

Analisa Kepentingan atribut data pada Klasifikasi Heregistrasi Mahasiswa STMIK Widya Pratama

Dadang Aribowo, Ivandari

STMIK Widya Pratama Pekalongan

E-mail: dadang.stmik.wp@gmail.com, ivandarialkaromi@gmail.com

RINGKASAN

Mahasiswa merupakan aset yang paling berharga dalam sebuah perguruan tinggi swasta (PTS). Karena sebagian besar pendapatan serta biaya operasional PTS didapatkan dari mahasiswa. Banyaknya mahasiswa yang melakukan heregistrasi jelas akan menjadi angin segar bagi lembaga. Dalam 5 tahun terakhir tercatat sekitar 20% calon mahasiswa STMIK Widya Pratama tidak melakukan heregistrasi. Data terakhir pada 31 Agustus 2018 tercatat ada 32,7% pendaftar belum melakukan heregistrasi. Penurunan jumlah mahasiswa dapat mempengaruhi stabilitas keuangan lembaga utamanya sekolah tinggi swasta. Analisa terhadap algoritma terbaik untuk klasifikasi heregistrasi mahasiswa pernah dilakukan dan membuktikan bahwa decision tree C45 merupakan algoritma dengan tingkat akurasi terbaik. Pengetahuan dini terhadap calon mahasiswa yang mungkin tidak akan melakukan heregistrasi dapat menjadi acuan lembaga untuk melakukan tindakan guna mempertahankan mahasiswa. Pencatatan data mahasiswa yang tersusun rapi dapat digunakan pihak manajemen untuk melakukan analisa terhadap karakteristik serta penyebab mahasiswa tidak melakukan heregistrasi. Penelitian ini akan melakukan analisa terhadap semua data dan atribut data yang ada. Metode yang digunakan dalam pembobotan adalah information gain yang telah terbukti dapat mengatasi dataset dengan jenis atribut yang banyak. Hasil dari penelitian ini menyimpulkan bahwa atribut pekerjaan orang tua merupakan atribut dengan tingkat kepentingan tertinggi. Sedangkan atribut status sipil merupakan atribut dengan tingkat kepentingan paling rendah.

Kata Kunci : heregistrasi mahasiswa, information gain, penerimaan mahasiswa baru

1. PENDAHULUAN

1.1 Latar Belakang Masalah

Sekolah Tinggi Manajemen Informatika dan Komputer (STMIK) Widya Pratama adalah satu satunya perguruan tinggi bidang komputer di Kota Pekalongan. STMIK Widya Pratama didirikan pada 2002 berdasarkan SK Mendiknas Nomor: 149/D/O/2002. Pada 2004 dilakukan penggabungan antara AMIK dan STMIK yang didasarkan pada SK Mendiknas Nomor: 75/D/O/2004. Sebagai perguruan tinggi swasta STMIK membutuhkan banyak dana operasional, salah satu penyokong dana operasional terbesar adalah dari mahasiswa.

Banyaknya mahasiswa menjadi salah satu tolok ukur dari perguruan tinggi swasta pada umumnya.

Setiap tahun STMIK Widya Pratama membuka pendaftaran mahasiswa baru. Publikasi dan promosi telah dilakukan dengan mengunjungi SMA/SMK atau sederajat untuk memberikan paparan terkait fasilitas dan gambaran umum kuliah di STMIK. Nyatanya setiap tahun selalu ada selisih yang signifikan terkait jumlah pendaftar dengan jumlah mahasiswa yang melakukan heregistrasi. Tabel 1 merupakan data yang diperoleh dari panitia penerimaan mahasiswa baru STMIK Widya Pratama Pekalongan.

