

Analisis Produktivitas Tumbuhan Buah Melalui *Feature Selection*

Miftahuddin, M. Subianto, Zumaidar*

Abstrak

Variabel dan seleksi fitur yang difokuskan dari penelitian ini diterapkan untuk data tumbuhan buah. Suatu objek perlu diketahui fitur-fiturnya agar dapat dikenali dan dibedakan dari objek lain. Fitur-fitur optimal yang dapat diketahui dari suatu objek akan mempermudah dan mempercepat proses identifikasi objek tersebut. Oleh karena itu perlu dilakukan seleksi fitur (*feature selection*). *Feature selection* adalah suatu metode penganalisaan data yang bertujuan untuk memilih fitur yang berpengaruh (fitur optimal) dan mengesampingkan fitur yang tidak berpengaruh. Ada beberapa algoritma *feature selection* yang dapat digunakan, salah satunya adalah *Relief*. *Relief* memanfaatkan teknik bobot (*weight*) untuk mengukur signifikansi fitur dalam konteks klasifikasi. Fitur yang memiliki nilai bobot di atas ambang batas (*threshold*) yang digunakan akan dipilih. Penelitian ini bertujuan untuk mendapatkan fitur optimal dari data tumbuhan buah di kabupaten Aceh Tengah dan Bener Meriah. Hasil menunjukkan untuk setiap data yang diuji hanya menghasilkan satu fitur optimal dengan nilai *threshold* berbeda. Hubungan yang terjadi antara jumlah data dan *threshold* terhadap fitur optimal adalah terbalik.

Kata Kunci: *Feature selection, algoritma Relief, threshold, fitur optimal.*

Abstract

Variable and feature selection have become the focus of this research in areas of application for fruit plants datasets. An object needs to be known its features to be recognizable and distinguished from other objects. Optimal features that can be known from an object will simplify and accelerate the process of identifying the object. Therefore needs the feature selection to select the influential features (optimal feature) and discard not effected features. Feature selection has varied method or algorithm, one of which is Relief. Relief utilizing value of weights to measure the significance features in classification. Features that have a weight value above the threshold used will be selected. The purpose of this study is to get optimal features of fruit plants datasets of Aceh Tengah and Bener Meriah. The result shows for any data that tested only produce one optimal feature with a different threshold value. And the relationship between the amount of data and the threshold toward optimal features is reversed.

Keywords: Feature selection, relief algorithm, threshold, optimal features.

*Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam,
Universitas Syiah Kuala, Nangroe Aceh Darussalam

1. Pendahuluan

Provinsi Aceh terletak di bagian barat Indonesia tepatnya di bagian ujung pulau Sumatera. Secara geografis Aceh terletak antara $2^{\circ} - 6^{\circ}$ LU dan $95^{\circ} - 98^{\circ}$ LS dengan ketinggian rata-rata 125 m di atas permukaan laut. Luas wilayah Aceh $57.365,57 \text{ km}^2$, atau 12,26% dari luas pulau Sumatera yang terdiri atas hutan bagian terluas dari Aceh dengan luas $39.615,76 \text{ km}^2$, lahan perkebunan $3.135,22 \text{ km}^2$ dan lahan pertambangan mempunyai luas $4,42 \text{ km}^2$ [1].

Provinsi Aceh merupakan provinsi yang sangat cocok untuk budidaya berbagai jenis komoditi perkebunan, salah satunya adalah tumbuhan buah. Secara tradisi Aceh di Istana Kerajaan masa Sultan Iskandar Thani terdapat suatu kebun bunga dan buah yang digelar Taman Ghairah yang kini dikenal Taman Putroe Phang. Di taman tersebut ditanam berpuluh jenis bunga dan buah-buahan. Hal ini menunjukkan adanya upaya pelestarian terhadap jenis-jenis tanaman baik bunga maupun buah yang dilakukan pemerintah. Perhatian pemerintah Aceh terhadap jenis buah-buahan telah dilakukan sejak lama, terutama di kabupaten Aceh Tengah dan Bener Meriah yang merupakan daerah perekonomian dengan mayoritas pekerjaan penduduknya bersumber dari pemanfaatan hasil alam. Melihat dari kondisi topografi daerah yang dikenal sebagai kawasan pegunungan itu, Aceh Tengah dan Bener Meriah sangat cocok untuk pengembangan tumbuhan buah dan memiliki prospek sangat baik. Ukuran perkebunan yang luas serta cuaca yang dingin sangat sesuai untuk tanaman budidaya buah-buahan seperti alpukat, jeruk, pisang, nenas, markisah dan sebagainya.

Kabupaten Aceh Tengah dengan ibukota Takengon terdiri 14 kecamatan dan 271 gampong/desa. Kabupaten ini berbatasan dengan: kabupaten Aceh Timur sebelah timur, kabupaten Gayo Lues sebelah selatan, kabupaten Nagan Raya dan Aceh Barat sebelah barat, Bireuen dan Pidie sebelah barat laut, serta berbatasan dengan Bener Meriah sebelah utara. Secara geografis kabupaten ini terletak pada $4^{\circ}10'33'' - 5^{\circ}57'50''$ LU dan $95^{\circ}15'40'' - 97^{\circ}20'25''$ BT. Luas daerah $4.318,39 \text{ km}^2$ dan tinggi rata-rata 200 -2.600 m di atas permukaan laut.

Sedangkan kabupaten Bener Meriah dengan ibukota Simpang Tiga Redelong terdiri atas 7 kecamatan, 13 mukim, dan 232 gampong/desa. Kabupaten Bener Meriah merupakan kabupaten pemekaran dari kabupaten Aceh Tengah yang terletak pada tinggi rata-rata 100 – 2.500 m di atas permukaan laut. Kabupaten ini memiliki luas daerah $1.888,70 \text{ km}^2$ dan terletak pada $4^{\circ}33'50'' - 4^{\circ}54'50''$ LU dan $96^{\circ}40'75'' - 97^{\circ}17'50''$ BT. Kabupaten ini masih tergolong muda sejak 2004 atau dibagian tengah provinsi Aceh yang berbatasan langsung dengan kabupaten Aceh Utara sebelah utara, kabupaten Aceh Timur sebelah timur, kabupaten Aceh Tengah sebelah selatan dan barat serta kabupaten Bireuen sebelah barat laut.

Adanya pasar bebas memungkinkan kedua kabupaten tersebut mengerahkan potensi dan kemampuan petani atau pedagang buah dalam rangka meningkatkan kemampuannya untuk bersaing, baik memproduksi dan memasarkan hasil tumbuhan buah. Dengan kata lain, harus mampu bersaing dalam perekonomian global yang kompetitif, karena kemampuan produksi dan pemasaran dilandaskan pada kemampuan menciptakan bibit tumbuhan buah yang unggul dalam perkembangan pertanian dan perkebunan. Karena itu, perlu dilakukan pengklasifikasian yang tepat terhadap fitur-fitur yang mempengaruhi produktivitas tumbuhan buah sehingga dapat diketahui faktor-faktor apa saja yang perlu diutamakan guna meningkatkan nilai ekonomi petani buah. Sementara itu banyak *software* yang dapat digunakan untuk menganalisis fenomena tersebut, di antaranya dengan menggunakan *software R*.

