

ANALISIS KLASIFIKASI REKAPITULASI PENGADUAN PELANGGAN UP3 PT. PLN SEMARANG MENGGUNAKAN ALGORITMA QUEST (*QUICK, UNBIASED, AND EFFICIENT STATISTICAL TREE*)

Sang Nur Cahya Widiutama¹, Budi Warsito^{2*}, Sudarno³

^{1,2,3} Departemen Statistika, Fakultas Sains dan Matematika, Universitas Diponegoro

*email: budiwrst2@gmail.com

ABSTRACT

Every company must have a way to solve the problems faced by its customers, PT. PLN Persero, the Indonesian national energy utility, must have a method to handle consumer complaints. PT. PLN Persero has a recovery time strategy for resolving consumer concerns, but it is not always effective in doing so. The QUEST algorithm (Quick, Unbiased, and Efficient Statistical Tree) approach is used to classify the problem of the recovery time policy failing on specific complaints. Classification of complaint data in order to obtain characteristics and factors as the main influence on the complaints and be able to provide new opinions for PT. PLN to address customer complaints. The QUEST method is a classification tree technique with two nodes per split that yields an unbiased variable. The QUEST method may be used with both category and numerical data. QUEST uses three stages to create a classification tree: picking the splitting variable, identifying the split point, and pausing the split. The classification tree generated has a tree depth of four layers and obtained three essential factors in the classification, namely weather, the number of customers experiencing the same event, and distance from the site. The classification tree accuracy level is 0.851 (or 85.1%), with a prediction error rate of 0.149 (or 14.9%).

Keywords: *binary classification tree, recovery time, QUEST algorithm.*

1. PENDAHULUAN

Energi Listrik saat ini merupakan salah satu kebutuhan pokok yang tidak terpisahkan dari kehidupan manusia dan memiliki sifat fleksibel serta dapat dengan mudah diubah menjadi energi lain sesuai kebutuhan manusia (Shanefield, 2001). Pada tahun 2010 Badan Pusat Statistik mencatat bahwa sebanyak 17,8% penduduk Indonesia merupakan pelanggan PT. PLN Persero. Menyadari pentingnya energi listrik dalam kehidupan manusia serta banyaknya pelanggan, PT. PLN Persero sebagai pemasok listrik Indonesia, berkomitmen untuk memberikan pelayanan terbaik kepada pelanggannya. Pelayanan Teknik adalah pelayanan terhadap masyarakat mengenai pengaduan masalah kelistrikan yang dihadapi pelanggan terhadap kondisi kelistrikannya dengan menghubungi *Call Center* 123 (Halo PLN).

Dalam pelayanan teknik, PT. PLN Persero memiliki parameter waktu dalam menangani suatu pengaduan masalah kelistrikan dengan batasan *recovery time*, *recovery time* mulai dihitung saat petugas PT. PLN Persero sampai dilokasi sesuai dengan pengaduan yang diajukan. Hal ini dilakukan agar kualitas pelayanan dari PT. PLN Persero tetap memenuhi harapan pelanggan tanpa mengurangi kualitas pelayanan. Kualitas layanan (*Service quality*)

adalah ukuran seberapa baik suatu layanan memenuhi kecocokan dengan harapan pelanggan (Lewis dan Booms,1983).

Dalam penelitian ini data rekapitulasi Unit Pelayanan Pelanggan (UP3) Semarang pada tahun 2019 digunakan sebagai data yang akan diteliti. Dari data didapatkan bahwa masih terdapat pengaduan pelanggan yang diselesaikan lebih dari batasan yang ditentukan. Data yang direkapitulasi oleh Unit Pelayanan Pelanggan (UP3) Semarang berupa data kategorik dan data numerik serta memiliki variabel faktor yang banyak sehingga metode yang cocok dalam pengklasifikasian penelitian ini adalah klasifikasi menggunakan metode Algoritma *QUEST*.

Algoritma *QUEST* adalah algoritma pohon klasifikasi yang menghasilkan variabel tak bias dan memiliki 2 simpul setiap penyekatan (Kim dan Loh,2001). *QUEST* dapat diterapkan pada data dengan variabel berbentuk kategorik maupun numerik. Variabel penyekat dapat dipilih dari variabel prediktor numerik maupun kategorik (Grabczewski,2014). Variabel prediktor numerik digunakan uji ANOVA F untuk menguji perbedaan rata-rata antar kelas pada variabel prediktor numerik X atau digunakan uji Levene F untuk menguji kesamaan varian antar kelas. Variabel kategorik digunakan uji chi-kuadrat (χ^2) untuk menguji kebebasan antar variabel respon Y dan variabel prediktor kategorik X .