Tabel 1. Data Penerimaan Mahasiswa Baru 5 tahun terakhir

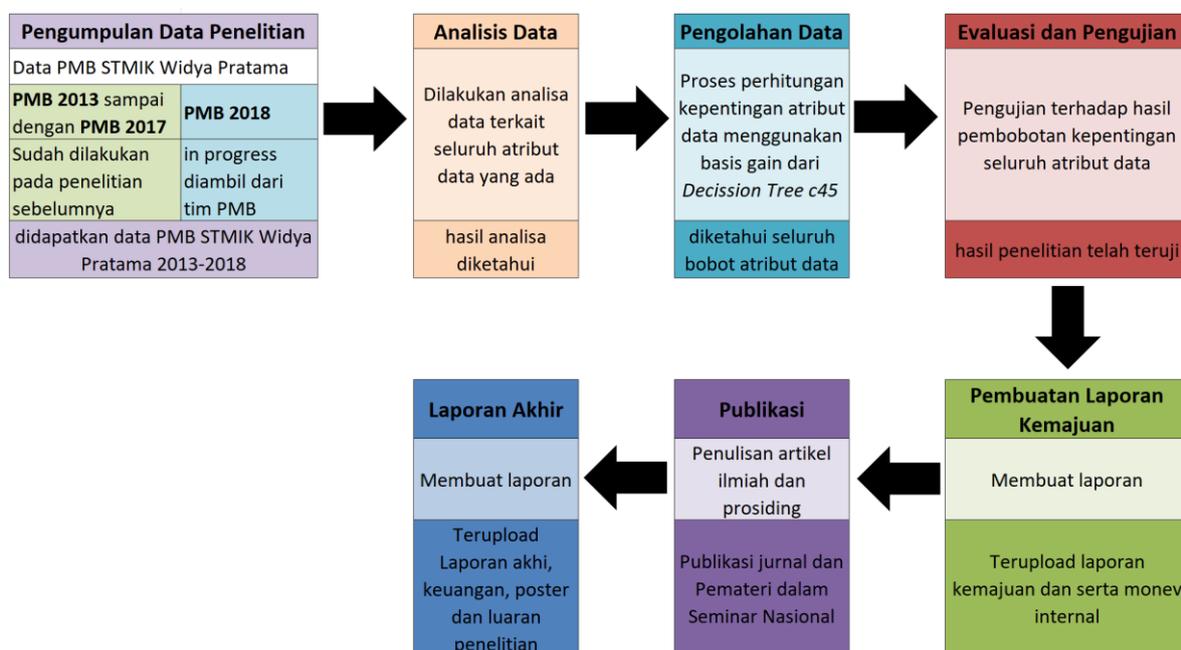
Tahun	Pendaftaran	Registrasi	Selisih	Prosentase	
				Registrasi	Tidak Registrasi
2013	705	513	192	73%	27%
2014	608	474	134	78%	22%

2015	667	500	167	75%	25%
2016	654	427	227	77%	23%
2017	506	428	78	84,6%	15,4%
2018	510	352	158	69%	31%

nurunan jumlah heregistrasi mahasiswa dapat mempengaruhi pendapatan utama lembaga utamanya perguruan tinggi swasta. Beberapa tindakan dilakukan untuk mengupayakan peningkatan jumlah mahasiswa baru. salah satunya dengan memberikan keringanan pembayaran ataupun memberikan beasiswa kepada calon mahasiswa berprestasi. Adanya selisih yang cukup signifikan antara jumlah pendaftar dengan jumlah mahasiswa heregistrasi menjadi satu masalah yang harus segera dipecahkan. Lembaga utamanya tim penerimaan mahasiswa baru harus berusaha lebih untuk meningkatkan jumlah mahasiswa atau minimal mempertahankan jumlah yang ada. Apabila kemungkinan calon mahasiswa yang tidak melakukan heregistrasi dapat diketahui lebih awal maka dapat dilakukan tindakan untuk mempertahankan calon mahasiswa tersebut (Kusrini & Taufiq, 2009).

Klasifikasi merupakan salah satu fungsi utama data mining (Witten, Frank, & Hall, 2011) Proses klasifikasi dapat dilakukan dengan berbagai

macam cara (Larose, 2005). Dalam proses klasifikasi dapat digunakan berbagai macam algoritma klasifikasi (Han & Kamber, 2006a). Dalam penelitian sebelumnya dilakukan komparasi antara beberapa algoritma untuk klasifikasi heregistrasi mahasiswa STMIK Widya Pratama. Hasil penelitian membuktikan bahwa algoritma C45 merupakan yang terbaik untuk klasifikasi heregistrasi calon mahasiswa. Dalam tahapan algoritma C45 digunakan perhitungan pembobotan untuk semua atribut dataset. Penelitian ini akan melakukan perhitungan untuk mengetahui tingkat kepentingan semua atribut data dalam dataset heregistrasi mahasiswa STMIK Widya Pratama. Hasil dari penelitian ini nantinya dapat digunakan sebagai analisa dalam penerimaan mahasiswa baru di STMIK. Harapan dari adanya analisa tersebut dapat dijadikan salah satu strategi untuk meningkatkan jumlah mahasiswa baru di STMIK Widya Pratama Pekalongan



