

## Penerapan algoritma Adaboost guna menentukan pola masuknya calon mahasiswa.

**Nisa Hanum Harani, Cahyo Prianto**

<sup>1</sup>Politeknik Pos Indonesia

Sariasih no 54, Bandung, 022-2009562, nisahanum@poltekpos.ac.id

<sup>2</sup>Politeknik Pos Indonesia

Sariasih no 54, Bandung, 022-2009562, cahyoprianto@poltekpos.ac.id

---

### ARTICLE INFO

Article history:

Received 15 September 2019

Received in revised form 02 January 2020

Accepted 29 July 2020

Available online 31 July 2020

---

### ABSTRACT

In general, the college admission process is done through registration, file selection, examinations, an announcement of the results of students who pass, and ends with re-registration. In this case, a problem was found where there is a significant decrease in the number of student who register with those who re-register. Things like this can reduce the balance between new students and students who meet the requirements, to make a decrease in the quality of higher education and affect accreditation. Based on these problems, a classification method was developed to look for patterns of students who would enter institutions and what factors influence students to re-register.

To improve the accuracy of the decision tree algorithm the author use adaptive boosting (adaboost) in finding factors that make prospective students continue to the re-registration process.

From the results of the study, the AdaBoost-based decision tree algorithm shows that the level of accuracy has an increase of 20%. The presentation of results is as follows, 61.4% (decision tree); 91.35% (decision tree + AdaBoost).

Keywords: higher education, decision tree, adaptive boosting

---

### 1. Pendahuluan

Informasi dalam bidang pendidikan umumnya semakin bertambah tiap tahunnya, hal ini dikarenakan adanya penerimaan siswa baru bagi sekolah – sekolah dan penerimaan mahasiswa baru bagi perguruan tinggi. Semakin bertambahnya siswa maupun mahasiswa baru akan membuat penumpukan data, sehingga pihak sekolah maupun universitas dan politeknik harus melakukan pengolahan dan penyusunan data agar terlihat lebih rapih serta mudah dicari ketika dibutuhkan kembali.

Penerimaan Mahasiswa Baru atau biasa kita sebut dengan “PMB” merupakan salah satu proses yang ada di instansi pendidikan seperti Universitas dan Politeknik yang berguna untuk menyaring calon mahasiswa baru. Mahasiswa yang hendak mendaftar perguruan tinggi biasanya

dapat melalui berbagai macam cara seperti menerima undangan dari pihak perguruan tinggi, penelusuran minat dan kemampuan (PMDK), dan ujian saringan yang dilakukan oleh pihak universitas dan politeknik. Pada umumnya proses penerimaan mahasiswa baru atau “PMB” dilakukan melalui tahapan pendaftaran atau registrasi, seleksi berkas, melakukan ujian, pengumuman hasil mahasiswa yang lolos, dan berujung dengan her-registrasi atau pendaftaran ulang. Pada kasus kali ini ditemukan sebuah permasalahan dimana terjadi penurunan jumlah mahasiswa yang daftar dibandingkan mahasiswa yang melakukan her-registrasi. Hal seperti ini dapat mengakibatkan tidak adanya keseimbangan antara mahasiswa baru dengan mahasiswa yang lolos, sehingga dapat membuat penurunan nilai kualitas perguruan tinggi tersebut dan berpengaruh terhadap akreditasi.