Tumbuhan buah merupakan hasil perkebunan yang banyak mendatangkan keuntungan di Indonesia. Hal ini didukung oleh sumber daya lahan dan potensi buah yang sangat beragam. Pengertian buah-buahan mengacu pada pengertian sehari-hari adalah bagian tanaman yang memiliki bau aromatis dan rasa manis. Tumbuhan buah juga memiliki peranan ekonomi sebagai bahan pangan atau bahan baku industri karena di dalamnya tersimpan berbagai macam produk metabolisme, mulai dari karbohidrat, protein, lemak, vitamin, mineral, alkaloid hingga terpenoid.

Keanekaragaman tumbuhan buah merupakan sumber genetik yang tidak akan pernah tergantikan jika hilang ataupun punah. Plasma nutfah ini menjadi bahan baku utama bagi pengembangan jenis baru atau varietas unggul buah-buahan di masa yang akan datang. Jenis-jenis tumbuhan buah liar memiliki beberapa kelebihan bagi pemilihan sifat guna penyilangan jenis.

Adapun rumusan masalah dalam penelitian ini adalah bagaimana produktivitas dan fenomena tumbuhan buah, sekaligus bagaimana pengaruh perubahan iklim terhadap tumbuhan buah di kabupaten Aceh Tengah dan Bener Meriah. Sedangkan tujuan penelitian adalah untuk menganalisis beberapa faktor ekonomi, yaitu analisa endogen (nilai penjualan, keuntungan *netto*, keuntungan *bruto*, besar biaya waktu tanam, besar biaya waktu pemeliharaan, besar biaya waktu panen) dan analisa eksogen (pengaruh cuaca dan iklim serta pola tanam) yang mempengaruhi produktivitas tumbuhan buah. Sehingga dapat diketahui produktivitas dan faktor-faktor mana yang perlu diutamakan dalam melihat produktivitas tumbuhan buah.

Selain dapat memberikan kemudahan dalam pengumpulan fitur-fitur yang akan dipilih sehingga dapat diketahui faktor-faktor yang paling mempengaruhi produktivitas tumbuhan buah di Aceh Tengah dan Bener Meriah, manfaat penelitian ini adalah dapat menjadi acuan atau bahan pertimbangan bagi para petani tumbuhan buah dalam menghadapi perubahan iklim, dan memberikan kemudahan dalam proses pengklasifikasian fitur yang memiliki dimensi besar dan juga dapat menentukan fitur terbaik (optimal) diantara fitur-fitur yang ada.

2. Landasan Teori

2.1. Pengertian Klasifikasi

Menurut [2], klasifikasi data merupakan langkah awal dalam analisis data. Klasifikasi data berarti memisah-misahkan sifat-sifat dari data yang heterogen ke dalam kelompok-kelompok yang homogen, sehingga sifat-sifat data yang menonjol mudah dilihat.

Adapun tujuan klasifikasi data adalah :

1. Mengelompokkan sifat-sifat yang sama ke dalam kelompok atau kelas tertentu.
2. Mempermudah untuk membandingkan.
3. Mengelompokkan informasi yang menonjol dan menghilangkan hal-hal tidak perlu.
4. Menunjukkan sifat-sifat yang menonjol sehingga mudah dilihat secara sekilas.
5. Mempermudah mengadakan perlakuan secara statistik terhadap data yang telah dikumpulkan.

2.2. Feature Selection

2.2.1 Definisi Feature Selection

Menurut [3], *Feature Selection* adalah suatu proses yang mencoba untuk menemukan sub himpunan dari himpunan fitur yang tersedia untuk meningkatkan aplikasi dari suatu algoritma pembelajaran. *Feature Selection* digunakan dibanyak area aplikasi sebagai alat untuk menghilangkan fitur yang tidak relevan dan atau fitur berlebihan. Sebuah fitur dikatakan tidak relevan jika memberikan sedikit informasi, sedangkan sebuah fitur dikatakan berlebihan jika informasi yang diberikan adalah informasi yang terkandung dalam fitur lain (tidak memberikan informasi baru).

2.2.2 Metode Feature Selection

Secara garis besar ada empat langkah dalam metode *Feature Selection* yaitu :

1. Prosedur generasi (pembangkitan), untuk menghasilkan calon subhimpunan berikutnya dapat dilakukan dengan beberapa cara yaitu : lengkap, heuristik dan acak.

2. Evaluasi fungsi, untuk mengevaluasi subhimpunan, dengan cara mengukur jarak, informasi, konsistensi, ketergantungan, dan mengukur tingkat kesalahan klasifikasi.
3. Kriteria penghentian, untuk memutuskan kapan harus berhenti, dengan cara melihat nilai ambang batas, diawali dengan sejumlah pengulangan dan sebuah ukuran subhimpunan fitur terbaik.
4. Prosedur validasi, untuk memeriksa apakah subhimpunan valid.

Secara jelas, keempat langkah tersebut diuraikan berikut.

Prosedur Generasi

Prosedur generasi merupakan prosedur pencarian yang pada dasarnya menghasilkan *subset* (subhimpunan) dari fitur-fitur untuk dievaluasi. Prosedur generasi dapat dimulai: (1) Tanpa fitur, (2). Dengan semua fitur, dan (3). Dengan sub himpunan acak dari fitur.

Dalam dua kasus pertama, fitur secara iteratif ditambahkan atau dihapus, sedangkan dalam kasus terakhir, fitur secara iteratif ditambahkan atau dihapus secara acak setelahnya.

Jika himpunan fitur asli berisi N jumlah fitur, maka jumlah calon bersaing untuk menjadi sub himpunan yang dihasilkan adalah 2^N . Ini merupakan jumlah besar bahkan untuk setengah dari jumlah N . Ada berbagai pendekatan untuk menyelesaikan masalah ini, yaitu:

(a) Lengkap

Urutan ruang pencarian prosedur generasi ini adalah $O(2^N)$, sebuah sub himpunan yang sedikit untuk dievaluasi. Sub himpunan fitur yang optimal sesuai dengan evaluasi fungsi, karena prosedur ini dapat dilakukan dengan cara mundur. Mundur dapat dilakukan dengan menggunakan berbagai teknik, seperti: *branch and bound*, pencarian pertama terbaik, dan balok pencarian.

(b) Heuristik

Dalam setiap pengulangan prosedur generasi ini, semua sisa fitur yang belum dipilih (ditolak) masih dipertimbangkan untuk pemilihan (penolakan). Ada banyak variasi untuk proses sederhana ini, tapi generasi subhimpunan pada dasarnya meningkat atau menurun. Urutan ruang pencarian adalah $O(N^2)$ atau kurang. Prosedur ini sangat sederhana untuk diterapkan dan sangat cepat dalam memperoleh hasil, karena ruang pencarian hanya kuadrat dari jumlah fitur.

(c) Acak

Prosedur generasi ini masih baru dalam penggunaannya dalam metode *Feature Selection* dibandingkan dengan dua kategori lainnya. Meskipun ruang pencarian adalah $O(2^N)$, tetapi metode ini biasanya mencari lebih sedikit jumlah subhimpunan daripada 2^N dengan menetapkan jumlah maksimum pengulangan yang mungkin. Optimalitas subhimpunan yang dipilih tergantung pada sumber daya yang tersedia. Setiap prosedur generasi acak akan memerlukan nilai-nilai dari beberapa parameter.

