

## MEMPREDIKSI HARGA JUAL RUMPUT LAUT KERING PADA TINGKAT PETANI DENGAN DATA MINING

W. Latuny

e-mail: [wlatuny@gmail.com](mailto:wlatuny@gmail.com)

Universitas Pattimura

### ABSTRAK

*Euचेuma cottonii* sebagai salah satu jenis rumput laut kering cottonii (*Dried Euचेuma Seaweed* (DES)) adalah salah satu komoditas perikanan budidaya utama di Indonesia. Petani lokal menanam, memanen, dan mengeringkan rumput laut, dan kemudian menjualnya ke pedagang. Harga jual rumput laut tergantung pada faktor internal dan eksternal. Untuk memaksimalkan keuntungan mereka, para petani harus memperkirakan perkembangan harga di masa mendatang dengan menggunakan faktor-faktor ini. Penelitian ini menyajikan metode berbasis data baru untuk memperkirakan harga DES di masa depan untuk membantu petani dalam membuat prediksi tersebut. Dalam metode kami, kami menerapkan *data mining* untuk tugas memprediksi harga jual rumput laut pada saat penjualan, delapan minggu ke depan. Algoritma *data mining*, yaitu, penggolong, membutuhkan atribut sebagai inputnya. Dalam percobaan kami, kami mengidentifikasi tiga faktor internal dan tiga faktor eksternal sebagai atribut masukan. Atribut internal faktor yang digunakan adalah: harga DES terkini, kadar kebersihan dari DES, dan kadar air dari DES. Tiga eksternal-faktor atribut semua berhubungan dengan cuaca dan suhu minimum, suhu maksimum, dan curah hujan yang. Semua atribut yang diukur setiap hari di hari d. Output dari classifier adalah prediksi, klasifikasi biner yang menunjukkan apakah rumput laut harga jual pada waktu d ditambah delapan minggu lebih besar atau lebih kecil dari harga pada saat itu d.

**Kata kunci:** Rumput laut kering, harga jual, *data mining*

### PENDAHULUAN

Prediksi harga jual untuk produk yang mudah rusak merupakan salah satu topik aktif penelitian (Li, Xu, & Li, 2010). Membuat prediksi jangka panjang yang akurat untuk harga pasar agro-produk merupakan tantangan yang cukup besar karena faktor *hard-to-predict* mempengaruhi (Li, et.al, 2010). Di Indonesia, akuakultur telah menjadi sector pertanian yang penting, karena kepentingan relative dari budaya konsumsi masyarakat desa (Xiaoshuan, Tao, Revell, & Zetian, 2005). Saat ini, budidaya merupakan alternative sumber pendapatan utama bagi banyak petani di Indonesia.

Untuk memaksimalkan keuntungan mereka, petani harus memperkirakan harga jual rumput laut beberapa minggu ke depan, karena waktu produksi rumput laut yang diperlukan. Perkiraan akurat dari harga jual masa depan, memungkinkan petani untuk membuat keputusan yang berkualitas tentang investasi dalam proses produksi mereka. Terkadang, perkiraan semacam itu sangat sulit dilakukan. Sebagai contoh: selama periode hujan, petani mengalami kesulitan untuk memperkirakan harga jual rumput laut di masa depan. Untuk rumput laut, kandungan air yang lebih rendah dan kemurnian konten yang lebih tinggi akan meningkatkan kualitas dan dapat meningkatkan harga jual. Namun, ada lebih banyak factor yang mempengaruhi harga. Pengembangan metode prediksi efektif yang mengandalkan faktor-faktor tersebut untuk

memperkirakan harga jual rumput laut akan sangat bermanfaat bagi para petani akuakultur Indonesia.

Statistik menawarkan berbagai macam metode untuk membuat prediksi (Zhang, Chen, & Wang, 2010). Misalnya, metode regresi linier multivariat telah digunakan untuk prediksi harga produk pertanian dan peternakan (Zhang, et al., 2010). Selama decade terakhir, metode penambangan data (*data mining*) telah menjadi semakin populer karena kekuatan prediktif dan kerangka evaluasi yang terkait (Li, et al., 2010). Berikut diberikan tiga contoh terbaru dari penerapan sukses penambangan data ke domain pertanian. Contoh pertama adalah karena Li, et.al. (2010), yang menggunakan *neural network classifier* pada tugas memprediksi harga pertanian. Hasilnya menunjukkan bahwa akurasi prediksi satu hari, satu minggu, dan satu bulan diperoleh dengan *neural network classifier* mengungguli yang hasil yang diperoleh dengan model *time series* tradisional. Contoh kedua penerapan penambangan data (*data mining*) yang diterapkan pada domain pertanian adalah prediksi curah hujan di India menggunakan *multilayer perceptron network* (Enireddy, Varma, Rao, & Satapati, 2010). Akurasi prediksi 94,3% diperoleh mengarah ke peningkatan kontrol sistem irigasi untuk produk pertanian. Contoh ketiga yaitu pengembangan sistem pendukung keputusan untuk meramalkan harga produk akuatik di China (Xiaoshuan, et.al., 2005). Sistem pendukung memberikan perkiraan

harga untuk ikan air tawar dalam beberapa skala waktu (setiap tahun, tiga bulanan, bulanan, dan mingguan). Sayangnya, para penulis tidak melaporkan akurasi prediksi mereka.

Uraian contoh di atas and kesuksesan *data mining* dalam memprediksi pertanian dan akuakultur memberikan inspirasi untuk menerapkan *data mining* untuk memprediksi harga jual rumput laut.

Dalam rangka untuk melakukan memprediksi, faktor-faktor mempengaruhi harga jual rumput laut perlu diidentifikasi. Ada dua jenis faktor yang mempengaruhi perkembangan harga agro-produk pada umumnya dan rumput laut khususnya: faktor internal dan faktor eksternal. Contoh faktor internal adalah: jumlah konsumsi per kapita produk, kualitas produk, dan harga jual terbaru (Zhang, *et al.*, 2010). Contoh untuk eksternal biaya transport dan kondisi cuaca. Secara khusus untuk rumput laut, faktor internal dan eksternal yang mempengaruhi harga jual telah diidentifikasi. Menurut *Food Agriculture Organization* (FAO) dan (Tiroba, 2006) dua factor internal mempengaruhi harga rumput laut: (1) kelembaban rumput laut ("kadar air") dan (2) kebersihan rumput laut isi ("kandungan kotor") (Trono, 1990). Kandungan air adalah faktor penting karena menentukan harga pasar rumput laut. Seperti disebutkan di atas, kadar air yang lebih rendah mengarah ke harga yang lebih tinggi. Kandungan kelembaban maksimum yang disarankan adalah sekitar 35% setelah penyimpanan (Pelinggon & Tito, 2009). Kandungan kotor juga penting karena intensitas pembersihan dan pengeringan mempengaruhi kualitas rumput laut. Pembersihan dilakukan dengan membuang sedimen dan partikel yang tidak diinginkan lainnya dengan membilas rumput laut beberapa kali saat masih di laut. Selanjutnya pembuangan hewan lainnya yang melekat (kerang, karang lunak, spons) dan bahan mengikat ('*tie-tie*') dilakukan di darat sebelum pengeringan. Selain dua faktor internal utama ini, harga jual terkini juga merupakan faktor internal ketiga yang mempengaruhi harga jual rumput laut (Pelinggon & Tito, 2009). Untuk faktor eksternal, ada tiga yang mempengaruhi harga rumput laut: (1) curah hujan, (2) suhu minimum, dan (3) suhu maksimum. Curah hujan mungkin memiliki efek yang merusak pada budidaya rumput laut dan tahap panen. Sebagai contoh, sentra produksi rumput laut utama yang terletak di provinsi Sulawesi Selatan dan Bali memiliki curah hujan yang cukup besar pada periode dari Juli hingga September. Curah hujan mempengaruhi siklus produksi rumput laut. Hujan yang ekstensif mengurangi tingkat salinitas dan meningkatkan suhu permukaan laut. Keduanya menghambat pertumbuhan rumput laut (Bank Indonesia, 2006) dan dapat menyebabkan penyakit, dikenal sebagai '*ice-ice*', yang mematikan untuk

rumpun laut. Suhu minimum dan maksimum mempengaruhi baik budidaya dan panen rumput laut. Rumput laut *cottonii* ditanam dan dipanen pada suhu minimum yang disukai dari 27 0 C dan suhu maksimum yang disukai dari 32 0 C. Setiap penyimpangan dari kisaran suhu ini dapat mempengaruhi rumput laut secara negatif. Meskipun, wilayah pesisir Indonesia memiliki suhu yang sebagian besar jatuh ke dalam kisaran yang diperlukan (Bank Indonesia, 2006) nilai suhu dapat mempengaruhi prediksi harga jual.

