

# PERBANDINGAN AKURASI MODEL REGRESI LOGISTIK UNTUK PREDIKSI KATEGORI IP MAHASISWA JALUR PRESTASI DENGAN NON JALUR PRESTASI

## *COMPARISON OF LOGISTIC REGRESSION MODEL ACCURATION TO PREDICT GPA OF STUDENTS THROUGH ACHIEVEMENT AND REGULAR ADMISSION*

R. Gunawan Santosa<sup>1</sup>, Antonius Rachmat Chrismanto<sup>2</sup>

<sup>1,2</sup>Prodi Teknik Informatika

Universitas Kristen Duta Wacana Yogyakarta

<sup>1</sup>gunawan@staff.ukdw.ac.id, <sup>2</sup>anton@ti.ukdw.ac.id

### Abstrak

Penelitian ini membandingkan akurasi prediksi menggunakan regresi logistik terhadap kategori Indeks Prestasi (IP) semester pertama (IPS1) mahasiswa jalur prestasi dengan mahasiswa non jalur prestasi di Fakultas Teknologi Informasi (FTI) Universitas Kristen Duta Wacana (UKDW). Variabel dependen pada jalur prestasi adalah kategori IP semester pertama sedangkan variabel independen yang digunakan adalah lokasi, kategori, status SMA, dan *level* tes kemampuan bahasa Inggris. Variabel untuk model regresi logistik untuk non jalur prestasi adalah sama seperti jalur prestasi namun ditambah empat variabel lain, yaitu hasil tes nilai numerik, verbal, spasial, dan analogi. Sebagai *dataset* untuk membuat model regresi logistik digunakan data mahasiswa FTI angkatan 2008-2015, sedangkan data ujinya menggunakan data angkatan 2016. Ketepatan atau akurasi kedua model dalam memprediksi tersebut diukur dengan menggunakan tabel *Crosstab*. Penelitian ini menghasilkan delapan model regresi logistik, yaitu model A-H. Nilai akurasi tertinggi untuk jalur prestasi adalah model F (*dataset* 2008-2013) dengan akurasi prediksi 73,73%. Hasil akurasi tertinggi untuk jalur non prestasi adalah dengan model D (*dataset* 2008-2011) dan E (*dataset* 2008-2012) dengan akurasi prediksi 56,76%. Dilihat dari segi akurasinya, model yang dihasilkan lebih baik digunakan untuk memprediksi jalur prestasi daripada jalur non prestasi.

**Kata Kunci:** regresi logistik, prediksi, jalur prestasi, jalur non prestasi.

### Abstract

*This research compares accuracy of prediction of Grade Point Average (GPA) of the first semester students who were admitted through Achievement Admission (AA) and Regular Admission (RA) using logistic regression in Faculty of Information Technology, Universitas Kristen Duta Wacana (UKDW), Yogyakarta. The dependent variables of logistic regression on AA was the category of first semester GPA, meanwhile the independent variables were location, category, high school status, and level of English proficiency. The variables of logistic regression on RA were the same as on AA along with four additional variables: the results of numeric, verbal, spatial and analogy test. The dataset used to create logistic regression model were FTI students, 2008 – 2015 batch, while the testing data were FTI students 2016 batch. This research produced 8 logistics regression model for AA and RA, model A – H. The accuracy of prediction was measured by using Crosstab table. The Crosstab table showed that the highest accuracy of AA was model F (dataset 2008-2013 with accuracy 73.73%), while the highest accuracy of RA was model D (dataset 2008-2011) and E (dataset 2008-2012) with accuracy 56.76%. From these accuracy results, our model was better to predict AA rather than RA.*

**Keywords:** *logistics regression, prediction, achievement admission, regular admission*

**Tanggal Terima Naskah** : 20 Oktober 2017  
**Tanggal Persetujuan Naskah** : 06 November 2017

## 1. PENDAHULUAN

Pada era komputerisasi saat ini, *data mining* telah digunakan untuk berbagai keperluan terutama di bidang industri dan organisasi secara umum [1]. *Data mining* telah sukses diterapkan dalam bidang analisis kegagalan atau analisis kerusakan, periklanan, *marketing*, dan analisis risiko pemberian kredit [2]. Akhir-akhir ini, *data mining* juga digunakan dalam bidang pendidikan, terutama dalam *educational data mining* [3]. *Data mining* digunakan untuk analisis dan prediksi data-data mahasiswa di perguruan tinggi yang belum digunakan sehingga data tersebut dapat menolong dalam pengambilan keputusan/kebijakan untuk meningkatkan performa mahasiswa tersebut.

Salah satu tolok ukur keberhasilan hasil belajar mahasiswa di perguruan tinggi dapat dilihat dari Indeks Prestasi Kumulatif (IPK). Semakin baik IPK berarti mahasiswa tersebut dapat menjalani proses belajar mengajar dengan baik. Di beberapa perguruan tinggi, seperti Universitas Kristen Duta Wacana Yogyakarta (UKDW), IPK mahasiswa memegang peran penting dalam ketepatan waktu penyelesaian studi mahasiswa. Pada penelitian terdahulu Santosa dan Setiadi pada tahun 2015 telah menyelidiki tentang faktor-faktor yang mempengaruhi IPK mahasiswa secara signifikan, namun belum melihat hubungan yang jelas antara IPK dan faktor-faktor tersebut. Ternyata Indeks Prestasi Semester satu (IPS1) mempunyai korelasi yang erat dengan indeks prestasi semester sesudahnya [4].

Di UKDW terdapat dua jalur proses seleksi penerimaan mahasiswa baru, yaitu jalur prestasi dan non jalur prestasi. Jalur prestasi adalah penerimaan mahasiswa baru melalui nilai rapor SMA, sedangkan non jalur prestasi adalah penerimaan mahasiswa baru melalui nilai tes numerik, verbal, spasial, dan analogi. Tulisan ini merupakan kelanjutan dari tulisan sebelumnya yang telah diterbitkan oleh Santosa dan Chrismanto pada tahun 2017. Pada penelitian tersebut didapat bahwa presentase keakuratan model regresi logistik untuk memprediksi kategori IPS1 mahasiswa angkatan 2015 melalui jalur prestasi di Fakultas Teknologi Informasi UKDW adalah 72% [5].

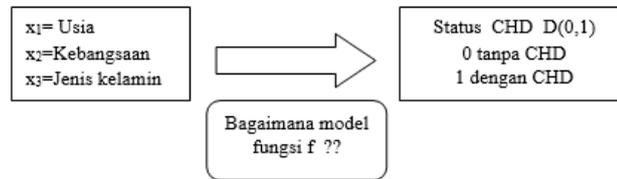
Pada penelitian ini akan dibandingkan akurasi model Regresi Logistik yang digunakan untuk memprediksi kategori IPS1 mahasiswa angkatan 2016 melalui jalur prestasi dengan akurasi model regresi logistik yang digunakan untuk memprediksi kategori IPS1 melalui non jalur prestasi. Di antara dua macam cara penerimaan mahasiswa baru angkatan 2016 tersebut, model regresi logistik mana yang mempunyai presentase keakuratan yang lebih tinggi dalam memprediksi kategori IPS1.

