validasi silang adalah metode statistik yang dapat digunakan untuk mengevaluasi kinerja suatu model atau algoritma di mana data dipecah menjadi dua himpunan bagian yaitu data proses pembelajaran dan data validasi / evaluasi.

Confusion Matrix

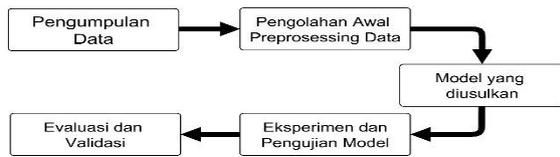
Evaluasi terhadap pengukuran kinerja model menggunakan *confusion matrix*, metode ini adalah alat yang berguna untuk menganalisis seberapa baik pengklasifikasian dapat mengidentifikasi fitur dari kelas yang berbeda [11]. *Confusion Matrix* dapat membantu menunjukkan detail kinerja klasifikasi dengan menyediakan sejumlah fitur kelas yang akurat dan tidak akurat. [12].

$$Accuracy = \frac{TP+TN}{(TP+FN)+(FP+TN)} \dots\dots\dots [2.1]$$

$$Precision = \frac{TP}{FP+FP} \dots\dots\dots [2.2]$$

$$Recall = \frac{TP}{TP+FN} \dots\dots\dots [2.3]$$

III. METODOLOGI



Gbr 1. Model Tahapan Penelitian

Langkah-langkah pemrosesan adalah sebagai berikut untuk mendapatkan hasil:

1. Pengumpulan data
2. *Preprocessing* data.
3. Penentuan model yaitu *Splite feature reduction model* dengan metode penentuan rangker berdasarkan nilai gain tertinggi.
4. Penerapan klasifikasi algoritma C4,5.
5. Analisis hasil dan evaluasi.

Pada proses akuisisi pengumpulan data mahasiswa pada database SIPT Universitas Pembangunan Pancabudi Dengan jumlah 3900 record, tahun ajaran 2017/2018, program studi manajemen dan ilmu komputer berupa data akademik dan data non- akademik. dan penelitian ini memiliki struktur kurikulum yang sama. dataset diperoleh dengan melakukan *quenbvry* data pada langkah *preprocessing* data, pertama-tama peneliti memeriksa elemen yang hilang, menerapkan analisis semantik, dan kemudian normalisasi. Untuk langkah klasifikasi,

IV. ANALISIS DAN HASIL

Dari proses perhitungan untuk mencari nilai gain rasio menghasilkan beberapa atribut untuk diurutkan berdasarkan nilai gain dan Pemilihan fitur menggunakan metode peringkat, pemilihan fitur dari nilai gain tertinggi ke nilai gain terendah, Setelah menerapkan Gain Ratio, penulis mendapatkan 20 fitur secara rangker. Pembagian fitur mengklasifikasikan 4 tingkat kelompok. split. Split 1 terdiri dari 8 fitur, Split 2 terdiri dari 12 fitur, Split 3 terdiri dari 16 fitur, dan Split 4 terdiri dari 20 fitur. Pada tahap *preprocessing*, fitur yang dipilih adalah fitur yang memiliki nilai gain di atas 0,01 Rasio perolehan adalah modifikasi dari perolehan informasi untuk mengurangi bias fitur yang memiliki banyak cabang. Berikut tabel pembagian fitur berdasarkan nilai gain.

TABEL I
PENENTUAN PERINGKAT BERDASARKAN NILAI GAIN RASIO TERTINGGI

NO	ATRIBUT	RANGKER	SPLI	SPLI	SPLI	SPLI
			T 4	T 3	T 2	T 1
1	SKS 2018 2	0.6034823				
2	STATUS AKADEMIK	0.4938867				
3	IPS 2018 1	0.4316499				
4	SKS 2017 2	0.4214887				
5	SKS 2018 1	0.3886703				
6	IPK	0.269066				
7	IPS 2017 2	0.2251757				
8	IPS 2017 1	0.1756792				
9	STATUS PERKAWINAN	0.0347005				
10	UMUR	0.026295				
11	JENISKELAMIN	0.0186973				
12	ASAL SEKOLAH	0.0088473				
13	SUMBER BIAYA	0.008818				
14	STATUS BEKERJA	0.0074187				
15	KOTA DOMISILI	0.0057709				
16	PRODI	0.0050674				
17	AGAMA	0.0038026				
18	JENIS PEKERJAAN ORTU	0.0028058				
19	GAJI ORTU	0.0024066				
20	STATUS DOMISILI	0.0019868				