### 1. Tujuan penelitian dan pertanyaan Penelitian

Tiga tujuan penelitian ini adalah: (1) untuk memprediksi apakah harga rumput laut pada tingkat petani meningkat atau menurun pada saat pengiriman; (2) untuk melakukan evaluasi perbandingan teknik *data mining* untuk prediksi, dan (3) untuk mengetahui faktor mana yang penting untuk mempengaruhi harga jual rumput laut menggunakan teknik data mining. Tujuan-tujuan ini mengarah pada perumusan tiga pertanyaan penelitian berikut.

*Pertanyaan penelitian 1: Sejauh mana harga jual rumput laut diprediksi 8 minggu ke depan tepat waktu?*

*Pertanyaan penelitian 2: Metode data mining apa yang cocok untuk memprediksi harga jual rumput laut?*

*Pertanyaan penelitian 3: Faktor apa yang menentukan harga jual rumput laut di masa yang akan ?*

Menjawab pertanyaan penelitian pertama mengharuskan untuk membuat prediksi harga rumput laut setelah jangka waktu yang sesuai dengan waktu produksi. Produksi rumput laut berlangsung dalam lima tahap: ( i ) menanam benih (2 hari), (ii) mengolah rumput laut (40-45 hari), (iii) panen (1-2 hari) , ( iv) pengeringan (3) -7 hari), dan (v) pengemasan dan pengiriman (4 hari). Total waktu produksi rumput laut sesuai dengan 50 hingga 60 hari, yang sesuai dengan beberapa laporan (FAO, 1988; Bank Sentral Indonesia, 2006; Prasodjo , 2008). Oleh karena itu, dilakukan prediksi harga rumput laut 50-60 hari (sekitar 8 minggu) ke depan dalam waktu.

Untuk menjawab pertanyaan penelitian kedua dengan memeriksa teknik-teknik *data mining* utama yang digunakan dan membandingkan kinerja prediksi teknik *data mining* tersebut. Kami akan memilih empat pengklasifikasi (*classifier*) berbeda yang sering digunakan dan terbukti berhasil dalam berbagai macam tugas. Selanjutnya, kami menentukan kinerja prediksi dari empat pengklasifikasi (*classifier*) dan menilai bagaimana setiap *classifier* dibandingkan yang berhubungan dengan akurasi prediksinya.

Pertanyaan penelitian ketiga ditujukan dengan menentukan faktor-faktor yang mempengaruhi harga jual rumput laut. Keenam atribut (yaitu, 3 faktor internal dan 3 faktor eksternal) diharapkan berkontribusi secara berbeda terhadap prediksi harga jual di masa mendatang. Kami akan menganalisa *classifier* yang terlatih untuk menentukan faktor mana yang merupakan kontributor utama untuk prediksi yang sukses.

## 2. Penelitian-penelitian sebelumnya yang berhubungan

Sepanjang diketahui, ini adalah salah satu studi pertama tentang aplikasi *data mining* pada domain akuakultur. Satu-satunya penelitian lain yang ditemukan adalah (Hu, *et al.*, 2005) yang menerapkan *data mining* untuk meramalkan harga jual jangka pendek produk akuatik. Penggunaan data mining di bidang pertanian bukanlah hal baru. Banyak penelitian telah dilakukan untuk memprediksi hasil produk pertanian (Ruß, Kruse, Schneider, Wagner, & Perner, 2008), cuaca (Baskar, Arockiam, Kumar, & Jeyasimman, 2010), atau indikator ekonomi (Li, *et al.*, 2010).

Dalam bagian ini ditinjau tiga studi yang relevan dengan penelitian ini. Selanjutnya, akan dijelaskan temuan tentang studi ini berikut ini. Dalam studi pertama, *data mining* telah diterapkan untuk memprediksi hasil gandum (Ruß, *et al.*, 2008). Para peneliti menggunakan 5241 catatan data yang menyebar ke 8 atribut. Untuk prediksi, penulis menggunakan *Artificial Neural Networks* (ANN). Sayangnya, hasilnya tidak terlalu bagus. Oleh karena itu, mereka menyarankan untuk menggunakan pengklasifikasi yang lebih kuat, yaitu, *Support Vector Machines* (SVMs) untuk memprediksi hasil gandum.

Dalam studi kedua, harga tomat di Cina diperkirakan (Li, *et al.*, 2010). Para penulis membandingkan *Artificial Neural Networks* (ANNs) dan *Auto-Regression Moving Average Analysis* (ARIMA). Perbandingan menunjukkan bahwa ANNs melakukan lebih baik daripada ARIMA dengan mencapai akurasi 80-90%. Meskipun hasil ini baik, penulis menyarankan untuk mengeksplorasi pengklasifikasi lain untuk lebih mengoptimalkan akurasi prediksi.

Studi ketiga adalah survei singkat tentang penerapan teknik penambangan data dalam domain pertanian (Baskar, *et al.*, 2010). Para penulis menemukan tiga pengklasifikasi, yaitu *Support Vector Machines* (SVMs), *Decision Tree*, dan *k-Nearest-Neighbor* untuk menghasilkan penampilan prediksi terbaik.

Ketiga penelitian ini menunjukkan bahwa teknik penambangan data akan dapat diterapkan dalam memprediksi keputusan masa depan dalam domain pertanian. Mengingat keberhasilan *data mining* untuk

memprediksi, peneliti sekarang mencoba untuk menerapkan dalam domain pertanian khususnya dalam domain akuakultur. Berikut akan dilihat deskripsi penelitian sebelumnya yang terkait menerapkan *data mining* untuk memprediksi produk akuakultur. Salah satu cabang produk pertanian adalah produk akuakultur yang merupakan budidaya organisme air, misalnya, ikan, *molluscs*, *krustasea*, tumbuhan laut, buaya, kura-kura dan amfibi (FAO, 2002). Pertanian menyiratkan beberapa bentuk intervensi dalam proses pemeliharaan untuk meningkatkan produksi seperti persediaan reguler, perlindungan dari predator, dan untuk menyiratkan kepemilikan individu atau perusahaan dari saham yang dibudidayakan (Organisasi Pertanian Pangan (FAO), 2002; Gordon, 2010). Akuakultur saat ini adalah salah satu sistem produksi pangan yang paling cepat berkembang di dunia. Oleh karena itu memprediksi harga akuakultur mungkin menjadi perhatian dalam penelitian pada domain akuakultur.

Penulis hanya dapat menemukan satu studi tentang prediksi harga dalam budidaya: Model hibrida untuk peramalan produk akuatik yang dijelaskan dalam pendahuluan (Hu, *et al.*, 2005). Model ini memprediksi harga jangka pendek produk akuatik dengan akurasi sekitar 70%.

## 3. Relevansi sosial dan ilmiah dari penelitian

Relevansi bagi masyarakat dalam penelitian ini adalah dapat memberikan aturan atau pedoman bagi petani rumput laut di Indonesia (dan mungkin di tempat lain) untuk memprediksi harga jual rumput laut. Prediksi yang tepat dapat meningkatkan kekuatan bisnis mereka yang berkontribusi pada ekonomi lokal. Relevansi ilmiah dari studi ini adalah bahwa dalam mengeksplorasi domain aplikasi baru dari *data mining* akan berkontribusi pada pengembangan metode untuk mengekstraksi pengetahuan tugas spesifik dari *classifier* yang terlatih.

## KAJIAN TEORI DAN METODE

### 1. Kajian Teori

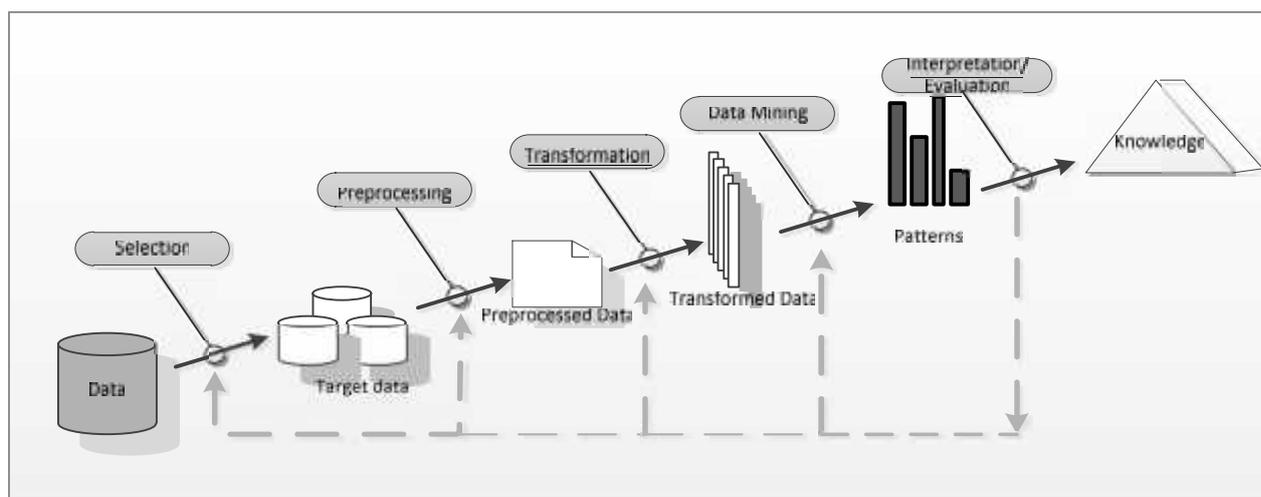
Bab ini menjelaskan kerangka penelitian *data mining* dan bagaimana hal itu dapat diterapkan pada tugas prediksi kami. Kerangka penelitian penambangan data berasal dari *Knowledge Discovery in Databases* (KDD) dan dijelaskan dalam subbagian 2.1. Model proses yang lebih spesifik untuk *data mining* dalam domain pertanian diuraikan dalam sub-bab 2.2. Akhirnya, relevansi model proses untuk tugas kita memprediksi harga rumput laut dibahas pada subbagian 2.3.