Gambar 1 Kerangka pemikiran penelitian

2. METODE PENELITIAN

Penelitian ini dilakukan dengan melewati beberapa tahapan. Tahapan yang dilakukan adalah sebagaimana berikut:

2.1 Tahapan Pengumpulan Data

Pengumpulan data dilakukan dengan mengambil data dari panitia penerimaan mahasiswa baru STMIK Widya Pratama Pekalongan. Data PMB

dari tahun 2013 sampai dengan 2017 telah didapatkan dan digunakan sebagai bahan penelitian sebelumnya. Untuk data PMB tahun 2018 sampai saat ini masih berkembang dikarenakan pada 30 Agustus ini proses penerimaan mahasiswa baru masih berlangsung dan baru akan ditutup pada 21 September mendatang. Dari data yang telah diperoleh diketahui terdapat 43 atribut data dengan satu atribut tujuan yaitu registrasi. Data yang diperoleh telah berbentuk file dengan ekstensi xlsx yang merupakan hasil download dari aplikasi PMB STMIK Widya Pratama Pekalongan.

2.2 Tahapan Analisis Data

Tahapan analisa data dilakukan untuk mempersempit atribut data yang ada. Dari 43 atribut data yang ada terdapat atribut dengan varian yang sangat banyak misalnya nomor KTP. Setiap *record* yang ada memiliki nomor KTP yang berbeda. Selain itu ada atribut alamat dan nama dan nama orang tua dengan kondisi yang hamper sama. Beberapa atribut sebagaimana tersebut diatas nantinya tidak akan digunakan dalam proses berikutnya. Dalam proses analisis data ini juga akan dilakukan perbaikan beberapa isian data yang tidak relevan atau kesalahan pengisian oleh panitia saat menginput data.

2.3 Tahap Pengolahan Data

Tahapan pengolahan akan dilakukan dengan menggunakan semua data yang ada yaitu data PMB dari tahun 2013 sampai tahun 2018. Proses pengolahan merupakan proses inti dari penelitian ini. Proses ini menggunakan bantuan aplikasi rapid miner untuk pengolahan data. Selain itu digunakan perhitungan algoritma *information gain* untuk mengetahui tingkat kepentingan seluruh atribut yang ada. Hasil akhir dari tahapan ini adalah diketahui kepentingan seluruh atribut yang ada.

2.4 Tahap Evaluasi dan Pengujian

Tahapan pengujian dilakukan untuk melakukan evaluasi terhadap hasil penelitian yang sudah dilakukan. Dalam tahapan ini akan dilakukan menggunakan aplikasi rapid miner serta menggunakan Microsoft excel. Hasil dari pengujian ini akan digunakan sebagai acuan dalam pengambilan keputusan manajemen kampus.

3. HASIL DAN PEMBAHASAN

3.1 Hasil Eksperimen

Penelitian mengenai heregistrasi mahasiswa pernah ada sebelumnya dengan menggunakan dataset pendaftaran tahun 2013 (Alkaromi, 2014). Dalam penelitian tersebut digunakan algoritma *information gain* untuk pemilihan fitur yang ada. Pada tahun setelahnya juga dilakukan optimasi untuk parameter k dalam algoritma KNN (Al Karomi, 2015). Dalam penelitian ini digunakan pengembangan data sesuai dengan update data terkini yaitu penerimaan mahasiswa baru dari tahun 2013/2014 sampai dengan tahun 2018/2019.