## 2. KONSEP DASAR

Regresi linier mempunyai keterbatasan, yaitu tidak dapat menangani variabel dependen yang bersifat dikotomi ataupun kategori, padahal banyak variabel dikotomi yang menarik untuk diteliti. Beberapa penelitian yang dapat dilakukan, diantaranya permasalahan pengambilan keputusan untuk membeli atau tidak membeli suatu produk, pada manajemen pengendalian kualitas digunakan untuk bisa melabelkan apakah sebuah

produk akan lolos atau gagal pada proses pengendalian mutu, pada sebuah perusahaan besar digunakan untuk menentukan seorang karyawan dapat dipromosikan atau tidak berdasarkan kriteria tertentu, pada bank dapat digunakan untuk mengelompokkan seorang yang akan mengambil kredit apakah pengajuan kreditnya diterima atau ditolak berdasarkan data-data yang ada, demikian pula pada perusahaan asuransi, dapat digunakan untuk mengelompokkan orang yang akan mengambil premi apakah diluluskan atau tidak berdasarkan keadaan orang maupun data riwayat hidupnya. Berdasarkan kasus-kasus tersebut, maka dikembangkanlah regresi logistik [6].

Pada awalnya regresi logistik dikembangkan dan diaplikasikan melalui riset epidemiologi bidang kedokteran. Model yang biasa digunakan untuk penelitian ini adalah apa hubungan antara satu atau lebih variabel independen (*exposure*) terhadap kondisi penyakit yang diderita oleh penderita. Kondisi penyakit yang diderita merupakan variabel bernilai biner, yaitu nilai 0 mewakili kondisi tidak mengalami sakit tertentu sedangkan nilai 1 mewakili mengalami sakit tertentu. Kleinbaum D.G. dan Klein M. pada tahun 2002 menyebutkan hasil dari model ini adalah status penderita yang diklasifikasikan sebagai 0 (“tanpa CHD”) dan 1 (“dengan CHD”) dimana dalam hal ini CHD merupakan kependekan dari *Coronary Heart Disease* [7]. Dalam penelitian tersebut variabel independen yang digunakan adalah  $x_1$  yang menunjukkan usia,  $x_2$  kebangsaan, dan  $x_3$  adalah jenis kelamin. Gambar 1 menunjukkan skema model regresi logistik pada penelitian tersebut.



Gambar 1. Skema Model Regresi Logistik

Terdapat dua kegunaan utama dari regresi logistik, yang pertama adalah untuk prediksi keanggotaan kelompok, karena regresi logistik dapat menghitung probabilitas suatu variabel tertentu apabila variabel data independen diketahui. Yang kedua regresi logistik juga memberikan pengetahuan tentang hubungan dan kekuatan diantara variabel-variabel. Beberapa asumsi penting mengenai regresi logistik adalah:

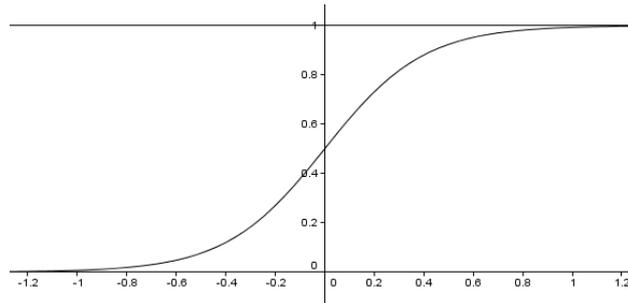
1. Regresi logistik tidak mengasumsikan hubungan linier antara variabel dependen dan variabel independen.
2. Variabel dependen harus mempunyai sifat dikotomi (2 kategori).
3. Variabel independen tidak perlu variabel yang bersifat interval, tidak perlu berdistribusi normal, tidak perlu bersifat linier, dan tidak perlu ada asumsi kesamaan variansi dalam setiap grup atau kategori.
4. Kategori (kelompok) harus saling eksklusif dan lengkap sebuah kasus hanya bisa menjadi anggota sebuah kelompok dan setiap kasus harus menjadi anggota di satu kelompok.
5. *Sample* yang lebih besar diperlukan pada regresi logistik karena perhitungan koefisien *likelihood* maksimum memerlukan validitas *sample* yang lebih besar [6].

Regresi Logistik adalah pendekatan model yang dapat digunakan untuk menggambarkan relasi beberapa variabel independen X terhadap variabel dependen D yang bernilai biner. Untuk para peneliti epidemiologi, maka regresi logistik ini adalah prosedur yang paling populer, bila data kondisi penyakit ini bersifat biner. Regresi logistik didasarkan pada fungsi logistik yang dapat dilihat pada persamaan (1) berikut:

$$f(z) = \frac{1}{1+e^{-z}} \dots\dots\dots(1)$$

dengan  $-\infty < z < \infty$

Bentuk grafik fungsi logistik adalah seperti pada Gambar 2 berikut ini:



Gambar 2. Grafik fungsi logistik

Terlihat pada Gambar 2 bahwa *range* nilai  $f(z)$  terletak di antara 0 sampai dengan 1 tanpa memandang nilai  $z$ . Untuk  $z \rightarrow -\infty$  maka  $f(-\infty) = 0$  dan untuk  $z \rightarrow \infty$  maka  $f(\infty) = 1$ . Karena nilai  $f(z)$  terletak di antara 0 sampai dengan 1 inilah maka alasan regresi logistik menjadi populer. Model ini dirancang untuk menggambarkan probabilitas yang selalu terletak antara 0 sampai dengan 1. Dalam istilah epidemiologi probabilitas seperti ini memberikan nilai risiko dari seorang penderita untuk mendapatkan penyakit tertentu [8]. Alasan lain mengapa model logistik menjadi populer adalah karena bentuk dari fungsi logistik berupa kurva S. Seperti dilihat pada Gambar 2, jika dimulai dari  $z = -\infty$  bergerak ke kanan maka nilai  $f(z)$  naik melengkung tajam mendekati  $z=0$  untuk sementara, kemudian menurun secara dramatis mendekati 1 dan akhirnya akan mendekati di sekitar 1. Hasil ini menggambarkan bentuk kurva S atau Sigmoid.