#### a. Penemuan pengetahuan dalam basis data

*Data mining* adalah aplikasi dari algoritma khusus untuk mengekstraksi pola dalam

data (Fayyad, Piatetsky -Shapiro, & Smyth, 1996). *Data mining* tidak hanya tentang menerapkan satu set algoritma, tetapi melibatkan prosedur untuk memecahkan masalah atau membuat keputusan (Liu, 2002). Prosedur ini melibatkan pemanfaatan berbagai teknik matematika dan komputer untuk menganalisis data yang relevan disimpan dalam database yang sangat besar, untuk mengklasifikasikan atau mengelompokkan data dengan menggunakan algoritma pengenalan pola, dan untuk menerapkan hasil klasifikasi atau pengelompokan ke masalah yang ditentukan (Liu, 2002). Prosedur ini telah disebut sebagai *Knowledge Discovery Database*

(KDD; (Fayyad, et al., 1996) . KDD melibatkan seluruh prosedur mengeksplorasi dan menemukan pengetahuan yang berharga dari data dimana *data mining* hanya satu langkah. Prosedur KDD terdiri dari lima langkah berikut: (1) pemilihan data, (2) *preprocessing* data yang dipilih, (3) transformasi data yang sudah diolah sebelumnya, (4) menerapkan *data mining* ke data yang diubah untuk mengekstrak dan menemukan pola dan hubungan, dan (5) menafsirkan dan mengevaluasi struktur yang ditemukan (Hand, Mannila, & Smyth, 2001) .



Gambar 2.1. Gambaran dari lima langkah yang menyusun prosedur KDD (direproduksi dari Fayyad.et.al, 1996 )

Pada Gambar 2.1 (direproduksi from Fayyad *et al.*, 1996), lima langkah dari prosedur KDD diilustrasikan. Biasanya, lima langkah dilakukan secara berurutan, tetapi seperti yang digambarkan oleh panah putus-putus pada Gambar 2.1, ketika langkah terakhir selesai, hasil evaluasi atau analisis umumnya menimbulkan peninjauan kembali salah satu langkah sebelumnya. Ini berarti bahwa KDD bukan prosedur satu-shot, tetapi prosedur literasi. Tergantung pada evaluasi atau analisis hasil, pilihan data yang berbeda dapat dipertimbangkan, cara yang berbeda dari *preprocessing* data dapat dilakukan, transformasi yang berbeda dapat digunakan, atau algoritma *data mining* yang berbeda dapat dipanggil.

Prosedur KDD (atau bagian dari itu) telah diterapkan untuk memecahkan berbagai macam masalah, misalnya, dalam bisnis (Seng & Chen, 2010) prediksi harga saham (Yakup, Melek Acar, & mer Kaan, 2011), dan meramalkan makanan yang mudah rusak di toko-toko (FL Chen & Ou, 2008; Philip, Wei, Andreas, & Salvatore, 1999) .

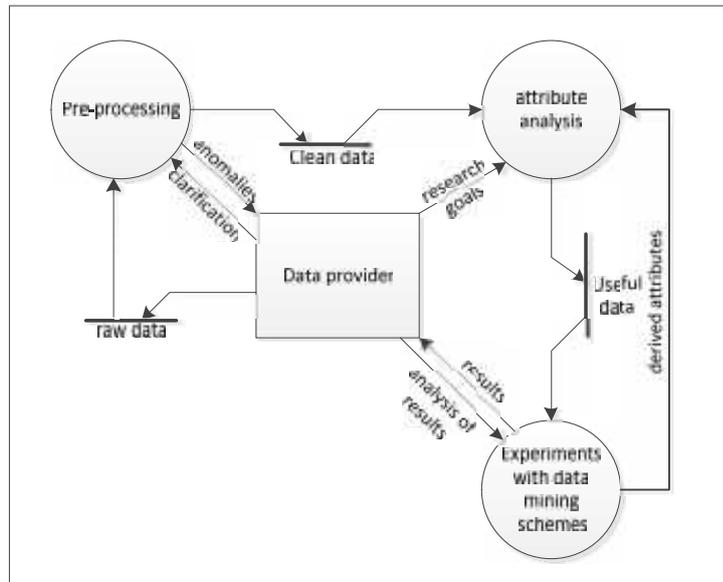
Inti dari prosedur KDD adalah penerapan metode penambangan data, yaitu langkah

keempat. Berikut diuraikan deskripsi langkah ini lebih rinci. Langkah-langkah lain akan dibahas dalam konteks tugas prediksi harga rumput laut. Karena prosedur KDD sangat umum, peneliti mengadopsi model proses yang lebih spesifik dari data mining yang diterapkan ke domain pertanian karena Cunningham dkk (1999).

#### b. Model proses data mining

Model proses *data mining* Cunningham (1999) menetapkan aplikasi penambangan data ke domain pertanian dan memiliki dua fitur penting: Interaksi dua arah antara penyedia data dan pakar data mining. Keduanya mengubah data mentah menjadi set input data set akhir ke algoritma pembelajaran mesin.

Pendekatan berulang. Proses pembelajaran mesin menetapkan data eksperimental dan mengambil beberapa urutan melalui model proses untuk menghasilkan kecocokan yang tepat antara representasi data dan algoritma penambangan data.



Gambar 2.2. Model proses untuk penambangan data (diagram direproduksi dari Cunningham, 1999).

Gambar 2.2 mengilustrasikan model proses di mana tahap-tahap utama diwakili oleh lingkaran. Persegi panjang di tengah mewakili penyedia data, yaitu sumber data. Penyedia data adalah orang atau organisasi yang menyediakan akses ke basis data dan memiliki pengetahuan tentang data dan domain. Tahap pertama dari model proses adalah tahap pra-pemrosesan di mana data mentah diubah menjadi representasi yang sesuai. Tahap kedua proses adalah analisis atribut data. Analisis diperlukan untuk memilih atribut yang tepat untuk menyelesaikan tugas yang ada. Apa yang tepat tergantung pada tujuan dari penyelidikan *data mining* yang sering dirumuskan oleh (atau diekstraksi dari) penyedia data. Tahap ketiga adalah evaluasi eksperimen dari algoritma *data mining* yang diterapkan pada data yang diwakili dalam hal atribut yang dipilih. Hasil eksperimen dinyatakan dalam ukuran akurasi atau ukuran kinerja lain (Seng & Chen, 2010). Pada setiap tahap dari model proses, ada kemungkinan pertukaran dengan penyedia data. Selama *preprocessing*, anomali dapat dideteksi dan dilaporkan ke penyedia data. Juga, *preprocessing* dapat menyebabkan permintaan untuk klarifikasi dari penyedia data. Misalnya, ketika karakteristik data tertentu sulit untuk menafsirkan penyedia data dapat memberikan penjelasan berdasarkan pengetahuan domain. Selama tahap percobaan, hasil (menengah) dapat dikomunikasikan kepada penyedia data, yang, pada gilirannya, dapat menafsirkan atau menganalisis hasil dalam hal domain aplikasi.

### 3. Relevansi model proses penambangan data

Model proses Cunningham (1999) memiliki relevansi dengan tugas kami dalam memprediksi

harga jual rumput laut, karena dua alasan: (1) model mengambil analisis atribut secara serius, dan (2) model menekankan pada penanganan data dan pertukaran informasi dengan penyedia data. Sementara sebagian besar skema penambangan data, termasuk KDD, mengambil atribut secara serius, model Cunningham menekankan kebutuhan dan keuntungan dari analisis atribut untuk membangun hubungannya dengan tugas yang ada. Dalam tesis ini, atribut akan dianalisis dengan melakukan analisis statistik dari data yang sudah diolah sebelumnya dan dengan memeriksa penggolong yang terlatih. Meskipun, dalam konteks penulisan ini, peneliti tidak dapat berkomunikasi secara langsung dengan penyedia data (yaitu, petani rumput laut), akan tetapi menurut peneliti model proses Cunningham sangat relevan untuk studi saat ini dan akan datang.