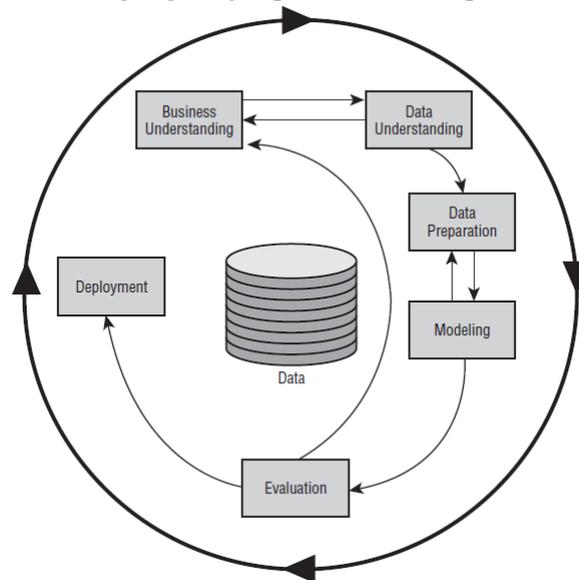
*Data mining* merupakan salah satu proses inti yang terdapat dalam KDD. Banyak orang memperlakukan *data mining* sebagai sinonim dari KDD, karena sebagian besar pekerjaan dalam KDD difokuskan pada *data mining*. Namun, langkah – langkah lain merupakan proses – proses penting yang menjamin kesuksesan dari aplikasi KDD [1]. *Decision Tree* adalah pohon dimana setiap cabangnya menunjukkan pilihan diantara sejumlah alternatif pilihan yang ada, dan setiap daunnya menunjukkan keputusan yang dipilih. *Decision tree* (pohon keputusan) biasa digunakan untuk mendapatkan informasi untuk tujuan pengambilan sebuah keputusan [2]. Algoritma *decision tree* (pohon keputusan) memiliki bias terhadap generalitas yang sangat cocok dalam fungsi yang besar, namun tidak pada fungsi yang kecil. Di sisi lain, algoritma *decision tree* adalah algoritma pencarian yang kuat, fleksibel yang cenderung mengatasi interaksi atribut lebih baik dari pada kebanyakan algoritma induksi aturan [3]. Pembelajaran *decision tree* adalah salah satu metode yang paling banyak digunakan dan praktis untuk inferensi induktif. *Decision tree* adalah metode untuk mendekati fungsi bernilai diskrit yang kuat untuk data yang mampu mempelajari ekspresi pemisahan [4]. *Decision Tree* (Pohon Keputusan) adalah sistem pendukung keputusan yang menggunakan grafik keputusan seperti pohon. *Decision tree* atau biasa disebut juga dengan pohon klasifikasi digunakan untuk mempelajari fungsi klasifikasi yang menyimpulkan nilai atribut dependen (variabel) dengan nilai atribut yang independen (input variabel). Ini memverifikasi masalah yang dikenal sebagai klasifikasi terawasi karena atribut dependen dan perhitungan *class* (nilai) telah diberikan [5]. *Decision tree* (Pohon keputusan) didasarkan pada probabilitas bersyarat dan menggunakan grafik seperti pohon dengan kemungkinan konsekuensi, hasil, biaya, peristiwa, dan utilitas. *Decision tree* banyak digunakan dalam riset operasi, analisis pemasaran, tren keuangan khususnya dalam pengambilan keputusan, guna mengidentifikasi jalur terpendek atau strategi yang membantu mencapai suatu tujuan. *Decision tree* mirip dengan bagan alir dimana simpul internal mewakili tes pada atribut atau parameter seperti apakah melempar koin akan menghasilkan kepala atau ekor, kemudian cabang menggambar hasil pengujian dan label kelas diwakili oleh simpul daun. Aturan klasifikasi diatur oleh jalur dari simpul akar ke simpul daun. Dalam proses pengambilan keputusan, *decision tree* digunakan sebagai alat visual dan analitis, untuk memprediksi nilai target [6]. Algoritma C4.5 adalah algoritma populer dalam klasifikasi data mining. Algoritma C4.5 menggunakan kriteria *information gain* untuk memilih atribut yang akan digunakan dalam pemisahan objek. Atribut yang mempunyai *information gain* paling tinggi dibandingkan dengan atribut pada data lain, maka atribut tersebutlah yang dipilih sebagai pemecahan. *Decision Tree* (Pohon Keputusan) sesuai digunakan untuk kasus – kasus dimana outputnya bernilai diskrit [7].

Permasalahan yang sering muncul dalam proses klasifikasi adalah ketidakseimbangan kelas (*class imbalance*) yang menyebabkan *overlapping* [8]. Ketidakseimbangan kelas dapat mempengaruhi kinerja klasifikasi, karena terjadi banyak *misclassification* [9]. Metode *ensemble* dapat menangani ketidakseimbangan dataset, yang palng populer adalah *boosting*. Algoritma *boosting* telah dilaporkan sebagai meta-teknik untuk mengatasi masalah *unbalance* dataset [10].

