

## ANALISA KEPENTINGAN ATRIBUT DATA PADA KLASIFIKASI HEREGISTRASI MAHASISWA STMIK WIDYA PRATAMA

Dadang Aribowo <sup>1)</sup>, Erni Rahmawatie <sup>2)</sup>, Ivandari <sup>3)</sup>, M. Adib Al Karomi <sup>3)</sup>

<sup>1</sup> STMIK Widya Pratama Pekalongan  
email: dadang.stmik@gmail.com

<sup>2</sup> STMIK Widya Pratama Pekalongan  
email: ernie.stmik@gmail.com

<sup>3</sup> STMIK Widya Pratama Pekalongan  
email: ivandarialkaromi@gmail.com

<sup>4</sup> STMIK Widya Pratama Pekalongan  
email: adib@alkaromi.com

### *Abstract*

*Students are the most valuable assets in a private tertiary institution. Because most of private tertiary institution revenue and operational costs are obtained from students. The number of students who carry out re-registration will obviously be a breath of fresh air for the institution. In the last 5 years it was noted that around 20% of prospective STMIK Widya Pratama student did not register. The latest data on 31 August 2018 recorded that 32.7% of registrants had not registered. The decline in the number of students can affect the financial stability of institutions primarily private colleges. Analysis of the best algorithm for the classification of student registration was carried out and proved that the decision tree C45 is the algorithm with the best accuracy. Early knowledge of prospective students who might not be able to register can become an institution's reference to take action to retain students. Neatly arranged student data can be used by management to analyze the characteristics and causes of students not to register. This research will conduct an analysis of all existing data and attribute data. The method used in weighting is information gain which has been proven to be able to handle datasets with many types of attributes. The results of this study conclude that the work attributes of parents are the attributes with the highest level of importance. While the civil status attribute is the attribute with the lowest level of importance.*

**Keywords:** student registration, information gain, acceptance of new students

### 1. PENDAHULUAN

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 1.1

merupakan data yang diperoleh dari panitia penerimaan mahasiswa baru STMIK Widya Pratama Pekalongan.

Tabel 1.1 Daftar jumlah pendaftar dan registrasi mahasiswa  
(sumber: PMB STMIK Widya Pratama Pekalongan)

Tahun	Pendaftaran	Registrasi	Selisih	Prosentase (%)	
				Registrasi	Tidak Registrasi
2013/2014	705	513	192	73%	27%
2014/2015	608	474	134	78%	22%
2015/2016	667	500	167	75%	25%
2016/2017	654	427	227	77%	23%
2017/2018	506	428	78	84,6%	15,4%
2018/2019	510	352	158	69%	31%

Penurunan 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.

## 2. KAJIAN LITERATUR

### 2.1 Kajian Pustaka

1. Penelitian yang dilakukan M. Adib Alkaromi (Alkaromi, 2014) ini melakukan klasifikasi heregistrasi calon mahasiswa dengan menggunakan dataset sebanyak 2389 record. Dataset yang dipakai merupakan data PMB dari tahun 2011 sampai dengan tahun 2013. Atribut yang ada dalam dataset tersebut sebanyak 44 dengan didalamnya satu atribut label dan satu atribut id. Kesemua atribut silakukan seleksi fitur dengan menggunakan information gain kemudian dilakukan klasifikasi dengan memanfaatkan algoritma K-NN. Dalam penelitian tersebut algoritma K-NN memperoleh tingkat akurasi sebesar 78,15% dengan menggunakan kesemua atributnya. Sedangkan setelah dilakukan seleksi fitur menggunakan information gain akurasi dari algoritma K-NN naik hingga mencapai 83,93%. Penelitian ini

menggunakan confusion matrix untuk evaluasi serta 10folds cross validation untuk proses validasinya.

