

# Klasifikasi Nilai Peminat SBMPTN (Seleksi Bersama Masuk Perguruan Tinggi Negeri) ITS dengan Pendekatan *Classification and Regression Trees (CART)*

Lely Dwi Bhakti Pratiwi, Wahyu Wibowo, dan Ismaini Zain  
 (1,2,3)Jurusan Statistika, FMIPA, Institut Teknologi Sepuluh Nopember (ITS)  
 Jl. Arief Rahman Hakim, Surabaya 60111 Indonesia  
*e-mail* ismaini\_z@statistika.its.ac.id

**Abstrak**— SBMPTN merupakan seleksi masuk Perguruan Tinggi Negeri yang dilakukan secara serentak melalui ujian tertulis. Institut Teknologi Sepuluh Nopember (ITS) juga menerapkan SBMPTN sebagai salah satu jalur penerimaan mahasiswa baru. Terdapat sebanyak 9720 peminat yang memilih ITS pada pilihan pertama dengan jumlah yang diterima sebanyak 807 dan sisanya adalah peminat yang tidak diterima. Sedikitnya peserta yang diterima menunjukkan ketatnya persaingan dalam penerimaan SBMPTN. Oleh karena itu, perlu diketahui karakteristik peminat ITS di SBMPTN berdasarkan faktor yang diduga mempengaruhinya serta bagaimana pengklasifikasian peminat ITS di SBMPTN berdasarkan status penerimaan. Sehingga diperoleh informasi yang dapat digunakan sebagai pertimbangan dalam SBMPTN tahun berikutnya. Metode yang umum digunakan dalam pengklasifikasian adalah metode analisis diskriminan dan regresi logistik. Namun, metode tersebut memiliki keterbatasan dalam hal pemenuhan asumsi dan kesederhanaan interpretasi sehingga dalam penelitian ini digunakan metode *Classification and Regression Trees (CART)* yang mampu mengatasi keterbatasan tersebut. Hasil klasifikasi *CART* menunjukkan bahwa variabel nilai Matematika Dasar merupakan variabel terpenting dalam pengklasifikasian penerimaan SBMPTN ITS dengan kombinasi data learning dan testing terbaik yaitu data learning 85% dan data testing 15%. Terdapat sebanyak 12 terminal nodes yang diklasifikasikan sebagai peminat yang tidak diterima dan sebanyak 13 terminal nodes yang diklasifikasikan sebagai peminat yang diterima dengan total accuracy rate (1-APER) untuk kombinasi data tersebut adalah sebesar 92,3% untuk data learning dan 91,4% untuk data testing.

**Kata Kunci**—*CART*, Data Learning, Data Testing, SBMPTN, Total Accuracy Rate.

## I. PENDAHULUAN

POLA penerimaan mahasiswa baru yang diselenggarakan pemerintah salah satunya adalah SBMPTN (Seleksi Bersama Masuk Perguruan Tinggi Negeri). SBMPTN merupakan seleksi yang dilaksanakan secara bersama dan serentak oleh seluruh Perguruan Tinggi Negeri melalui ujian tertulis. Pendaftar SBMPTN di seluruh Perguruan Tinggi Negeri pada tahun 2014 sebanyak 664.509 dan hanya sebanyak 104.862 peserta yang dinyatakan lulus. Kurang lebih sekitar 15% saja peserta ujian yang lulus ujian SBMPTN [1]. Sesuai dengan peraturan perundangan, Institut Teknologi

Sepuluh Nopember (ITS) yang merupakan salah satu Perguruan Tinggi Negeri juga menerima mahasiswa baru melalui jalur SBMPTN. Berdasarkan data peminat SBMPTN tahun 2013, peminat ITS tercatat sebanyak 20.278 peserta untuk semua jurusan, sedangkan daya tampung hanya sebanyak 958 peserta untuk semua jurusan [2]. Hal tersebut menunjukkan peminat ITS dalam SBMPTN cukup tinggi dengan daya tampung yang terbatas, sehingga dapat memicu ketatnya persaingan diantara siswa-siswi di seluruh Indonesia untuk dapat masuk ke ITS. Ketatnya persaingan dalam SBMPTN juga ditunjukkan dengan sedikitnya jumlah peserta yang lulus ujian. Sedikitnya jumlah peserta yang diterima dapat dilihat dari jumlah peminat yang jauh lebih banyak jika dibandingkan dengan daya tampung yang ditetapkan.

Faktor yang mempengaruhi diterima atau tidaknya peserta ujian SBMPTN adalah nilai Verbal, nilai Numerikal, nilai Figural, nilai Matematika Dasar, nilai Bahasa Indonesia, nilai Bahasa Inggris, nilai Matematika IPA, nilai Fisika, nilai Kimia, dan nilai Biologi [3]. Selain itu, pemilihan jurusan juga berpengaruh dalam penerimaan. Jika jurusan yang dipilih mempunyai peminat yang cukup banyak, maka tingkat persaingan akan lebih ketat mengingat sedikitnya daya tampung yang disediakan. Berdasarkan hal tersebut, ingin dilakukan pengklasifikasian terhadap peminat ITS menurut status penerimaan (diterima atau tidak diterima) dengan didasarkan pada faktor-faktor yang diduga mempengaruhinya.