Berdasarkan permasalahan tersebut, maka dibuatlah sebuah metode klasifikasi guna mencari pola mahasiswa yang akan benar – benar masuk perguruan tinggi (PT). Data diambil dari kasus PMB yang terjadi di perguruan tinggi (PT) yang ada di Indonesia. Data tersebut akan diprediksi menggunakan data mining dengan metode klasifikasi *decision tree* dengan pendekatan *boosting*. Data tersebut akan diolah menggunakan sebuah *software* yang dinamakan *RapidMiner Studio* dengan melalui 6 tahap proses CRISP-DM pada data mining. Sehingga terciptalah sebuah hasil prediksi yang akurat dan dapat membantu pihak PMB guna menyelesaikan permasalahan tersebut.

## 2. Metode Penelitian

Proses penelitian dilaksanakan dengan mengikuti prosedur standar dalam penggalian sebuah data yang biasa disebut dengan *CRISP-DM* (*Cross Industry Standard Process for Data Mining*)[11]. CRISP-DM terdiri dari enam tahap seperti yang terlihat dalam gambar berikut :



Gambar 1 CRISP-DM

### 1. Pemahaman Bisnis

Pemahaman bisnis meliputi penetapan tujuan, penilaian situasi terkini, penetapan tujuan *data mining* dan pengembangan rencana. Dalam penelitian ini, peneliti telah merumuskan bahwa tujuan dari penelitian ini adalah untuk mengetahui pola mahasiswa baru yang telah lolos dan melakukan her-registrasi dan juga untuk mengetahui hal apa saja yang mempengaruhi dan menjadi nilai timbang untuk mereka meyakinkan diri dalam melakukan her-registrasi.

### 2. Pemahaman Data

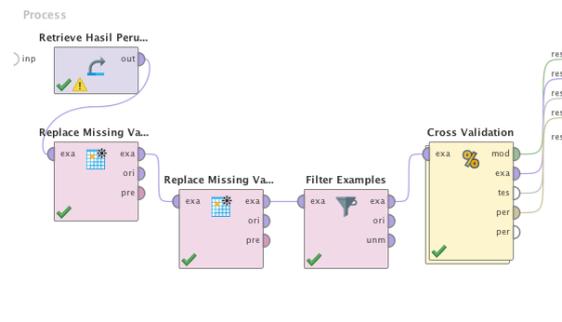
Pada tahap ini melakukan pemahaman terhadap data yang dibutuhkan, untuk kemudian mengambil data yang relevan dan memiliki keterkaitan dengan tujuan penelitian yang tentunya telah disesuaikan dan diproses sedemikian rupa untuk memberikan penelitian yang baik. Dataset mahasiswa yang didapatkan berjumlah 3300 *record* . Terdapat 9 atribut yang dikelola yaitu : penghasilan orang tua, provinsi asal , perguruan tinggi, jalur pendaftaran, pilihan 1, pilihan 2, pilihan 3, jurusan dan keterangan.

### 3. Persiapan data

Pada tahap ini peneliti mengolah data yang didapat. Melakukan pembersihan terhadap data, melakukan integrasi data, melakukan pemilihan data ( training / testing ), dan transformasi pada data. *Tools* yang digunakan dalam menunjang pengolahan datanya ialah *microsoft excel dan RapidMiner*.

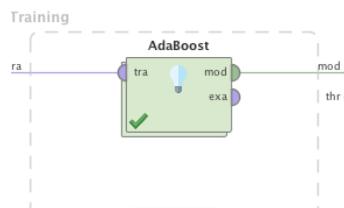
#### 4. Pembuatan Model

Model yang dihasilkan harus melalui proses validasi, pada penelitian ini digunakan teknik *cross validation*. Pada proses ini, data dibagi menjadi dua bagian, data training berfungsi untuk membuat model dengan implementasi pohon keputusan berbasis Adaboost. Setelah model terbentuk, data desting digunakan untuk menguji performa dari model yang telah dibuat.



