

KLASIFIKASI JENIS BERAS PECAH KULIT MENGUNAKAN METODE POHON KLASIFIKASI

Dedy Sugiarto, Lukmanul Hakim, Ryan Brilliant

Fakultas Teknologi Industri
Universitas Trisakti
email: dedy@trisakti.ac.id
Jl. Kyai Tapa No. 1, Jakarta 11440, Indonesia

Abstrak

Penelitian ini bertujuan untuk menghasilkan pohon klasifikasi jenis beras pecah kulit dengan dua kelas yaitu beras pecah kulit pandanwangi dan beras pecah kulit muncul. Atribut yang digunakan untuk memprediksi kelas adalah beberapa parameter uji yaitu kadar air, proporsi patahan, proporsi butir mengapur, proporsi butir kuning dan proporsi butir gabah. Data didapatkan dari hasil pengujian bahan baku di sebuah laboratorium beras perusahaan pengolahan beras modern. Statistik uji t digunakan untuk mendapatkan gambaran awal perbedaan rata-rata antar grup dan metode pohon klasifikasi yang menggunakan package rpart dari R digunakan untuk mendapatkan pohon dan hasil prediksi. Hasil penelitian menunjukkan pohon klasifikasi memiliki tingkat akurasi sebesar 81.25%, sensitifitas 100% dan yaitu spesifisitas 80% dengan pemisah utama adalah proporsi butir kapur.

Kata kunci: pohon klasifikasi, beras pecah kulit, kadar air, biji kapur, butir patah

1 PENDAHULUAN

PT. XYZ adalah sebuah perusahaan penggilingan beras (*Rice Milling Unit*) yang mengolah beras pecah kulit (beras setengah jadi) menjadi beras dengan kualitas premium. Pengujian mutu bahan baku beras pecah kulit (PK) senantiasa dilakukan dengan mengambil sampel pada setiap kali bahan baku tiba. Parameter uji yang diperiksa adalah kadar air, butir patah (broken), butir kuning, butir mengapur dan butir gabah. Parameter uji bersifat semakin rendah semakin baik (*smaller is better*) namun tetap ditetapkan batas maksimumnya. Jenis bahan baku yang digunakan antara lain beras pecah kulit pandanwangi (PK PW) yang berasal dari wilayah Bandung dan selatan Jawa Barat serta beras pecah kulit Muncul yang berasal dari wilayah Cilamaya dan pantura Jawa Barat. Beras jadi yang berasal dari PK Muncul akan diberikan merek Rojolele dan yang berasal dari PK PW akan diberikan merek Pandanwangi.

Pengetahuan terkait berbagai atribut yang dapat digunakan untuk memprediksi jenis beras pecah kulit merupakan salah satu pengetahuan kunci yang dibutuhkan perusahaan untuk dapat menjaga mutu produknya. Salah satu metode prediksi dalam *data mining* yang digunakan adalah metode pohon keputusan (*decision tree*). Metode ini merupakan salah satu metode data mining yang banyak digunakan oleh peneliti karena mudah dalam interpretasi model yang dihasilkan (Sartono dan Safitri, 2010). *Decision tree* seringkali disebut dengan *classification tree* jika variabel dependennya bertipe kategorik. Sedangkan jika variabel

dependennya bertipe numerik metode ini seringkali disebut sebagai metode *regression tree* (Mardika et al,2016; Kayri dan Kayri, 2015). Oleh karena itu tujuan penelitian ini adalah untuk menghasilkan pohon klasifikasi jenis beras pecah kulit dengan dua kelas yaitu PK PW dan PK Muncul.

2 METODOLOGI PENELITIAN

Data pada penelitian ini adalah data hasil pengujian mutu beras pecah kulit yang diperoleh dari sebuah pabrik pengolahan beras modern di Pasar Induk Beras Cipinang Jakarta. Terdapat 5 parameter uji yang diteliti dengan dua variable grup yaitu beras PK Muncul dan beras PK PW. Jumlah baris data yang didapatkan adalah sebanyak 48 data hasil pengujian beras PK PW dan 14 data hasil pengujian beras PK Muncul. Langkah pertama sebelum dilakukan pembuatan pohon keputusan adalah dengan menggunakan pengujian rata-rata antar jenis beras menggunakan uji t pada taraf signikan 0.05.

Selanjutnya untuk pembuatan pohon keputusan, data dibagi menjadi dua bagian yaitu data *testing* dan data *training*. Data *training* yang digunakan untuk membuat model sebanyak 75% dan jumlah data *testing* yang digunakan untuk melihat kebaikan model adalah sebanyak 25%. Metode yang digunakan pada penelitian ini yaitu *decicion tree* menggunakan algoritma *classification and regression trees* (CART). CART menggunakan *Gini index* untuk mengubah atribut numeric menjadi kategorik 2 level (Gupta et al, 2017). Pada dasarnya dalam membuat sebuah pohon klasifikasi, CART bekerja dalam lima langkah utama (Mardika et al,2016) yaitu :

1. *Tree building process* (pembuatan pohon klasifikasi yaitu pemecahan simpul induk menjadi dua buah simpul anak melalui aturan pemecahan (*splitting rules*) tertentu dan dilakukan secara berulang (*recursive*).
2. *Class assignment* yaitu identifikasi simpul-simpul yang terbentuk pada suatu kelas tertentu melalui aturan pengidentifikasian (*assignment rules*).
3. *Stopping the trees building process* yaitu penghentian pembuatan pohon klasifikasi
4. *Pruning the tree* yaitu proses pemangkasan atau pemotongan menjadi pohon yang lebih kecil (T).
5. *Optimal tree selection* yaitu penentuan pohon klasifikasi yang optimal.