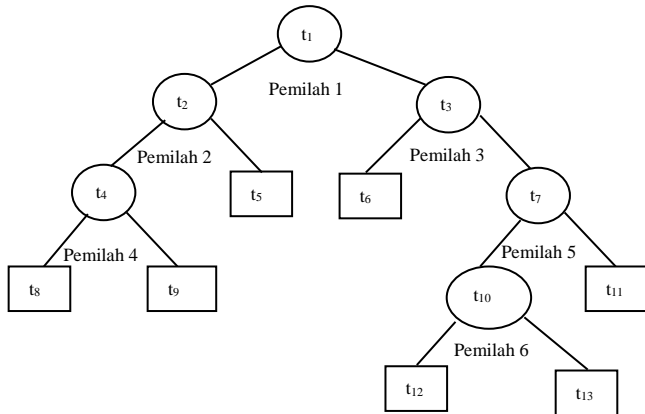
Metode yang umum digunakan dalam pengklasifikasian adalah metode analisis diskriminan dan regresi logistik [4]. Analisis diskriminan mensyaratkan terpenuhinya asumsi multivariat normal untuk variabel prediktor dan varians yang homogen pada setiap kelas dalam variabel respon [5]. Sementara regresi logistik memiliki kelebihan yaitu nilai *odds ratio* yang menunjukkan seberapa besar pengaruh variabel prediktor suatu kategori referensi pada suatu variabel respon. Kelemahan yang dihasilkan dari analisis menggunakan regresi logistik berupa nilai probabilitas yang dianggap kurang praktis sehingga cenderung sulit untuk diinterpretasikan [6]. Oleh karena itu, digunakan metode *Classification and Regression Trees (CART)* yang mempunyai kelebihan yaitu mampu bekerja pada dimensi data yang besar dan struktur data yang kompleks, tidak terikat oleh asumsi kenormalan ataupun variansi homogen, dapat mengetahui interaksi antar variabel prediktor dan hasil klasifikasi yang diperoleh lebih mudah

dipahami serta mudah diinterpretasikan [7]. Berdasarkan kekurangan metode parametrik yang telah dijelaskan, dilakukan pengklasifikasian peminat ITS di SBMPTN 2014 dengan menggunakan metode CART.

## II. TINJAUAN PUSTAKA

### A. Classification and Regression Tress (CART)

Analisis CART merupakan teknik klasifikasi dengan menggunakan algoritma partisi rekursif biner. Istilah *binary* dapat diartikan bahwa setiap kelompok data yang terkumpul dalam suatu ruang yang disebut simpul/*node* hanya dapat dibagi menjadi dua kelompok. Dengan demikian, setiap *node* dapat dibagi menjadi dua kelompok yang disebut dengan simpul anak (*child nodes*). Istilah *recursive* mengacu pada proses partisi biner yang dapat dilakukan secara berulang-ulang, sehingga setiap simpul anak yang dihasilkan dari partisi simpul awal dapat dipartisi kembali menjadi dua simpul anak lagi dan begitu seterusnya sampai pada batas kriteria tertentu. Istilah *partitioning* berarti bahwa proses klasifikasi pada suatu kelompok data dilakukan dengan cara dipartisi atau membagi data tersebut menjadi beberapa bagian [7].



Gambar 1 Ilustrasi Pohon Klasifikasi

Gambar 1 adalah ilustrasi pohon klasifikasi. Simpul/*node* awal yang dinotasikan dengan  $t_1$  merupakan simpul terpenting dalam menduga kelas amatan dan biasa disebut dengan *parent node*. Simpul dengan notasi  $t_2, t_3, t_4, t_7,$  dan  $t_{10}$  merupakan simpul dalam atau *internal nodes*, dan simpul akhir atau biasa disebut dengan *terminal nodes* dinotasikan dengan  $t_5, t_6, t_8, t_9, t_{11}, t_{12},$  dan  $t_{13}$  dimana setelah itu tidak dapat dilakukan partisi atau pemilahan lagi. Setiap simpul/*node* memiliki kedalaman (*depth*) masing-masing. Misalkan,  $t_1$  berada pada kedalaman 1,  $t_2$  dan  $t_3$  berada pada kedalaman 2, dan seterusnya. Terdapat tiga tahapan dalam pengklasifikasian dengan menggunakan metode CART yaitu pembentukan pohon klasifikasi, pemangkasan pohon klasifikasi, dan penentuan pohon klasifikasi optimum.

#### Pembentukan Pohon Klasifikasi

Tahapan dalam pembentukan pohon klasifikasi ada tiga tahap, yaitu.