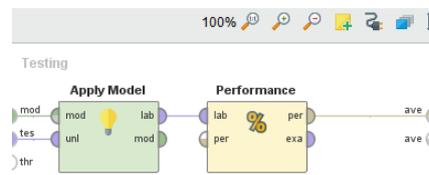
Gambar 2 Pembuatan Model

Operator “Validation” ( Cross Validation ). Operator ini digunakan untuk membagi data menjadi 2 bagian yang nantinya akan diproses di sub proses dan berada pada proses validasi itu sendiri.



Gambar 3 Operator Adaboost

Operator “ Cross Validation”, pada bagian training digunakan operator “ Adaboost” . Operator ini bertugas untuk melakukan proses algoritma boosting.



#### 5. Evaluasi

Pada tahap evaluasi, akan diketahui apakah hasil daripada tahap pemodelan dapat menjawab tujuan yang telah ditetapkan pada fase pertama. Tujuan dari fase pertama adalah memahami pola dan mencari faktor yang paling berpengaruh dalam menentukan apakah seorang mahasiswa akan melakukan re-registrasi.

#### 6. Pelaksanaan

Model yang telah dibentuk kemudian diuji akurasi, setelah akurasi dapat mendekati kenyataan dilapangan, maka model tersebut dipercaya untuk menjadi tolak ukur pengambilan keputusan pada sebuah perguruan tinggi.

### 3. Hasil dan Pembahasan

Setelah menghasilkan model akurasi yang baik, maka model tersebut sudah dianggap merepresentasikan keadaan general pada dunia nyata, hasil dan luaran yang dicapai adalah sebagai berikut.

#### 3.1. Hasil

Model yang dihasilkan harus melalui tahap pengujian, pengujian dilakukan untuk mengukur kinerja algoritma yang sudah ditentukan. Pada uji coba ini, diberikan data latih untuk membentuk klasifikasi pohon keputusan. Teknik pengujian yang digunakan adalah *cross validation*, teknik validasi ini merupakan teknik yang penting dalam ilmu *data science*. Berangkat dari sangat dibutuhkannya tolak ukur untuk memvalidasi stabilitas model pembelajaran, dilihat dari seberapa baik model tersebut digeneraliskan ke data yang baru. Perlu dipastikan model yang dibuat telah mendapatkan pembelajaran dari data yang benar, jauh dari noise atau dengan kata lain rendahnya bias dan varians.

##### A. Analisis Kinerja pengklasifikasi Pohon Keputusan (DT)

Dalam proses ini dipilih model algoritma pohon keputusan dan tools yang dipakai adalah RapidMiner. Dalam operator “Decision Tree” terdapat beberapa pengaturan yang harus dilakukan. Untuk memeriksa “purity” sebuah atribut yang dipakai adalah *Information Gain*. Kedalaman Pohon yang paling ideal adalah 10 dengan melakukan proses pruning dengan confidence 0.5

Tabel 1 Pengujian Pohon Keputusan

Kriteria	Kedalaman Pohon	Pruning	Confidence	Pre Pruning	Min Gain	Akurasi
Information Gain	6	ya	0.5	tidak		63.97 %
Information Gain	7	ya	0.5	tidak		62.75 %
Information Gain	8	ya	0.5	tidak		62.21 %
Information Gain	10	ya	0.5	tidak		62.21 %
Information Gain	10	tidak	0.5	tidak		61.65 %
Information Gain	10	ya	0.5	ya	0.1	70.72 %
Information Gain	10	ya	0.5	ya	0.5	54.85 %

Dari hasil tabel 1 akurasi tertinggi yang di dapat dari implementasi pohon keputusan adalah 70.72 %.

Tabel 2 Confusion Matrix DT

		Aktual	
		Positif	Negatif
	Positif		
	Negatif		

	Positif	1079	778
Prediksi	negatif	768	1466

Tabel 2 menunjukkan bahwa perhitungan confusion Matrix nya menghasilkan presisi sebesar 58.10 % untuk nilai prediksi positif dan 65.62% untuk nilai prediksi negatif.