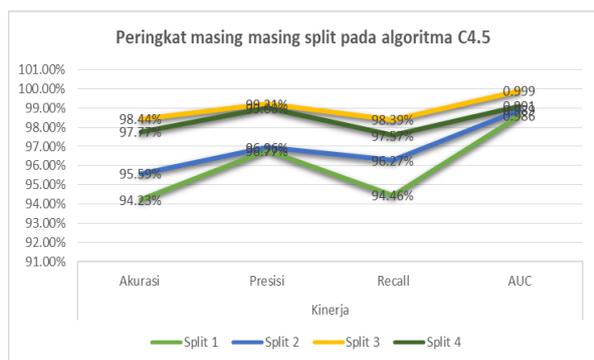
TABEL II
HASIL KINERJA DENGAN PROSES SFRM DENGAN KLASIFIKASI ALGORITMA C4.5

Validasi / Confidence	TP	FP	TN	FN	Akurasi	Presi	Recall	AUC (optimis tic)	AUC (pessim istic)
Split 1	2406	84	1269	141	94.23%	96.77%	94.46%	0.986	0.981
Split 2	2452	77	1276	95	95.59%	96.96%	96.27%	0.989	0.985
Split 3	2506	20	1333	41	98.44%	99.21%	98.39%	0.999	0.999
Split 4	2486	25	1327	62	97.77%	99.00%	97.57%	0.997	0.995

Pada split 3 proses pengukuran *SFRM* terdapat 3900 jumlah data pengukuran pada kelas tepat waktu (*Yes*) 2506 data yang diprediksi benar (TP) dan 20 data prediksi salah (FP). Kelas cenderung *student loses* (*No*). Diperoleh 1333 data diprediksi benar (TN). serta 41 data diprediksi tidak sesuai (FN) pada kelas *student loses* pada Split 3. Dengan nilai AUC 0,999.

Hasil Analisis Kinerja Klasifikasi C4.5 Dengan Confusion Matrix

Pada Tabel 2. diperoleh akurasi tertinggi pada Split 3 dengan 98,44%. Jadi, Split 3 dengan 18 fitur pilihan yaitu fitur yang paling berpengaruh terhadap klasifikasi untuk meningkatkan akurasi algoritma C4.5 dengan menerapkan model pengurangan fitur split. Berikut grafik peningkatan akurasi di setiap pemisahan pada masing-masing split,



Gbr 2. hasil akurasi dan peringkat masing masing split dengan algoritma C4.5

Dari gambar 2 menunjukkan model algoritma C.45 dengan penanganan (*SFRM*) lebih baik dalam penanganan ketidakseimbangan kelas pada dataset mahasiswa dengan tingkat diagnosa Nilai AUC adalah klasifikasi (*Excellent Classification*). Hal ini menunjukkan bahwa dengan menerapkan *SFRM* pada pengklasifikasi C4.5 dapat menangani permasalahan ketidakseimbangan kelas yang menjadi kelemahan pada algoritma C4,5 pada dataset mahasiswa, dan dari hasil pengukuran tersebut menunjukkan split 3 adalah klasifikasi nilai tertinggi dari split yang lainnya, hasil akurasi telah meningkat dibandingkan dengan split lainnya keakuratan Dataset mahasiswa dengan menggunakan algoritma C4.5 Hasil lebih baik pada Split 3 dengan 98,44%. Kemudian, peningkatan akurasi dengan menerapkan *SFRM* Hasil akurasi setiap pemisahan setelah diterapkan Algoritma C4.5.