2. TINJAUAN PUSTAKA

2.1. Pelanggan

Greenberg (2010) berpendapat bahwa pelanggan atau *costumer* adalah individu atau kelompok yang terbiasa membeli sebuah produk atau jasa berdasarkan keputusan mereka atas pertimbangan manfaat maupun harga yang kemudian melakukan hubungan dengan perusahaan melalui telepon, surat, dan fasilitas lainnya untuk mendapatkan suatu penawaran dari perusahaan.

2.2. Energi Listrik

Shanefield (2001) berpendapat bahwa energi listrik yang yang mengalir pada perlatan listrik memiliki satuan untuk arus listrik yaitu *ampere* (A) dan tegangan listrik dengan satuan *volt* (V) dengan kebutuhan konsumsi daya listrik dengan satuan *Watt* (W). Marsudi (2006) berpendapat bahwa penyediaan energi listrik oleh PT. PLN dibagi atas tiga proses penyaluran energi listrik yaitu pembangkit, transmisi, dan distribusi. Aliran listrik yang paling dekat dengan pelanggan PT. PLN dalam penyaluran energi listrik ke rumah atau bangunan adalah distribusi.

2.3. Uji Chi-Kuadrat

Lestari (2005) berpendapat uji Chi-kuadrat (χ^2) untuk memeriksa kebebasan digunakan untuk memutuskan antara dua variabel kategorik dalam suatu kelompok saling bebas atau tidak. Chi-kuadrat memiliki asumsi-asumsi sebagai berikut:

- a. Data terdiri dari sebuah data acak sederhana berukuran n dari suatu populasi yang diminati.
- b. Hasil-hasil pengamatan dalam data dapat diklasifikasikan secara silang (*cross-classified*) menurut variabel-variabel yang diamati.

Dalam Dalil Kaidah Penggandaan Khusus menurut Walpole (1992) dari frekuensi sel yang teramati (O_{ij}) dan frekuensi sel yang diharapkan (E_{ij}) tersebut dapat dihitung suatu statistik uji Chi-kuadrat (χ^2) yang mencerminkan perbedaan antara keduanya, yang dirumuskan sebagai berikut:

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^c \left[\frac{(O_{ij} - E_{ij})^2}{E_{ij}} \right] \quad (1)$$

Dalam pengambilan keputusan, H_0 ditolak dengan taraf signifikansi α jika nilai statistik uji χ^2 lebih besar dari pada nilai $\chi_{\alpha, (r-1)(c-1)}^2$.

2.4. Uji Anova

Uji ANOVA F digunakan untuk membandingkan nilai tengah dari dua atau lebih kelompok data yang saling bebas. Ukuran data antar masing masing kelompok tidak harus sama, tetapi perbedaan yang besar dalam ukuran data dapat mempengaruhi hasil uji perbandingan nilai tengah (Walpole, 1992). Persamaan yang digunakan saat menghitung nilai didefinisikan sebagai berikut:

$$F = \frac{S_1^2}{S_2^2} \quad (2)$$

Dalam pengambilan keputusan H_0 ditolak dengan taraf signifikansi α jika nilai statistik uji F lebih besar daripada nilai $F_{\alpha, (K-1), (N-K)}$.

2.5. Uji Levene

Menurut Brown (1974), uji Levene F digunakan untuk menguji kesamaan varian variabel dari dua kelompok atau lebih. Bila σ_k^2 adalah nilai varian variabel dari kelompok ke- k ($k = 1, 2, \dots, K$) maka hipotesis yang digunakan dalam uji ini adalah

$$H_0 : \sigma_1^2 = \sigma_2^2 = \dots = \sigma_K^2$$

$$H_1 : \text{ada } \sigma_k^2 \neq \sigma_{k'}^2, k \neq k', (k, k' = 1, 2, \dots, K).$$

Untuk mendapatkan statistik uji Levene F, data ditransformasikan dulu menjadi simpangan mutlaknya terhadap rata-rata data, yaitu:

$$y_{ki} = |x_{ki} - \bar{x}_k| \quad (3)$$

dengan:

$$i = 1, 2, \dots, n_k, k = 1, 2, \dots, K.$$

y_{ki} : hasil transformasi data pengamatan ke- i dari kelompok ke- k

\bar{x}_k : rata-rata data kelompok ke- k

x_{ki} : pengamatan ke- i dari kelompok ke- k .

Dalam pengambilan keputusan, H_0 ditolak dengan taraf signifikansi α jika nilai statistik uji F lebih besar daripada nilai $F_{\alpha, (K-1), (N-K)}$.

2.6. Analisis Diskriminan Kuadratik

Analisis diskriminan bertujuan untuk membentuk fungsi diskriminan yang mampu membedakan kelompok. Analisis dilakukan berdasarkan suatu perhitungan statistik terhadap objek-objek yang telah diketahui dengan jelas pengelompokannya.