Pada penelitian ini akan diterapkan metode regresi logistik untuk *educational data mining*. Beberapa literatur dan penelitian yang berkaitan dalam bidang *educational data mining* adalah:

- Aher, Sunita, dan Lobo pada tahun 2011 melakukan penelitian yang berkaitan dengan analisis dan performa siswa di tahun terakhir. Mereka menggunakan klasifikasi dan klusterisasi untuk mengelompokkan siswa-siswa tersebut [9].
- Youping dan Gong pada tahun 2010 melakukan penelitian yang berkaitan dengan evaluasi siswa sekolah menengah atas dan mempelajari keefektifannya [10].
- Liu, Zhiwu, dan Zhang pada tahun 2010 melakukan penelitian dengan membangun model peramalan tanda pada nilai siswa untuk mengidentifikasi kebiasaan belajar yang negatif atau sifat-sifat negatif yang dimiliki oleh siswa [11].
- Osmanbegovic, Edin, dan Mirza Suljic pada tahun 2012 melakukan penelitian tentang pencarian atau penemuan model prediksi untuk kinerja akademik yang *user friendly* untuk para profesor dan juga untuk para pengguna yang *non-expert*. Mereka menggunakan Uji Chi-Square, *R-test*, *Info Gain and Ratio Test*, Naïve Bayes dan *Decicion Tree* [12].
- Sukanya, Biruntha, Karthik, dan Kalaikumaran pada tahun 2012 melakukan penelitian tentang analisis dan pertolongan (pembimbingan) bagi siswa dengan kemampuan akademik yang rendah di pendidikan sekolah menengah atas. Mereka menggunakan metode klasifikasi Bayesian [13].

Model regresi logistik dinyatakan dalam fungsi  $f$  dan sebagai fokus utama adalah  $z$  yang dapat dituliskan sebagai kombinasi linier dari  $x_1, x_2, x_3, \dots, x_k$ , dan dapat dilihat pada persamaan (2) dan (3) berikut:

$$z = \alpha + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k \dots\dots\dots(2)$$

sehingga  $\hat{f}(z) = \frac{1}{1 + e^{-(\alpha + \sum \beta_i x_i)}} \hat{f}(z) = \frac{1}{1 + e^{-(\alpha + \sum \beta_i x_i)}} \dots\dots\dots(3)$

$$\hat{f}(z) = \frac{1}{1 + e^{-(\alpha + \sum \beta_i x_i)}} \hat{f}(z) = \frac{1}{1 + e^{-(\alpha + \sum \beta_i x_i)}}$$

Nilai  $\alpha$  dan  $\beta_i$  merupakan parameter yang tidak diketahui dan akan diestimasi dari *data set* yang berupa *sample* dengan menggunakan *Maksimum Likelihood Estimator* (MLE) [14] sehingga parameter  $\alpha$  dan  $\beta_i$  akan diestimasi oleh  $\hat{\alpha}$  dan  $\hat{\beta}_i$ . Ukuran ketepatan hasil prediksi dengan model regresi logistik diukur menggunakan *Crosstab* (tabulasi silang) antara data hasil prediksi dan nilai data sebenarnya seperti pada Tabel 1 berikut.

Tabel 1. Tabel *Crosstab*

		Nilai Data Hasil Prediksi		Total Baris
		Kategori 1	Kategori 2	
Nilai Data Observasi Sebenarnya	Kategori 1	$n_{11}$	$n_{12}$	$b_1$
	Kategori 2	$n_{21}$	$n_{22}$	$b_2$
Total Kolom		$k_1$	$k_2$	$n$

Keterangan:

- Nilai  $n_{11}$  adalah banyaknya data hasil prediksi yang masuk kategori 1 dan kenyataan pada data observasi sesungguhnya nilai tersebut merupakan kategori 1.
- Nilai  $n_{12}$  adalah banyaknya data hasil prediksi yang masuk kategori 1 tapi kenyataan pada data observasi sesungguhnya nilai tersebut merupakan kategori 2.
- Nilai  $n_{21}$  adalah banyaknya data hasil prediksi yang masuk kategori 2 tapi kenyataan pada data observasi sesungguhnya nilai tersebut merupakan kategori 1.
- Nilai  $n_{22}$  adalah banyaknya data hasil prediksi yang masuk kategori 2 dan kenyataan pada data observasi sesungguhnya nilai tersebut merupakan kategori 2.

Ketepatan hasil prediksi dengan menggunakan model regresi logistik dapat dihitung dengan banyaknya data hasil prediksi yang masuk kategori 1 dan kenyataan pada data observasi sesungguhnya nilai tersebut merupakan kategori 1 ditambah dengan banyaknya data hasil prediksi yang masuk kategori 2, serta kenyataan pada data observasi sesungguhnya nilai tersebut merupakan kategori 2, kemudian dibagi dengan jumlah seluruh data, seperti yang dapat dilihat pada persamaan (4) [15]:

$$\text{Presentasi kecocokan} = (n_{11} + n_{22})/n \dots \dots \dots (4)$$

### 3. METODE PENELITIAN

Pada penelitian ini variabel independen yang akan digunakan untuk memprediksi IPS1 adalah lokasi sekolah menengah atas, yaitu SMA di Jawa dan SMA di luar Jawa. Kategori sekolah menengah atas, yaitu Sekolah Menengah Atas (SMA) dan Sekolah Menengah Kejuruan (SMK). Demikian pula status SMA, yaitu SMA/SMK Negeri atau SMA/SMK Swasta. *Level* bahasa Inggris dari calon mahasiswa juga digunakan sebagai faktor yang mempengaruhi IP mahasiswa, yaitu *level* 1, 2, 3, dan ESP (*English for Special Purpose*).

Tahapan-tahapan penelitian yang dilakukan adalah sebagai berikut:

1. Peneliti mengambil data berupa data mahasiswa UKDW dari Puspindika yang meliputi data mahasiswa Fakultas Teknologi Informasi, yang terdiri dari program studi SI dan program studi TI. Sebagai batasan hanya akan diambil data mahasiswa SI dan TI sebanyak delapan tahun akademik yang berurutan, yaitu dari angkatan 2008 sampai dengan angkatan 2016. Data dari Puspindika meliputi kategori SMA, status SMA, lokasi SMA, dan IPS1.
2. Peneliti juga mengambil data kemampuan tes masuk calon mahasiswa baru yang ada di unit Humas UKDW. Data ini meliputi hasil nilai potensi akademik, yaitu hasil tes kemampuan *spatial*, hasil tes kemampuan verbal, hasil tes kemampuan numerik, dan hasil tes kemampuan analogi.

