

# Pemodelan Sistem Prediksi Tanaman Pangan Menggunakan Algoritma Decision Tree

## Crop Prediction System Using Decision Tree Algorithm

<sup>1</sup>Fergie Joanda Kaunang, <sup>2</sup>Reymon Rotikan, <sup>3</sup>Gleadies Stella Tulung

<sup>1,3</sup>Progam Studi Teknik Informatika, Universitas Klabat, Airmadidi

<sup>2</sup>Program Studi Sistem Informasi, Universitas Klabat, Airmadidi

e-mail: <sup>1</sup>fergie@unklab.ac.id, <sup>2</sup>reymonr@unklab.ac.id, <sup>3</sup>gleadiestulung@gmail.com

### Abstrak

*Pertanian sebagai salah satu sektor industri menjadi bagian pekerjaan yang menunjang pemenuhan kebutuhan makanan pokok masyarakat seperti tanaman pangan. Cuaca yang berubah-ubah dan tidak menentu dapat mempengaruhi hasil panen terlebih khusus dari segi jumlah hasil produksi. Hal ini membuat cuaca menjadi salah satu faktor penentu hasil produksi dari tanaman pangan. Memprediksi hasil panen tanaman pangan dengan baik dapat membantu para pemangku kepentingan baik petani ataupun mereka yang bekerja dalam industri pengolahan hasil tanaman pangan. Dewasa ini, Data Mining dan Machine Learning adalah dua topik berkaitan yang sering digunakan dalam berbagai bidang kehidupan termasuk bidang pertanian. Penelitian ini menggunakan teknik Data Mining dan Machine Learning untuk membuat suatu model sistem prediksi tanaman pangan di provinsi Sulawesi Utara berdasarkan iklim/cuaca. Menggunakan algoritma Decision Tree J48, penelitian ini memberikan hasil yang dapat dipertimbangkan dalam pengambilan keputusan bagi para pemangku kepentingan di bidang pertanian. Hasil yang diperoleh juga menunjukkan pengaruh parameter iklim/cuaca terhadap jumlah hasil produksi tanaman pangan di provinsi Sulawesi Utara. Penelitian ini tidak menggunakan parameter lain yang dapat mempengaruhi hasil produksi tanaman pangan seperti kondisi tanah ataupun harga jual di pasaran, mengingat dapat terjadinya perubahan terhadap harga yang tidak menentu. Penelitian ini diharapkan dapat memberikan kontribusi terhadap bidang pertanian terlebih khusus kepada para petani tanaman pangan dalam menopang pemenuhan kebutuhan pokok masyarakat.*

**Keywords : Machine Learning, Data Mining, Algoritma Decision Tree J48, Sistem Prediksi, Tanaman Pangan**

### Abstract

*Agriculture as one of the industrial sector becomes part of the work that support the fulfillment of basic food needs of the community such as food crops. The unpredicted change of the weather can affect the amount of the harvest. Therefore, the climatic parameters become one of the factor that impact the crop yield. Predicting the crop yield well may help the farmers and those who work in agriculture industry to make the right decision. Data Mining and Machine Learning has been an emerging issue nowadays for they can be applied in any sector including agricultural sector. Using Data Mining and Machine Learning techniques, we developed a model to predict crop yield in North Sulawesi province based on climatic parameters. This study used Decision Tree J48 algorithm and generates results that can be taken into consideration to make a decision. This study also shows how climatic characteristics impacts the crop yield in North Sulawesi. Other parameter or agro-input like soil condition as well as market price are not considered in this study due to its unpredicted change day by day.*

However, we expected that this study may be useful to the farmers as well as those who work in the related field.