Evaluasi Fungsi

Evaluasi fungsi mengukur kebaikan sub himpunan yang dihasilkan oleh beberapa prosedur generasi, dan nilai ini dibandingkan dengan yang terbaik sebelumnya. Jika ditemukan yang lebih baik, maka sub himpunan terbaik sebelumnya digantikan. Ada beberapa cara dalam melakukan evaluasi fungsi yaitu :

(a) Ukuran Jarak

Juga dikenal sebagai keterpisahan, perbedaan, atau diskriminasi ukuran. Untuk dua kelas, fitur X adalah fitur yang lebih disukai dari fitur Y apabila X menginduksi perbedaan yang lebih besar antara kedua kelas probabilitas kondisional dari Y dan jika perbedaan adalah nol, maka X dan Y tidak dapat dibedakan (sama). Sebagai contoh jarak *Euclidean* yang merupakan metode pengukuran jarak di antara dua objek berdasarkan akar jumlah kuadrat jarak kedua objek. Rumus umum jarak *Euclidean* yaitu, jika X berkoordinat (x_1, x_2, \dots, x_n) dan objek Y berkoordinat (y_1, y_2, \dots, y_n) , maka jarak *Euclidean* kedua objek tersebut menurut [5] adalah

$$d(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_n - y_n)^2}$$

(b) Informasi Ukuran

Langkah ini biasanya untuk menentukan kualitas informasi dari suatu fitur. Fitur X lebih disukai dari pada fitur Y jika kualitas informasi dari fitur X adalah lebih besar dari pada yang dari fitur Y (misalnya, ukuran entropi). *Entropy* dapat diukur dengan rumusan:

$$Entropy(D_i) = - \sum_{j=1}^k Pr_i(c_j) \log_2 Pr_i(c_j),$$

dimana $Pr_i(c_j)$ adalah rasio data dengan kategori c_j dalam kelas D_i . Total *entropy* dari seluruh kelas (kelompok) menurut [5] dihitung dengan :

$$Entropy_{total}(D_i) = \sum_{i=1}^k \frac{|D_i|}{|D|} \times entropy(D_i),$$

dengan D_i adalah jumlah data kelas i dan D adalah total jumlah data.

(c) Ukuran Ketergantungan

Ukuran ketergantungan sama halnya dengan mengukur korelasi. Koefisien adalah ukuran ketergantungan klasik dan dapat digunakan untuk menemukan korelasi antara fitur dan kelas. Jika korelasi fitur X dengan kelas C lebih tinggi dari pada korelasi fitur Y dengan C , maka fitur X lebih disukai dari pada Y . Nilai ini untuk menentukan ketergantungan fitur pada fitur lain dan menunjukkan tingkat redundansi dari fitur.

(d) Ukuran Konsistensi

Ini adalah karakteristik yang berbeda dari ukuran lain, karena ketergantungannya pada kumpulan data pelatihan dan penggunaan *min-feature bias* dalam memilih sebuah subhimpunan dari fitur.

(e) Ukuran Tingkat Kesalahan Klasifikasi

Metode-metode yang menggunakan jenis evaluasi fungsi ini disebut "metode *wrapper*", yaitu klasifikasi adalah evaluasi fungsi. Sebagai fitur yang dipilih menggunakan klasifikasi yang kemudian akan digunakan dalam memperkirakan label kelas dari contoh yang tak terlihat, memiliki tingkat akurasi sangat tinggi namun komputasi cukup mahal.

Kriteria Penghentian

Prosedur generasi dan evaluasi fungsi dapat mempengaruhi pilihan untuk kriteria penghentian. Kriteria penghentian berdasarkan prosedur generasi meliputi:

1. Apakah yang sudah ditetapkan sejumlah fitur yang dipilih dan

2. Apakah yang sudah ditetapkan mencapai jumlah pengulangan.

Sedangkan kriteria penghentian berdasarkan evaluasi fungsi dapat berupa:

1. Apakah tambahan (atau penghapusan) dari setiap fitur tersebut tidak menghasilkan *subset* (subhimpunan) yang lebih baik dan
2. Apakah bagian yang optimal menurut beberapa evaluasi fungsi diperoleh.

Prosedur Validasi

Proses validasi bukan merupakan bagian dari proses *Feature Selection* itu sendiri, namun sebuah *Feature Selection* harus divalidasi yaitu dengan cara melakukan pengulangan terhadap evaluasi fungsi sub himpunan dari fitur sampai kriteria penghentian terpenuhi [3].

2.3. Relief Feature Selection

2.3.1 Algoritma Relief

Relief adalah algoritma *Feature Selection* yang terinspirasi oleh *instance-based learning*, yang memanfaatkan teknik bobot untuk mengukur signifikansi fitur. Misal diberikan data pelatihan δ , besar sampel m dan batas relevansi τ , *Relief* akan mendeteksi himpunan fitur secara statistik yang relevan dengan konsep target. τ adalah *threshold* (ambang batas) relevan yang berada pada interval $0 \leq \tau \leq 1$. Dengan asumsi bahwa skala dari setiap fitur nominal atau numerik (*integer* atau *real*) maka selisih nilai fitur antara dua contoh X dan Y didefinisikan oleh fungsi *difference* sebagai berikut :

- a. Jika x_k dan y_k nominal, $\text{diff}(x_k, y_k) = 0$, jika x_k dan y_k sama; $\text{diff}(x_k, y_k) = 1$, jika x_k dan y_k berbeda.
- b. Jika x_k dan y_k numerik, $\text{diff}(x_k, y_k) = \frac{(x_k - y_k)}{\text{nu}_k}$, dimana, nu_k adalah unit normalisasi untuk menormalkan nilai *difference* (selisih) ke dalam interval $[0, 1]$.

Relief Feature Selection merupakan *Feature Selection* yang masuk ke dalam kelompok *subset selection* tipe filter. Secara garis besar ada dua kelompok besar dalam pelaksanaan *Feature Selection* yaitu *Ranking Selection* dan *Subset Selection*. *Ranking Selection* adalah metode yang secara khusus memberikan ranking pada setiap fitur yang ada dan mengesampingkan fitur yang tidak memenuhi standar tertentu. Kelompok ini menentukan tingkat ranking secara bebas antara satu fitur dengan fitur yang lainnya. Fitur yang mempunyai ranking tinggi akan digunakan dan yang rendah akan di kesampingkan. Sedangkan *Subset Selection* adalah metode pemilihan yang mencari suatu himpunan dari fitur yang dianggap sebagai fitur yang optimal [4].

2.3.2 Threshold τ (Ambang Batas)

Threshold merupakan nilai batas relevan untuk pemilihan fitur optimal. Nilai *threshold* berada pada interval 0 sampai 1 ($0 \leq \tau \leq 1$), dan penggunaannya bersifat independen tergantung pada pengguna. Dalam algoritma *Relief*, *threshold* akan dibandingkan dengan nilai *weight* (bobot) dari suatu fitur. Apabila suatu fitur memiliki nilai bobot lebih besar dari *threshold* yang digunakan maka fitur tersebut merupakan fitur optimal sedangkan jika nilai bobot fitur lebih kecil sama dengan (\leq) dari *threshold* maka fitur tersebut tidak akan dipilih karena tidak termasuk dalam kategori fitur optimal [6,7].

3. Metode Penelitian

3.1. Sumber Data

Data yang digunakan berasal dari data sekunder Dinas Tanaman Pangan dan Holtikultura Kabupaten Aceh Tengah dan Bener Meriah serta data primer yaitu data tumbuhan buah dengan beberapa faktor yang mempengaruhinya dalam riset unggulan strategis nasional (Rusnas) batch II Tahun 2009 oleh Subianto dkk. pada tahun 2009.