Dari tabel 1 terlihat selama 6 tahun terakhir. Jumlah pendaftar di STMIK Widya Pratama adalah 3650 orang dengan jumlah mahasiswa yang melakukan heregistrasi sebanyak 2694. Artinya terdapat selisih 956 mahasiswa yang tidak melakukan heregistrasi. Jumlah selisih yang terlalu banyak tentunya dapat mempengaruhi pemasukan dari STMIK Widya Pratama selaku perguruan tinggi swasta. Dalam dataset mentah terdapat 60 atribut dan tidak semuanya dapat digunakan dalam klasifikasi. Atribut yang tidak dapat digunakan dalam proses klasifikasi antara lain atribut dengan varian yang unik atau semua isian nya berbeda. Atribut dengan varian unik misalnya NIK, Nama dan nomor handphone. Selain itu ada juga atribut dengan varian yang sama yang tidak pula dapat digunakan, misalnya atribut kewarganegaraan

Analisis Data

3.2 Analisis Data

Dataset yang diperoleh dari tim penerimaan mahasiswa baru STMIK Widya Pratama Pekalongan sejumlah 3650 record. Data ini merupakan gabungan data dari tahun 2013 sampai dengan tahun 2018. Dilakukan penggabungan data dari keseluruhan data sehingga didapatkan 45 atribut. Beberapa atribut dianggap tidak relevan karena memiliki varian yang berbeda atau dapat dikategorikan sebagai atribut unik sehingga proses berikutnya atribut ini tidak digunakan. Beberapa atribut unik yang dihilangkan antara lain, nama, alamat, nomor KTP, nomor hp dan beberapa atribut lain. Setelah dilakukan penyesuaian atribut didapatkan 19 atribut yang akan digunakan dalam proses selanjutnya. Metadata dari kesemua atribut yang akan digunakan dapat dilihat pada tabel 2 berikut.

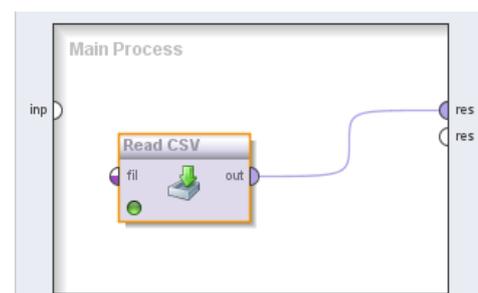
Tabel 2 Metadata dataset heregistrasi calon mahasiswa setelah dilakukan manual seleksi fitur

Role	Name	Type	Statistics	Range	Missings
id	no_d aftar	polyn omin al	mode = SIMBANG WETAN (2), least = 13- 010-0001 (1)	DESA LANGAK KEC BODEH PEMALANG (2), KARANGANYAR (2), PEKALONGAN (2), RT 27 RW 06 (2), SIMBANG WETAN (2), 13-010-0001 (1), 13-010-0006 (1), 13-010-0027 (1), 13-010-0035 (1), 13-010-0046 (1), 13-010-0067 (1), 13-010- 0083 (1), 13-010-0094 (1), 13-010-0095 (1), 13- 010-0126 (1), 13-010-0181 (1), 13-010-0190 (1), 13-010-0191 (1), 13-010-0194 (1), 13-010-0195 (1), 13-010-0212 (1), 13-010-0213 (1), 13-010- 0231 (1), 13-010-0232 (1), 13-010-0240 (1), 13- 010-0300 (1), ... and 2277 more ... , 3327125708000006 (1), 3327126712980003 (1), 332713100400001 (1), 3327135903990001 (1), 33326112904990001 (1), 336051505000001 (1), 3375011005970006 (1), 3375012701990004 (1), 337501270199004 (1), 3375022108860007 (1), 3375022701010005 (1), 3375025509980003 (1), 3375025612970003 (1), 3375025804000004 (1), 3375031302770003 (1), 337503411299004 (1), 3375035109940003 (1), 3375036812990004 (1), 3375043108990001 (1), 3375045507000007 (1), 3375045905000003 (1), 3603026901000002 (1), 52364284629561 (1), ;PEMALANG;-; ;PEMALANG;1996-05- 11;D3;0;REGULER;110;0;ISLAM;BELUM BEKERJA;LAKI-LAKI;WNI;TIDAK MENIKAH;SMA NEGERI 1 BODEH;2013;DN-03 MA 0058371;SUHARSO - DIYAHTININGSIH; (1), KEC. SRAGI (1), KEC. SRAGI (1)	0
label	regist rasi	bino minal	mode = Ya (1814), least = Tidak (503)	Tidak (503), Ya (1814)	15
regul ar	kota_ kec	polyn omin al	mode = PEKALONGAN (1173), least = BANYUMAS (1)	KAB PEKALONGAN (144), PEKALONGAN (1173), PEMALANG (254), BATANG (518), KAB PEMALANG (19), SEMARANG (2), YOGYAKARTA (2), BREBES (3), BANYUMAS (1), KAB BATANG (13), TEGAL (21), PALEMBANG (1), JAKARTA (2), SURABAYA (1), WIRADESA (6), DEMAK (2), KARANGANYAR (3), KEDUNGWUNI (6), DORO (1), KENDAL (2), BEKASI (1), PURBALINGGA (1), PADANG PANJANG (1), MANADO (1), WONOKERTO (1), REMBANG (1), MAGELANG (1), PATI (1), BANJARNEGARA (4), COMAL (2), AMBON (2), BANDUNG (1), PEKALONGAN KABUPATEN (86), PONTIANAK (1), BATANG KABUPATEN (14), ACEH (5), KLATEN (1), WONOGIRI (1), BOYOLALI (1), CURUP (1), BLORA (2), BOGOR (1), SEKADAU (1), PEMALANG KABUPATEN (8), TANGERANG (1), TRENGGALEK (1), TEGAL KABUPATEN (2), JAKARTA TIMUR (1), SUKOHARJO (1)	13
regul ar	Jenja ng	bino minal	mode = S1 (1897), least = D3 (422)	D3 (422), S1 (1897)	13
regul	Jenja	bino	mode = S1 (476), least =	D3 (288), S1 (476)	1568