Software yang digunakan pada analisis ini yaitu *software* R dengan bantuan beberapa *package* seperti *package* (rpart), *package* (rpart.plot) dan yang terakhir *package* (caret). Rpart dapat digunakan untuk algoritma CART (Loh, 2008). Masing-masing fungsi dari *package* tersebut diantaranya *package* rpart digunakan untuk membuat model pada klasifikasi, *package* (rpart.plot) digunakan untuk membuat pohon keputusan sedangkan yang terakhir *package* (caret) digunakan untuk melihat hasil dari prediksi baik itu akurasi, sensitifitas dan spesifisitas. Berikut ini adalah rumus dari kebaikan model, akurasi, sensitifitas dan spesifisitas (Han dan Kember, 2006).

Table 1: Tabel Confusion Matrix

Aktual	Prediksi	
	Ya	Tidak
Ya	TP	FN
Tidak	FP	TN

Ket : TP (*True Positive*), TN (*True Negative*), FP (*False Positive*), FN (*False Negative*)

$$Akurasi = \frac{TP + TN}{TP + FN + FP + TN} \quad (1)$$

$$\text{Sensitifitas} = \frac{TP}{TP + FN} \quad (2)$$

$$\text{Spesifisitas} = \frac{TN}{TN + FP} \quad (3)$$

3 HASIL DAN PEMBAHASAN

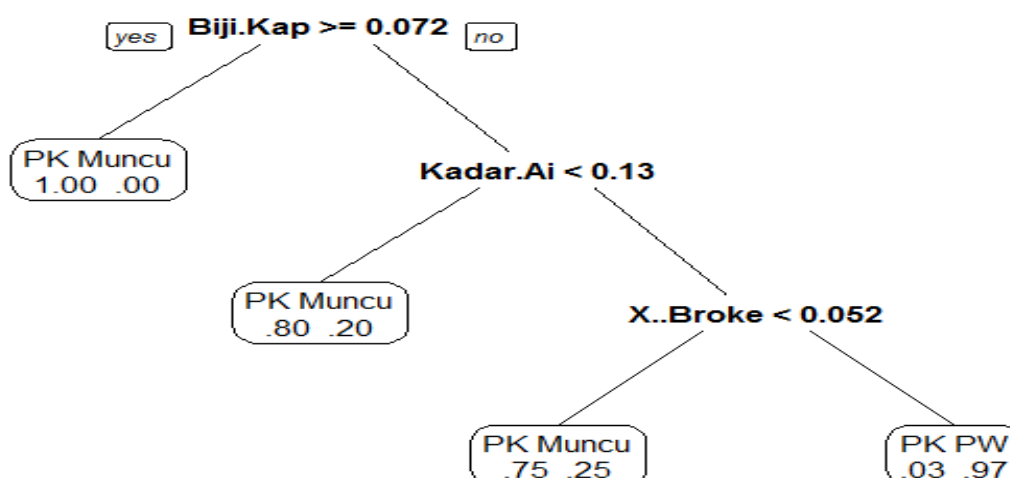
Hasil pengujian statistik perbedaan rata-rata berbagai parameter uji yaitu kadar air, persentase patahan, persentase biji kapur, persentase gabah dan persentase biji kuning dapat dilihat pada Tabel 2. Hasil uji menggunakan uji-t dengan taraf nyata 5% atau 0.05. Hasil pengujian dilihat berdasarkan nilai p-value dimana kedua grup jenis beras dinyatakan berbeda signifikan nilai rata-ratanya bila p-value kurang dari taraf nyata. Berdasarkan pengujian tersebut dapat diketahui bahwa parameter uji sampel bahan baku yang signifikan adalah persentase patahan (X..Broken) dan persentase biji kapur (Biji.Kapur). Hal ini menunjukkan secara statistic, kedua parameter tersebut dapat menjadi pembeda antar grup atau jenis bahan baku beras.