3.2. Identifikasi Fitur

Terdapat 18 fitur penjelas dan 1 fitur respon yang digunakan pada penelitian ini. Fitur yang terdapat dalam penelitian merupakan item dari setiap pertanyaan dalam kuesioner dan merupakan faktor-faktor yang diduga memiliki pengaruh terhadap produktivitas tumbuhan buah. Jika suatu gejala diduga dipengaruhi oleh beberapa fitur kualitatif maka dibutuhkan pengklasifikasian kelompok (kategori). Setiap fitur memiliki 2 sampai 5 kategori, seperti yang terlihat pada tabel berikut.

Tabel 1. Fitur-fitur yang Digunakan dalam Analisa Algoritma *Relief*.

Fitur	Keterangan	Kategori
F_1	Keuntungan <i>netto</i> per musim	1. \leq Rp. 100.000 (1)
		2. Rp. 100.000 – Rp. 500.000 (2)
		3. Rp. 500.000 – Rp. 1.000.000 (3)
		4. Rp. 1.000.000 – Rp. 5.000.000 (4)
		5. $>$ Rp. 5.000.000 (5)
F_2	Keuntungan <i>bruto</i> per musim	1. \leq Rp. 100.000 (1)
		2. Rp. 100.000 – Rp. 500.000 (2)
		3. Rp. 500.000 – Rp. 1.000.000 (3)
		4. Rp. 1.000.000 – Rp. 5.000.000 (4)
		5. $>$ Rp. 5.000.000 (5)
F_3	Besar biaya waktu tanam	1. \leq Rp.700.000 (1); 2. $>$ Rp. 700.000 (2)
F_4	Besar biaya waktu pemeliharaan	1. \leq Rp. 400.000 (1); 2. $>$ Rp. 400.000 (2)
F_5	Besar biaya waktu panen	1. \leq Rp. 230.000 (1); 2. $>$ Rp. 230.000 (2)
F_6	Waktu tanam	1. \leq 6 hari (1); 2. $>$ 6 hari (2)
F_7	Waktu pemeliharaan	1. \leq 12 bulan (1); 2. $>$ 12 bulan (2)
F_8	Waktu panen	1. \leq 4 Bulan (1); 2. $>$ 4 Bulan (2)
F_9	Jenis Tumbuhan Buah	1. Alpukat (1); 2. Jeruk (2); 3. Pisang (3)
F_{10}	Perkiraan umur tumbuhan buah	1. \leq 5 tahun (1); 3. 10 – 20 tahun (3)
		2. 5 – 10 tahun (2) 4. $>$ 20 tahun (4)
F_{11}	Jumlah Tenaga Kerja	1. \leq 2 orang (1); 2. $>$ 2 orang (2)
F_{12}	Umur Petani	1. \leq 30 tahun (1); 2. 31 – 40 tahun (2)
		3. $>$ 40 tahun (3)
F_{13}	Luas Lahan (ha)	1. \leq 1/8 ha (1); 4. 1/2 – 3/4 ha (4)
		2. 1/8 – 1/4 ha (2); 5. $>$ 1 ha (5)
		3. 1/4 – 1/2 ha (3)

F_{14}	Pengaruh Cuaca dan Iklim terhadap bunga	1. Bunga rontok (1); 2. Bunga tidak rontok(2)
F_{15}	Pengaruh Cuaca dan Iklim terhadap Ukuran buah	1. Buah menjadi kecil (1) 2. Buah tidak menjadi kecil (2)
F_{16}	Pengaruh Cuaca dan Iklim terhadap Gagal Buah	1. Buah tidak jadi (1); 2. Buah jadi (2)
F_{17}	Pengaruh Cuaca dan Iklim terhadap Buah	1. Buah berjatuhan (1); 2. Buah tidak berjatuhan (2)
F_{18}	Pengaruh Cuaca dan Iklim terhadap Kuantitas hasil panen	1. Hasil panen buah berkurang (1) 2. Hasil panen buah tidak berkurang (2)
F_{19}	Nilai Penjualan per musim	1. \leq Rp. 5.000.000 (1); 2. $>$ Rp. 5.000.000 (2)

3.3. Metode Pengambilan dan Pengolahan Data

Metode pengambilan data menggunakan metode PRA (*Participatory Rural Appraisal*) dan RRA (*Rapid Rural Appraisal*) terhadap pemanfaatan jenis tumbuhan buah dari kedua kabupaten. Sedangkan metode pengolahan data berupa metode statistika deskriptif dan inferensia dengan *Feature Selection* menggunakan algoritma *Relief* dengan tahapan:

1. Melabel data menjadi *binary classification*: kelas positif dan negatif.
2. Menentukan *threshold* yang dipakai 0; 0,1; 0,2; 0,3; 0,4; 0,5; 0,6; 0,7; 0,8; 0,9 & 1
3. Mengimport data *Excel* dengan *R.2.6.1*. Data yang dipanggil merupakan data buah di Aceh Tengah, Bener Meriah dan data gabungan dari kedua kabupaten tersebut setelah diklasifikasi sebelumnya.
4. Langkah 2 dan 3 diulang sebanyak 10 kali untuk setiap *threshold* yang digunakan dari masing-masing data. Dari banyaknya pengulangan yang dilakukan, berapa nilai *weight* dari tiap fitur.
5. Membandingkan hasil nilai bobot dari tiap fitur tersebut dengan nilai *threshold* yang digunakan dan mendapatkan simpulan fitur terbaik (*optimal*). Suatu fitur dikatakan *optimal* apabila nilai bobotnya lebih besar dari nilai *threshold* yang digunakan.

3.4. Deskripsi R dan Analisis

Untuk membantu perhitungan dalam analisa data digunakan *Tinn-R* versi 1.19.2.3 dan *R* versi *R.2.6.1*. Analisis luaran *R* dikaitan dengan kondisi di lapangan dengan menggambarkan diagram alur produktivitas ketiga sampel tumbuhan buah yang diujicobakan berdasarkan tinjauan analisa endogen dan eksogen.

R project pertama kali dikembangkan oleh Robert Gentleman dan Ross Ihaka yang bekerja di departemen statistik Universitas Auckland tahun 1995. Sejak saat itu *software* ini mendapat sambutan yang luar biasa dari kalangan statistikawan, *industrial engineering*, peneliti, *programmer*, dan sebagainya. Saat ini *R* ditangani oleh *R Core Development Team R*. *Software* ini merupakan suatu fasilitas perangkat lunak versi *open source* untuk perhitungan statistik (model linier dan nonlinier, uji statistik klasik, analisis *time-series*, klasifikasi, *clustering* dan sebagainya) dan peragaan grafik. *R* memiliki kemampuan menganalisis data dengan sangat efektif

dan dilengkapi dengan operator pengolahan *array* dan matriks. *R* juga mampu menampilkan grafik yang sangat canggih demikian pula untuk peragaan datanya¹.

4. Hasil dan Pembahasan

Dari hasil riset diperoleh 39 jenis tumbuhan buah dengan 23 famili. Untuk Aceh Tengah dan Bener Meriah, Rutaceae (Jeruk-jerukan) dan Solanaceae (Terong-terongan) adalah kelompok tumbuhan yang banyak jenisnya dibandingkan lainnya. Diduga jenis-jenis tersebut adalah yang paling mudah hidup di kabupaten tersebut karena lingkungannya sesuai dibandingkan jenis dari famili lainnya. Keberadaan jenis ini (lokal) akan semakin bertambah/berkurang sejalan dengan penggunaan bibit tanaman/varietas baru (non lokal) yang dilakukan oleh petani.