3. Peneliti juga mengambil data kemampuan bahasa Inggris calon mahasiswa FTI UKDW dari unit Pusat Pelatihan Bahasa (PPB) UKDW.
4. Langkah berikutnya, peneliti melakukan pemisahan data yang berasal dari penerimaan jalur prestasi dan jalur reguler (non jalur prestasi). Perbedaan antara dua jalur ini adalah pada proses penerimaannya, pada jalur prestasi tidak mengikuti tes kemampuan akademik (tes kemampuan numerik, verbal, spasial, dan analogi) sedangkan pada jalur non prestasi harus mengikuti empat tes kemampuan akademik tersebut.
5. Langkah berikutnya adalah data *cleaning*, yaitu membersihkan data dari data yang tidak berguna, data kosong, dan data *double*.
6. Setelah data dibersihkan dilakukan integrasi data, yaitu proses integrasi dari tiap kolom (*field*) dan baris (*row*).
7. Langkah berikutnya, data tersebut diolah dengan menggunakan regresi logistik. Model regresi logistik sebagai fokus utama adalah  $z$  yang dapat dituliskan sebagai kombinasi linier dari  $x_1, x_2, x_3, \dots, x_k$ , yang dapat dilihat pada persamaan (2) dan (3). Nilai  $\alpha$  dan  $\beta_i$  merupakan parameter yang tidak diketahui dan akan diestimasi dari data *sample* dengan menggunakan *Maksimum Likelihood Estimator* (MLE) sehingga  $\alpha$  dan  $\beta_i$  akan diestimasi oleh  $\hat{\alpha}$  dan  $\hat{\beta}_i$ .
8. Langkah berikutnya adalah mencari delapan model Regresi Logistik untuk kasus penerimaan mahasiswa baru melalui jalur prestasi dimana:
  - Model A adalah model regresi logistik dengan menggunakan data mahasiswa angkatan 2008
  - Model B adalah model regresi logistik dengan menggunakan data mahasiswa angkatan 2008 sampai dengan angkatan 2009
  - Model C adalah model regresi logistik dengan menggunakan data mahasiswa angkatan 2008 sampai dengan angkatan 2010
  - Model D adalah model regresi logistik dengan menggunakan data mahasiswa angkatan 2008 sampai dengan angkatan 2011
  - Model E adalah model regresi logistik dengan menggunakan data mahasiswa angkatan 2008 sampai dengan angkatan 2012
  - Model F adalah model regresi logistik dengan menggunakan data mahasiswa angkatan 2008 sampai dengan angkatan 2013
  - Model G adalah model regresi logistik dengan menggunakan data mahasiswa angkatan 2008 sampai dengan angkatan 2014
  - Model H adalah model regresi logistik dengan menggunakan data mahasiswa angkatan 2008 sampai dengan angkatan 2015
9. Langkah berikutnya adalah mencari delapan model regresi logistik untuk kasus penerimaan mahasiswa baru melalui non jalur prestasi dimana:
  - Model A adalah model regresi logistik dengan menggunakan data mahasiswa angkatan 2008
  - Model B adalah model regresi logistik dengan menggunakan data mahasiswa angkatan 2008 sampai dengan angkatan 2009
  - Model C adalah model regresi logistik dengan menggunakan data mahasiswa angkatan 2008 sampai dengan angkatan 2010
  - Model D adalah model regresi logistik dengan menggunakan data mahasiswa angkatan 2008 sampai dengan angkatan 2011
  - Model E adalah model regresi logistik dengan menggunakan data mahasiswa angkatan 2008 sampai dengan angkatan 2012

- Model F adalah model regresi logistik dengan menggunakan data mahasiswa angkatan 2008 sampai dengan angkatan 2013
- Model G adalah model regresi logistik dengan menggunakan data mahasiswa angkatan 2008 sampai dengan angkatan 2014
- Model H adalah model regresi logistik dengan menggunakan data mahasiswa angkatan 2008 sampai dengan angkatan 2015

Digunakan model kumulatif dari model A sampai dengan model H agar dapat dilihat perubahan variabel dependen dalam mempengaruhi variabel independen.

10. Dengan menggunakan delapan model Regresi Logistik (model A sampai dengan model H) dilakukan untuk memprediksi kategori IPS1 mahasiswa baru angkatan 2016 yang penerimaannya melalui jalur prestasi dan non jalur prestasi.
11. Mengukur ketepatan hasil prediksi regresi logistik dari ketujuh model tersebut terhadap data observasi sebenarnya dengan menggunakan *crossstab* (tabulasi silang) untuk penerimaan melalui jalur prestasi dan non jalur prestasi.
12. Membandingkan hasil prediksi delapan model regresi logistik untuk jalur prestasi dengan delapan model regresi untuk jalur non prestasi dari sisi keakuratannya.

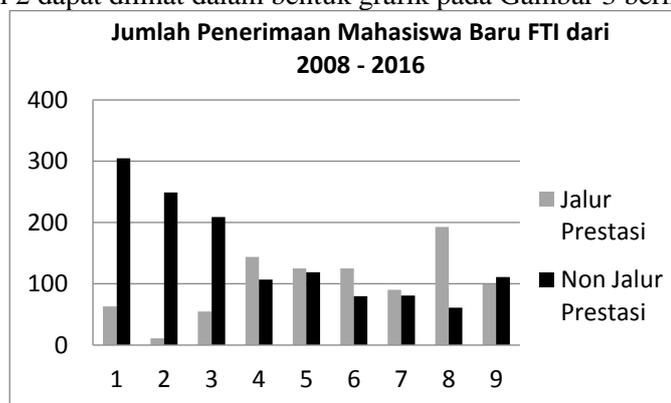
#### 4. HASIL DAN PEMBAHASAN

Setelah dilakukan proses data *cleaning* untuk *dataset* yang telah diambil, maka dapat diperoleh jumlah kumulatif mahasiswa FTI yang akan digunakan untuk membuat regresi logistik yang ditabelkan sebagai Tabel 2 dan 3 berikut ini.

Tabel 2. Jumlah mahasiswa Fakultas Teknologi Informasi dari tahun 2008 sampai dengan tahun 2016

No	Angkatan	Total	Jalur Prestasi	Non Jalur Prestasi
1	2008	368	63	305
2	2009	260	11	249
3	2010	264	55	209
4	2011	251	144	107
5	2012	244	125	119
6	2013	205	125	80
7	2014	171	90	81
8	2015	254	193	61
9	2016	210	99	111

Data pada Tabel 2 dapat dilihat dalam bentuk grafik pada Gambar 3 berikut.



Gambar 3. Grafik Jumlah Penerimaan Mahasiswa Baru FTI dari tahun 2008–2016

Tabel 3. Jumlah kumulatif mahasiswa Fakultas Teknologi Informasi dari tahun 2008 sampai dengan tahun 2015

No	Angkatan	Total	Jalur Prestasi	Jalur Non Prestasi
1	2008	368	63	305
2	2008 sampai dengan 2009	628	74	554
3	2008 sampai dengan 2010	892	129	763
4	2008 sampai dengan 2011	1143	273	870
5	2008 sampai dengan 2012	1387	398	989
6	2008 sampai dengan 2013	1592	523	1069
7	2008 sampai dengan 2014	1763	613	1150
8	2008 sampai dengan 2015	2017	806	1211
9	2016	210	99	111

Karena data variabel dependen harus data dikotomi D(0,1) maka Indeks Prestasi Semester 1 (IPS1) dijadikan data biner, yaitu  $0 < \text{IPS1} < \text{rata-rata IPS1}$  dikodekan menjadi "0" dan  $(\text{rata-rata IPS1} + 0,0001) < \text{IPS1} < 4,000$  dikodekan menjadi "1", dalam hal ini rata-rata diambil berdasarkan seluruh *sample*. Jadi kode "0" berarti IPS1 yang di bawah rata-rata dan kode "1" berarti IPS1 di atas rata-rata. Tahap berikutnya adalah *data transformation* (data ditransformasikan atau diubah menjadi bentuk yang cocok untuk diolah atau dianalisis). Data kategori SMA ("1" untuk SMA dan "2" untuk SMK), status SMA ("1" untuk Negeri dan "2" untuk Swasta), lokasi SMA ("1" untuk Jawa dan "2" untuk Luar Jawa) dan *level* bahasa Inggris (*level\_1* dikodekan menjadi "1", *level\_2* dikodekan menjadi "2", *level\_3* dikodekan menjadi "3", dan ESP dikodekan menjadi "4") transformasi dari data alfa numerik menjadi data numerik karena variabel independent pada regresi logistik membutuhkan data bertipe numerik.