Tabel 2: Hasil Uji Perbedaan Rata-rata (Taraf nyata 0.05)

	Kadar.Air	X..Broken	Biji.Kapur	Biji.Gabah	Biji.Kuning
Mean_PK Muncul	0.133	0.077	0.057	0.013	0.000
Mean_PK PW	0.137	0.101	0.046	0.015	0.001
p-value	0.142	0.017*	0.049*	0.386	0.11
Signifikansi	Tidak	Ya	Ya	Tidak	Tidak

Hasil pohon klasifikasi yang menggunakan data training (75%) data juga menunjukkan hal yang hamper senada dimana pemisah utama pohon klasifikasi adalah persentase biji kapur (Gambar 1). Jika biji kapur nilainya lebih besar dari 0.072 dengan kadar air lebih besar dari 0.13 dan Broken lebih besar dari 0.052 maka jenis berasnya adalah PK PW sebesar 97% dan sisanya sebanyak 3% jenis beras PK Muncul. Jika biji kapur nilainya lebih besar dari 0.072 dengan kadar air lebih besar dari 0.13 dan broken kurang dari 0.052 maka jenis berasnya adalah PK Muncul sebesar 75% dan sisanya sebanyak 25% jenis beras PK PW. Jika Biji Kapur nilainya lebih besar dari 0.0071 dan nilai kadar airnya kurang dari 0.13 maka jenis berasnya PK Muncul sebesar 80% sisanya sebesar 20% jenis beras PK PW. Jika biji kapurnya kurang dari .071 secara keseluruhan jenis berasnya PK Muncul sebanyak 100%. sedangkan jenis beras PK PW tidak ada dengan kriteria tersebut

Setelah melihat pola yang terbentuk dari pohon klasifikasi, langkah selanjutnya yaitu melihat hasil prediksi dari model yang dibuat terhadap data testing (Gambar 2). Secara keseluruhan metode ini mampu memprediksi kelas testing dengan akurasi 81.25%. Selain itu metode decision tree mampu menebak dengan baik secara keseluruhan jenis beras PK PW dilihat dari nilai sensitifitas dimana nilai yang didapatkan sebesar 100%. Hasil tebakan yang terakhir yaitu metode decision tree hanya mampu menebak jenis beras PK Muncul sebesar 80% dapat dilihat dari nilai spesifisitas yang dihasilkan. Angka-anangka tersebut berada diatas 80% hal ini membuktikan bahwa pada kasus beras pecah kulit metode decision tree bekerja dengan baik.



Gambar 1: Pohon Klasifikasi

```

pred.respon PK Muncul PK PW
PK Muncul    1      3
PK PW        0     12

      Accuracy : 0.8125
      95% CI   : (0.5435, 0.9595)
No Information Rate : 0.9375
P-Value [Acc > NIR] : 0.9849

      Kappa : 0.3333
McNemar's Test P-Value : 0.2482

      Sensitivity : 1.0000
      Specificity : 0.8000
      Pos Pred Value : 0.2500
      Neg Pred Value : 1.0000
      Prevalence : 0.0625
      Detection Rate : 0.0625
      Detection Prevalence : 0.2500
      Balanced Accuracy : 0.9000

      'Positive' Class : PK Muncul
  
```

Gambar 2: Output R untuk ketepatan klasifikasi

4 KESIMPULAN

Hasil uji statistik menunjukkan terdapat dua parameter uji yang signifikan sebagai pembeda beras pecah kulit PW dan Muncul yaitu proporsi butir kapur dan proporsi butir patah. Beras PK PW memiliki rata-rata proporsi butir kapur yang secara signifikan lebih rendah dibandingkan beras PK Muncul. Sedangkan untuk parameter butir patah, beras PK PW memiliki nilai rata-rata yang lebih tinggi dibandingkan PK Muncul. Pohon klasifikasi yang didapatkan menunjukkan parameter butir kapur merupakan pemisah utama diikuti oleh kadar air dan butir patah. Hasil pengujian dengan menggunakan 25% data menunjukkan nilai akurasi sebesar 81.25 % yang berarti sudah cukup baik untuk dapat digunakan untuk memprediksi jenis beras pecah kulit.

Referensi

- Mardika Z.W, Mukid M.A, Yasin H. (2016). Pembentukan pohon klasifikasi biner dengan algoritma *classification and regression trees* : Studi kasus: kredit macet di PD. BPR-BKK Purwokerto Utara). *Jurnal Gaussian*, 5(3), 583-592
- Gupta, B., Rawat, A, Jain A, Arora A, Dhami, N. (2017). Analysis of Various Decision Tree Algorithms for Classification in Data Mining. *International Journal of Computer Applications*, 163(8), 15-19
- Han J, Kamber M. (2006). *Data mining concepts and techniques second edition*. San Francisco: Diane Cerra.
- Kayri M, Kayri I. (2015). The Comparison of Gini and Twoing Algorithms in Terms of Predictive Ability and Misclassification Cost in Data Mining : An Empirical Study. *International Journal of Computer Trends and Technology (IJCTT)*, 27 (1), 21-30
- Loh, W. (2008). *Classification and Regression Tree Methods (In Encyclopedia of Statistics in Quality and Reliability)*. London : Wiley
- Sartono, B., Syafitri, U.D. (2010). Metode pohon gabungan: solusi pilihan untuk mengatasi kelemahan pohon regresi dan klasifikasi tunggal. *Forum Statistika dan Komputasi*. 15, 1-7