#### B. Analisis Kinerja pengklasifikas DT berbasis Adaboost

Dalam proses ini dipilih model algoritma pohon keputusan dan tools yang dipakai adalah RapidMiner. Dalam operator “AdaBoost” kita akan melatih algoritma Decision Tree (Pohon Keputusan ) untuk belajar dengan bobot yang telah ditentukan secara otomatis oleh rapidminer. Kedalaman Pohon yang paling ideal adalah 10 dengan melakukan proses pruning dengan confidence 0.5 .

Kriteria	Kedalaman Pohon	Pruning	Confidence	Pre Pruning	Min Gain	Akurasi
Information Gain	6	ya	0.5	tidak		87.69 %
Information Gain	7	ya	0.5	tidak		90.62%
Information Gain	8	ya	0.5	tidak		91.23 %
Information Gain	10	ya	0.5	tidak		91.23 %
Information Gain	10	tidak	0.5	tidak		91.35 %
Information Gain	10	ya	0.5	ya	0.1	70.93 %
Information Gain	10	ya	0.5	ya	0.5	54.84 %

Dari hasil tabel 5.2 maximal akurasi yang bisa didapatkan dengan menambahkan adaboost sebagai pembelajaran model klasifikasi adalah 91.35.

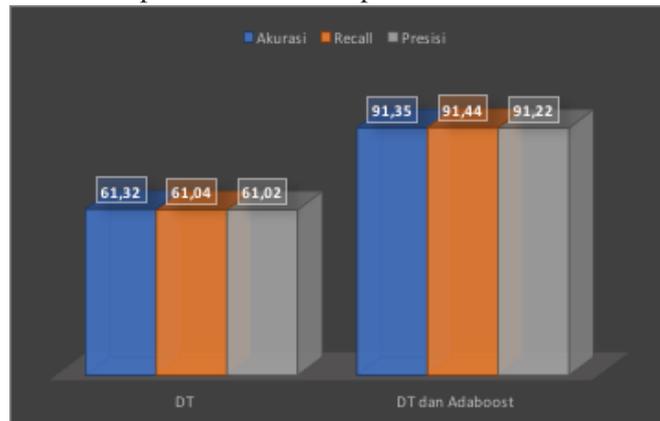
Tabel 3 Confusion Matrix DT berbasis Adaboost

		Aktual	
		Positif	Negatif
Prediksi	Positif	1707	217
	negatif	142	2027

Tabel 3 menunjukkan bahwa perhitungan confusion Matrix nya menghasilkan presisi sebesar 88.72 % untuk nilai prediksi positif dan 93.45% untuk nilai prediksi negatif.

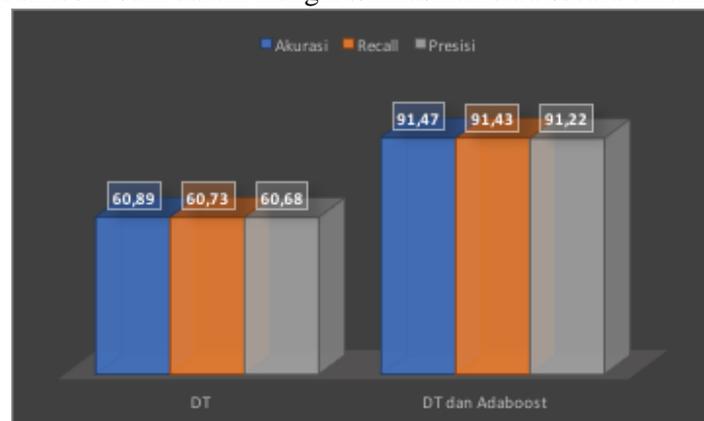
Berdasarkan pengukuran kinerja, diperoleh informasi tingkat kinerja meliputi kemampuan model classifier dalam mengklasifikasikan data secara umum (*akurasi*), besarnya presentase classifier dalam menebak dengan tepat kelas true positive dengan melihat perbandingan true positive yang dapat diprediksi dengan penjumlahan true positive dan false positive yang ada (*presisi*), serta sensitivitas (*recall*) yang merupakan ukuran tingkatan presentase classifier dapat mengenal

contoh kelas positif (tepat waktu) berdasarkan jumlah true positive yang dapat diprediksi dengan benar berdasarkan contoh kelas positif dan RMSE pada dataset PMB



Gambar 4 Grafik Perbandingan Kinerja 5-fold dan Confidence Level 0,5

Pada gambar 4 terlihat bahwa kinerja model DT dan Adaboost masih menghasilkan akurasi yang lebih baik dibandingkan dengan model yang lain, hal ini menunjukkan bahwa model tersebut memiliki kemampuan lebih baik dalam mengklasifikasikan data secara umum.