Tabel 2. Tumbuhan Buah Potensial di Daerah Kajian (Sept – Nov) Tahun 2009.

No	Kab. Bener Meriah	Kab. Aceh Tengah
1	DURIAN / <i>Durian</i>	ALPUKAT / <i>Avocado</i>
2	JAMBU BIJI / <i>Guava</i>	DURIAN / <i>Durian</i>
3	PEPAYA / <i>Papaya</i>	JAMBU BIJI / <i>Guava</i>
4	PISANG / <i>Banana</i>	JERUK KEPROK
5	NENAS / <i>Pineapple</i>	JERUK SIAM
6	ALPUKAT / <i>Avocado</i>	JERUK LAINNYA
7	JERUK / <i>Orange</i>	NENAS / <i>Pineapple</i>
8	MARKISAH / <i>Marquisa</i>	PEPAYA / <i>Papaya</i>
9	MANGGA/ <i>Mangosteen</i>	PISANG / <i>Banana</i>
10	NANGKA/ <i>Jackfruit</i>	MARKISAH / <i>Marquisa</i>
11	JAMBU AIR	NANGKA / <i>Jackfruit</i>

Untuk Aceh Tengah, luas tanam tahun 2008 (1757 ha), dengan 11 jenis tumbuhan buah potensial, total produktivitas 1.170 ton/ha. Perkembangan 2005-2008, luas tanam naik 22,93%, luas panen turun 0,38%, total produksi turun 3,36%.

Adanya keterkaitan perubahan iklim, demografi dan daya dukung lingkungan menyebabkan perubahan siklus tanam, pemeliharaan/perawatan, dan panen tumbuhan buah. Dari hasil riset diketahui perubahan tersebut tidak hanya berdampak pada volume hasil panen (kuantitas) juga berpengaruh pada kualitas hasil panen. Pengaruh iklim menurut para petani terhadap tumbuhan buah antara lain 88 petani (39,64%) menyatakan bunga rontok/berguguran, 33 petani (14,86%) menyatakan buah menjadi kecil, 21 petani (9,46%) menyatakan buah tidak jadi/gagal menjadi buah, 94 petani (42,34%) menyatakan buah berjatuhan, dan 102 petani (45,95%) menyatakan hasil panen buah berkurang. Bila hal ini berlanjut maka akan mempengaruhi produksi tumbuhan buah dan nilai ekonomi tumbuhan buah (rendah).

4.1 Identifikasi dan Kategori

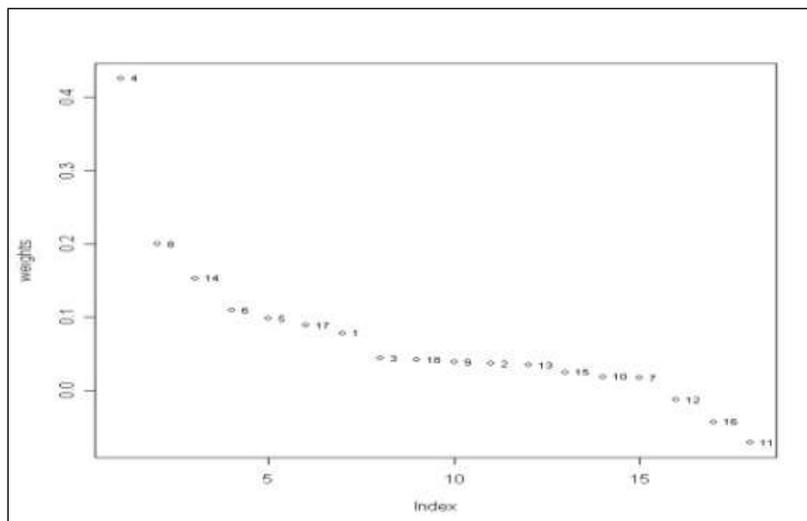
Petani tumbuhan buah yang menjadi responden untuk kabupaten Aceh Tengah berjumlah 53 petani (43,80%) dan Bener Meriah berjumlah 68 petani (56,20%), dengan total keseluruhan responden 121 petani. Dilihat dari segi jenis tumbuhan buah yang ditanam petani, maka tumbuhan buah jeruk 47,1%, alpukat 40,5% dan pisang 12,4%. Dalam penelitian ini analisis faktor endogen yang mempengaruhi produktivitas tumbuhan buah dilihat pada nilai penjualan, keuntungan *bruto*, keuntungan *netto*, besar biaya waktu tanam, besar biaya waktu pemeliharaan

¹ <http://www.r-project.org>

dan besar biaya waktu panen. Sedangkan faktor eksogen adalah pengaruh cuaca dan iklim terhadap bunga, buah jadi, ukuran buah, gagal buah, pola waktu tanam, waktu pemeliharaan, waktu panen, perkiraan umur tumbuhan buah, jumlah tenaga kerja, jenis tumbuhan buah, umur petani dan luas lahan.

4.2 Pengujian Data Buah Kabupaten Aceh Tengah

Data yang pertama diuji adalah data buah kabupaten Aceh Tengah, terdapat 19 fitur untuk setiap data yang diuji dan menggunakan *threshold* yang dimulai dari 0 sampai dengan 1. Gambar 1 menyajikan hasil analisa algoritma *Relief* secara keseluruhan dari semua *threshold* yang digunakan. Tampak bahwa nilai bobot tertinggi dari data buah kabupaten Aceh Tengah adalah pada persekitaran 0,4, yang artinya hanya sampai *threshold* 0,4 saja fitur optimal akan diperoleh. Sedangkan untuk *threshold* $\geq 0,5$ tidak terdapat fitur optimal karena tidak ada nilai bobot yang melebihi, dan fitur yang memiliki nilai bobot tertinggi adalah fitur 4 (besar biaya waktu pemeliharaan).



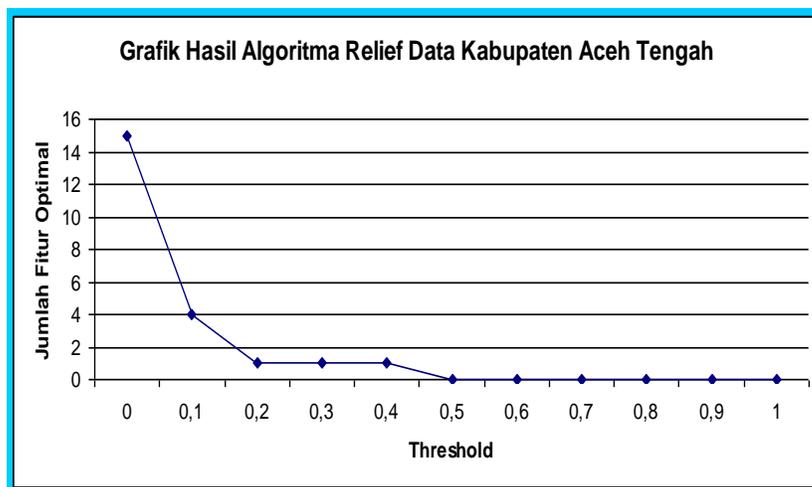
Gambar 1. Hasil Fitur Kab. Aceh Tengah.