##### a. Pemilihan Pemilahan

Pada tahap ini dilakukan pemilahan pada sampel data *learning* berdasarkan pada aturan pemilahan dan kriteria *goodness of split*. Hasil dari pemilahan berupa himpunan bagian yang harus bersifat lebih homogen dibandingkan

dengan simpul induknya. Tingkat heterogenitas tersebut dapat diukur dengan menggunakan nilai *impurity* Indeks Gini. Penggunaan Indeks Gini dalam pemilihan pemilahan memiliki kelebihan yaitu proses perhitungan yang sederhana dan relatif cepat, serta mudah dan sesuai untuk diterapkan dalam berbagai kasus [8]. Pemilahan terbaik dipilih berdasarkan nilai penurunan tingkat heterogenitas yang paling tinggi dari semua kemungkinan pemilahan yang dilakukan. Fungsi Indeks Gini dituliskan dalam persamaan (1).

$$i(t) = \sum_{j \neq k} p(j|t)p(k|t) \tag{1}$$

dengan  $p(j|t)$  adalah proporsi kelas  $j$  pada simpul  $t$  dan  $p(k|t)$  adalah proporsi kelas  $k$  pada simpul  $t$ . Pengevaluasian pemilahan  $s$  pada simpul  $t$  dilakukan berdasarkan kriteria *goodness of split* ( $\phi(s,t)$ ) sesuai dengan persamaan (2).

$$\phi(s,t) = \Delta i(s,t) = i(t) - p_L i(t_L) - p_R i(t_R) \tag{2}$$

dengan  $i(t_L)$  adalah *impurity* pada simpul kiri dan  $i(t_R)$  adalah *impurity* pada simpul kanan dimana pemilahan yang menghasilkan *goodness of split* tertinggi merupakan pemilahan yang terbaik.

##### b. Penentuan Terminal Node

Suatu simpul  $t$  dikatakan sebagai *terminal node* ketika tidak terdapat penurunan heterogenitas yang berarti, atau hanya terdapat satu pengamatan pada tiap simpul anak atau adanya batasan minimum  $n$ . Pengembangan pohon akan berhenti ketika dalam simpul terdapat pengamatan sejumlah kurang dari 5 [8].

##### c. Penandaan Label Kelas

Penandaan label kelas pada *terminal node* didasarkan pada aturan jumlah terbanyak seperti ditunjukkan dalam persamaan (3).

$$p(j_0|t) = \max_j p(j|t) = \max_j \frac{N_j(t)}{N(t)} \tag{3}$$

dengan  $N_j(t)$  adalah banyaknya pengamatan kelas  $j$  pada *terminal node*  $t$ , dan  $N(t)$  merupakan jumlah total pengamatan dalam *terminal node*  $t$ .

#### Pemangkasan Pohon Klasifikasi

Hasil pohon klasifikasi yang terbentuk dari pemilahan biasanya cukup besar. Pohon yang berukuran besar tersebut dapat mengakibatkan adanya kasus *overfitting* yaitu nilai prediksi melebihi nilai sebenarnya. Sedangkan apabila diberikan batasan dalam proses pemilahan, namun masih layak dilakukan pemilahan maka akan terjadi kasus *underfitting*. Sehingga dilakukan pemangkasan (*pruning*) untuk mendapatkan ukuran pohon yang layak dengan ukuran *cost complexity pruning minimum* seperti ditunjukkan pada persamaan (4).

$$R_\alpha(T) = R(T) + \alpha |\tilde{T}| \tag{4}$$

#### Penentuan Pohon Klasifikasi Optimal

Penduga pengganti yang dapat digunakan dalam menentukan pohon klasifikasi optimal yaitu penduga sampel uji (*test sample estimate*). Penduga sampel uji digunakan

ketika data berukuran besar. Persamaan (5) adalah persamaan yang digunakan untuk menghitung *test sample estimate*

$$R^s(T_i) = \frac{1}{N_2} \sum_{(x_n, j_n) \in L_2} X(d(x_n) \neq j_n) \tag{5}$$

dengan  $N_2$  adalah jumlah pengamatan dalam  $L_2$  dan  $X(.)$  bernilai 0 jika pernyataan dalam tanda kurung salah dan bernilai 1 jika pernyataan dalam tanda kurung benar. Pohon klasifikasi yang optimum adalah pohon  $T^*$  yang memiliki nilai penduga sampel uji minimum.

Perhitungan nilai variabel terpenting untuk pohon klasifikasi optimal diawali dengan menghitung nilai *goodness of split* untuk masing-masing variabel prediktor. Nilai *goodness of split* yang dihasilkan diurutkan mulai dari tertinggi sampai terendah. Kemudian dilakukan penjumlahan dari keseluruhan *goodness of split*. Untuk menghitung skor tertinggi dalam variabel terpenting yaitu menjumlahkan *goodness of split* tertinggi sampai terendah kemudian dibagi dengan penjumlahan keseluruhan *goodness of split*. Sedangkan, untuk menghitung skor variabel terpenting kedua adalah dengan menjumlahkan nilai *goodness of split* urutan kedua sampai terendah kemudian dibagi dengan penjumlahan keseluruhan *goodness of split*. Begitu seterusnya hingga skor variabel terpenting yang terakhir.