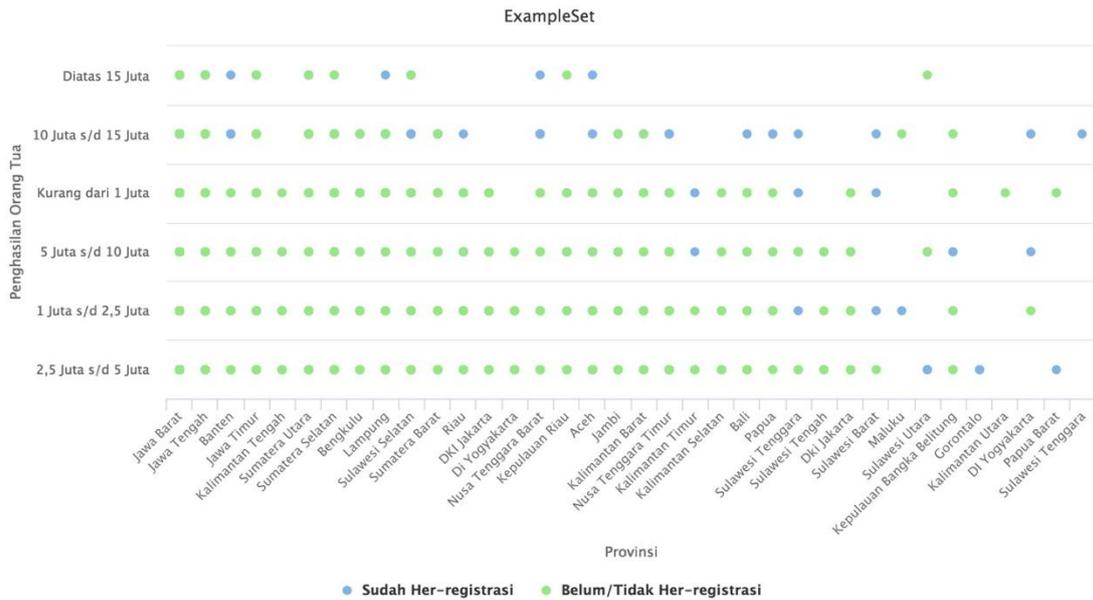


Gambar 5 Grafik Perbandingan Kinerja 10 Fold dan Confidence Level 0,25

Pada Gambar 5 terlihat bahwa kinerja model pohon keputusan berbasis adboost masih menghasilkan akurasi dan presisi yang lebih baik dibandingkan dengan model yang menggunakan pohon keputusan saja.