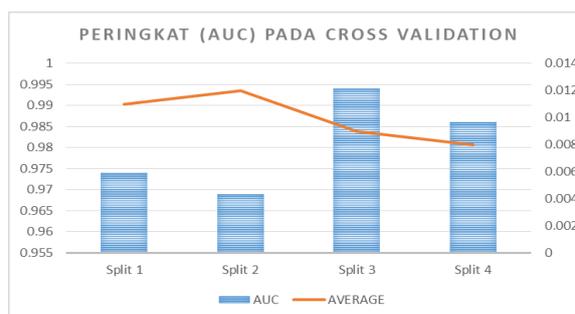
Hasil Kinerja Proses Pengujian Dengan Cross Validation

TABEL III
HASIL KINERJA PROSES PENGGUJIAN DENGAN CROSS VALIDATION DENGAN SFRM

Cros validati on dan C4,5	TP	FP	TN	FN	Akurasi / micro average	Presisi / micro average	Recall / micro average	AUC (optimistic) / micro average	AUC / micro average	AUC (pessimistic) / micro average
split 1	2378	99	1254	169	92.95% +/- 1.08%	95.47% +/- 1.42%	93.68% +/- 1.52%	0.974 +/- 0.012	0.966 +/- 0.014	0.959 +/- 0.016
split 2	2385	123	1230	162	92.69% +/- 1.19%	95.11% +/- 1.29%	93.64% +/- 1.44%	0.969 +/- 0.010	0.957 +/- 0.012	0.949 +/- 0.014
split 3	2466	55	1298	81	96.51% +/- 1.04%	97.83% +/- 1.08%	96.82% +/- 1.30%	0.994 +/- 0.003	0.973 +/- 0.011	0.968 +/- 0.013
split 4	2454	70	1282	94	95.79% +/- 0.98%	97.23% +/- 0.80%	96.31% +/- 1.16%	0.986 +/- 0.008	0.958 +/- 0.019	0.956 +/- 0.019

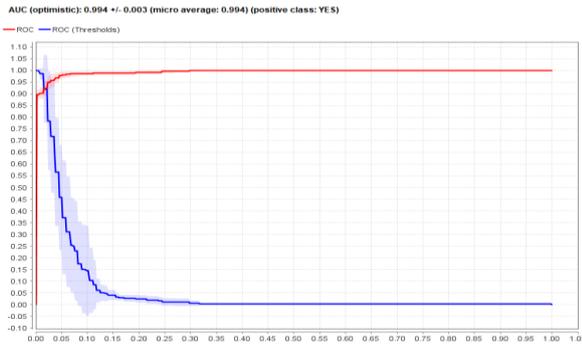
Pada split1 proses pengujian *SFRM cross validation* terdapat 3900 jumlah data, terdapat kelas tepat waktu (*Yes*) 2378 data yang diprediksi benar (TP) dan 99 data prediksi salah (FP). Kelas cenderung *student loses* (*No*). Diperoleh 1254 data diprediksi negatif (TN). serta 169 data diprediksi tidak sesuai (FN) pada kelas *student loses*. Dengan nilai AUC 0.974 +/- 0.012 *micro average*. Pada split2 proses pengujian *SFRM cross validation*

terdapat 3900 jumlah data, terdapat kelas tepat waktu (*Yes*) 2385 data yang diprediksi benar (TP) dan 123 data prediksi salah (FP). Kelas cenderung *student loses* (*No*). Diperoleh 1230 data diprediksi negatif (TN). serta 162 data diprediksi tidak sesuai (FN) pada kelas *student loses*. Dengan nilai AUC 92.69% +/- 1.19% *micro average*. Pada split 3 proses pengujian *SFRM cross validation* terdapat 3900 jumlah data, terdapat kelas tepat waktu (*Yes*) 2466 data yang diprediksi benar (TP) dan 55 data prediksi salah (FP). Kelas cenderung *student loses* (*No*). Diperoleh 1298 data diprediksi negatif (TN). serta 81 data diprediksi tidak sesuai (FN) pada kelas *student loses*. Dengan nilai AUC 0.994 +/- 0.003 *micro average*.