**B. Ukuran Ketepatan Klasifikasi**

Terdapat dua cara dalam mengukur ketepatan klasifikasi, yaitu dengan perhitungan *apparent error rate* (APER), dan *total accuracy rate* (1-APER). *Apparent error rate* (APER) adalah proporsi observasi yang diprediksi secara tidak benar (ukuran kesalahan klasifikasi total) dan *total accuracy rate* (1-APER) adalah proporsi observasi yang diprediksi secara benar (ukuran ketepatan klasifikasi total). Berikut disajikan *crosstab* untuk menghitung ketepatan klasifikasi yang ditunjukkan dalam Tabel 1.

**Tabel 1.** *Crosstab* Ketepatan Klasifikasi

Kelas Observasi	Kelas Prediksi Y		Total
	1	2	
Y			
1	$n_{11}$	$n_{12}$	$n_{1.}$
2	$n_{21}$	$n_{22}$	$n_{2.}$
Total	$n_{.1}$	$n_{.2}$	$N$

$$APER = \frac{n_{21} + n_{12}}{N} = \frac{\text{jumlah prediksi salah}}{\text{jumlah total prediksi}}$$

$$1 - APER = \frac{n_{11} + n_{22}}{N} = \frac{\text{jumlah prediksi benar}}{\text{jumlah total prediksi}}$$

**C. Seleksi Bersama Masuk Perguruan Tinggi Negeri (SBMPTN)**

Seleksi Bersama Masuk Perguruan Tinggi Negeri (SBMPTN) merupakan pola seleksi yang dilaksanakan secara bersama oleh seluruh Perguruan Tinggi Negeri dalam satu sistem yang terpadu dan diselenggarakan secara serentak melalui ujian tertulis. Ujian SBMPTN terdiri atas ujian tertulis dan ujian keterampilan. Ujian tertulis berlaku bagi semua peserta, sedangkan ujian keterampilan hanya berlaku bagi peserta yang memilih program studi bidang Ilmu Seni dan Keolahragaan [9].

**III. METODOLOGI PENELITIAN**

**A. Sumber Data**

Data yang digunakan dalam penelitian ini adalah data sekunder yang diambil dari Biro Administrasi Akademik dan Kemahasiswaan (BAAK) ITS yaitu data biodata peminat ITS di Seleksi Bersama Masuk Perguruan Tinggi Negeri (SBMPTN) 2014. Banyak data yang digunakan dalam penelitian ini sebanyak 9720 peminat dengan jumlah peminat yang diterima sebanyak 807 dan sisanya adalah peminat yang tidak diterima.

**B. Variabel Penelitian**

Variabel yang digunakan dalam penelitian ini berupa faktor yang diduga mempengaruhi penerimaan peminat ITS di Seleksi Bersama Masuk Perguruan Tinggi Negeri (SBMPTN) 2014. Adapun variabel-variabel yang digunakan dalam analisis CART dan variabel yang digunakan dalam analisis statistika deskriptif disajikan pada Tabel 2 dan Tabel 3.

**Tabel 2.** Variabel Penelitian yang Digunakan dalam Analisis CART

Variabel	Skala
Status Penerimaan (Y)	Nominal
Nilai Verbal (X <sub>1</sub> )	Rasio
Nilai Numerikal (X <sub>2</sub> )	Rasio
Nilai Figural (X <sub>3</sub> )	Rasio
Nilai Matematika Dasar (X <sub>4</sub> )	Rasio
Nilai Bahasa Indonesia (X <sub>5</sub> )	Rasio
Nilai Bahasa Inggris (X <sub>6</sub> )	Rasio
Nilai Matematika IPA (X <sub>7</sub> )	Rasio
Nilai Fisika (X <sub>8</sub> )	Rasio
Nilai Kimia (X <sub>9</sub> )	Rasio
Nilai Biologi (X <sub>10</sub> )	Rasio
Pilihan Jurusan (X <sub>11</sub> )	Nominal

**Tabel 3.** Variabel Penelitian yang Digunakan dalam Analisis Statistika Deskriptif

Variabel	Skala
Asal SMA (Z <sub>1</sub> )	Nominal
Asal Daerah (Z <sub>2</sub> )	Nominal
Jenis Kelamin (Z <sub>3</sub> )	Nominal

**C. Langkah-langkah Penelitian**

Langkah-langkah analisis yang digunakan untuk menjawab tujuan dari penelitian ini antara lain sebagai berikut.