ar	ng2	minal	D3 (288)		
regul ar	Kelas	polyn omin al	mode = REGULER (2181), least = KARYAWAN (10)	REGULER (2181), ALIH JENJANG (93), TRANSFER (35), KARYAWAN (10)	13
regul ar	prodi	polyn omin al	mode = 240 (1389), least = 120 (178)	110 (232), 240 (1389), 230 (520), 120 (178)	13
regul ar	prodi 2	polyn omin al	mode = 0 (1555), least = 240 (100)	120 (119), 110 (177), 240 (100), 230 (368), 0 (1555)	13
regul ar	agam a	polyn omin al	mode = ISLAM (2249), least = HINDU (2)	ISLAM (2249), KATHOLIK (37), KRISTEN (31), HINDU (2)	13
regul ar	st_ke rja	bino minal	mode = BELUM BEKERJA (1788), least = BEKERJA (461)	BELUM BEKERJA (1788), BEKERJA (461)	83
regul ar	jnkel	bino minal	mode = LAKI-LAKI (1511), least = PEREMPUAN (793)	LAKI-LAKI (1511), PEREMPUAN (793)	28
regul ar	stssip il	bino minal	mode = TIDAK MENIKAH (2167), least = MENIKAH (76)	TIDAK MENIKAH (2167), MENIKAH (76)	89
regul ar	krjort u	polyn omin al	mode = LAIN-LAIN (568), least = GURU (1)	LAIN-LAIN (568), SWASTA (355), WIRUSAHA (560), POLRI/TNI (39), PNS (474), BURUH (8), PETANI (2), 0 (279), TIDAK BEKERJA (11), LAINNYA (21), GURU (1)	14
regul ar	pend _ortu	polyn omin al	mode = SMA (655), least = D1 (2)	SD (549), SMA (655), SMP (336), TIDAK SEKOLAH (222), S1 (223), D3 (31), S2 (21), D1 (2), 0 (279)	14
regul ar	ST_TE ST	bino minal	mode = TEST (1450), least = BEBAS TEST (806)	TEST (1450), BEBAS TEST (806)	76
regul ar	PEKE RJAA N	polyn omin al	mode = BELUM BEKERJA (1181), least = POLRI/TNI (1)	BELUM BEKERJA (1181), SWASTA (132), PNS (531), WIRUSAHA (30), LAIN-LAIN (122), POLRI/TNI (1), LAINNYA (43), WIRASWASTA (3)	289
regul ar	KD_K ONSE NTRA SI	polyn omin al	mode = 24001 (759), least = 11002 (39)	11001 (192), 24001 (759), 23002 (196), 24002 (629), 12001 (180), 23001 (323), 11002 (39)	14
regul ar	KD_K ONSE NTRA SI2	polyn omin al	mode = 0 (1554), least = 24002 (35)	12001 (119), 11002 (39), 24001 (65), 11001 (138), 23001 (240), 23002 (128), 24002 (35), 0 (1554)	14
regul ar	KETL ULUS	bino minal	mode = lls (2317), least = tls (1)	tls (1), lls (2317)	14