Taksiran dari skor diskriminan kuadratik:

$$d_k^Q(\mathbf{x}^*) = -\frac{1}{2} \ln |\mathbf{S}_k| - \frac{1}{2} (\mathbf{x}^* - \bar{\mathbf{x}}_k)' (\mathbf{S}_k)^{-1} (\mathbf{x}^* - \bar{\mathbf{x}}_k) + \ln p_k \quad (4)$$

Kelompokkan \mathbf{x}^* ke- k jika:

$$\text{Skor kuadratik } d_k^Q(\mathbf{x}^*) = \max \{(d_1^Q(\mathbf{x}^*), d_2^Q(\mathbf{x}^*), \dots, d_k^Q(\mathbf{x}^*)\} \quad (5)$$

Kuantitas data yang relevan untuk kelompok ke- k adalah $\bar{\mathbf{x}}_k$: vektor nilai rata-rata data dari kelompok ke- k , \mathbf{S}_k : matriks varian-kovarian data dari kelompok ke- k , n_k : ukuran data dari kelompok ke- k .

Fungsi diskriminan yang dibangun dengan asumsi bahwa kelompok-kelompok memiliki matriks ragam yang sama dinamakan fungsi diskriminan linier, sedangkan fungsi yang dibangun tanpa asumsi tersebut dinamakan fungsi diskriminan kuadrat (Rachmatin dan Kania, 2009).

2.7. Pohon Klasifikasi

Menurut Rokach dan Maimon (2008) pohon keputusan adalah model prediksi yang digunakan untuk memberikan gambaran bentuk pengklasifikasian dan model regresi. Pohon keputusan memiliki dua hal yang berbeda, yaitu pohon regresi dan pohon klasifikasi. Pada pohon klasifikasi data kategorik termasuk dalam variabel respon, sedangkan pada pohon regresi data numerik menjadi variabel respon. Pohon klasifikasi adalah pengklasifikasi nonparametrik yang membuat pohon keputusan hierarki dengan membagi data antar kelas kriteria pada langkah yang diberikan seperti sebuah perintah “*If-then*” yang ditetapkan pada variabel prediktor, menjadi dua anak simpul (*child node*) atau lebih secara terus menerus, dari simpul akar yang berisi seluruh data (Maroco *et al*, 2011)

2.7.1. Algoritma Quest

Pohon klasifikasi *QUEST* dibentuk dari pemilihan variabel penyekat, pemilihan titik sekat untuk variabel yang dipilih, dan proses pemberhentian. Pemilihan variabel penyekat digunakan untuk menentukan simpul yang akan dibentuk. Variabel penyekat dapat dipilih dari variabel prediktor numerik maupun kategorik (Loh dan Shih, 1997).

2.7.1.1. Pemilihan Variabel Penyekat

Pemilihan variabel penyekat pada *QUEST* adalah memilih nilai *p-value* terkecil dari hasil uji setiap variabel prediktor terhadap variabel respon dan dibandingkan dengan koreksi Bonferroni. Koreksi Bonferroni diusulkan untuk menggunakan tingkat kepercayaan lebih kecil ketika beberapa uji dilakukan (Grabczewski, 2014). Ketika n uji dilakukan maka masing masing uji menggunakan tingkat kepercayaan $\frac{\alpha}{n}$, dengan α adalah tingkat signifikansi. Berikut adalah langkah-langkah pemilihan variabel penyekat:

1. Setiap variabel prediktor X yang terdapat dalam masing-masing simpul dilakukan:
 - a. Uji ANOVA F, untuk menguji perbedaan rata-rata antar kelas pada setiap variabel prediktor numerik X .
 - b. Uji χ^2 untuk menguji kebebasan antar variabel respon Y dan variabel prediktor kategorik X .
 - c. Dari uji ANOVA F dan χ^2 akan didapatkan *p-value* dari analisis menggunakan SPSS.