Pada penerimaan mahasiswa melalui jalur prestasi melibatkan variabel status SMA ( $x_1$ ), lokasi SMA ( $x_2$ ), kategori SMA ( $x_3$ ), *level* bahasa Inggris ( $x_4$ ), sedangkan untuk non jalur prestasi masih ditambah lagi variabel hasil nilai Numerik ( $x_5$ ), hasil nilai Verbal ( $x_6$ ), hasil nilai Spasial ( $x_7$ ), dan hasil nilai Analogi ( $x_8$ ). Oleh karena itu, model regresi logistik jalur prestasi hanya mempunyai empat variabel independen  $x_1, x_2, x_3$  dan  $x_4$ , untuk non jalur prestasi mempunyai delapan variabel independen, yaitu  $x_1$  sampai dengan  $x_8$ . Seperti pada kasus regresi biasa model regresi logistik yang sudah ada dapat disederhanakan dengan uji signifikansi koefisien beta pada model. Berikut ini merupakan bentuk model-model hasil komputasi sebelum dan sesudah disederhanakan dengan uji signifikansi. Uji signifikansi yang digunakan dalam penelitian ini adalah uji yang berdasarkan [16]:

$$H_0 : \beta_i = 0 \text{ vs } H_a : \beta_i \neq 0 \text{ dengan statistik uji } t = \frac{\hat{\beta}_i}{s(\hat{\beta}_i)}$$

dimana  $\hat{\beta}_i$  adalah dugaan parameter  $\beta_i$  dan  $s(\hat{\beta}_i)$  adalah standar deviasi dugaan parameter  $\beta_i$

Alasan digunakan uji signifikansi  $\beta_i$  adalah:

- Melihat atau mempertimbangkan koefisien regresi yang benar-benar berpengaruh terhadap variabel dependen.
- Mengurangi banyak variabel independen pada model sehingga akan didapat model regresi logistik yang lebih ringkas tanpa mengabaikan akurasi model tersebut.

Model A adalah model regresi logistik dengan menggunakan data mahasiswa angkatan 2008, untuk jalur prestasi adalah:

$$P(\mathbf{x}) = 1/(1 + \exp\{-[-21,331 - 0,126(x_1) - 0,583(x_2) + 20,219(x_3) + 1,403(x_4)]\})$$

dan setelah disederhanakan modelnya menjadi:

$$P(\mathbf{x}) = 1/(1 + \exp\{-[-2,113 + 1,438(x_4)]\})$$

untuk non jalur prestasi adalah:

$$P(\mathbf{x}) = 1/(1+\exp\{-[0,760 - 0,260(x_1) - 0,591(x_2) + 0,253(x_3) + 0,475(x_4) + 0,003(x_5) + 0,001(x_6) - 0,001(x_7) - 0,002(x_8)]\})$$

dan setelah disederhanakan modelnya menjadi:

$$P(\mathbf{x}) = 1/(1+\exp\{-[0,700 - 0,528(x_2) + 0,461(x_4)]\})$$

Model B adalah model regresi logistik dengan menggunakan data mahasiswa angkatan 2008 – 2009, untuk jalur prestasi adalah:

$$P(\mathbf{x}) = 1/(1+\exp\{-[-22,137 + 0,561(x_1) - 0,243(x_2) + 20,680(x_3) + 0,831(x_4)]\})$$

dan setelah disederhanakan modelnya menjadi:

$$P(\mathbf{x}) = 1/(1+\exp\{-[-0,700 + 0,760(x_4)]\})$$

untuk non jalur prestasi adalah:

$$P(\mathbf{x}) = 1/(1+\exp\{-[-0,154 - 0,135(x_1) - 0,647(x_2) - 0,127(x_3) + 0,804(x_4) + 0,000(x_5) + 0,000(x_6) - 0,001(x_7) + 0,002(x_8)]\})$$

dan setelah disederhanakan modelnya menjadi:

$$P(\mathbf{x}) = 1/(1+\exp\{-[-0,461 - 0,611(x_2) + 0,833(x_4)]\})$$

Model C adalah model regresi logistik dengan menggunakan data mahasiswa angkatan 2008 – 2010, untuk jalur prestasi adalah:

$$P(\mathbf{x}) = 1/(1+\exp\{-[-1,033 - 0,031(x_1) - 1,116(x_2) + 1,321(x_3) + 0,869(x_4)]\})$$

dan setelah disederhanakan modelnya menjadi:

$$P(\mathbf{x}) = 1/(1+\exp\{-[-2,519 - 0,378(x_2)]\})$$

untuk non jalur prestasi adalah:

$$P(\mathbf{x}) = 1/(1+\exp\{-[-0,131 - 0,146(x_1) - 0,592(x_2) - 0,215(x_3) + 0,644(x_4) + 0,000(x_5) + 0,002(x_6) + 0,000(x_7) + 0,002(x_8)]\})$$

dan setelah disederhanakan modelnya menjadi:

$$P(\mathbf{x}) = 1/(1+\exp\{-[-0,251 - 0,603(x_2) + 0,726(x_4)]\})$$

Model D adalah model regresi logistik dengan menggunakan data mahasiswa angkatan 2008 – 2011, untuk jalur prestasi adalah:

$$P(\mathbf{x}) = 1/(1+\exp\{-[-0,573 - 0,134(x_1) - 1,363(x_2) + 0,066(x_3) + 0,812(x_4)]\})$$

dan setelah disederhanakan modelnya menjadi :

$$P(\mathbf{x}) = 1/(1+\exp\{-[-0,390 - 1,337(x_2) + 0,813(x_4)]\})$$

untuk non jalur prestasi adalah :

$$P(\mathbf{x}) = 1/(1+\exp\{-[-0,140 - 0,133(x_1) - 0,620(x_2) - 0,011(x_3) + 0,477(x_4) + 0,000(x_5) + 0,002(x_6) + 0,001(x_7) + 0,002(x_8)]\})$$

dan setelah disederhanakan modelnya menjadi :

$$P(\mathbf{x}) = 1/(1+\exp\{-[-0,184 - 0,662(x_2) + 0,575(x_4)]\})$$

Model E adalah model regresi logistik dengan menggunakan data mahasiswa angkatan 2008 – 2012, untuk jalur prestasi adalah :