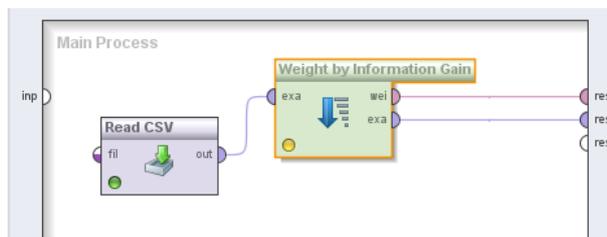
3.3 Pengolahan Data

Proses pengolahan data dilakukan dengan memanfaatkan aplikasi rapid miner. Dalam aplikasi ini tahap pertama yang dilakukan adalah melakukan import konfigurasi data dengan menentukan atribut yang akan digunakan. Proses ini akan memungkinkan luaran yaitu dataset beserta metadata dari dataset tersebut. Gambar 2 merupakan lembar kerja utama pada rapid miner.



Gambar 2 Tampilan read dataset

Dari gambar 2 akan menghasilkan tampilan data beserta atributnya apabila aplikasi dijalankan. Proses ini tidak memakan waktu yang lama karena sebenarnya hanya proses membaca dataset. Selanjutnya data yang telah terbaca oleh aplikasi dapat diolah dan digunakan algoritma information gain. Algoritma ini membaca data lalu memperhitungkan nilai entropy setiap varian data dan akhirnya dapat menghitung nilai gain untuk setiap atribut data yang dihitung. Gambar 3 merupakan tampilan proses aplikasi, dilanjutkan gambar 4 adalah hasil perhitungan dari aplikasi rapid miner untuk semua dataset dengan menggunakan information gain.



Gambar 3 Proses pembobotan dengan information gain

attribute	weight
stssipil	0
Kelas	0.007
KETLULUS	0.010
Jenjang	0.017
st_kerja	0.020
ST_TEST	0.054
prodi	0.065
KD_KONSENTRASI	0.092
agama	0.110
PEKERJAAN	0.114
prodi2	0.118
KD_KONSENTRASI2	0.127
jnkel	0.330
Jenjang2	0.379
pend_orstu	0.442
kota_kec	0.594
krjortu	1

Gambar 4 Hasil pembobotan information gain

Dari hasil penelitian sesuai gambar 4 dapat diketahui bahwa atribut dengan bobot gain terbesar adalah atribut kerjaortu. Sedangkan beberapa atribut lain tingkat kepentingannya dibawah dengan bobot sebagaimana diatas. Tingkat kepentingan atribut terkecil adalah status sipil.

4. KESIMPULAN DAN SARAN

4.1 Kesimpulan

Setelah dilakukan proses perhitungan menggunakan rapid miner didapatkan hasil sebagaimana tabel 3. Dari tabel 3 dapat diketahui bahwa status sipil merupakan atribut dengan

tingkat kepentingan terkecil yaitu 0. Diikuti dengan kelas dengan tingkat kepentingan 0,007 dan seterusnya. Sedangkan atribut dengan tingkat kepentingan tertinggi adalah atribut pekerjaan orang tua dengan tingkat kepentingan sebesar 1.

Tabel 3 Hasil Penelitian

Atribut	Tingkat Kepentingan
stssipil	0.0
Kelas	0.007098885648491337
KETLULUS	0.009858348118777867
Jenjang	0.01695386590071956
st_kerja	0.01996439934408597
ST_TEST	0.05446792978601089
prodi	0.06501923575617813
KD_KONSENTRASI	0.0918255552773498
agama	0.10951712162921348
PEKERJAAN	0.11360989474793413
prodi2	0.11768288622175352
KD_KONSENTRASI2	0.12665107469492465
jnkel	0.33029345341925725
Jenjang2	0.3785081321283303
pend_orstu	0.44193358495962637
kota_kec	0.5941005152885248
krjortu	1.0

4.2 Saran

Hasil dari penelitian ini dapat menunjukkan tingkat kepentingan atribut data heregistrasi mahasiswa STMIK Widya Pratama Pekalongan. Dengan adanya analisis kepentingan atribut ini tahap selanjutnya dapat dilakukan klasifikasi menggunakan metode terbaik dengan memanfaatkan hasil penelitian ini. Hasil akhir penelitian ini belum dapat digunakan untuk mengambil keputusan tanpa dilakukan klasifikasi lebih lanjut.