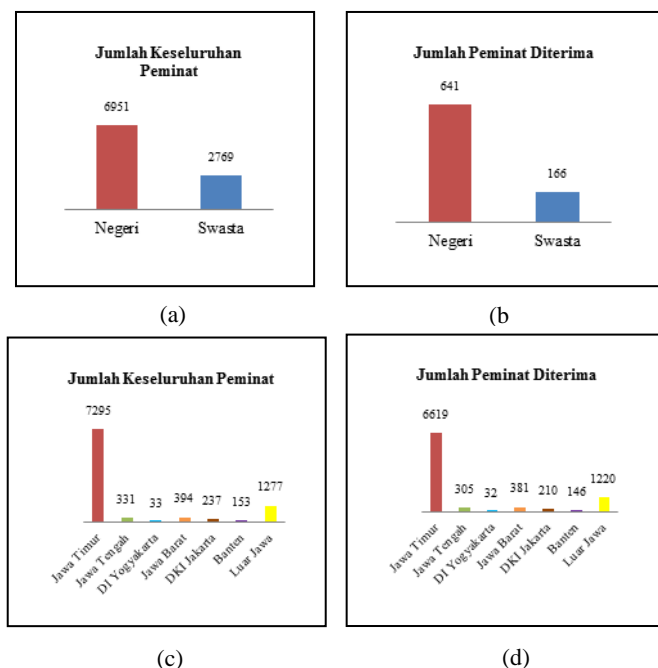
- Langkah dalam analisis statistika deskriptif adalah sebagai berikut.
  - Membuat diagram batang jumlah peminat berdasarkan asal daerah dan asal SMA.
  - Membuat diagram batang antara pilihan jurusan dengan jenis kelamin dan asal daerah peminat serta membuat tabel jumlah peminat berdasarkan asal daerah dan jurusan.
  - Menghitung nilai *mean* dan standar deviasi nilai ujian SBMPTN peminat ITS berdasarkan asal SMA dan jenis kelamin.

2. Langkah-langkah yang dilakukan dalam analisis CART adalah sebagai berikut.
  - a. Membagi data menjadi data *learning* dan data *testing* dengan proporsi sebesar 75%:25%, 80%:20%, 85%:15%, 90%:10%, dan 95%:5%.
  - b. Membentuk pohon klasifikasi maksimal
  - c. Melakukan pemangkasan pohon (*pruning*)
  - d. Menentukan pohon klasifikasi optimal
  - e. Menghitung ketepatan klasifikasi dari pohon klasifikasi.
  - f. Membandingkan hasil ketepatan klasifikasi semua pohon klasifikasi yang terbentuk dari setiap kombinasi data *learning* dan data *testing*.

IV. HASIL DAN PEMBAHASAN

A. Karakteristik Peminat ITS di SBMPTN 2014

Penyelenggaraan SBMPTN 2014 dilakukan secara serentak di seluruh Indonesia, oleh karena itu peminat ITS di SBMPTN 2014 juga berasal dari seluruh wilayah di Indonesia. Gambar 2 adalah jumlah peminat ITS berdasarkan asal SMA dan asal Daerah. Gambar 2(a) menunjukkan jumlah seluruh peminat ITS yaitu sebanyak 9720 peminat dengan peminat yang berasal dari SMA Negeri sebanyak 6951 peminat dan jumlah peminat yang berasal dari SMA Swasta sebanyak 2769 peminat. Jumlah peminat yang diterima berdasarkan asal SMA ditunjukkan pada Gambar 2(b). Terlihat bahwa jumlah peminat yang diterima dan berasal dari SMA Negeri lebih banyak jika dibanding peminat dari SMA Swasta yaitu sebanyak 641 peminat sedangkan yang berasal dari SMA swasta sebanyak 166 peminat.



Gambar 2. Jumlah Peminat Berdasarkan Asal SMA dan Asal Daerah

Gambar 2(c) merupakan jumlah peminat ITS berdasarkan asal daerah peminat, terlihat bahwa peminat asal Jawa Timur merupakan peminat paling banyak dengan jumlah 7295 peminat. Peminat kedua terbanyak berasal dari luar Jawa

yaitu sebanyak 1277 peminat dan peminat ketiga terbanyak adalah Jawa Barat dengan 394 peminat. Sedangkan, peminat paling sedikit berasal dari DI Yogyakarta yaitu sebanyak 33 peminat. Sedikitnya peminat dari DI Yogyakarta dapat mejadi salah satu perhatian ITS jika melakukan *roadshow* dalam rangka memperkenalkan jurusan yang ada di ITS pada siswa SMA agar kedepannya peminat ITS dari daerah tersebut semakin banyak. Gambar 2(d) menunjukkan jumlah peminat yang diterima berdasarkan asal daerah peminat.

Tabel 4. Nilai Peminat yang Diterima Berdasarkan Asal SMA dan Jenis Kelamin

Nilai	Asal SMA				Jenis Kelamin			
	Negeri		Swasta		Laki-Laki		Perempuan	
	Mean	St.dev	Mean	Stdev	Mean	Stdev	Mean	Stdev
Verbal	624,93	71,81	611,41	74,43	622,33	74,47	621,79	68,57
Numerikal	682,15	46,26	683,17	51,57	688,23	47,03	670,61	45,92
Figural	648,56	66,86	642,10	62,40	651,82	64,63	638,05	67,79
MatDas	781,53	160,45	757,83	173,50	785,73	168,54	758,49	151,25
B.Indonesia	630,05	104,75	605,37	99,12	616,07	100,48	642,77	108,83
B.Inggris	680,57	137,60	691,80	150,26	684,52	148,74	679,59	121,80
MatIPA	621,63	123,71	621,19	122,22	628,57	129,82	607,49	108,06
Fisika	711,53	111,16	697,68	118,59	719,88	117,56	686,30	99,09
Kimia	682,78	141,43	667,75	142,15	683,67	145,57	671,74	133,29
Biologi	594,81	103,81	604,87	101,46	600,89	104,95	588,84	99,76