$$P(\mathbf{x}) = 1/(1+\exp\{-[-0,896 - 0,230(x_1) - 1,105(x_2) + 0,145(x_3) + 0,640(x_4)]\})$$

dan setelah disederhanakan modelnya menjadi :

$$P(\mathbf{x}) = 1/(1+\exp\{-[-0,579 - 1,029(x_2) + 0,639(x_4)]\})$$

untuk non jalur prestasi adalah :

$$P(\mathbf{x}) = 1/(1+\exp\{-[-0,339 - 0,160(x_1) - 0,580(x_2) + 0,043(x_3) + 0,515(x_4) + 0,001(x_5) + 0,001(x_6) + 0,001(x_7) + 0,003(x_8)]\})$$

dan setelah disederhanakan modelnya menjadi :

$$P(\mathbf{x}) = 1/(1+\exp\{-[-0,080 - 0,636(x_2) + 0,617(x_4)]\})$$

Model F adalah model regresi logistik dengan menggunakan data mahasiswa angkatan 2008 – 2013, untuk jalur prestasi adalah :

$$P(\mathbf{x}) = 1/(1+\exp\{-[-0,662 - 0,183(x_1) - 1,096(x_2) - 0,108(x_3) + 0,799(x_4)]\})$$

dan setelah disederhanakan modelnya menjadi :

$$P(\mathbf{x}) = 1/(1+\exp\{-[-0,187 - 1,063(x_2) + 0,808(x_4)]\})$$

untuk non jalur prestasi adalah :

$$P(\mathbf{x}) = 1/(1+\exp\{-[-0,038 - 0,162(x_1) - 0,611(x_2) + 0,091(x_3) + 0,466(x_4) + 0,000(x_5) + 0,001(x_6) + 0,000(x_7) + 0,004(x_8)]\})$$

dan setelah disederhanakan modelnya menjadi :

$$P(\mathbf{x}) = 1/(1+\exp\{-[-0,137 - 0,004(x_8)]\})$$

Model G adalah model regresi logistik dengan menggunakan data mahasiswa angkatan 2008 – 2014 , untuk jalur prestasi adalah :

$$P(\mathbf{x}) = 1/(1+\exp\{-[0,514 - 0,196(x_1) - 1,009(x_2) - 0,197(x_3) + 0,894(x_4)]\})$$

dan setelah disederhanakan modelnya menjadi :

$$P(\mathbf{x}) = 1/(1+\exp\{-[-0,076 - 0,988(x_2) + 0,908(x_4)]\})$$

untuk non jalur prestasi adalah :

$$P(\mathbf{x}) = 1/(1+\exp\{-[-0,275 - 0,130(x_1) - 0,529(x_2) + 0,051(x_3) + 0,449(x_4) + 0,000(x_5) + 0,001(x_6) + 0,002(x_7) + 0,003(x_8)]\})$$

dan setelah disederhanakan modelnya menjadi :

$$P(\mathbf{x}) = 1/(1+\exp\{-[-0,229 - 0,540(x_2) + 0,467(x_4) - 0,004(x_8)]\})$$

Model H adalah model regresi logistik dengan menggunakan data mahasiswa angkatan 2008 – 2015, untuk jalur prestasi adalah :

$$P(\mathbf{x}) = 1/(1+\exp\{-[-0,381 + 0,001(x_1) - 0,189(x_2) - 0,509(x_3) + 0,934(x_4)]\})$$

dan setelah disederhanakan modelnya menjadi :

$$P(\mathbf{x}) = 1/(1+\exp\{-[-0,581 - 0,593(x_3) + 0,954(x_4)]\})$$

untuk non jalur prestasi adalah :

$$P(\mathbf{x}) = 1/(1+\exp\{-[-0,286 - 0,112(x_1) - 0,442(x_2) - 0,031(x_3) + 0,461(x_4) + 0,000(x_5) + 0,001(x_6) + 0,001(x_7) + 0,003(x_8)]\})$$

dan setelah disederhanakan modelnya menjadi :

$$P(\mathbf{x}) = 1/(1+\exp\{-[-0,293 - 0,472(x_2) + 0,483(x_4) - 0,004(x_8)]\})$$

Hasil rangkuman koefisien regresi logistik untuk jalur prestasi dan non jalur prestasi setelah disederhanakan atau setelah uji signifikansi dan nilai kecocokkan model dengan data sesungguhnya dapat diringkas dalam Tabel 4 dan Tabel 5 sebagai berikut:

Tabel 4. Rangkuman Koefisien Regresi Logistik untuk jalur prestasi setelah Uji Signifikansi

Tahun	$\alpha$	$\beta_1$	$\beta_2$	$\beta_3$	$\beta_4$	% cocok
2008	-2,113	0	0	0	1,438	73,0%
2008 sd 2009	-0,700	0	0	0	0,760	71,6%
2008 sd 2010	2,519	0	-1,378	0	0	70,5%
2008 sd 2011	0,390	0	-1,337	0	0,813	69,6%
2008 sd 2012	0,579	0	-1,029	0	0,639	69,1%
2008 sd 2013	0,187	0	-1,063	0	0,808	70,9%
2008 sd 2014	-0,076	0	-0,988	0	0,908	71,6%
2008 sd 2015	-0,581	0	0	-0,593	0,954	71,2%
<b>Rata-Rata</b>						<b>70,94%</b>

Dari Tabel 4, terdapat karakteristik yang dijumpai, yaitu:

- Variabel yang berpengaruh pada model A dan B adalah *level* bahasa Inggris  
 Pada model A dan B, *level* bahasa Inggris terlihat menaik, yang berarti *level* bahasa Inggris semakin meningkat maka kategori IPS1 mahasiswa juga semakin meningkat.
- Variabel yang berpengaruh pada model C adalah lokasi SMA

Pada model C, lokasi SMA (Jawa/Luar Jawa) terlihat menurun, yang berarti lokasi SMA Jawa memiliki kategori IPS1 yang lebih tinggi dibandingkan dengan lokasi SMA luar Jawa.

- Variabel yang berpengaruh pada model D sampai dengan G adalah lokasi SMA dan Bahasa Inggris.

Pada model D, E, F, dan G, lokasi SMA (Jawa/Luar Jawa) terlihat menurun, yang berarti lokasi SMA Jawa memiliki kategori IPS1 yang lebih tinggi dibandingkan dengan lokasi SMA luar Jawa, namun *level* Bahasa Inggris terlihat menaik, yang berarti *level* bahasa Inggris semakin meningkat maka kategori IPS1 mahasiswa juga semakin meningkat.