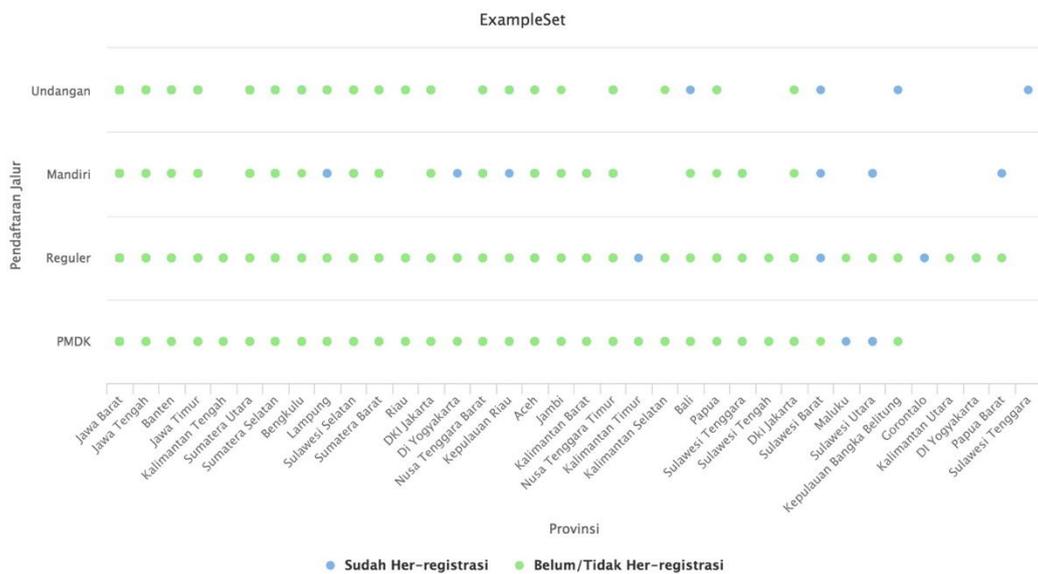
### 3.2. Pembahasan

Berdasarkan hasil pembelajaran menggunakan metode algoritma adaboost pada algoritma klasifikasi pohon keputusan di dapat akurasi yang sudah tinggi sebanyak 91.23 %. Model sudah dikatakan akurat, tahap selanjutnya adalah visualisasi hasil analisa. Dalam rapidminer visualisasi data menggunakan teknik *Scatter* dengan matriks 2 dimensi dan informasi yang ditampilkan berbeda-beda tergantung kebutuhan.



Gambar 6 Penghasilan Orang Tua Menurut Sebaran Daerah Asal Mahasiswa

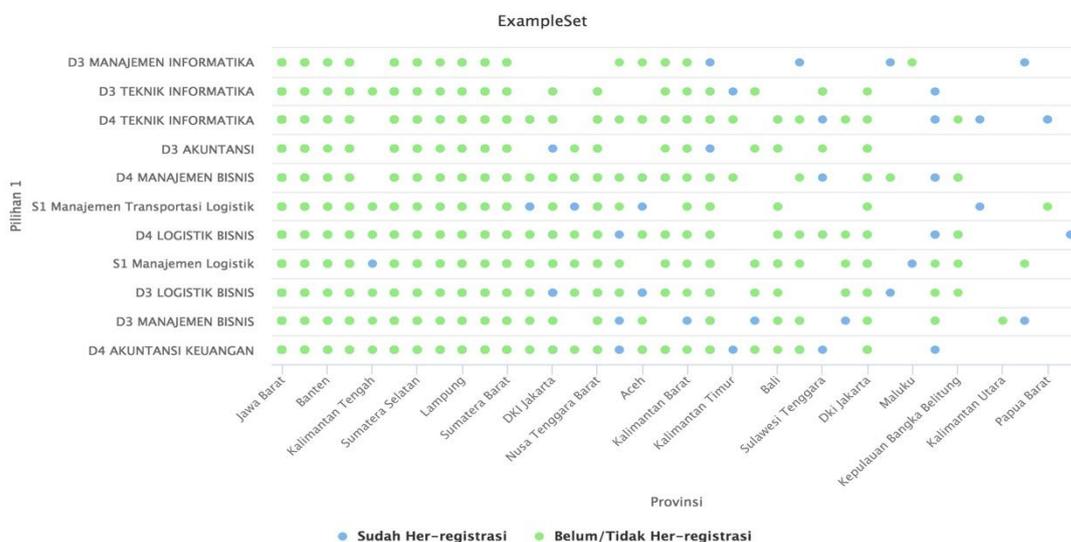
Dapat dilihat dari gambar 6 bahwa penghasilan orang tua mahasiswa yang paling banyak adalah kisaran 2.5 juta s/d 5 juta dan mahasiswa lebih banyak berasal dari Jawa Barat. Dengan sumbu X adalah provinsi daerah asal mahasiswa dan sumbu Y adalah range penghasilan orang tua. Titik biru adalah klasifikasi keterangan sudah melakukan registrasi seangkan titik hijau adalah klasifikasi keterangan belum/tidak melakukan registrasi. Perdasarakan gambar tersebut, penghasilan orang tua kisaran 2.5 juta s/d 5 juta jika berasal dari daerah Sumatra Utara , Gorontalo dan Papua Barat kemungkinan besar akan melakukan her-registrasi.