2. Variabel dengan p -value terkecil dipilih.
3. Jika p -value terkecil kurang dari $\frac{\alpha}{M}$, dengan M adalah banyaknya variabel prediktor. Maka variabel prediktor X dipilih sebagai variabel penyekat untuk simpul dan dilanjutkan ke tahap 5. Jika tidak, dilanjutkan ke tahap 4.
4. Jika p -value terkecil yang didapat $\geq \frac{\alpha}{M}$, maka:
 - a. Untuk setiap variabel prediktor numerik X , dilakukan uji Levene F untuk menguji kesamaan varian.
 - b. Didapatkan p -value dari setiap variabel diuji menggunakan uji Levene F dengan bantuan aplikasi SPSS.
 - c. Variabel prediktor dengan p -value terkecil dipilih.
 - d. Jika p -value terkecil kurang dari nilai uji Levene dengan taraf $\frac{\alpha}{M+M1}$, dengan M_1 adalah banyaknya variabel prediktor numerik, maka variabel prediktor X dipilih menjadi variabel penyekat untuk simpul. Jika tidak, maka penyekatan simpul berhenti.
5. Hasil dari tahap ke 3 dan tahap ke 4 yaitu variabel penyekat yang diperoleh untuk simpul dan dilakukan permisalan terhadap variabel penyekat dengan X^* .
 - a. Jika X^* merupakan variabel numerik, maka dilanjutkan ke analisis diskriminan kuadratik, kemudian dilanjutkan ke penentuan titik sekat.
 - b. Untuk X^* merupakan variabel kategorik, maka dilakukan transformasi ke dalam variabel *dummy* lalu diproyeksikan ke dalam koordinat diskriminan terbesarnya sesuai dengan pada subab 2.7.1.2 poin b.

2.7.1.2. Algoritma Menentukan Titik Sekat

Proses penentuan titik sekat pada *QUEST* dibedakan menjadi dua, yaitu proses penentuan titik sekat dengan variabel prediktor numerik dan proses penentuan titik sekat dengan variabel prediktor kategorik.

a. Menentukan Titik Sekat dengan Variabel Prediktor Numerik

Misalkan variabel respon adalah Y memiliki dua kelas, yaitu $j = A, B$. Misalkan pula variabel X dipilih untuk menyekat simpul t , maka langkah berikutnya adalah menentukan titik sekat.

1. Misalkan \bar{x}_A dan s_A^2 menyatakan rata-rata dan varian data kelas A , sedangkan \bar{x}_B dan s_B^2 adalah rata-rata dan varian data kelas B .
2. Untuk menentukan titik sekat yang merupakan suatu angka yang dapat membagi atau mempartisi suatu data ke dalam dua data baru yang direpresentasikan oleh dua simpul yang berbeda maka digunakan bantuan persamaan kuadrat $ax^2 + bx + c = 0$ yang diperoleh dari penggunaan fungsi \ln pada dua sisi dari persamaan yang diberikan Loh dan shih (1997) sebagai berikut:

$$\frac{p(A|t)}{S_A} \cdot \Phi\left\{\frac{x - \bar{x}_A}{s_A}\right\} = \frac{p(B|t)}{S_B} \cdot \Phi\left\{\frac{x - \bar{x}_B}{s_B}\right\} \quad (6)$$

dengan $\phi(x) = \frac{1}{\sqrt{2\pi}} \cdot \exp\left(-\frac{x^2}{2}\right)$ menunjukkan fungsi densitas normal standar, maka:

$$a = S_A^2 - S_B^2 \quad (7)$$

$$b = 2(\bar{x}_A S_B^2 - \bar{x}_B S_A^2) \quad (8)$$

$$c = (\bar{x}_B S_A)^2 - \left(\bar{x}_A S_B \right)^2 + 2S_A^2 S_B^2 \cdot \ln\left(\frac{p(A|t) \cdot S_B^2}{p(B|t) \cdot S_A^2} \right) \quad (9)$$

3. Sebuah simpul disekat pada $x = d$, dengan d didefinisikan sebagai berikut:

a. Jika $a = 0$, maka

$$d = \begin{cases} \frac{\bar{x}_A + \bar{x}_B}{2} - \frac{S_A^2 \cdot \ln\left(\frac{p(A|t) \cdot S_B}{p(B|t) \cdot S_A}\right)}{(\bar{x}_A - \bar{x}_B)}, & \bar{x}_A \neq \bar{x}_B \\ \bar{x}_A, & \bar{x}_A = \bar{x}_B \end{cases} \quad (10)$$

b. Jika $a \neq 0$, maka

(i) Untuk $b^2 - 4ac < 0$, maka $d = \frac{(\bar{x}_A + \bar{x}_B)}{2}$

(ii) Untuk $b^2 - 4ac \geq 0$, maka:

ditentukan bahwa d akar dari $\frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$ yang lebih dekat dengan rata-rata kelas sehingga menghasilkan dua simpul tak kosong.