Gbr 3. Garafik peningkatan kinerja AUC pada pengujian cross validation menggunakan SFRM

Pada grafik diatas, AUC merupakan ukuran numerik untuk membedakan kinerja model dalam hasil pengukuran pada klasifikasi, nilai AUC yang tinggi menunjukkan kinerja yang lebih baik (Weiss, 2017). Dan seberapa sukses dalam memisahkan pengamatan positif dan negative (Branco, Paula, 2016) AUC menyediakan ukuran tunggal kinerja klasifikasi untuk mengevaluasi model mana yang lebih baik, pada gambar 3 dan dari hasil pengukuran tersebut menunjukkan split 3 adalah klasifikasi nilai tertinggi dari split yang lainnya, hasil nilai AUC paling tinggi dibandingkan dengan Split lainnya, dari pengujian *cross validation* model algoritma C4.5 dengan menerapkan seleksi fitur dengan *SFRM* pada Split 3 nilai AUC: 0.994 *micro average* 0.003 dengan Hasil terbaik diperoleh di Split 3 dengan 96,85%. setelah dilakukan pengujian *cross validation* pada Algoritma C4.5, Akan tetapi pada pengujian ini terjadi penurunan akurasi di setiap split sebesar 1,5%.



Gbr 4. ROC pada cross validation Split 3

Gambar 4. Menunjukkan kurva ROC kinerja model algoritma C4.5 dengan proses *SFRM* Split 3, pada pengujian 10-fold Cross validation maka nilai AUC (Area Under Curve) sebesar 0,994. *Micro average* 0.003 Maka tingkat akurasi didiagnosa sebagai klasifikasi sangat baik (*Excelance classification*), dan bias lebih mengecil ketika dengan proses *SFRM* dan Split 3 merupakan pengujian dengan nilai terbaik diantara split yang lainnya.

V. PENUTUP

Berdasarkan hasil penelitian yang dilakukan, dapat diambil kesimpulan sebagai berikut :

1. Penanganan terhadap Pengaruh ketidak seimbangan kelas Menggunakan Pendekatan *SFRM* dianggap mampu mengatasi permasalahan kelemahan klasifikasi algoritma C4,5. Untuk mengurangi pengaruh ketidakseimbangan kelas adalah Terdapat model prediksi *Student loses* yang memiliki kinerja lebih baik dari klasifikasi algoritma C4.5, sehingga kinerja klasifikasi algoritma C4.5 masih bisa ditingkatkan untuk memperbaiki model prediksi.
2. Menerapkan pendekatan *Splice Feature Reduction Model* dapat meningkatkan kinerja pengklasifikasi C4.5 dengan diagnosa klasifikasi sangat baik (*Excellence Classification*) dalam penanganan ketidakseimbangan kelas, selama proses validasi data training dan data testing dilakukan dengan metode 10-fold Cross-Validation dan nilai AUC sangat baik (*Excellence Classification*) pada Split3, dan terjadi penurunan akurasi sebesar 1,5% pada setiap split pada saat dilakukan pengujian dengan metode *10-fold Cross Validation*.
3. Perbandingan rating klasifikasi Split3 dengan klasifikasi menghasilkan akurasi tertinggi 98,44% dan nilai AUC 0,994% merupakan nilai klasifikasi paling tinggi di antara split 1, split 2, split 4.
4. Faktor rendahnya kemampuan akademik, merupakan faktor paling berpengaruh dan diikuti faktor asal sekolah, domisili saat menempuh studi dan faktor pembiayaan adalah faktor yang mempengaruhi *student loses*