- Variabel yang berpengaruh pada model H adalah kategori SMA dan *level* bahasa Inggris. Pada model H, seperti pada grafik Lampiran A Gambar 8, kategori SMA (SMA/SMK) *level* bahasa Inggris terlihat menaik, yang berarti *level* bahasa Inggris semakin meningkat maka kategori IPS1 mahasiswa juga semakin meningkat

Hal ini menunjukkan bahwa ada perubahan karakteristik pada model regresi menurut tahun angkatan. Dari Tabel 4 terlihat pula bahwa rata-rata presentase kecocokannya adalah 70,94%. Presentase tertinggi adalah 73,00% (dengan model A), dan presentase terendah adalah 69,1% (dengan model E). Dari Tabel 5, terdapat karakteristik yang dijumpai, yaitu:

- Variabel yang berpengaruh pada model A sampai dengan model E adalah lokasi SMA dan *level* bahasa Inggris

Pada model A, B, C, D, dan E, lokasi SMA (Jawa/Luar Jawa) terlihat menurun, yang berarti lokasi SMA Jawa memiliki kategori IPS1 yang lebih tinggi dibandingkan dengan lokasi SMA luar Jawa, namun *level* Bahasa Inggris terlihat menaik, yang berarti *level* bahasa Inggris semakin meningkat maka kategori IPS1 mahasiswa juga semakin meningkat.

- Variabel yang berpengaruh pada model F sampai dengan model H adalah lokasi SMA, *level* bahasa Inggris, dan kemampuan analogi

Pada model F, G, dan H, lokasi SMA (Jawa/Luar Jawa) terlihat menurun, yang berarti lokasi SMA Jawa memiliki kategori IPS1 yang lebih tinggi dibandingkan dengan lokasi SMA luar Jawa, namun *level* Bahasa Inggris dan kemampuan analogi terlihat menaik, yang berarti *level* bahasa Inggris dan kemampuan analogi semakin meningkat maka kategori IPS1 mahasiswa juga semakin meningkat.

Dari Tabel 5 terlihat pula bahwa rata-rata persentase kecocokannya adalah 65,8%, persentase tertinggi adalah 69,50% (dengan model A) dan persentase terendah adalah 62,8% (dengan model C). Selanjutnya akan digunakan kedelapan model pada Tabel 4 dan Tabel 5 untuk memprediksi kategori IPS1 mahasiswa FTI pada penerimaan mahasiswa melalui jalur prestasi dan persentase kebenarannya dengan tabel *crosstab* pada Tabel 6.

Tabel 5. Rangkuman Koefisien Regresi Logistik untuk non jalur prestasi setelah Uji Signifikansi

Tahun	$\alpha$	$\beta_1$	$\beta_2$	$\beta_3$	$\beta_4$	$\alpha$	$\beta_1$	$\beta_2$	$\beta_3$	% cocok
2008	0,700	0	-0,528	0	0,461	0	0	0	0	69,5%
2008 sd 2009	-0,611	0	-0,611	0	0,833	0	0	0	0	64,4%
2008 sd 2010	-0,251	0	-0,603	0	0,726	0	0	0	0	62,8%
2008 sd 2011	0,184	0	-0,662	0	0,575	0	0	0	0	66,1%
2008 sd 2012	0,080	0	-0,636	0	0,617	0	0	0	0	65,7%
2008 sd 2013	-0,137	0	-0,594	0	0,472	0	0	0	0,004	66,3%
2008 sd 2014	-0,229	0	-0,540	0	0,467	0	0	0	0,004	66,1%
2008 sd 2015	-0,293	0	-0,472	0	0,483	0	0	0	0,004	65,6%
<b>RATA-RATA</b>										<b>65,8%</b>

Tabel 6. Crosstab prediksi kategori IPS1 mahasiswa angkatan 2016 melalui jalur prestasi dengan menggunakan Model A sampai dengan Model H

Model A		Nilai Data Hasil Prediksi		Total Baris	Model B		Nilai Data Hasil Prediksi		Total Baris
		0	1				0	1	
Nilai Data Obeservasi Sebenarnya	0	0	39	39	Nilai Data Obeservasi Sebenarnya	0	0	39	39
	1	0	60	60		1	0	60	60
Total Kolom		0	99	99	Total Kolom	0	99	99	

Akurasi Model A = 60,60%, Akurasi Model B = 60,60%.

Model C		Nilai Data Hasil Prediksi		Total Baris	Model D		Nilai Data Hasil Prediksi		Total Baris
		0	1				0	1	
Nilai Data Obeservasi Sebenarnya	0	0	39	39	Nilai Data Obeservasi Sebenarnya	0	32	7	39
	1	0	60	60		1	20	40	60
Total Kolom		0	99	99	Total Kolom	52	47	99	

Akurasi Model C = 60,60%, Akurasi Model D = 72,72%

Model E		Nilai Data Hasil Prediksi		Total Baris	Model F		Nilai Data Hasil Prediksi		Total Baris
		0	1				0	1	
Nilai Data Obeservasi Sebenarnya	0	32	7	39	Nilai Data Obeservasi Sebenarnya	0	22	17	39
	1	20	40	60		1	9	51	60
Total Kolom		52	47	99	Total Kolom	31	68	99	

Akurasi Model E = 72,72%, Akurasi Model F = 73,73%

Model G		Nilai Data Hasil Prediksi		Total Baris	Model H		Nilai Data Hasil Prediksi		Total Baris
		0	1				0	1	
Nilai Data Obeservasi Sebenarnya	0	32	7	39	Nilai Data Obeservasi Sebenarnya	0	23	16	39
	1	20	40	60		1	15	45	60
Total Kolom		52	47	99	Total Kolom	38	61	99	

Akurasi Model G = 72,72%, Akurasi Model H = 68,68%

Selanjutnya digunakan kedelapan model pada Tabel 5 untuk memprediksi kategori IPS1 mahasiswa FTI pada penerimaan mahasiswa melalui non jalur prestasi dan persentase kebenarannya dengan tabel *crosstab* pada Tabel 7.

Tabel 7. Crosstab prediksi kategori IP semester 1 mahasiswa angkatan 2016 melalui non jalur prestasi dengan menggunakan Model A sampai dengan Model H

Model A		Nilai Data Hasil Prediksi		Total Baris	Model B		Nilai Data Hasil Prediksi		Total Baris
		0	1				0	1	
Nilai Data Observasi Sebenarnya	0	0	52	52	Nilai Data Observasi Sebenarnya	0	43	9	52
	1	0	59	59		1	42	17	59
Total Kolom		0	111	111	Total Kolom		85	26	111

Akurasi Model A = 53,15%, Akurasi Model B = 54,05%.