Berdasarkan Tabel 4, Peminat yang berasal dari SMA Negeri dan Swasta mempunyai rata-rata paling tinggi yaitu nilai Matematika Dasar dengan masing-masing nilai sebesar 781,53 dan 757,83. Sedangkan rata-rata nilai terendah yang adalah Biologi sebesar 594,8 untuk peminat dari SMA Negeri dan sebesar 604,87 untuk peminat dari SMA Swasta. Rata-rata nilai peminat dari SMA Negeri dan Swasta tidak menunjukkan adanya perbedaan yang cukup tinggi. Jika dilihat dari jenis kelamin, rata-rata nilai tertinggi untuk peminat berjenis kelamin laki-laki dan perempuan adalah Matematika Dasar yaitu sebesar 785,73 dan 758,49. Rata-rata nilai terendah adalah nilai Biologi sebesar 600,89 untuk peminat berjenis kelamin laki-laki dan 588,84 untuk peminat berjenis kelamin perempuan. Berdasarkan rata-rata nilai tertinggi dan terendah dapat dikatakan tidak terlalu ada perbedaan nilai antara peminat berjenis kelamin laki-laki dan perempuan

B. Klasifikasi Peminat dengan Menggunakan Analisis CART

Pengklasifikasian peminat ITS di SBMPTN 2014 didasarkan pada status penerimaan peserta yaitu diterima atau tidak diterima. Data yang digunakan dalam pengklasifikasian peminat ITS di SBMPTN 2014 berdasarkan status penerimaan sebanyak 9.720 pengamatan. Berikut disajikan penjelasan secara lebih rinci dari masing-masing tahapan dalam anaisis CART dengan menggunakan kombinasi data *learning* dan *testing* sebesar 85%:15%.

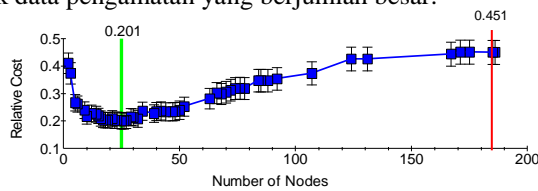
Pembentukan Pohon Klasifikasi Maksimal

Tahapan dalam pembentukan pohon klasifikasi maksimal dimulai dengan melakukan pemilahan terhadap semua kemungkinan variabel pemilah dan *threshold* dengan menggunakan Indeks Gini. Indeks Gini yang telah didapatkan kemudian digunakan untuk menghitung nilai *goodness of split* dari masing-masing pemilah. Pemilah yang menghasilkan

*goodness of split* tertinggi merupakan pemilah yang terbaik. Proses pemilahan dilakukan hingga terbentuk pohon klasifikasi maksimal (sudah tidak dapat dilakukan pemilahan lagi). Pohon klasifikasi maksimal yang terbentuk mempunyai ukuran yang besar dengan jumlah *terminal node* atau simpul akhir sebanyak 186 *nodes* dan tingkat kedalaman (*depth*) sebesar 25. Variabel prediktor yang mejadi pembentuk dari pohon klasifikasi sebanyak 11 variabel dengan variabel yang menjadi pemilah utama (*parent node*) adalah variabel nilai Matematika Dasar.

**Pemangkasan Pohon (Pruning)**

Pohon klasifikasi maksimal yang dihasilkan mempunyai ukuran yang cukup besar, untuk memudahkan proses analisis dilakukan pemangkasan pada pohon klasifikasi maksimal. Pada pemangkasan pohon maksimal di peneliti ini digunakan metode *test sample estimate* karena lebih sesuai digunakan untuk data pengamatan yang berjumlah besar.

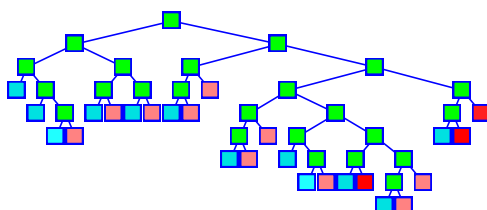


**Gambar 3.** Plot *Relative Cost* dan Banyaknya *Terminal Nodes*

Hasil dari pemangkasan pohon maksimal dengan menggunakan metode *test sample estimate* ditunjukkan pada Gambar 3. Garis merah menunjukkan pohon klasifikasi maksimal dan garis berwarna hijau menunjukkan pohon kasifikasi optimal. Pohon maksimal memiliki jumlah *terminal nodes* sebanyak 186 dengan nilai *complexity parameter* sebesar 0,00, *test set relative cost* sebesar 0,451 ± 0,044 dan *resubtitution relative cost* sebesar 0,017.

**Pemilihan Pohon Klasifikasi Optimal**

Pemangkasan pohon klasifikasi maksimal secara iteratif akan menghasilkan pohon klasifikasi optimal. Pohon klasifikasi optimal yang terbentuk terdiri dari 25 *terminal nodes* dan menghasilkan *complexity parameter* sebesar 0,001, *test set relative cost* sebesar 0,201 ± 0,029 serta *resubtitution relative cost* sebesar 0,108. Konstruksi pohon klasifikasi optimal ditunjukkan pada Gambar 4. simpul berwarna merah adalah *terminal node* yang diklasifikasikan sebagai peminat diterima. Sedangkan simpul berwarna biru adalah *terminal node* yang diklasifikasikan sebagai peminat tidak diterima. Semakin kuat warna yang dihasilkan pada setiap simpul maka perbedaan proporsi antara kategori status penerimaan semakin besar.