Untuk mendapatkan penyekatan biner, *QUEST* hanya menggunakan satu dari dua akar sebagai titik sekat yaitu akar nilainya mendekati rata-rata data dari setiap kelas (Loh dan shih, 1997).

b. Menentukan Titik Sekat dengan Variabel Prediktor Kategorik

Sebuah variabel prediktor kategorik dengan kelas $\{b_1, b_2, \dots, b_L\}$ dilakukan permisalan variabel prediktor tersebut dengan X . X ditransformasikan ke dalam variabel numerik ξ untuk setiap kelas. Dilakukan langkah langkah sebagai berikut:

1. Setiap nilai kategori dari variabel prediktor X pada simpul terpilih ditransformasikan kedalam vektor *dummy* L dimensi, $v = (v_1, v_2, \dots, v_L)'$,

$$\text{dengan: } v_l = \begin{cases} 1 & x = b_l \\ 0 & x \neq b_l \end{cases}$$

dengan:

$$l = 1, 2, \dots, L$$

2. Menghitung rata-rata keseluruhan kelas j dari v .

$$\bar{v} = \frac{\sum_{l=1}^L f_l v_l}{N_t} \quad (11)$$

$$\bar{v}^{(j)} = \frac{\sum_{l=1}^L n_l v_l}{N_{j,t}} \quad (12)$$

ditentukan matriks $L \times L$ antar kelas (B) dan total kelas (T)

$$\mathbf{B} = \sum_{j=1}^J N_{j,t}(\bar{\mathbf{v}}^{(j)} - \bar{\mathbf{v}})(\bar{\mathbf{v}}^{(j)} - \bar{\mathbf{v}})' \quad (13)$$

$$\mathbf{T} = \sum_{l=1}^L f_l(\mathbf{v}_l - \bar{\mathbf{v}})(\mathbf{v}_l - \bar{\mathbf{v}})' \quad (14)$$

3. Dilakukan *singular value decomposition* (SVD) pada \mathbf{T} untuk memperoleh $\mathbf{T} = \mathbf{Q}\mathbf{D}\mathbf{Q}'$, dengan \mathbf{Q} adalah matriks orthogonal $L \times L$, $\mathbf{D} = \text{diag}(d_1, d_2, \dots, d_l)$, dengan $d_1 \geq d_2 \geq \dots \geq d_l \geq 0$

Misalkan $\mathbf{D}^{-\frac{1}{2}} = \text{diag}(d_1^*, d_2^*, \dots, d_l^*)$,

dengan $d_l^* = d_l^{-\frac{1}{2}}$, jika $d_l > 0$, dan 0 untuk lainnya.

4. Sederhanakan nilai tunggal pada $\mathbf{D}^{-\frac{1}{2}}\mathbf{Q}'\mathbf{B}\mathbf{Q}\mathbf{D}^{-\frac{1}{2}}$ untuk menentukan vektor eigen \mathbf{a} yang berhubungan dengan nilai eigen yang terbesar.
5. Koordinat diskriminan terbesar dari \mathbf{v} dapat diproyeksikan dengan persamaan berikut:

$$\boldsymbol{\xi} = \mathbf{a}'\mathbf{D}^{-\frac{1}{2}}\mathbf{Q}'\mathbf{v} \quad (15)$$

QUEST mengaplikasikan algoritma pemilihan titik sekat untuk variabel numerik yang dikasus ini diwakili $\boldsymbol{\xi}$ untuk menentukan titik sekat.

2.7.1.3. Proses Pemberhentian Penyekatan Simpul

Proses penyekatan dilakukan secara berulang sampai tidak mungkin untuk dilanjutkan. Menurut Rokach dan Maimon (2008), penyekatan diberhentikan karena:

Proses penyekatan dilakukan secara berulang sampai tidak mungkin untuk dilanjutkan. Menurut Rokach dan Maimon (2008), penyekatan diberhentikan karena:

1. Pada simpul hanya terdapat kasus yang berasal dari salah satu kelas variabel respon.
2. Kedalaman pohon maksimal telah tercapai
3. Jumlah kasus disimpul terminal kurang dari jumlah minimal kasus untuk menjadi orang tua simpul (*parent node*).
4. Jika simpul diseekat, jumlah kasus dalam satu atau lebih anak simpul (*child node*) akan kurang dari jumlah minimal kasus untuk anak simpul, maka proses penyekatan simpul berhenti.
5. Jika semua variabel prediktor mempunyai nilai signifikansi lebih dari nilai alpha (α) yang ditentukan, maka simpul tidak dapat diseekat.

2.8 Ketepatan Pohon Klasifikasi

Sebuah sistem yang melakukan klasifikasi diharapkan dapat melakukan klasifikasi semua set data dengan benar, namun kinerja suatu sistem klasifikasi ini tidaklah sepenuhnya terhindar dari kesalahan. Bentuk kesalahannya adalah dalam mengklasifikasikan objek baru ke dalam suatu kelas (*misclassification*).