## 2. Metode

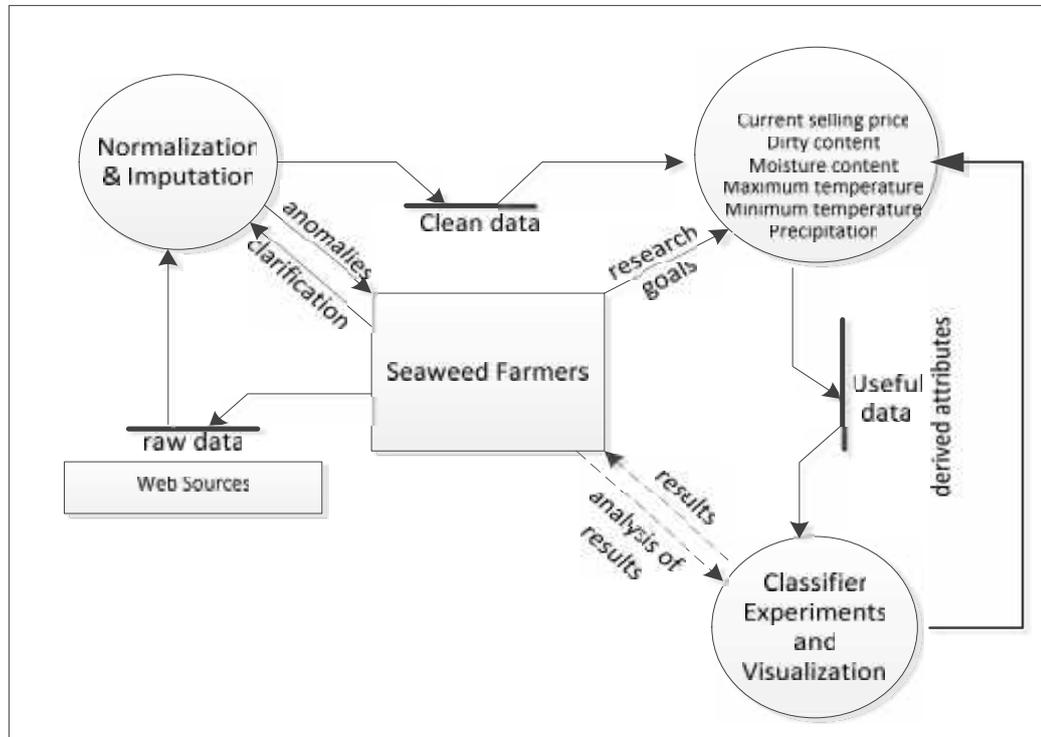
Bab ini mengambil model proses Cunningham (1999) dan merumuskannya menjadi model proses untuk prediksi harga jual rumput laut.

### a. Model proses *data mining* untuk prediksi harga jual rumput laut

Model proses *data mining* prediksi harga jual rumput laut adalah contoh spesifik dari model proses Cunningham yang kami jelaskan pada bab sebelumnya. Model spesifik diilustrasikan pada Gambar 3.1. Setiap tahapan (lingkaran) telah diberi label sesuai dengan langkah-langkah yang diambil dalam prediksi harga jual rumput laut. Tahap pertama (*preprocessing*) diberi label dengan "normalisasi & imputasi" untuk menunjukkan *preprocessing* data, yaitu, normalisasi data untuk membatasi berbagai

kemungkinan nilai dan imputasi untuk menangani nilai-nilai yang hilang dalam data. Tahap kedua ("analisis atribut") diberi label oleh enam atribut yang akan dianalisis dan digunakan untuk tugas klasifikasi. Tahap ketiga ("Eksperimen dengan skema *data mining*") diberi label "Eksperimen Pengklasifikasian dan Visualisasi". Pada tahap ini,

eksperimen penambangan data aktual akan dilakukan dan hasilnya akan divisualisasikan dalam upaya untuk mendapatkan pemahaman yang lebih mendalam tentang data. Percobaan akan dilakukan dengan menggunakan metode klasifikasi standar yang akan dijelaskan dalam subbagian berikutnya.

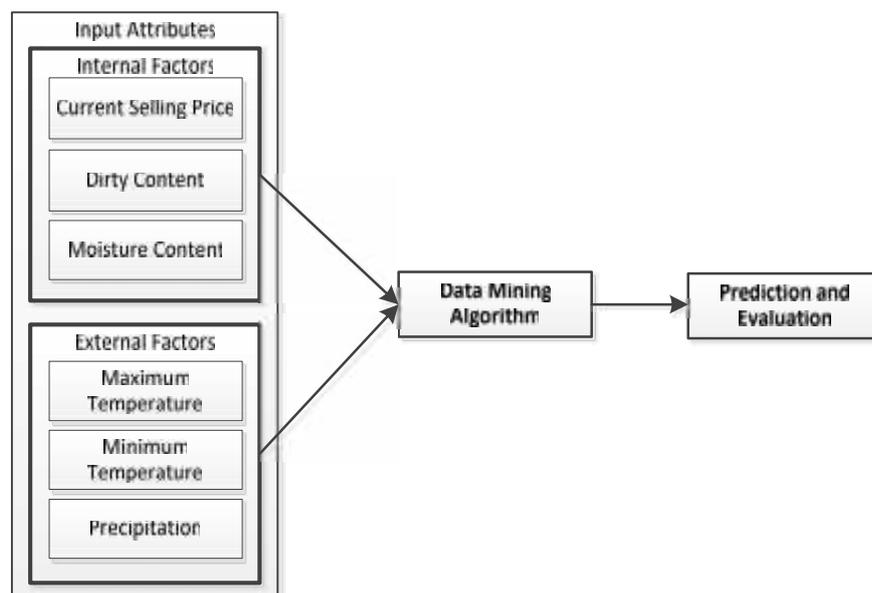


Gambar 3.1 Model proses data harga jual rumput laut

### Metode klasifikasi

Metode ini terdiri dari tiga komponen, diwakili pada Gambar 3.2 oleh tiga kotak pembatas. Kotak di sebelah kiri menunjukkan atribut input ke algoritma penambangan data. Dua jenis atribut diwakili oleh kotak terpisah berlabel "faktor

internal" dan "faktor eksternal". Kotak di tengah mewakili algoritma penambangan data. Dibutuhkan ttribut sebagai input dan menghasilkan prediksi sebagai output. Kotak di sebelah kanan menunjukkan prediksi dan evaluasi (yaitu perbandingan dengan harga aktual).



Gambar 3.2 Ilustrasi metode klasifikasi untuk prediksi harga jual eaweed

Subbagian berikut membahas tiga komponen secara lebih rinci.

#### b. Atribut prediksi

Sebagaimana dijelaskan pada bab 1, atribut yang dipilih sebagai input ke algoritma penambangan data adalah id yang diperoleh dari studi pustaka. Kami secara singkat merangkum enam atribut di bawah ini.

Harga jual rumput laut saat ini (CP) didefinisikan sebagai harga rumput laut yang dijual oleh petani kepada pedagang pada saat itu (Bank Indonesia, 2006).

Rumput laut kotor s konten (DC) didefinisikan sebagai tingkat kotoran dalam rumput laut, misalnya, pasir, kotoran, rafia, ikan, kerang, batu, potongan karang, atau garam berlebihan. Kotoran mempengaruhi harga yang dibayarkan untuk rumput laut. Biasanya, kotoran kurang dari 3% dari isi rumput laut (Tiroba, 2006).

Rumput Laut kadar air s (MC) didefinisikan sebagai tingkat kelembaban di rumput laut. Kandungan air yang lebih rendah memungkinkan untuk mengompres lebih baik dari rumput laut ke dalam bal. Kandungan kelembaban sangat penting. Kandungan air yang direkomendasikan adalah sekitar 35% setelah penyimpanan (Pelinggon & Tito, 2009).

Suhu maksimum (MaxT) didefinisikan sebagai suhu maksimum pada hari tertentu. Suhu maksimum mempengaruhi baik fotosintesis rumput laut dan ing kering rumput laut setelah panen. Untuk rumput laut, suhu tidak boleh melebihi 32 0C. (Bank Indonesia, 2006; Y.-c. Chen, Rogoff, & Rossi, 2010). Suhu minimum (MinT) didefinisikan sebagai suhu minimum pada hari tertentu. Untuk rumput laut, suhu tidak boleh lebih rendah dari 27 0 C. (Bank Indonesia, 2006).

Presipitasi (P) didefinisikan sebagai kepadatan hujan pada hari tertentu. Hujan menurunkan salinitas air laut yang dapat menyebabkan penyakit 'ice-ice' mempengaruhi panen dan pengeringan rumput laut (Bank Indonesia, 2006).

#### c. Algoritma data mining

Algoritma atau pengelompokan data mining mengambil atribut sebagai input dan mengubah nilainya menjadi nilai prediksi. Penentuan algoritma penambangan data yang paling tepat harus sangat ditentukan secara empiris. Menurut "tidak ada teorema makan siang gratis" yang terkenal (Wolpert & Macready, 1997), kinerja algoritme tergantung pada domain aplikasi dan tidak ada satu pun "pengklasifikasi terbaik". Oleh karena itu, kita akan mengkaji empat pengklasifikasi perwakilan yang sering menghasilkan pertunjukan yang berbeda: (1) *Decision Tree classifier*, (2) *naïve Bayes classifier*, (3) *k-Nearest Neighbor classifier*, dan (4) *Support*

*Vector Machine classifier*. Semua *classifier* ini semuanya tersedia di WEKA, perangkat lunak *data mining* yang dapat diakses secara bebas (Witten, dkk, 2011).