2. Penelitian yang dilakukan Devi Sugianti (Sugianti, 2012) di STMIK Widya Pratama Pekalongan menghasilkan akurasi sebesar 78% menggunakan algoritma Bayesian Classification untuk memprediksi heregistrasi mahasiswa. Penelitian tersebut menggunakan data PMB STMIK tahun 2011 dengan 913 record. Atribut yang digunakan adalah: kota asal, program studi, status daftar dan gelombang. Dalam penelitian tersebut menggunakan dua class dalam label atau atribut tujuan, yaitu "Registrasi" dan "Tidak Registrasi". Perhitungan manual menggunakan rumus algoritma Bayes dari data sampel dan data baru yang belum diketahui dijelaskan secara terperinci. Namun tidak dilakukan evaluasi terhadap algoritma seperti halnya dengan menggunakan confusion matrix. Dalam penelitian ini juga tidak dilakukan perhitungan validasi seperti halnya menggunakan 10folds cross validation. Sehingga tidak dapat dilihat perhitungan performa dari algoritma Bayes untuk dataset tersebut.
3. Dalam penelitian klasifikasi yang dilakukan oleh Tacbir Hendro Pudjianto, Faiza Reinaldi dan Age Teogunadi pada tahun 2011 ini digunakan algoritma ID3 (Pudjianto, Renaldi, & Teogunadi, 2011). Data yang digunakan adalah data Penerimaan Mahasiswa Baru UNJANI pada 5 tahun terakhir. Atribut yang digunakan dalam penelitian tersebut adalah: Kode Progdi, Jenis Kelamin, Agama, Gol.Darah, Pekerjaan, Penghasilan, Jurusan SLTA, Asal Sekolah, Kode Progdi1, Kode Progdi2, Kode Progdi3, Gelombang, Registrasi. Atribut terakhir yaitu registrasi dijadikan sebagai atribut tujuan atau label dengan 2 kemungkinan yaitu registrasi atau tidak registrasi. Dari hasil penelitian yang telah dilakukan didapatkan tingkat akurasi dari algoritma ID3 untuk dataset tersebut diatas adalah 61,89%. Hasil dari penelitian tersebut berikutnya dijadikan sebagai acuan untuk membangun sebuah sistem pendukung keputusan yang nantinya dapat digunakan untuk pihak kampus. Sistem tersebut nantinya juga akan digunakan sebagai alat ukur ketercapaian target penerimaan mahasiswa baru yang ditetapkan oleh pihak manajemen.
4. Dalam penelitian yang dilakukan Kusri dkk (Kusri, Hartati, Wardoyo, & Harjoko, 2009). menerapkan penggunaan algoritma K-Nearest Neighbour untuk kemungkinan heregistrasi mahasiswa STMIK AMIKOM Yogyakarta. Sebelumnya Kusri juga telah melakukan penelitian yang sama dengan menggunakan algoritma C4.5. Penelitian tersebut dilakukan karena dari data sebelumnya diketahui hanya sekitar 75% dari calon mahasiswa pendaftar yang melakukan heregistrasi. Data yang digunakan dalam penelitian tersebut diambil dari panitia PMB STMIK AMIKOM Yogyakarta. Atribut yang digunakan dalam penelitian ini antara lain: NEM, Jenis Kelamin, Jurusan, Gelombang, Pilihan1, Catatan, Nilai dan Jurusan Lulus. Hasil dari penelitian ini menyatakan bahwa algoritma K-Nearest Neighbour tidak lebih baik dari algoritma C4.5 untuk klasifikasi heregistrasi calon mahasiswa STMIK AMIKOM Yogyakarta. Dalam proses klasifikasi menggunakan K-Nearest Neighbour diperlukan sebuah proses pembobotan untuk setiap atribut yang ada serta dilakukan perhitungan kedekatan seluruh data training dengan data testing. Proses tersebut jelas akan memerlukan banyak waktu untuk perhitungan terlebih lagi jika atribut yang digunakan semakin banyak serta dataset semakin besar.

## 2.2 Kajian Teori

### 1. Data Mining

Data Mining atau sering juga disebut Knowledge Discovery in Database (KDD) adalah sebuah bidang ilmu yang banyak membahas tentang pola sebuah data. Serangkaian proses guna mendapatkan pengetahuan atau pola dari kumpulan data

disebut dengan data mining (Ian H Witten, Eibe Frank, Mark A Hall, 2011). Data mining merupakan sebuah proses ekstraksi untuk mendapatkan suatu informasi yang sebelumnya tidak diketahui dari sebuah data (Witten et al., 2011). Data mining dapat menganalisa kasus lama untuk menemukan pola dari data dengan menggunakan teknik pengenalan pola seperti statistik dan matematika (Larose, 2005). Kumpulan data yang besar bisa jadi tidak berarti apabila nilai informasi didalamnya kurang atau tidak dapat diambil sebuah pengetahuan.. Data mining menjawab masalah ini dengan menganalisa data yang besar tersebut kemudian membuat sebuah aturan, pola, ataupun model tertentu untuk mengenali data baru yang tidak berada dalam baris data yang tersimpan (Prasetyo, 2012).