Menurut Prasetyo (2012), matriks konfusi merupakan tabel pencatat hasil kerja klasifikasi. Tabel 5 merupakan matriks konfusi yang melakukan klasifikasi masalah biner. Setiap sel f_{ij} dalam matriks menyatakan jumlah *record* (data) dari kelas I yang hasil prediksinya masuk ke kelas j . Misalnya, sel f_{11} adalah jumlah data dalam kelas A yang secara benar dipetakan ke kelas

A, dan f_{10} adalah data dalam kelas A yang dipetakan secara salah ke kelas B. Maka dapat dihitung tingkat akurasi dan tingkat kesalahan prediksi sebagai berikut:

$$1. \text{ Akurasi} = \frac{f_{11}+f_{00}}{f_{11}+f_{10}+f_{01}+f_{00}} \quad (16)$$

$$2. \text{ Tingkat salah prediksi} = \frac{f_{10}+f_{01}}{f_{11}+f_{10}+f_{01}+f_{00}} \quad (17)$$

3. METODOLOGI PENELITIAN

3.1. Jenis dan Sumber Data

Data yang digunakan dalam proses penelitian pembentukan pohon klasifikasi dengan algoritma *QUEST* adalah data pengaduan pelanggan yang direkapitulasi oleh Unit Pelaksana Pelayanan Pelanggan (UP3) Semarang PT. PLN tahun 2019. Data yang digunakan terdiri dari 311 pengaduan pelanggan terselesaikan melebihi batas *recovery time* dan 219 pengaduan pelanggan terselesaikan kurang dari batas *recovery time*.

3.2. Langkah-Langkah Analisis Data

Tahapan analisis data yang dilakukan menggunakan metode algoritma *QUEST* adalah sebagai berikut:

1. Memasukkan data Status sebagai variabel respon (Y) dan Waktu (X_1), Dampak kerusakan (X_2), Jumlah mengalami (X_3), Kelompok penyebab (X_4), Cuaca (X_5) serta jarak (X_6) sebagai variabel prediktor.
2. Pengujian masing masing variabel prediktor (X) berdasarkan jenis data (numerik atau kategorik). Uji Anova dilakukan untuk data numerik dan Uji Chi-kuadrat dilakukan untuk data kategorik, lalu dilakukan perbandingan nilai p -value terhadap koreksi Befferoni.
3. Untuk variabel prediktor (X) dengan p -value lebih besar dari koreksi Befferoni dilanjutkan ke Uji Levene F dan dilakukan perbandingan hasil p -value dari Uji Levene F dengan taraf tertentu.
4. Setelah didapatkan variabel penyekat (X^*) dilakukan analisis diskriminan kuadratik untuk variabel penyekat numerik. Untuk variabel penyekat kategorik dilakukan transformasi (X^*) menjadi koordinat diskriminasi terbesar terlebih dahulu sebelum dilakukan analisis diskriminasi kuadratik.
5. Lalu dilakukan perhitungan nilai d sebagai nilai akar persamaan kuadrat, jika nilai d sama atau mendekati \bar{x} , maka d akan menjadi titik penyekat.
6. Proses Pemilihan variabel penyekat dan titik penyekat dilakukan berulang-ulang sampai proses pembentukan pohon klasifikasi berhenti dengan peraturan pemberhentian yang diberlakukan.
7. Interpretasi pohon klasifikasi yang terbentuk.

4. HASIL DAN PEMBAHASAN

Variabel respon (Y) terdapat dua kelas yaitu A dan B. Kelas A menyatakan data pengaduan pelanggan yang terselesaikan kurang dari *recovery time* dan B menyatakan data pengaduan pelanggan yang terselesaikan lebih dari *recovery time*. Jika banyaknya kasus dari masing masing kelas dinyatakan dalam notasi maka:

$$N_A = 219, N_B = 311, \text{ dan } N = 530$$

dengan:

N_A = banyaknya kasus pengaduan pelanggan yang terselesaikan kurang dari *recovery time*.

N_B = banyaknya kasus pengaduan pelanggan yang terselesaikan lebih dari *recovery time*.

N = banyaknya kasus keseluruhan.

Variabel prediktor (X) terdiri dari 6 variabel dengan rincian sebagai berikut:

X_1 = Waktu saat pelanggan membuat pengaduannya dengan empat kategori yaitu pagi hari (pukul 06.00 – 12.00), siang hari (pukul 12.01 – 17.00), sore hari (pukul 17.01 – 20.00), dan malam hari (20.01 – 05.59).

X_2 = Dampak kerusakan yang didapati saat penyelesaian pengaduan dengan empat kategori yaitu tidak ada kerusakan, kerusakan ringan, kerusakan sedang, dan kerusakan berat.