Dari hasil analisa algoritma *Relief* pada Tabel 3 dapat dilihat bahwa untuk *threshold* 0, ada 15 fitur optimal yang diperoleh dan fitur yang memiliki nilai bobot paling tinggi adalah fitur 4 atau besar biaya waktu pemeliharaan. Untuk *threshold* 0,1 ada 5 fitur yang diperoleh, namun yang menjadi fitur optimal hanya 4 fitur yaitu fitur 4, 8, 14 dan fitur 6. Sedangkan fitur 5 karena nilai bobotnya lebih kecil dari *threshold* yang digunakan ($0,0981132 < 0,1$) maka tidak termasuk dalam kategori fitur optimal. Dan untuk *threshold* 0,2, 0,3 dan 0,4 hanya menghasilkan satu fitur optimal yaitu fitur 4. Sehingga dapat dikatakan bahwa fitur 4 (besar biaya waktu pemeliharaan) merupakan fitur terbaik (optimal) untuk data buah Kabupaten Aceh Tengah. Hal ini berarti besar biaya waktu pemeliharaan sangat mempengaruhi hasil buah yang ditanam sehingga akan berpengaruh pula pada nilai penjualan atau pendapatan yang akan diperoleh petani.

Gambar 2 menunjukkan bahwa tingginya penurunan jumlah fitur optimal yang terjadi antara *threshold* 0 dengan 0,1, berawal berjumlah 15 fitur optimal menjadi 4 fitur optimal. Untuk *threshold* 0,1 ke 0,2, berkurang 3 fitur optimal (dari 4 menjadi 1 fitur optimal). Sedangkan antara *threshold* 0,2, 0,3 dan 0,4 tidak terjadi penurunan stasioner. Namun untuk *threshold* 0,5 sampai 1 tidak ada fitur optimal yang dihasilkan.

Tabel 3. Hasil Analisa Data untuk Setiap *Threshold*.

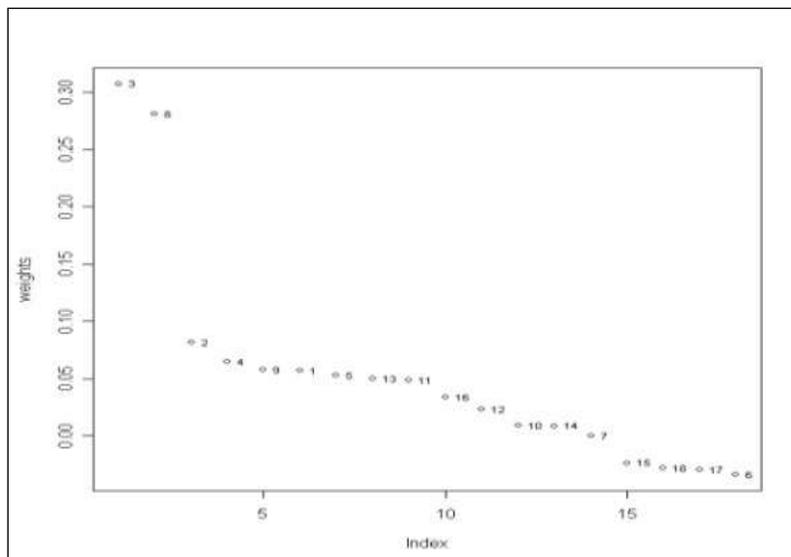
<i>Threshold</i> = 0				<i>Threshold</i> = 0,1			
No	Feature	Frequency	Weight	No	Feature	Frequency	Weight
1	4	10	0,4264151	1	4	10	0,4264151
2	8	10	0,2000000	2	8	10	0,2000000
3	14	10	0,1528302	3	14	9	0,1528302
4	6	10	0,1094340	4	6	6	0,1094340
5	5	9	0,0981132	5	5	7	0,0981132
6	17	9	0,0886793	<i>Threshold</i> = 0,2			
7	1	10	0,0779481	1	4	10	0,4264151
8	3	6	0,0433962	<i>Threshold</i> = 0,3			
9	18	9	0,0415094	1	4	9	0,4264151
10	9	10	0,0386793	<i>Threshold</i> = 0,4			
11	2	10	0,0365566	1	4	7	0,4264151
12	13	10	0,0341719				
13	15	6	0,0245283				
14	10	10	0,0184486				
15	7	5	0,0169811				

Gambar 3. Hasil Data untuk Setiap *Threshold*.

4.3 Pengujian Data Buah Kabupaten Bener Meriah

Sama halnya dengan pengujian pada data di Aceh Tengah, dengan 19 fitur dan menggunakan 11 buah *threshold* yang dimulai dari 0 sampai dengan 1. Gambar 3 menampilkan hasil analisa algoritma *Relief* untuk semua *threshold* yang digunakan. Tampak bahwa nilai bobot tertinggi dari data buah di Bener Meriah berada pada persekitaran 0,3, artinya fitur optimal akan diperoleh hanya sampai nilai *threshold* 0,3. Untuk *threshold* \geq 0,4 tidak terdapat fitur optimal

karena tidak ada nilai bobot yang melebihinya. Fitur yang memiliki nilai bobot tertinggi fitur 3 adalah besar biaya waktu tanam.

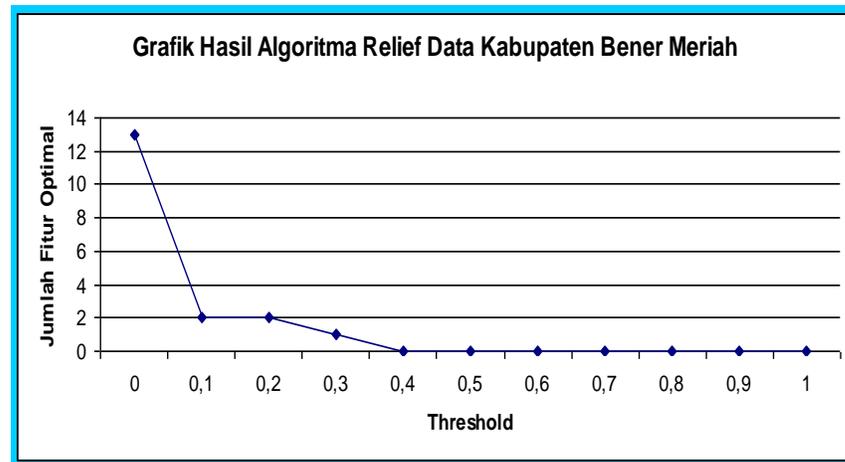


Gambar 3. Hasil Fitur di Kab. Bener Meriah.

Dari hasil analisa algoritma *Relief* pada Tabel 4 dapat dilihat bahwa untuk *threshold* 0, ada 13 fitur optimal yang diperoleh dan fitur yang memiliki nilai bobot tertinggi adalah fitur 3 (besar biaya waktu tanam). Untuk *threshold* 0,1 dan 0,2 ada dua fitur yang diperoleh yaitu fitur 3 (besar biaya waktu tanam) dan fitur 8 (waktu panen). Nilai bobot fitur dari kedua *threshold* ini adalah sama, yang berbeda terletak pada nilai frekuensinya. Untuk *threshold* 0,3 hanya menghasilkan satu fitur optimal yaitu fitur 3 (besar biaya waktu tanam). Sehingga dapat dikatakan bahwa fitur 3 merupakan fitur terbaik untuk data buah di Bener Meriah. Hal ini berarti besar biaya waktu tanam sangat mempengaruhi hasil buah yang ditanam sehingga akan berpengaruh pula pada nilai penjualan atau pendapatan yang akan diperoleh petani buah.