Data mining memiliki beberapa fungsi. Fungsi utama data mining adalah: Estimasi, Prediksi, Klustering, Klasifikasi, dan Asosiasi. Berdasarkan metode pembelajarannya fungsi data mining dibagi menjadi 2 (Santosa, 2007) yaitu Supervised Learning, Unsupervised Learning. Supervised learning harus memiliki data sampel atau sering disebut juga dengan data training. Sedangkan dalam unsupervised learning tidak membutuhkan data training. Klasifikasi merupakan fungsi data mining yang tergolong dalam supervised learning.

## **2. Information gain**

Information gain adalah metode seleksi fitur yang banyak dipakai oleh peneliti (Deng & Runger, 2012). Dengan menggunakan information gain batas dari kepentingan sebuah atribut dapat ditentukan (Novakovic, 2010). Nilai information gain merupakan nilai entropy sebelum pemisahan dikurangi dengan nilai entropy setelah pemisahan. Pengukuran nilai ini digunakan untuk pembobotan atribut yang nantinya akan digunakan dalam tahap klasifikasi atau tidak. Atribut yang memenuhi kriteria pembobotan tertentu nantinya akan digunakan dalam proses klasifikasi sebuah algoritma.

Secara umum tahapan information gain dilakukan dengan cara sebagai berikut:

1. Menghitung nilai information gain untuk semua atribut dalam dataset original.
2. Menentukan batas (threshold) yang dikehendaki. Dengan adanya batasan ini memungkinkan atribut yang berbobot sama dengan batas atau lebih besar dari batas akan dipertahankan dan digunakan dalam tahap klasifikasi. Kemudian atribut dengan bobot dibawah batas akan dibuang dan tidak digunakan dalam tahap klasifikasi.
3. Dataset diperbarui dengan hanya menggunakan atribut terpilih.

## **3. METODE PENELITIAN**

### **3.1 Metode 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.

### **3.2 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.

### 3.3 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.

### 3.4 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 pembuatan jurnal dan publikasi pada prosiding seminar nasional.

## 4. HASIL PENELITIAN

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. Tabel 2 merupakan rekapitulasi data heregistrasi mahasiswa STMIK Widya Pratama.

Tabel 2 Rekapitulasi data heregistrasi mahasiswa  
(sumber: PMB STMIK Widya Pratama Pekalongan)

Tahun	Pen daftara n	Regist rasi	Selisih	Prosentase (%)	
				Registrasi	Tidak Registrasi
2013/2014	705	513	192	73%	27%
2014/2015	608	474	134	78%	22%
2015/2016	667	500	167	75%	25%
2016/2017	654	427	227	77%	23%
2017/2018	506	428	78	84,6%	15,4%
2018/2019	510	352	158	69%	31%
Jumlah	3650	2694	956		

Dalam tabel 2 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.

Setelah dilakukan pemrosesan dataset menggunakan aplikasi rapid miner, muncul metadata sebagaimana tampil pada tabel 2. Tabel 2 merupakan meta data dari dataset mentah atau dataset yang semua atributnya masih murni. 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 isianya 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.

## 5. SIMPULAN

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
Jenang	0.01695386590071956
st_kerja	0.01996439934408597
ST_TEST	0.05446792978601089
Prodi	0.06501923575617813
KD_KONSENTRASI	0.09182555552773498
agama	0.10951712162921348
PEKERJAAN	0.11360989474793413
prodi2	0.11768288622175352
KD_KONSENTRASI2	0.12665107469492465
Jnkel	0.33029345341925725
Jenang2	0.3785081321283303
pend_ortu	0.44193358495962637
kota_kec	0.5941005152885248
krjortu	1.0

## 6. REFERENSI

- 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 Decision 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.