#### *Decision tree classifier*

*Decision tree classifier* membagi ruang lingkup yang ditutupi oleh nilai-nilai atribut secara rekursif untuk memaksimalkan kinerja klasifikasi (Hand, et.al., 2001). *Decision tree* terdiri dari hirarki simpul yang membentuk pohon terbalik. Klasifikasi dimulai di bagian atas pohon terbalik, pada simpul akar. Klasifikasi contoh hasil dengan kerja ing bawah pohon dan memilih cabang yang tepat pada setiap node dengan menguji kondisi pada salah satu atribut. Pengujian berlanjut sampai daun yang berhenti dari pohon tercapai. Daun terminal memiliki label kelas yang ditetapkan untuk instance yang mencapai salah satu daun terminal. Varian *Decision tree* berbeda dalam urutan di mana atribut individu diuji. Ukuran informasi-teoretis sering digunakan untuk memilih atribut yang paling informatif terlebih dahulu. Dalam percobaan kami, kami akan menggunakan versi WEKA dari pohon keputusan C4.5 (Quinlan, 1993) yang disebut pohon keputusan J.48.

The *k*-Nearest Neighbor (*k*-NN) classifier *k*-Nearest Neighbor (*k*-NN) classifier menyimpan sejumlah label *instances* (*training set*) mengklasifikasi *instances* baru, dengan memberikan mereka label kelas yang sesuai dengan mayoritas label kelas dari *k* terdekat contoh yang disimpan (Witten, Frank, & Hall, 2011; Wu, et.al., 2007). Di dalam WEKA, the classifier *k*-NN disebut Instance Based (IBk) classifier Pengklasifikasi berbasis misalnya kadang-kadang disebut *lazy classifier*, karena mereka tidak menciptakan model melainkan bergantung pada contoh *training set* (Witten, et al., 2011).

#### The naïve Bayes classifier

*The naïve Bayes classifier* mengestimasi probability kelas, diberikan sebuah *instance*, juga dikenal sebagai probabilitas kondisional. Menggunakan contoh dalam *training set*, pengelompokan naïve Bayes menghitung probabilitas bersyarat dari kombinasi nilai-nilai kelas. Dengan menggunakan asumsi "naif" tentang independensi atribut, ia menerapkan aturan Bayes untuk menghitung probabilitas kelas yang diberikan contoh baru. Aturan keputusan digunakan untuk mengubah output probabilistik menjadi klasifikasi diskrit. Meskipun atribut umumnya tidak independen (misalnya, suhu minimum dan maksimum tentu tergantung), pengelompokan naïve Bayes melakukan cukup baik pada berbagai macam tugas (Witten, dkk, 2011).

The Support Vector Machine classifier Support Vector Machine (SVM) mengklasifikasikan kasus dengan mengubah mereka ke dalam ruang representasi nonlinear. Transformasi seperti ini sangat berguna jika kelas *instances* dari kelas berbeda tidak terpisah secara linier, yaitu, jika mereka tidak dapat dipisahkan oleh garis atau bidang. Asalkan transformasi yang sesuai dipilih, masalah klasifikasi tersebut dapat menjadi terpisah secara linier dalam ruang yang diubah (Witten, dkk., 2011). Transformasi ke ruang nonlinier dilakukan dengan menggunakan fungsi-fungsi kernel. Kernel fungsi dasar radial memungkinkan transformasi yang efektif. Dalam WEKA, SVM dihitung menggunakan Sequential Minimal Optimization (SMO).

### 5. Prediksi dan evaluasi

Komponen ketiga dari metode penimbangan data kami menyangkut prediksi dan komponen evaluasi. Sebagaimana dinyatakan dalam Bab 1, prediktif didefinisikan sebagai harga jual rumput laut delapan minggu sebelumnya. Evaluasi prediksi dilakukan dengan membandingkan harga yang diprediksi pada waktu  $t$  dengan harga aktual pada waktu  $t$  ditambah delapan minggu. Untuk memperkirakan bagaimana penggolong kami melakukan pada data yang tidak terlihat, kami akan menggunakan prosedur cross-validasi standar (Witten, et.al., 2011). Selain itu, kami akan melakukan analisis terhadap beberapa *classifier* yang

terlatih untuk menentukan kontribusi relatif dari atribut terhadap prediksi.

## HASIL DAN PEMBAHASAN

Peneliti mulai dengan memvisualisasikan kumpulan data dalam hal atribut ( 5 .1). Selanjutnya, Peneliti menyajikan hasil klasifikasi untuk empat *classifier* (5, 2). Akhirnya, Peneliti menganalisis model yang diinduksi untuk mengidentifikasi atribut yang paling informatif untuk prediksi (5, 3).

### 1. Hasil Klasifikasi

Hasil klasifikasi yang diperoleh untuk empat pengklasifikasi disajikan pada Tabel 5.1 . Untuk setiap penggolong, persentase dengan benar diklasifikasikan contoh terdaftar. Selain itu, matriks kebingungan diberikan. Classifier No. 0 adalah penggolong referensi WEKA yang disebut zeroR . Ini memberikan kinerja dasar yang sebenarnya dengan memprediksi kelas mayoritas dalam data pelatihan. The  $k$  -Nearest Neighbour Classifier ( $k = 27$ ) memberikan kinerja terbaik dari empat classifier dievaluasi, meskipun ini mungkin mencerminkan lebih pas karena optimasi parameter. Penggolong mencapai skor 186 dari 275 instance yang diklasifikasikan dengan benar instance  $s$  (Akurasi prediksi 67 % ). Tiga pengklasifikasi lainnya mencapai kinerja yang sedikit lebih buruk. Keputusan J48 mencapai kinerja prediksi 64%, dan Naïve Bayes dan SVM keduanya memprediksi harga dengan akurasi 65%.

**Table 5.1 Overview of the classification results**

No	Classifier technique	Percentage number of correctly classified		Confusion matrix		
0	ZeroR	52%	144	a=0 b=1	0 0	131 144
1	Naïve Bayes	65%	180	a=0 b=1	75 39	56 105
2	Ibk k=27	67%	186	a=0 b=1	85 43	46 101
3	J48 (pruning 0.1-0.5)	64%	177	a=0 b=1	76 43	55 101
4	SVMs (C 0.01-0.1)	65%	180	a=0 b=1	67 31	64 113

Hasil ini menunjukkan bahwa, berdasarkan kumpulan data Peneliti, pengklasifikasi standar dapat memprediksi arah harga di masa mendatang (naik

atau turun) dengan akurasi setidaknya 64%. Semua empat pertunjukan secara signifikan berbeda dari kesempatan yang ditentukan oleh uji Binomial.

## 2. Analisis hasil

Hasil Peneliti jelas menunjukkan bahwa kinerja prediksi minimal 65% dapat dicapai tanpa optimasi parameter apa pun. Sensitivitas parameter dari instance-based ( $k$ -NN) dan penggolongan Support Vector Machine diilustrasikan dalam Gambar 5.1. Grafik yang ditampilkan di sebelah kiri Gambar 5.1 menunjukkan kinerja prediksi (persentase contoh dengan benar diklasifikasikan) sebagai fungsi dari parameter kompleksitas  $k$ . Untuk nilai kecil  $k$  ( $k < 15$ ), kinerjanya bervariasi dari sekitar 57% ( $k = 2$ ) hingga sekitar 66% ( $k = 5$ ). Untuk nilai yang lebih besar dari kompleks untuk data.

$k$  ( $k > 17$ ), kinerjanya bervariasi sekitar 64% dan mencapai maksimumnya 67% pada  $k = 27$ . Untuk nilai-nilai besar seperti  $k$ , penggolong mengambil suara mayoritas lebih dari 27 label kelas sehingga merapikan kesalahan karena kebisingan. Ini mungkin mencerminkan sifat berisik data Peneliti. Grafik di sebelah kanan Gambar 5.1 menunjukkan kinerja SVM sebagai fungsi dari parameter kompleksitas  $C$ . Berbeda dengan IBK, dalam hal ini meningkatkan kompleksitas mengarah pada peningkatan kecil dalam kinerja. Rupanya, berbeda dengan classifier IBK, classifier SVM berhasil pas model yang

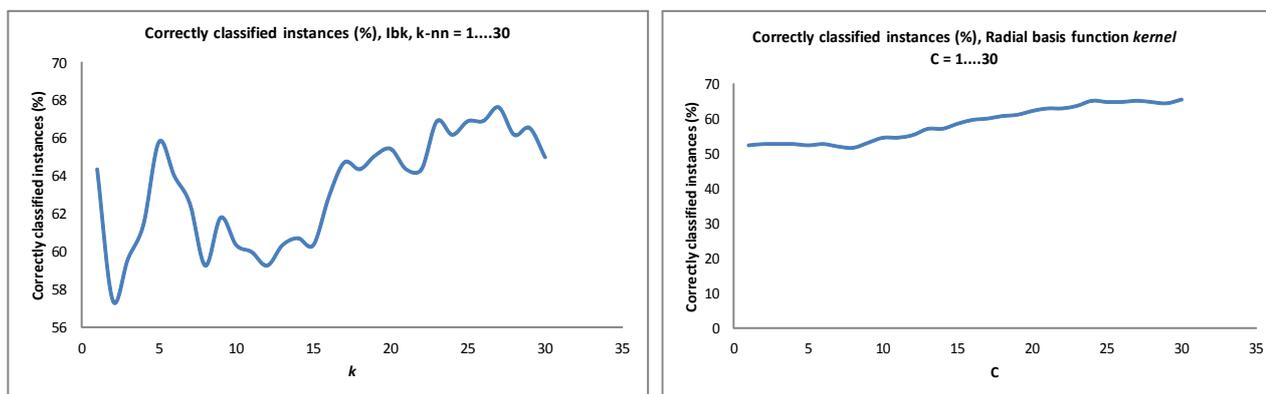


Figure 5.1 Percentage of correctly classified instances for (left) IBK, with  $k$  ranging from 1 to 30 and (right) SVMs, with the complexity  $C$  varying from 1 to 30.