Gambar 7 Pendaftaran Jalur Masuk menurut Asal Provinsi

Dapat dilihat dari gambar 6 bahwa mahasiswa terbanyak masuk melalui jalur PMDK . Dengan sumbu X adalah provinsi daerah asal mahasiswa dan sumbu Y adalah Jalur Masuk . Jalur mandiri

memiliki kemungkinan mahasiswa terbanyak untuk melakukan re-registrasi yaitu berasal dari daerah Lampung, Yogyakarta, Sulawesi Barat, Sulawesi Utara dan Papua Barat.



Gambar 8 Pilihan Jurusan Berdasarkan Provinsi Asal

Berdasarkan Gambar 8 dengan sumbu X adalah provinsi daerah asal dan sumbu Y adalah Pilihan pertama dapat diperoleh informasi bahwa sebagai berikut, Mahasiswa yang berasal dari Jaerah DKI Jakarta dan memilih prodi D3 Akuntansi sebagai pilihan pertama kemungkinan besar akan melakukan her-registrasi, berbanding terbalik dengan mahasiswa yang berasal dari daerah Jawa Barat mereka memiliki kemungkinan besar tidak akan melakukan her-registrasi di Jurusan manapun.

**4. Kesimpulan**

Adapun beberapa kesimpulan yang dihasilkan dalam penelitian ini untuk menentukan pola masuk nya mahasiswa baru adalah sebagai berikut :

- A. Faktor yang mempengaruhi mahasiswa melakukan her-registrasi berdasarkan data PMB adalah penghasilan orang tua. Penghasilan orang tua yang menjadi root dalam model , dianggap dapat mengklasifikasikan data PMB dengan baik dengan menggunakan information gain. Information Gain dipakai untuk menentukan berapa banyak informasi yang bisa diberikan oleh variable penghasilan orang tua.
- B. Penerapan Metode klasifikasi pohon keputusan (decission tree) dan adaboost dapat meningkatkan akurasi hingga 91,35%. Model kombinasi ini dianggap paling akurat jika dibandingkan dengan metode klasifikasi yang hanya menggunakan algoritma pohon keputusan saja (61,4%). Hasil akurasi menunjukan bahwa model yang dihasilkan dapat melakukan prediksi dengan tepat dalam menentukan pola mahasiswa yang akan benar – benar masuk Perguruan Tinggi (PT)

**Daftar Pustaka**

[1] R. A. Kaufman, *Educational system planning*. Englewood Cliffs : Prentice-Hall, 1972.  
 [2] M. Somantri, *PERENCANAAN PENDIDIKAN*. IPB Press, 2014.  
 [3] Y. H. Maturbongs and R. Satria, “PERENCANAAN STRATEGIS SISTEM INFORMASI PADA INSTITUSI PENDIDIKAN TINGGI,” vol. 2011, no. Snati, pp. 17–18, 2011.  
 [4] N. A. Ismail, “A Case Analysis of Information Technology Management in a Public,” pp. 505–511, 2020.

- 
- [5] M. A. Al Karomi, "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," no. 1, pp. 28–33, 2015.
- [6] D. T. Larose and C. D. Larose, "An Introduction to Data Mining," 2014.
- [7] J. Han and M. Kamber, *Data Mining: Concepts and Techniques*, Second. San Francisco: Morgan Kaufmann.
- [8] M. D. Borah, R. Jindal, D. Gupta, and G. C. Deka, "Application of knowledge based decision technique to Predict student enrollment decision," pp. 180–184, 2011.
- [9] P. Pumpuang, "Comparisons of Classifier Algorithms : Bayesian Network , C4 . 5 , Decision Forest and NBTree for Course Registration Planning Model of Undergraduate Students," pp. 3647–3651, 2008.
- [10] D. Aribowo, A. Ekyanto, and H. Setiadi, "Analisa Komparasi Algoritma Data Mining untuk Klasifikasi Heregistrasi Calon Mahasiswa STMIK Widya Pratama," vol. XIII, no. 2, pp. 1–6, 2018.
- [11] F. A. Hermawati, *Data Mining*, Yogyakarta: Penerbit Andi, 2013.