5. DAFTAR PUSTAKA

Al Karomi, M. A. (2015). Optimasi Parameter K pada Algoritma KNN untuk Klasifikasi heregistrasi mahasiswa Program Studi Teknik Informatika STMIK Widya Pratama Jl . Patriot 25 Pekalongan Email :

- adib.comp@gmail.com. *IC-TECH*, X(285), 5.
- Alkaromi, M. A. (2014). Information Gain untuk Pemilihan Fitur pada Klasifikasi Heregistrasi Calon Mahasiswa dengan Menggunakan K-NN.
- Amancio, D. R., Comin, C. H., Casanova, D., Travieso, G., Bruno, O. M., Rodrigues, F. a., & Costa, L. D. F. (2013). A systematic comparison of supervised classifiers. Retrieved from <http://arxiv.org/abs/1311.0202v1>
- Cover, T. M., & Hart, P. E. (1967). Nearest Neighbor Pattern Classification, *I*.
- Gorunescu, F. (2011). *Data Mining: Concept, Models and Techniques* (vol 12). Berlin: Heidelberg: Springer Berlin Heidelberg.
- Han, J., & Kamber, M. (2006a). *Data Mining: Concepts and Techniques Second Edition*. Elsevier. Elsevier.
- Han, J., & Kamber, M. (2006b). *Data Mining: Concepts and Techniques Second Edition*, 40(6), 9823. [https://doi.org/10.1002/1521-3773\(20010316\)40:6<9823::AID-ANIE9823>3.3.CO;2-C](https://doi.org/10.1002/1521-3773(20010316)40:6<9823::AID-ANIE9823>3.3.CO;2-C)
- Ian H Witten. Eibe Frank. Mark A Hall. (2011). *Data Mining 3rd*.
- Jiawei Han and Micheline Kaber. (2006). *Data Mining: Concepts and Techniques*. University of Illinois at Urbana-Champaign.
- Kusrini, Hartati, S., Wardoyo, R., & Harjoko, A. (2009). Perbandingan metode nearest neighbor dan algoritma c4.5 untuk menganalisis kemungkinan pengunduran diri calon mahasiswa di stmik amikom yogyakarta, *10*(1).
- Kusrini, & Taufiq, L. E. (2009). *Algoritma Data Mining*. Yogyakarta: Andi Offset.
- Larose, D. T. (2005). *Discovering Knowledge in Data: an Introduction to Data Mining*. John Wiley & Sons.
- Prasetyo, E. (2012). *Data Mining Konsep dan Aplikasi menggunakan Matlab*. Yogyakarta: Andi Offset.
- Pudjianto, T. H., Renaldi, F., & Teogunadi, A. (2011). Penerapan data mining untuk menganalisa kemungkinan pengunduran diri calon mahasiswa baru.
- Ragab, A. H. M., Noaman, A. Y., Al-Ghamdi, A. S., & Madbouly, A. I. (2014). A Comparative Analysis of Classification Algorithms for Students College Enrollment Approval Using Data Mining. *Proceedings of the 2014 Workshop on Interaction Design in Educational Environments - IDEE '14*, 106–113. <https://doi.org/10.1145/2643604.2643631>
- Santosa, B. (2007). *Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis* (Edisi Pert). Yogyakarta: Graha Ilmu.
- Sugianti, D. (2012). Algoritma Bayesian Classification untuk Memprediksi Heregistrasi Mahasiswa Baru di STMIK Widya Pratama, (2), 1–5.
- Susanto, S., & Suryadi, D. (2010). *Pengantar Data Mining: Menggali Pengetahuan dari Bongkahan Data*. Yogyakarta: Andi Offset.
- Widiastuti, D. (2007). Analisa Perbandingan Algoritma SVM, Naïve Bayes, dan Decission Tree dalam Mengklasifikasikan Serangan (Attack) pada Sistem Pendeteksi Intrusi. *Jurusan Sistem Informasi Universitas Gunadarma*, 1–8.
- Witten, I. H., Frank, E., & Hall, M. A. (2011). *Data Mining: Practical Machine Learning Tools and Techniques 3rd Edition*. Elsevier.
- Wu, X. (2009). *The Top Ten Algorithms in Data Mining*. (V. Kumar, Ed.). New York: Taylor & Francis Group, LLC.