## 3. Visualisasi kumpulan data

Hasil klasifikasi Peneliti menunjukkan bahwa beberapa tingkat akurasi prediksi dimungkinkan dengan menerapkan penambahan data ke basis data faktor internal dan eksternal. Peneliti sekarang beralih ke analisis data dan atribut dalam upaya untuk menilai kekuatan prediktif dari atribut individu. Secara khusus, tujuan Peneliti adalah untuk mengekstrak aturan atau pedoman yang dapat digunakan oleh petani rumput laut untuk memprediksi harga masa depan produk mereka. Menggunakan modul visualisasi WEKA, Peneliti menciptakan histogram nilai atribut untuk dua kelas yang sedang dipertimbangkan. Gambar 5.2 menunjukkan enam histogram, satu untuk setiap atribut dan histogram tambahan yang mewakili jumlah instance per kelas. Dalam setiap plot, ketinggian setiap batang mewakili frekuensi nilai atribut (kisaran). Warna masing-masing bar mewakili kehadiran relatif dari nilai atribut tertentu di setiap

kelas. Warna yang direpresentasikan adalah s "Penurunan harga" (0) dan "kenaikan harg" biru (1). Histogram untuk kelas (kanan bawah) menunjukkan 144 instance untuk kelas "penurunan harga" dan 131 instance untuk "peningkatan harga". Untuk setiap nilai atribut, rentang relatif merah dan biru merupakan indikasi seberapa baik kelas dapat dipisahkan berdasarkan nilai atribut tersebut. Jika bilah benar-benar merah, nilai atribut terkait selalu menunjukkan "penurunan harga". Jika kedua bagian merah dan biru yang ditumpuk kira-kira sama ukurannya, nilai atribut yang terkait tidak memberikan banyak informasi pada kelas terkait. Pemeriksaan dekat histogram mengungkapkan bahwa atribut Beli (histogram kiri atas) memberikan sebagian besar petunjuk untuk klasifikasi. Untuk nilai kecil atau besar atribut ini, warna bar dominan adalah merah atau biru, masing-masing. Untuk atribut lainnya, isyaratnya jauh lebih halus.

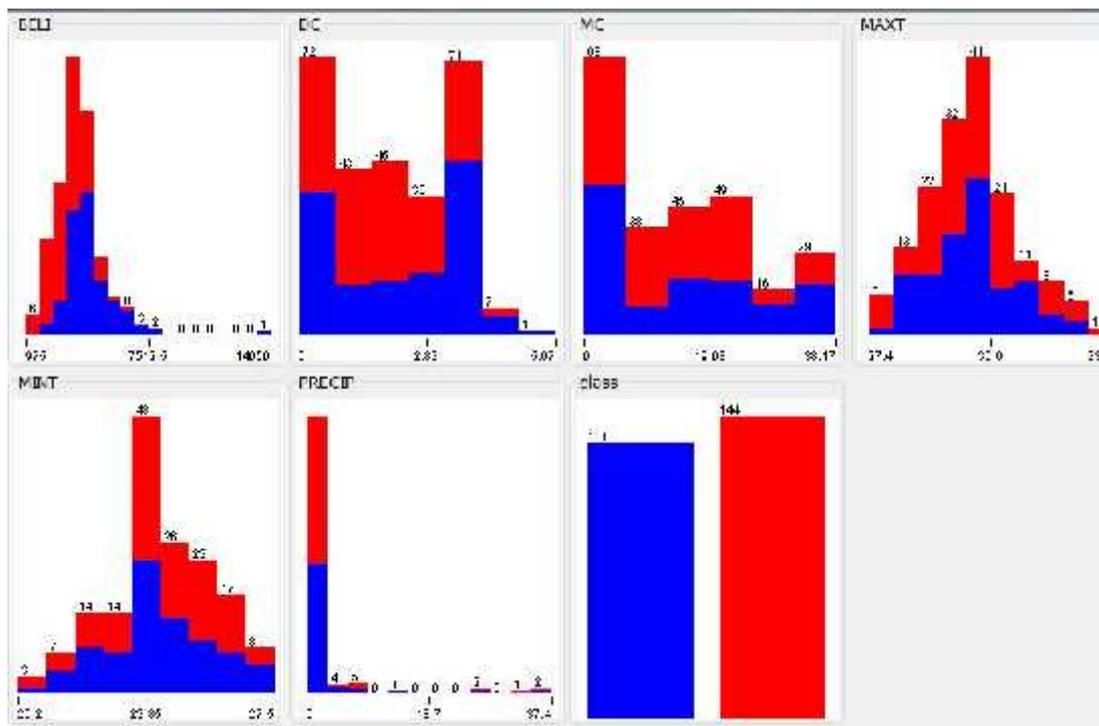


Figure 5.2 classification histograms of the six attributes and the two classes.

Karena pemeriksaan visual dari histogram atribut mengungkapkan kepentingan relatif dari atribut Beli (harga saat ini), Peneliti memeriksa atribut ini secara lebih rinci. Gambar 5.3 adalah plot pencar dari harga di hari  $d$  versus harga di hari  $d + 50-60$  hari, yaitu, atribut dan nilai-nilai prediksi yang diinginkan untuk semua kasus Peneliti. Penurunan dan kenaikan harga diwakili oleh tanda hijau - dan biru +, masing-masing. Penting untuk berkomentar bahwa pemisahan yang jelas dari tanda - dan + adalah hasil langsung dari prosedur pelabelan

Peneliti. Dengan demikian, itu tidak mengungkapkan apa pun tentang atribut.

Seperti dapat dilihat dari plot, nilai harga saat ini lemah terkait dengan nilai harga masa depan. Secara kasar, harga yang lebih rendah dari harga (Beli) pada hari ke- $d$ , memprediksi kenaikan harga pada hari  $d + 50-60$ , dan nilai yang lebih tinggi dari harga memprediksi penurunan harga di masa depan. Relasi lemah ini (yang juga terlihat pada histogram Beli pada Gambar 5.2) dieksploitasi oleh algoritma penambangan data Peneliti

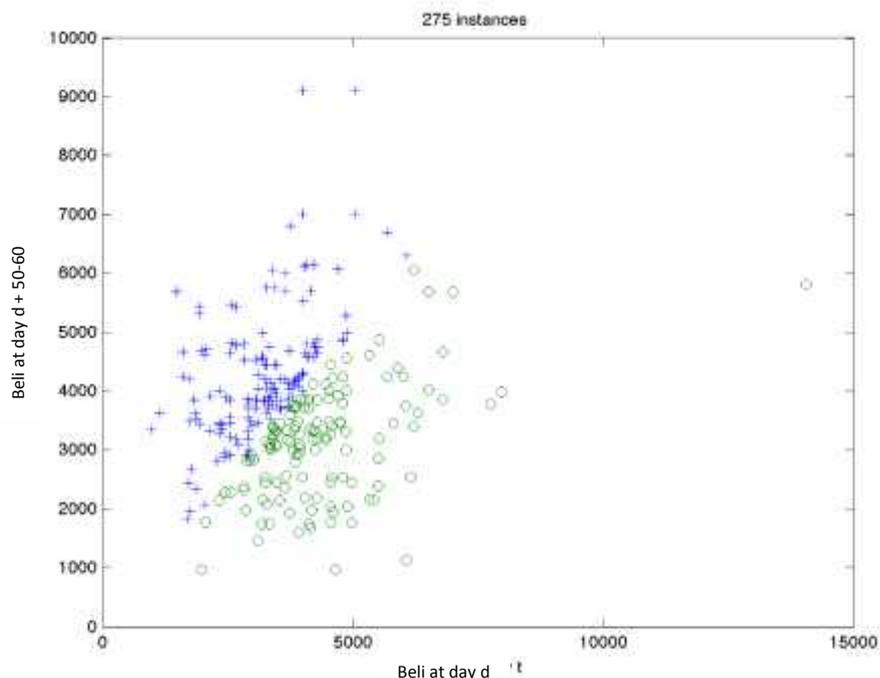


Figure 5.3 Beli at day  $d$  versus Beli at day  $d+50-60$  days.