X_3 = Jumlah pelanggan yang padam atau mengalami hal yang sama dengan pengaduan merupakan data numerik

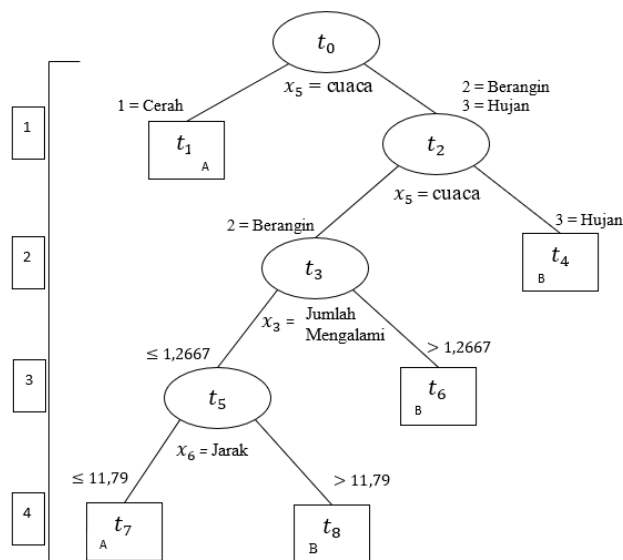
X_4 = Kelompok penyebab pengaduan terjadi yang diketahui oleh petugas dengan lima kategori yaitu kesalahan pemeliharaan, kesalahan operasional, kesalahan konstruksi, tersentuh benda asing, dan mutu material tidak standart.

X_5 = Cuaca saat pengaduan akan ditangani oleh petugas dengan tiga kategori yaitu cuaca cerah, cuaca berangin, dan cuaca hujan.

X_6 = Jarak lokasi pengaduan dengan kantor wilayah PT. PLN merupakan data numerik

4.1. Pembentukan Pohon Klasifikasi

Pada proses penyekatan dilakukan dengan banyaknya kasus dalam *Parent node* paling sedikit 100 kasus, sedangkan banyaknya kasus pada *Child node* paling sedikit 50 kasus, dan untuk nilai batas pengambilan keputusan signifikansi α (α) = 0,05. Setelah penyekatan terhenti karena telah memenuhi aturan pemberhentian yang berlaku, maka pohon klasifikasi sudah terbentuk secara utuh, berikut gambar pohon klasifikasi biner yang terbentuk:



Gambar 1. Pohon Klasifikasi yang Terbentuk

Notasi t menunjukkan simpul. Kedalaman pohon yang terbentuk adalah 4 level, dengan t_0 adalah simpul akar yang menunjukkan kedalaman 0 level. Notasi A menyatakan kelas pengaduan pelanggan yang diselesaikan kurang dari *recovery time* dan B menyatakan kelas pengaduan pelanggan yang diselesaikan lebih dari *recovery time*. Dari gambar 1 ditunjukkan bahwa simpul t_4, t_6 , dan t_8 dirediksi sebagai kelas pengaduan pelanggan yang diselesaikan lebih dari *recovery time*, dan untuk t_1 dan t_7 diprediksi sebagai kelas pengaduan pelanggan yang diselesaikan kurang dari *recovery time*.

4.2. Interpretasi Pohon Klasifikasi

Dari pohon klasifikasi dapat diketahui variabel yang menjadi faktor penting dalam mengklasifikasikannya yaitu x_5 (Cuaca saat pengaduan akan ditangani oleh petugas), x_3 (Jumlah rumah yang padam atau mengalami hal yang sama dengan pengaduan) dan x_6 (Jarak lokasi pengaduan dengan kantor wilayah PT. PLN). Pohon klasifikasi yang terbentuk dihasilkan suatu pohon dengan aturan klasifikasi yang digunakan untuk memprediksi pengaduan pelanggan akan diselesaikan kurang atau lebih dari *recovery time* terdapat 5 simpul akhir, sehingga didapat aturan klasifikasi pada tiap kelasnya. Berdasarkan klasifikasi dapat dijelaskan faktor faktor yang terdapat pada masing masing kelas yaitu:

1. Klasifikasi pertama

Klasifikasi pertama pada pengaduan diprediksi akan diselesaikan kurang dari *recovery time* dengan penyelesaian pengaduan dalam kondisi cuaca berangin, jumlah pelanggan yang mengalami hal yang sama paling banyak 1 pelanggan, dan jarak lokasi pengaduan kurang dari 11,79 KM.

2. Klasifikasi kedua

Klasifikasi kedua pada pengaduan diprediksi akan diselesaikan lebih dari *recovery time* dengan penyelesaian pengaduan dalam kondisi cuaca berangin, jumlah pelanggan yang mengalami hal yang sama paling banyak 1 pelanggan, dan jarak lokasi pengaduan lebih dari 11,79 KM.