Tabel 4. Hasil analisa data untuk setiap *threshold*

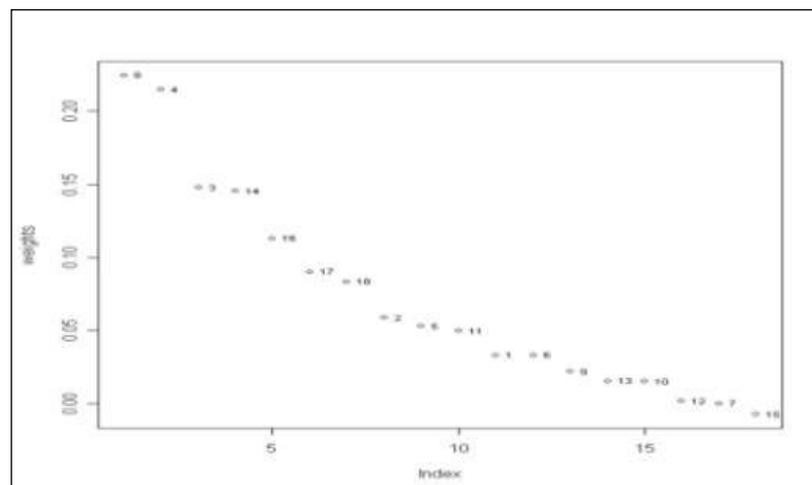
<i>Threshold = 0</i>				<i>Threshold = 0,1</i>			
No	Feature	Frequency	Weight	No	Feature	Frequency	Weight
1	3	10	0,3073529	1	3	10	0,3073529
2	8	10	0,2808824	2	8	10	0,2808824
3	2	10	0,0816176	<i>Threshold = 0,2</i>			
4	4	9	0,0647059	1	3	9	0,3073529
5	9	10	0,0577206	2	8	9	0,2808824
6	1	10	0,0572610	<i>Threshold = 0,3</i>			
7	5	8	0,0529412	1	3	5	0,3073529
8	13	10	0,0498366				
9	11	9	0,0485294				
10	16	9	0,0338235				
11	12	7	0,0231618				
12	10	9	0,0093137				
13	14	5	0,0088235				

Gambar 4. Hasil Analisa Data Setiap *Threshold*.

Gambar 4 menunjukkan bahwa tingginya penurunan jumlah fitur optimal yang terjadi antara *threshold* 0 dengan 0,1, yang awalnya berjumlah 13 fitur optimal menjadi 2 fitur optimal. Untuk *threshold* 0,1 ke 0,2, tidak terjadi penurunan (stasioner). Sedangkan untuk *threshold* 0,2 ke 0,3 dan 0,3 ke 0,4 jumlah fitur optimal berkurang 1, dan untuk *threshold* 0,4 sampai 1 tidak ada fitur optimal yang dihasilkan.

4.4 Pengujian Data Gabungan di Kedua Kabupaten

Data ketiga yang diuji adalah data buah gabungan dari kedua kabupaten. Dengan menggunakan *threshold* yang juga dimulai dari 0 sampai 1 dan dengan 19 fitur diuji. Gambar 5 menyajikan hasil analisa algoritma *Relief* secara keseluruhan dari semua *threshold* yang digunakan. Tampak bahwa nilai bobot tertinggi dari data buah gabungan dari kedua kabupaten pada persekitaran 0,2, yang artinya hanya sampai *threshold* 0,2 saja fitur optimal akan diperoleh. Sedangkan untuk *threshold* $\geq 0,3$ tidak terdapat fitur optimal karena tidak ada nilai bobot yang lebihnya dan fitur yang memiliki nilai bobot tertinggi adalah fitur 8 (waktu panen).



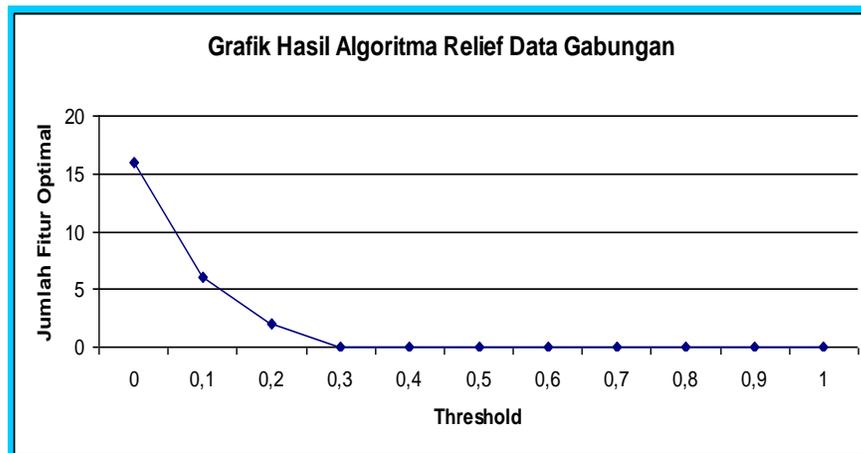
Gambar 6. Hasil Fitur Data Gabungan Kedua Kabupaten.

Dari hasil analisa algoritma *Relief* pada Tabel 5 dapat dilihat bahwa untuk *threshold* 0, ada 16 fitur optimal yang diperoleh dan fitur yang memiliki nilai bobot paling tinggi adalah fitur 8 (waktu panen). Untuk *threshold* 0,1 ada 6 fitur yang diperoleh namun yang menjadi fitur optimal hanya 5 fitur yaitu fitur 8, 4, 3, 14 dan fitur 16. Sedangkan fitur 17 (pengaruh cuaca dan iklim terhadap buah) karena nilai bobotnya lebih kecil dari *threshold* yang digunakan ($0,0900826 < 0,1$) maka tidak termasuk dalam kategori fitur optimal, dan fitur yang memiliki nilai bobot paling tinggi adalah fitur 8 (waktu panen), seperti halnya pada *threshold* 0. Selanjutnya untuk *threshold* 0,3 hanya menghasilkan dua fitur optimal yaitu fitur 8 (waktu panen) dan fitur 4 (besar biaya waktu pemeliharaan). Sehingga dapat dikatakan bahwa fitur 8 merupakan fitur terbaik atau optimal untuk data gabungan ini. Hal ini berarti waktu panen sangat mempengaruhi hasil buah yang dihasilkan sehingga akan berpengaruh pula pada nilai penjualan atau pendapatan yang akan diperoleh petani.

Tabel 5. Hasil Analisa Data untuk Setiap *Threshold*.

<i>Threshold</i> = 0				<i>Threshold</i> = 0,1			
No	Feature	Frequency	Weight	No	Feature	Frequency	Weight
1	8	10	0,2247934	1	8	10	0,2247934
2	4	10	0,2148760	2	4	10	0,2148760
3	3	10	0,1479339	3	3	9	0,1479339
4	14	10	0,1454546	4	14	8	0,1454546
5	16	10	0,1132231	5	16	5	0,1132231
6	17	10	0,0900826	6	17	5	0,0900826
7	18	10	0,0834711	<i>Threshold</i> = 0,2			
8	2	10	0,0586777	1	8	6	0,2247934
9	5	8	0,0528926	2	4	5	0,2148760
10	11	9	0,0495868				
11	1	10	0,0331095				
12	6	7	0,0330579				
13	9	10	0,0219008				
14	13	10	0,0153352				
15	10	10	0,0150597				
16	12	7	0,0018595				

Gambar 7 menunjukkan bahwa tingginya penurunan jumlah fitur optimal yang terjadi antara *threshold* 0 dengan 0,1, berawal berjumlah 16 fitur optimal menjadi 5 fitur optimal. Untuk *threshold* 0,1 ke 0,2, berkurang 3 fitur optimal (dari 5 fitur optimal menjadi 2 fitur optimal). Sedangkan untuk *threshold* 0,2 ke 0,3 jumlah fitur optimal berkurang 1, dan untuk *threshold* 0,3 sampai 1 tidak ada fitur optimal yang dihasilkan

Gambar 7. Hasil Analisa Data untuk Setiap *Threshold*.