REFERENSI

- [1]. BAN-PT. (2015). Akreditasi Institusi Perguruan Tinggi. In *Pedoman Penyusunan Borang*. ban-pt.depdiknas.go.id
- [2]. Dikti, P. D. dan I. I. (2017). statistik 2017. In *Statistik Pendidikan Tinggi Higher Education Statistical Year Book 2017* (Vol. 38, Issue 8).
- [3]. de O. Santos, K. J., Menezes, A. G., de Carvalho, A. B., & Montesco, C. A. E. (2019). Supervised Learning in the Context of Educational Data Mining to Avoid University Students Dropout. *2019 IEEE 19th International Conference on Advanced Learning Technologies (ICALT)*, 2161-377X, 207–208. <https://doi.org/10.1109/icalt.2019.00068>
- [4]. Nagy, M., & Molontay, R. (2018). Predicting Dropout in Higher Education Based on Secondary School Performance. *INES 2018 - IEEE 22nd International Conference on Intelligent Engineering Systems, Proceedings*, 000389–000394. <https://doi.org/10.1109/INES.2018.8523888>
- [5]. Tasnim, N., Paul, M. K., & Sattar, A. H. M. S. (2019). Identification of Drop Out Students Using Educational Data Mining. *2nd International Conference on Electrical, Computer and Communication Engineering, ECCE 2019*, 1–5. <https://doi.org/10.1109/ECACE.2019.8679385>
- [6]. Mayra, A., & Mauricio, D. (2018). Factors to predict dropout at the universities: A case of study in Ecuador. *IEEE Global Engineering Education Conference, EDUCON, 2018-April*, 1238–1242. <https://doi.org/10.1109/EDUCON.2018.8363371>
- [7]. Nugroho, M. F., & Wibowo, S. (2017). Fitur Seleksi Forward Selection Untuk Menentukan Atribut Yang Berpengaruh Pada Klasifikasi Kelulusan Mahasiswa Fakultas Ilmu Komputer UNAKI Semarang Menggunakan Algoritma Naive Bayes. *Jurnal Informatika Upgris*, 3(1), 63–70. <https://doi.org/10.26877/jiu.v3i1.1669>
- [8]. Hajrahnur, S., Nasrun, M., Setianingsih, C., & Murti, M. A. (2018). Classification of posts Twitter traffic jam the city of Jakarta using algorithm C4.5. *2018 International Conference on Signals and Systems, ICSigSys 2018 - Proceedings*, 294–300. <https://doi.org/10.1109/ICSIGSYS.2018.8372776>
- [9]. Liang, S., Ma, A., Yang, S., Wang, Y., & Ma, Q. (2018). A Review of Matched-pairs Feature Selection Methods for Gene Expression Data Analysis. *Computational and Structural Biotechnology Journal*, 16, 88–97. <https://doi.org/10.1016/j.csbj.2018.02.005>
- [10]. Wang, X., Zhou, C., & Xu, X. (2019). Application of C4.5 decision tree for scholarship evaluations. *Procedia Computer Science*, 151(2018), 179–184. <https://doi.org/10.1016/j.procs.2019.04.027>
- [11]. Sudarto, Zarlis, & Pahala, S. (2016). Integrasi Density Based Feature Selection dan Adaptive Boosting (Adaboost) dalam Menangani Ketidakseimbangan Kelas. *JSM (Jurnal SIFO Mikroskil)*, 17(2), 193–206.
- [12]. Muslim, M. A., Nurzahputra, A., & Prasetyo, B. (2018). Improving accuracy of C4.5 algorithm using split feature reduction model and bagging ensemble for credit card risk prediction. *2018 International Conference on Information and Communications Technology, ICOIACT 2018, 2018-Janua(1996)*, 141–145. <https://doi.org/10.1109/ICOIACT.2018.8350753>
- [13]. Jiang, W., Chen, Z., Xiang, Y., Shao, D., Ma, L., & Zhang, J. (2019). SSEM: A Novel Self-Adaptive Stacking Ensemble Model for Classification. *IEEE Access*, 7, 120337–120349. <https://doi.org/10.1109/access.2019.2933262>
- [14]. Gao, C., & Grzymala-Busse, J. W. (2019). Mining incomplete numerical data sets using C4.5 preceded by multiple scanning. *ICNC-FSKD 2018 - 14th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery*, 877–884. <https://doi.org/10.1109/FSKD.2018.8687148>
- [15]. Nugroho, H. W., Adji, T. B., & Setiawan, N. A. (2018). Performance Improvement of C4 . 5 Algorithm using Difference Values Nodes in Decision Tree. *2018 6th*

International Conference on Cyber and IT Service Management (CITSM), Citsm, 1–6.

<https://doi.org/10.1109/CITSM.2018.8674250>

- [16]. Dabab, M., Freiling, M., Rahman, N., & Sagalowicz, D. (2018). A decision model for data mining techniques. *PICMET 2018 - Portland International Conference on Management of Engineering and Technology: Managing Technological Entrepreneurship: The Engine for Economic Growth, Proceedings, 1–8.*
<https://doi.org/10.23919/PICMET.2018.8481953>