#### 4. Analisis Classifier

Pemeriksaan kumpulan data memberikan wawasan tentang informasi yang tersembunyi di masing-masing atribut. Pengklasifikasi dapat memanfaatkan kombinasi atribut dan karena itu memberikan wawasan tambahan. Berikut ini, Peneliti menganalisis penggolong yang terlatih dengan mempertimbangkan struktur mereka (untuk pohon keputusan) dan cara mereka membagi ruang atribut.

Pohon keputusan yang terlatih memberikan wawasan tambahan ke kontribusi relatif dari atribut ke prediksi. Gambar 5.4 adalah ilustrasi pohon yang diinduksi dari data Peneliti. Pohon berkinerja terbaik ini dipangkas untuk memasukkan tiga node, dua di antaranya menguji kondisi pada Beli . Titik ketiga mengeksplorasi informasi yang terkandung dalam atribut konten kotor.

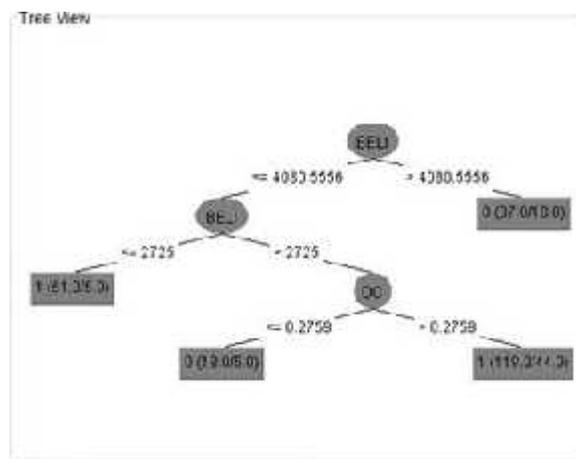


Figure 5.4 Decision tree induced from our data.

Pemeriksaan partisi dari ruang atribut yang direntang oleh dua atribut yang paling informatif, Beli dan Konten Kotor, telah dilakukan dengan bantuan modul visualisasi Weka. Gambar 5.5 menunjukkan visualisasi dari partisi ruang Beli –Dirty Content (sumbu horizontal adalah Beli , sumbu vertikal adalah Konten Kotor) oleh masing-masing penggolong (parameter-dioptimalkan) Peneliti. Warna-warna mewakili label kelas. Merah mewakili "penurunan harga", hijau mewakili "kenaikan harga". Warna menengah mewakili perkiraan kelas campuran (probabilistik). Untuk semua pengklasifikasi, transisi vertikal berwarna merah-hijau yang jelas terlihat. Sekali lagi, transisi ini mencerminkan nilai ambang untuk Beli : nilai kecil (bagian kiri dari ruang

atribut) mengarah pada peningkatan dan nilai besar (bagian kanan dari ruang atribut) mengarah pada penurunan. Empat pengklasifikasi berbeda dalam rincian partisi ruang atribut. Semua pengklasifikasi, kecuali untuk pohon keputusan, membuat prediksi yang tampaknya juga bergantung pada atribut Konten Kotor, sebagaimana tercermin dalam batas yang dimiringkan (seperti untuk SVM dan Naïve Bayes) atau menjadi lebih kompleks ( IBk ). Hasil kuantitatif Peneliti (yang tercantum dalam Tabel 5.1) entah bagaimana mencerminkan kompleksitas batas-batas. Batas yang paling sederhana (pohon keputusan J48) menghasilkan kinerja terendah, sedangkan batas yang lebih rumit (pengklasifikasi lainnya) menghasilkan kinerja yang lebih baik.

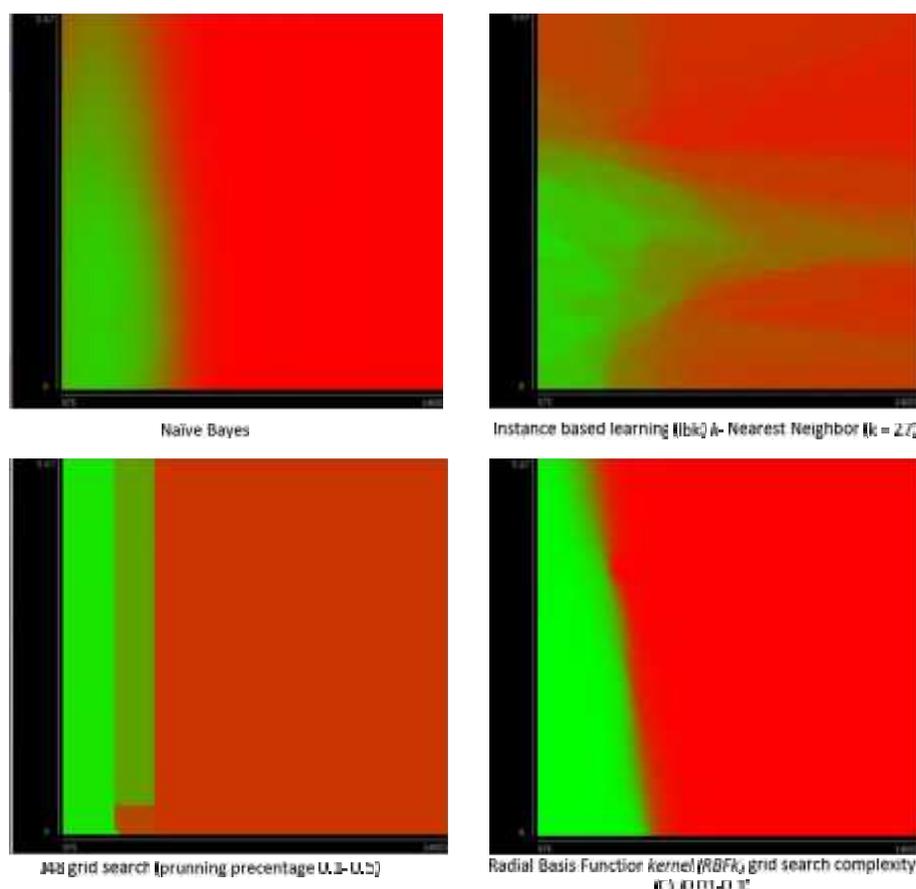


Figure 5.5 the boundary of class probability estimation. The classifiers have been parameter-optimized by hand (IBk) or by means of grid search (J48 and SVMs).

## KESIMPULAN

Bab ini menjelaskan diskusi umum, kesimpulan, dan penelitian masa depan.

### 1. Diskusi Umum

Dengan menerapkan model Cunningham yang diadaptasi (1999), Peneliti berhasil membuat prediksi yang wajar tentang arah harga jual masa depan dan dalam mengidentifikasi atribut yang paling informatif. Namun, prediksi ini agak sepele karena harga jual yang tinggi saat ini lebih cenderung turun dan harga jual yang rendah saat ini lebih cenderung naik. Hasil Peneliti tampaknya mencerminkan prinsip umum ini, daripada didasarkan pada pengetahuan khusus rumput laut. Di satu sisi, fakta bahwa pengklasifikasi Peneliti berhasil mengeksploitasi prinsip mendukung keefektifannya sebagai mesin inferensi statistik. Di sisi lain, itu belum mengungkapkan manfaat sebenarnya untuk domain yang sedang dipertimbangkan. Dalam pengertian itu, penelitian Peneliti harus dianggap sebagai percobaan percontohan untuk menentukan kelayakan penerapan penambahan data. Pelajaran utama yang dapat diambil dari percobaan percontohan adalah bahwa kita perlu memasukkan lebih banyak pengetahuan tentang domain di tangan dan bahwa tugas belajar harus disempurnakan. Lebih banyak pengetahuan membutuhkan interaksi intensif dengan penyedia

data (lihat di bawah). Tugas belajar itu sendiri dapat disempurnakan dengan mendefinisikan ulang label. Daripada memprediksi jika harga naik atau turun, kita bisa menggunakan harga sebenarnya sebagai label dan sebagai target untuk prediksi. Analisis lebih lanjut harus menunjukkan apakah tingkat harga absolut atau tingkat harga relatif lebih disukai.

Peneliti menganggap penyelidikan Peneliti sebagai studi awal percontohan yang mempersiapkan untuk penelitian yang lebih besar di masa depan diarahkan pada segmen yang lebih luas dan lainnya dari ekonomi Indonesia. Hasil yang Peneliti peroleh dalam studi percontohan Peneliti memperkuat keyakinan Peneliti bahwa penelitian ini akan berhasil. Secara khusus, Peneliti mempertimbangkan hal berikut sebagai empat poin kuat dari pendekatan Peneliti.