3. Klasifikasi ketiga

Klasifikasi ketiga pada pengaduan diprediksi akan diselesaikan lebih dari *recovery time* dengan penyelesaian pengaduan dalam kondisi cuaca berangin dan jumlah pelanggan yang mengalami hal yang sama lebih dari 1 pelanggan.

4. Klasifikasi keempat

Klasifikasi keempat pada pengaduan diprediksi akan diselesaikan lebih dari *recovery time* dengan penyelesaian pengaduan dalam kondisi cuaca hujan.

5. Klasifikasi kelima

Klasifikasi kelima pada pengaduan diprediksi akan diselesaikan kurang dari *recovery time* dengan penyelesaian pengaduan dalam kondisi cuaca cerah.

4.3. Ketepatan Pohon Klasifikasi

Uji ketepatan pohon klasifikasi dalam mengklasifikasikan data dilakukan menggunakan matriks konfusi sebagai berikut:

$$1. \text{ Akurasi} = \frac{202 + 249}{530} = 0,851$$

$$2. \text{ Tingkat salah prediksi} = \frac{17+62}{530} = 0,149$$

Sehingga didapatkan tingkat akurasi pohon klasifikasi dalam mengklasifikasi data sebesar 0,851 atau 85,1% dengan tingkat kesalahan memprediksi sebesar 0,149 atau 14,9%. Hal ini menunjukkan bahwa pohon klasifikasi yang terbentuk memiliki akurasi yang tinggi sehingga klasifikasi yang dibentuk menunjukkan hal yang benar.

5. KESIMPULAN

Variabel yang menjadi faktor penting dalam pengklasifikasian yaitu x_5 (Cuaca saat pengaduan akan ditangani oleh petugas), x_3 (Jumlah rumah yang padam atau mengalami hal yang sama dengan pengaduan) dan x_6 (Jarak lokasi pengaduan dengan kantor wilayah PT. PLN). Ketiga variabel tersebut dipilih menjadi faktor penting dalam pengklasifikasian karena variabel tersebut menjadi penyusun pohon klasifikasi sebagai variabel penyekat dan titik sekat terpilih serta membentuk aturan tiap kelompok pada hasil klasifikasi. Klasifikasi menghasilkan 5 kelas klasifikasi dengan tingkat akurasi pohon klasifikasi yang terbentuk sebesar 0,851 atau 85,1% dengan tingkat kesalahan memprediksi sebesar 0,149 atau 14,9%.

DAFTAR PUSTAKA

- Brown, M. B., Forsythe, A. B. 1974. Robust Tests for the Equality of Variances. *Journal of the American Statistical Association*, Vol. 69, Hal: (346), 364–367
- Grabczewski, K. 2014. *Meta-Learning in Decision Tree Induction*. Switzerland: Springer International Publishing.
- Kim, H., Loh. W.-Y. 2001. Classification Tree with Unbiased Multiway Splits. *Am. Statist. Assoc.* Vol.96, Hal: 590-604
- Lestari, R. D. 2005. Identifikasi Rumah Tangga Miskin di Kota Padang Menggunakan Metode CHAID. *Skripsi*. Institut Teknologi Bandung. Bandung
- Lewis, R.C., Booms, B.H. (1983). The marketing aspects of service quality. *Proceeding Marketing: American Marketing Association*, Chicago.
- Loh, W.-Y., Shih, Y.-S. 1997. Split Selection Methods for Classification Trees, *Statistica Sinica* Vol.7, Hal: 815-840
- Marsudi, D. 2006. *OPERASI SISTEM TENAGA LISTRIK*. Edisi Ke 2. Yogyakarta: Graha Ilmu
- Maroco, J., Silva, D., Rodrigues, A., Guerreiro, M., Santana, I., dan de Mendonca, A. 2011. Data Mining Methods in The Prediction of Dementia. *BMC Research Notes*. Vol.4, Hal: 299.
- Prasetyo, E. 2012. *Data Mining: Konsep dan Aplikasi menggunakan MATLAB*. Yogyakarta: C.V. Andi Offset.
- Rachmatin, D., Kania. S. 2009. Penerapan Prosedur Lachenbruch Pada Kasus Quadratic Discriminant Analysis. *Prosiding Seminar Nasional 2009*, Universitas Pendidikan Indonesia.
- Rokach, L., Maimon, O. 2008. *Data Mining with Decision Trees "Theory and Application"*. USA: World Scientific Publishing Co. Pte. Ltd.
- Shanefield, D. J. 2001. *Industrial Electronics for Engineers, Chemists, and Technicians*. New York: William Andre Publishing.
- Walpole, R. E. 1992. *Pengantar Statistika*. Edisi Ke 3. Jakarta: PT. Gramedia Pustaka Utama.