Model C		Nilai Data Hasil Prediksi		Total Baris	Model D		Nilai Data Hasil Prediksi		Total Baris
		0	1				0	1	
Nilai Data Observasi Sebenarnya	0	43	9	52	Nilai Data Observasi Sebenarnya	0	26	26	52
	1	42	17	59		1	22	37	59
Total Kolom		85	26	111	Total Kolom		48	63	111

Akurasi Model C = 54,05%, Akurasi Model D = 56,76%

Model E		Nilai Data Hasil Prediksi		Total Baris	Model F		Nilai Data Hasil Prediksi		Total Baris
		0	1				0	1	
Nilai Data Observasi Sebenarnya	0	26	26	52	Nilai Data Observasi Sebenarnya	0	36	16	52
	1	22	37	59		1	33	26	59
Total Kolom		48	63	111	Total Kolom		69	42	111

Akurasi Model E = 56,76%, Akurasi Model F = 55,86%

Model G		Nilai Data Hasil Prediksi		Total Baris	Model H		Nilai Data Hasil Prediksi		Total Baris
		0	1				0	1	
Nilai Data Observasi Sebenarnya	0	36	16	52	Nilai Data Observasi Sebenarnya	0	32	20	52
	1	33	26	59		1	31	28	59
Total Kolom		69	42	111	Total Kolom		63	48	111

Akurasi Model G = 55,86%, Akurasi Model H = 54,05%

Tabel 8. Persentase kebenaran prediksi kategori IP semester 1 mahasiswa angkatan 2016 melalui jalur prestasi dengan menggunakan Model A sampai dengan Model H

Model Prediksi	Model A	Model B	Model C	Model D	Model E	Model F	Model G	Model H
Persentase Kebenaran	60,60%	60,60%	60,60%	72,72%	72,72%	73,73%	72,72%	68,68%

Dari Tabel 8 terlihat bahwa rata-rata persentase kebenaran untuk prediksi Kategori IPS1 Mahasiswa Angkatan 2016 melalui Jalur Prestasi dengan menggunakan Model A sampai dengan Model H adalah 67,80 % dengan standar deviasi 6,14.

Tabel 9. Persentase kebenaran prediksi kategori IPS1 mahasiswa angkatan 2016 melalui non jalur prestasi dengan menggunakan Model A sampai dengan Model H

Model Prediksi	Model A	Model B	Model C	Model D	Model E	Model F	Model G	Model H
Persentase Kebenaran	53,15%	54,05%	54,05%	56,76%	56,76%	55,86%	55,86%	54,05%

Dari Tabel 9 terlihat bahwa rata-rata persentase kebenaran untuk prediksi kategori IPS1 mahasiswa angkatan 2016 melalui non jalur prestasi dengan menggunakan Model A sampai dengan Model H adalah 55,07 % dengan standar deviasi 1,04.

## 5. KESIMPULAN

Dari penelitian yang telah dilakukan, didapat beberapa kesimpulan sebagai berikut:

1. Hasil akurasi rata-rata model prediksi kategori IPS1 mahasiswa jalur prestasi setelah uji signifikansi adalah 70,94%
2. Hasil akurasi rata-rata model prediksi kategori IPS1 mahasiswa non jalur prestasi setelah uji signifikansi adalah 65,8%
3. Hasil akurasi rata-rata prediksi kategori IPS1 mahasiswa angkatan tahun 2016 jalur prestasi adalah 67,80%
4. Hasil akurasi rata-rata prediksi kategori IPS1 mahasiswa angkatan tahun 2016 jalur non prestasi adalah 55,07%
5. Hasil akurasi tertinggi untuk memprediksi kategori IPS1 mahasiswa angkatan tahun 2016 jalur prestasi adalah dengan model F dengan akurasi 73,73%
6. Hasil akurasi tertinggi untuk memprediksi kategori IPS1 mahasiswa angkatan tahun 2016 non jalur prestasi adalah dengan model D dan E dengan akurasi 56,76%
7. Persentase akurasi model regresi logistik dalam memprediksi kategori IPS1 mahasiswa baru melalui jalur prestasi lebih tinggi daripada persentase akurasi model regresi Logistik dalam memprediksi kategori IPS1 mahasiswa baru melalui non jalur prestasi.

## REFERENSI

- [1]. J. Han and M. Kamber. 2011. *Data Mining : Concepts and Techniques*, Morgan Kaufmann Publishers.
- [2]. O. P. Rud. 2001 *Data Mining Cookbook: Modeling Data for Marketing, Risk, and Customer Relationship Management*, John Wiley & Sons Inc.
- [3]. P. Thakar, A. Mehta and Manisha. 2015. *Performance Analysis and Prediction in Educational Data Mining : A Research Travelogue. International Journal of Computer Applications*, vol. 110, nr 15, pp. 60-68,.
- [4]. Santosa, R.G and H. Setiadi. 2015. *Analisis Faktorial untuk Uji Pengaruh Beberapa Faktor terhadap Indeks Prestasi Mahasiswa Fakultas Teknologi Informasi UKDW*. Yogyakarta: Fakultas Teknologi Informasi UKDW.
- [5]. Santosa, R.G and A. R. Chrismanto. 2017. *Logistic Regression Model For Predicting First Semester Student Gpa Category Based On High School Academic Achievement*. ResearchersWorld: Journal of Arts, Science & Commerce, vol. VII, nr 2(1), April 2017, pp. 58-66.
- [6]. D. R. Cook and E. J. Snell. 1989. *Analysis of Binary Data, 2nd red.* Chapman and Hall/CRC.

- [7]. Kleinbaum, D. G. and M. Klein. 2010. *Logistic Regression: A Self-Learning Text, 3rd red.* New York: Springer-Verlag.
- [8]. D. W. Hosmer, S. Lemeshow And R. X. Sturdivant. 2013. *Applied Logistic Regression, 3rd red.* John Weley and Sons, Inc, 2013.
- [9]. S. B. Aher and L. L. M. R. J., 2011. *Data mining in Educational System Using Weka*”, in *International Conference on Emerging Technology Trends (ICEIT) 2011*.
- [10]. Xiangjua, G. Y. Bian n. 2010. Hangzhou: *The Application Of Data Mining Technology In Analysis Of College Student's Performance*, in *The 2nd International Conference on Information Science and Engineering*.
- [11]. Liu, Z and X. Zhang 2010. *Prediction and Analysis for Students' Marks Based on Decision Tree Algorithm*. Shenyang: in *Intelligent Networks and Intelligent Systems (ICINIS)*, 2010 3rd International Conference on IEEE.
- [12]. Osmanbegović, E. and M. Suljic. *Data Mining Approach For Predicting Student Performance*, in *Economic Review - Journal of Economics and Busines*, vol. X, nr 1, May 2012, pp. 3-12.
- [13]. S. M., B. S. and K. S. *Data Mining : Performance Improvement in Education Sector using Classification and Clustering Algorithm*, in *International Conference on Computing (ICCE) 2012*.
- [14]. Bhattacharyya, G. K. And R. A. Johnson. 1997. *Statistical Concepts and Methods*. John Wiley & Sons, Inc.
- [15]. Kantardzic, M. 2011. *Data Mining: Concepts, Models, Methods, and Algorithms, 2nd ed.* Wiley-IEEE Press.
- [16]. Rawling, J.O., Pantula S.G. and Dickey D.A. *Applied Regression Analysis A Research Tool*, Springer-Verlage Inc, 1998 .