**Gambar 4.** Konstruksi Pohon Klasifikasi Optimal

Variabel terpenting yang paling menentukan pemilahan menjadi dua anak simpul berdasarkan konstruksi pohon

klasifikasi optimal adalah nilai Matematika Dasar dengan besar kontribusi 100. Nilai variabel lain yang berkontribusi dalam pemilahan pemilah adalah nilai Fisika dengan besar kontribusi sebesar 75,64, Sedangkan, untuk variabel prediktor lainnya hanya memberikan kontribusi dengan skor dibawah 50. Skor dari variabel lain yang berkontribusi dalam pembentukan pohon klasifikasi selengkapnya disajikan pada Tabel 5.

**Tabel 5.** Skor Variabel Terpenting dari Pohon Klasifikasi Optimal

Variabel	Skor
MATDAS	100,00
FISIKA	75,64
NUMERIKA	36,65
KIMIA	36,61
BING	24,37
FIGURAL	21,16
MATIPA	18,89
JURUSAN	8,69
BINDO	5,85
BIOLOGI	4,81
VERBAL	3,48

Karakteristik simpul terminal (*terminal node*) dalam Gambar 4 yang memiliki warna paling kuat (simpul yang memiliki perbedaan proporsi antara kategori paling besar) dijelaskan lebih rinci pada Tabel 6.

**Tabel 6.** Karakteristik Peminat ITS di SBMPTN 2014

Peminat yang Tidak Diterima (simpul akhir 23, kelas 1:100,00%, kelas 2:0,00%)	
▪	Nilai Matematika Dasar > 616,919
▪	Nilai Fisika > 531,696
▪	Nilai Kimia ≤ 599,281
▪	Nilai Numerikal ≤ 574,796
▪	Nilai Matematika IPA ≤ 698,262
Peminat yang Diterima (simpul akhir 24, kelas 1:0,00%, kelas 2:100,00%)	
▪	Nilai Matematika Dasar > 616,919
▪	Nilai Fisika > 531,696
▪	Nilai Kimia > 599,281
▪	Nilai Numerikal ≤ 574,796
▪	Nilai Matematika IPA > 698,262

Keterangan : Kelas 1 : Peminat yang Tidak Diterima  
Kelas 2 : Peminat yang Diterima

**Hasil Ketepatan Klasifikasi**

Pengukuran ketepatan klasifikasi penerimaan mahasiswa baru ITS jalur SBMPTN 2014 menggunakan perhitungan nilai *Apparent Error Rate (APER)* dan *Total Accuracy Rate (1-APER)*. Perhitungan selengkapnya disajikan pada Tabel 7.

**Tabel 7.** Tingkat Akurasi Pohon Klasifikasi Optimal

Kelas Observasi		Kelas Prediksi		Total	APER (%)	1-APER (%)
		1	2			
Data Learning	1	6.963	620	7583	7,7	92,3
	2	18	661	679		
Data Testing	1	1.219	111	1330	8,6	91,4
	2	15	113	128		

Keterangan : Kelas 1 adalah peminat yang tidak diterima  
Kelas 2 adalah peminat yang diterima

Dari 8.262 peminat ITS yang dijadikan data *learning* didapatkan hasil sebanyak 18 peminat yang salah diklasifikasikan sebagai peminat yang tidak diterima dan sebanyak 620 peminat yang salah diklasifikasikan sebagai peminat yang diterima. Sehingga, diperoleh nilai kesalahan klasifikasi (APER) sebesar 7,7 persen dan ukuran ketepatan klasifikasi total atau *total accuracy rate* (1-APER) untuk data *learning* sebesar 92,3 persen.

Jika hasil klasifikasi divalidasi dengan data *testing*, maka diperoleh hasil jumlah peminat yang salah diklasifikasikan menjadi peminat yang tidak diterima sebanyak 15 peminat sedangkan jumlah peminat yang salah diklasifikasikan sebagai peminat yang diterima yaitu sebanyak 111 peminat. Oleh karena itu, diperoleh nilai kesalahan klasifikasi (APER) untuk data *testing* sebesar 8,6 persen dan *total accuracy rate* (1-APER) adalah sebesar 91,4 persen.

### C. Pemilihan Kombinasi Data Learning dan Testing Terbaik

Analisis CART yang telah dijelaskan sebelumnya juga dilakukan pada beberapa kombinasi data *learning* dan *testing* lainnya antara lain 75%:25%, 80%:20%, 90%:10%, dan 95%:5%. Hasil ketepatan klasifikasi (1-APER) untuk masing-masing kombinasi data *learning* dan data *testing* disajikan pada Tabel 8.