4.5 Perbandingan Jumlah Fitur Optimal untuk Setiap *Threshold*

Dari ketiga data yang diuji dilakukan perbandingan terhadap jumlah fitur optimal yang dihasilkan untuk setiap *threshold*-nya, seperti yang diberikan pada Tabel 6. Untuk *threshold* 0 dan 0,1, yang memperoleh fitur optimal terbanyak adalah data gabungan sedangkan untuk *threshold* 0,2 fitur terbanyak diperoleh dari data kabupaten Bener Meriah dan gabungan. Namun untuk *threshold* 0,3 dan 0,4, data gabungan tidak lagi memperoleh fitur optimal.

Tabel 6. Perbandingan Jumlah Fitur Optimal untuk Setiap *Threshold*.

<i>Threshold</i> = 0			<i>Threshold</i> = 0.1		
Aceh Tengah	Bener Meriah	Gabungan	Aceh Tengah	Bener Meriah	Gabungan
1	1	1	4	3	3
2	2	2	6	8	4
3	3	3	8	-	8
4	4	4	14	-	14
5	5	5	-	-	16
6	8	6	<i>Threshold</i> = 0.2		
7	9	8	4	3	4
8	10	9	-	8	8
9	11	10	<i>Threshold</i> = 0.3		
10	12	11	4	3	-
13	13	12	<i>Threshold</i> = 0.4		
14	14	13	4	-	-
15	16	14			
17	-	16			
18	-	17			
-	-	18			

Sementara itu dari hasil uji coba terhadap tiga jenis tumbuhan buah tersebut, setelah melalui pemilihan peubah penjelas berpengaruh dengan uji independensi dan uji parameter model

maka diperoleh tiga peubah yang mempengaruhi produktivitas tumbuhan buah (nilai penjualan) yaitu: keuntungan bruto per musim (X_2), besar biaya waktu pemeliharaan (X_4) dan pengaruh iklim dan cuaca terhadap buah jadi (X_{17}). Faktor-faktor berpengaruh tersebut merupakan faktor yang diutamakan agar mencapai produktivitas maksimal. Model regresi logistik yang diperoleh setelah pencocokan model dengan peubah berpengaruh adalah

$$\ln \frac{(p)}{(1-p)} = -3.228 + 5.628 X_2 + 2.293 X_4 - 2.033 X_{17}$$

Melalui hasil penelitian ini diketahui bahwa produktivitas tumbuhan buah di kedua kabupaten terlihat adanya hubungan antara algoritma *relief* dengan model regresi logistik terutama pada fitur besar biaya pada waktu pemeliharaan. Sedangkan keuntungan bruto per musim dan pengaruh iklim dan cuaca terhadap buah jadi hanya terlihat pada *threshold* 0. Walaupun demikian faktor-faktor yang berpengaruh dari masing-masing metode tersebut menjadi perhatian penting dalam meningkatkan produktivitas tumbuhan buah di masa mendatang.

5. Kesimpulan

Dalam produktivitas tumbuhan buah di masing-masing kabupaten memiliki kesamaan dan perbedaan faktor-faktor berpengaruh sebagai karakteristik faktor. Hal ini terlihat pada *threshold* yang digunakan dan bobot yang dihasilkan. Fitur optimal untuk data buah di Aceh Tengah adalah fitur 4 (besar biaya waktu pemeliharaan) dengan nilai *threshold* 0,4 dan bobot 0,4264151. Fitur optimal untuk data buah di Bener Meriah adalah fitur 3 (besar biaya waktu tanam) dengan nilai *threshold* 0,3 dan bobot sebesar 0,3073529. Fitur optimal untuk data buah gabungan dari kedua kabupaten adalah fitur 8 (waktu panen) dengan *threshold* 0,2 dan nilai bobot 0,2247934.

Untuk *threshold* 0 ada 10 fitur optimal yang sama dari ketiga data yang diuji yaitu: Fitur 1 (keuntungan *netto* per musim); Fitur 2 (keuntungan *bruto* per musim); Fitur 3 (besar biaya waktu tanam); Fitur 4 (besar biaya waktu pemeliharaan); Fitur 5 (besar biaya waktu panen); Fitur 8 (waktu panen); Fitur 9 (jenis tumbuhan buah); Fitur 10 (perkiraan umur tumbuhan buah); Fitur 13 (luas lahan); Fitur 14 (pengaruh cuaca dan iklim terhadap bunga).

Untuk *threshold* 0,1 dan 0,2 hanya fitur 8 (waktu panen) yang merupakan fitur optimal yang sama dari ketiga data yang diuji. Jumlah data dan *threshold* memiliki hubungan linier yaitu semakin besar jumlah data yang diuji, maka akan semakin sedikit fitur optimal yang diperoleh begitu pula dengan *threshold*, semakin tinggi nilai *threshold* yang digunakan maka akan semakin sedikit pula fitur optimal yang akan diperoleh, hal ini dikarenakan *threshold* merupakan nilai batas penentuan (ambang batas) suatu fitur dikatakan optimal atau tidak. Oleh karena itu pengembangan lebih lanjut dampak faktor berpengaruh terhadap produktivitas tumbuhan buah agar lebih dicermati dengan analisa *feature* yang lain.

6. Ucapan Terima Kasih

Peneliti menyampaikan terima kasih sebesar-besarnya kepada Martina N, S.Si, Helmi, S.Si dan rekan-rekan di Dinas Pertanian dan Holtikultura di Kabupaten Aceh Tengah dan Bener Meriah serta DP2M Dikti yang telah mendanai penelitian ini melalui Hibah Penelitian Rusnas tahun 2009.

Daftar Pustaka

- [1] Badan Pusat Statistik Aceh, 2009. *Geografis*. Sumber: <http://aceh.bps.go.id/>, [diakses 16 Maret 2010].
- [2] Boedijoewono N., 1987. *Pengantar Statistika Ekonomi dan Bisnis*. Unit Penerbit dan Percetakan STIM YKPN, Yogyakarta.
- [3] Dash M. dan Liu H., 1997. Feature Selection for Classification. *Intelligent Data Analysis*,1(1-4) : 131-156.
- [4] Guyon I. dan Elisseeff A., 2003. An Introduction to Variable and Feature Selection. *Journal of Machine Learning Research* 3, 1157-1182.
- [5] Han J. dan Micheline K., 2006. *Data Mining: Concepts and Technique*. Morgan Kaufmann Publishers, San Francisco.
- [6] Kira K. dan Rendell L.A., 1992a. A Practical Approach to Feature Selection. In *Proceedings of the Ninth International Workshop on Machine Learning*, Morgan Kaufmann Publishers Inc., 249-256.
- [7] Kira K. dan Rendell L.A., 1992b. The Feature Selection Problem: Traditional Methods and A New Algorithm. In *Proceeding of Tenth National Conference on Artificial Intelligence*, MIT Press, 129-134.