1. Akurasi prediksi minimal 65%.
2. Akurasi prediksi tidak bergantung secara kritis pada tipe penggolong.
3. Prediksi bersifat transparan dan dapat diubah menjadi aturan main-jempol yang bisa dimengerti.
4. Pendekatan ini menggarisbawahi kelayakan penerapan penambahan data ke domain akuakultur.

Titik lemah dari pendekatan Peneliti mendorong Peneliti untuk memperbaiki metode yang diikuti dalam pekerjaan Peneliti di masa depan.

Keterbatasan pertama dan utama dari studi saat ini adalah bahwa Peneliti tidak memiliki akses ke penyedia data. Dalam model Cunningham, komunikasi dengan penyedia data sangat penting. Dalam kasus Peneliti, kontak langsung dengan petani rumput laut akan memberikan umpan balik langsung pada relevansi dan kelayakan hasil Peneliti. Dalam penelitian Peneliti di masa depan, Peneliti akan memastikan bahwa kontak yang kuat dengan penyedia data telah ditetapkan.

Keterbatasan kedua adalah terlalu tinggi dari akurasi prediksi karena optimasi parameter. Meskipun dalam penelitian Peneliti hasil yang diperoleh dengan dan tanpa optimasi parameter tidak berbeda terlalu banyak, dalam studi masa depan kita harus memasukkan optimisasi parameter dalam skema validasi silang untuk menghindari perkiraan yang terlalu tinggi dari kinerja prediksi.

Keterbatasan ketiga dan terakhir menyangkut kualitas data yang tidak pasti. Peneliti tidak memiliki sarana untuk mengkonfirmasi validitas data yang diperoleh melalui dua situs web. Dalam studi masa depan kita, kita perlu memastikan bahwa integritas data dijamin dengan memperoleh akses ke basis data resmi (misalnya pemerintah).

## 2. Kesimpulan

Dalam pengantar Peneliti menyatakan tiga pertanyaan penelitian. Peneliti sekarang menyimpulkan dengan mengulangi dan menjawab masing-masing pertanyaan ini secara bergiliran.

a. *Sampai sejauh mana harga jual rumput laut diprediksi 8 minggu ke depan tepat waktu?*

Menggunakan data mining adalah mungkin untuk memprediksi harga jual rumput laut dengan akurasi di atas tingkat kesempatan.

b. *Metode penambangan data apa yang cocok untuk prediksi harga jual rumput laut?*

Dari empat pengklasifikasi yang diperiksa, semuanya memiliki kinerja yang sama.

c. *Pertanyaan penelitian 3: Faktor apa saja yang menentukan harga jual rumput laut di masa depan ?*

Tiap faktor prediksi utama dari harga jual rumput laut adalah harga jual saat ini. Selain itu, konten kotor berkontribusi pada prediksi, meskipun pada tingkat lebih rendah.

## 3. Saran untuk penelitian masa depan

Mengingat diskusi dan kesimpulan Peneliti, Peneliti menyarankan tiga arah berikut dari penelitian masa depan.

- Mendefinisikan ulang tugas belajar dan menggabungkan lebih banyak pengetahuan domain.
- Penerapan model Cunningham ke set data yang lebih besar (lebih banyak catatan dan lebih banyak atribut).
- Melakukan evaluasi komparatif yang lebih halus dari algoritma penambangan data.

Membangun berdasarkan penelitian Peneliti dengan mengeksplorasi penggunaan penimbangan data di berbagai domain ekonomi dalam kerjasama erat dengan penyedia data yang mapan.

## DAFTAR PUSTAKA

- Bank Indonesia. (2006). Pola pembiayaan usaha kecil (PPUK). Jakarta: Direktorat kredit, BPR dan UMKM Bank Indonesia.
- Barrow, M. (2009). *Statistics for Economics, Accounting and Business Studies* (fifth ed.). London: Pearson Education Limited.
- Baskar, S. S., Arockiam, L., Kumar, A. V., & Jeyasimman, L. (2010). Brief survey of Data Mining Techniques to Agriculture. *Agricultural Journal*, 5, 116-118. doi: 10.3923/aj.2010.116.118
- Chen, F. L., & Ou, T. Y. (2008). *A Neural-Network-Based Forecasting Method for Ordering Perishable Food in Convenience Stores*. Paper presented at the Proceedings of the 2008 Fourth International Conference on Natural Computation - Volume 02.
- Chen, Y.-c., Rogoff, K. S., & Rossi, B. (2010). Predicting agri-commodity prices: an asset pricing approach (D. o. Economics, Trans.) (Vol. I, pp. 45). Washington: University of Washington.
- Enireddy, V., Varma, K. V. S. R. P., Rao, S. P., & Satapati, R. (2010). Prediction of rainfall using backpropagation neural network model. *International Journal on Computer Science and Engineering (IJCSE)*, 02 1119-1121.
- FAO. (2002). CWP Handbook of Fishery statistical standards Rome: FAO Fisheries and Aquaculture.
- Fayyad, U., M., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery: an overview. In M. F. Usama, P.-S. Gregory, S. Padhraic & U. Ramasamy (Eds.), *Advances in knowledge discovery and data mining* (pp. 1-34): American Association for Artificial Intelligence.
- Food Agriculture Organization (FAO). (2002). CWP Handbook of Fishery statistical standards Rome: FAO Fisheries and Aquaculture.

- Gordon, D. V. (2010). *Price modelling in the Canadian fish supply chain with forecast and simulation of the ex-vessel price of fish*. Paper presented at the Food Agriculture Workshop (FAO) on Supply Chain, Tokyo.
- Hand, D., Mannila, H., & Smyth, P. (2001). *Principles of data mining*. Massachusetts: Massachusetts Institute of Technology.
- Hu, T., Zhang, X., Hou, Y., Mu, W., Fu, Z., Wang, L., et al. (2005). A Hybrid Model for Forecasting Aquatic Products Short-Term Price Integrated Wavelet Neural Network with Genetic Algorithm  
Advances in Natural Computation (Vol. 3611, pp. 429-429): Springer Berlin / Heidelberg.
- Li, G.-q., Xu, S.-w., & Li, Z.-m. (2010). Short-Term Price Forecasting For Agro-products Using Artificial Neural Networks. *Agriculture and Agricultural Science Procedia*, 1(0), 278-287.
- Liu, Y. (2002). A framework of data mining application process for credit scoring. Retrieved from [www.wi.wiso.uni-goettingen.de](http://www.wi.wiso.uni-goettingen.de) website:
- Pelinggon, R. E., & Tito, O. D. (2009). Enhancing the demands of AFNR graduates through curricular intervention using modular approach with high S & T *Seaweed production*
- Philip, K. C., Wei, F., Andreas, L. P., & Salvatore, J. S. (1999). Distributed Data Mining in Credit Card Fraud Detection. *IEEE Intelligent Systems*, 14(6), 67-74. doi: 10.1109/5254.809570
- Quinlan, R. J. (1993). C4.5 Programs for Machine Learning.
- Ruß, G., Kruse, R., Schneider, M., Wagner, P., & Perner, P. (2008). Data Mining with Neural Networks for Wheat Yield Prediction  
Advances in Data Mining. Medical Applications, E-Commerce, Marketing, and Theoretical Aspects (Vol. 5077, pp. 47-56): Springer Berlin / Heidelberg.
- Seng, J.-L., & Chen, T. C. (2010). An analytic approach to select data mining for business decision. *Expert Systems with Applications*, 37(12), 8042-8057.
- Tiroba, G. (2006). *A practical guide for seaweed farmers, Seaweed extension officers, Buying agents, Fisheries officers and Exporters.*: CoSPSI.
- Trono, G. C. J. (1990). *Lesson from the history of seaweed culture in the Philippines and the trend of seaweed farming in Southeast Asia*. Paper presented at the The regional workshops on seaweed culture and marketing, Fiji.
- Witten, I. A., Frank, E., & Hall, M. A. (2011). Data Mining practical machine learning tools and technique
- Wolpert, D. H., & Macready, W. G. (1997). No Free Lunch Theorems for Optimization. *IEEE Transaction on Evolutionary*, 1, 67-82.
- Wu, X., Kuma, V., Quinlan, R. J., Ghosh, J., Yang, Q., Motoda, H., et al. (2007). The top 10 on Data Mining Algorithm. Retrieved from <http://www.cs.uvm.edu/~icdm/algorithms/10Algorithms-08.pdf> doi:DOI 10.1007/s10115-007-0114-2
- Xiaoshuan, Z., Tao, H., Revell, B., & Zetian, F. (2005). A forecasting support system for aquatic products price in China. *Expert Systems with Applications*, 28(1), 119-126.
- Yakup, K., Melek Acar, B., & mer Kaan, B. (2011). Predicting direction of stock price index movement using artificial neural networks and support vector machines: The sample of the Istanbul Stock Exchange. *Expert Syst. Appl.*, 38(5), 5311-5319. doi: 10.1016/j.eswa.2010.10.027
- Zhang, W., Chen, H., & Wang, M. (2010). *A forecast model of agricultural and livestock product price*. Paper presented at the International federation for information processing (IFIP), Beijing.