Berdasarkan Tabel 8 dapat diketahui bahwa kombinasi data *learning* dan data *testing* yang menghasilkan ketepatan klasifikasi paling besar adalah kombinasi data *learning* 85 persen dan data *testing* 15 persen. *Total accuracy rate* yang dihasilkan untuk data *testing* adalah 91,4 persen dimana nilai tersebut merupakan nilai paling tinggi jika dibandingkan dengan kombinasi data lainnya. Begitu pula dengan *total accuracy rate* untuk data *learning* yang merupakan nilai tertinggi jika dibandingkan dengan kombinasi data lain yaitu sebesar 92,3 persen.

**Tabel 8.** Perbandingan *Total Accuracy Rate* Pohon Klasifikasi Optimal Beberapa Kombinasi Data

Kombinasi Data Learning dan Testing	<i>Total Accuracy Rate</i>		Jumlah Terminal Nodes
	(1-APER)		
	Data Learning	Data Testing	
75%:25%	90,4	89,4	20
80%:20%	89,0	88,5	13
85%:15% *	92,3	91,4	25
90%:10%	89,1	86,8	17
95%:5%	89,7	89,9	15

\* kombinasi data *learning* dan *testing* terpilih

Kombinasi data *learning* 85 persen dan data *testing* 15 persen menghasilkan nilai ketepatan klasifikasi paling tinggi sehingga, hasil analisis CART yang digunakan dalam menjelaskan status penerimaan mahasiswa baru ITS jalur SBMPTN 2014 adalah hasil klasifikasi yang menggunakan kombinasi data *learning* 85 persen dan data *testing* 15 persen dimana kombinasi data tersebut merupakan kombinasi data terbaik.

## V. KESIMPULAN DAN SARAN

Berdasarkan penelitian yang telah dilakukan, didapatkan hasil karakteristik nilai peminat yang diterima baik yang berasal dari SMA Negeri maupun SMA Swasta serta jenis kelamin laki-laki maupun perempuan mempunyai kesamaan dalam hal rata-rata nilai tertinggi dan rata-rata nilai terendah dimana rata-rata nilai tertinggi adalah nilai Matematika Dasar dan rata-rata terendah adalah nilai Biologi. Pengklasifikasian peminat berdasarkan status penerimaan dengan metode CART menggunakan kombinasi data *learning* dan data *testing* sebesar 85 persen dan 15 persen. Hal tersebut dikarenakan kombinasi data *learning* dan *testing* tersebut menghasilkan nilai *total accuracy rate* (1-APER) yang paling tinggi yaitu sebesar 92,3 persen untuk data *learning* dan sebesar 91,4 persen untuk data *testing*.

Adapun saran yang untuk penelitian selanjutnya yaitu dibutuhkan suatu metode yang dapat mengklasifikasikan peminat ITS sesuai dengan pilihan pada SBMPTN (pilihan 1, pilihan 2 atau pilihan 3). Selain itu, bagi calon mahasiswa baru yang ingin masuk ITS melalui jalur SBMPTN perlu melakukan persiapan yang matang dikarenakan tingkat persaingan yang ketat dengan kuota yang diterima sangat sedikit serta memilih pilihan jurusan yang tepat sesuai dengan keinginan dan kemampuan.

## DAFTAR PUSTAKA

- [1] DIKTI. (2014). *Kabar Dikti*. Dipetik Januari 7, 2015, dari Kementerian Pendidikan dan Kebudayaan Direktorat Jenderal Pendidikan Tinggi: dikti.go.id
- [2] SMITS. (2014). *Seleksi Masuk ITS Program Sarjana*. Dipetik April 6, 2015, dari SMITS: <http://smits.its.ac.id/sarjana/#snmptn>
- [3] Anggraeni, Y. (2015). *Pemodelan Regresi Logistik Biner Terhadap Peminat ITS Di Seleksi Bersama Masuk Perguruan Tinggi Negeri (SBMPTN) 2014*. Institut Teknologi Sepuluh Nopember, Statistika, Surabaya.
- [4] Wibowo, Wahyu, "Perbandingan Hasil Klasifikasi Analisis Diskriminan dan Regresi Logistik Pada Pengklasifikasian Data Respon Biner," *KAPPA*, vol. 3, no. 1, 2002.
- [5] Johnson, R. A., & Wichern, D. W. (2007). *Applied Multivariate Statistical Analysis* (6th ed.). New Jersey: Prentice Hall.
- [6] Webb, P., & Yohannes, Y. (1999). *Classification and Regression Trees, CART: A User Manual for Identifying Indicators of Vulnerability to Famine and Chronic Food Insecurity*. Washington, D.C.: International Food Policy Research Institute.
- [7] Lewis, R. J. (2000). An Introduction to Classification and Regression Tree (CART) Analysis. *Annual Meeting of the Society for Academic Emergency Medicine*.
- [8] Breiman, L., Friedman, J. H., Olshen, R., & Stone, C. (1993). *Classification and Regression Trees*. New York: Chapman Hall.
- [9] SBMPTN. (2014). *Informasi Umum*. Dipetik Januari 7, 2015, dari Laman Resmi SBMPTN 2014: sbmptn.or.id