**Keywords:** *Machine Learning, Data Mining, Decision Tree J48 Algorithm, Prediction System, Crop*

## 1. PENDAHULUAN

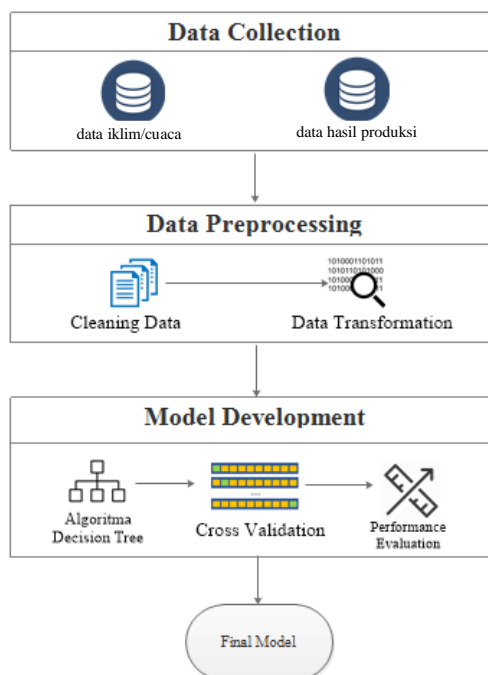
Indonesia merupakan negara agraris dimana pertanian memegang peran penting dikarenakan kaya sumber daya alam, seperti tanaman pangan dan hortikultura [1, 2, 3]. Sebagai salah satu sektor industri, pertanian menjadi bagian pekerjaan juga pemenuhan kebutuhan masyarakat seperti kebutuhan makanan pokok. Lahan yang subur menjadi faktor penunjang perkembangan pertumbuhan tanaman dan Provinsi Sulawesi Utara memiliki lahan yang cukup luas dan subur untuk area pertanian dalam hal ini tanaman pangan [4]. Tanaman pangan merupakan kebutuhan pokok untuk memenuhi ketercukupan nutrisi tubuh manusia seperti karbohidrat, protein, mineral dan vitamin yang bermanfaat untuk keberlangsungan hidup dan kesehatan manusia [5, 6]. Dengan demikian hasil panen dari tanaman pangan mempengaruhi sebagian besar hidup masyarakat. Iklim dan lingkungannya serta naik-turunnya harga pasar menjadi beberapa contoh faktor yang dapat berdampak terhadap hasil panen dari tanaman pangan tersebut [7]. Perkiraan hasil panen adalah salah satu persoalan paling kritis yang ditemui dalam sektor pertanian, hal ini dikarenakan kurangnya pengetahuan para petani tentang panen yang melimpah, ketidakpastian kondisi cuaca hingga naik-turunnya harga pasar yang berpengaruh terhadap hasil produksi tanaman [8, 9]. Adapun data tanaman pangan yang digunakan dalam penelitian ini meliputi jagung, kacang tanah, padi, ubi jalar dan ubi kayu.

Pada penelitian "*Prediction of Crop Yield using Regression Analysis*", digunakan *Regression Analysis* untuk mencari hubungan antara 3 variabel (*Area Under Cultivation* (AUC), *Annual Rainfall* (AR) dan *Food Price Index* (FPI)) yang dipertimbangkan dan mengidentifikasi efektifitas dalam memprediksi hasil panen dan menemukan jumlah relative antara variabel tidak bebas dan variabel mandiri. *Area Under Cultivation* (AUC), *Annual Rainfall* (AR) dan *Food Price Index* (FPI) merupakan faktor yang berkontribusi terhadap hasil panen. Penelitian ini menggunakan padi sebagai objek, dimana padi merupakan tanaman umum yang dipanen di sebagian besar daerah di India [8].

Pada penelitian "*Machine Learning Approach For Forecasting Crop Yield Based On Climatic Parameters*", digunakan teknik data mining yaitu decision tree algoritma C 4.5 dalam memprediksi hasil panen berdasarkan parameter input iklim untuk mengetahui parameter iklim yang paling berpengaruh terhadap hasil panen tanaman di beberapa daerah di Madhya Pradesh. Tanaman yang dipilih meliputi: *Soybean, Maize, Paddy* dan *Wheat* (kedelai, jagung, padi dan gandum). Parameter iklim yang digunakan yaitu: *Rainfall, Maximum & Minimum temperature, Potential Evapotranspiration, Cloud cover, Wetday frequency* (Curah Hujan, Suhu Maksimum & Minimum, Potensi kekurangan air tanah, tutupan Awan, frekuensi Wetday) [10].

## 2. METODE PENELITIAN

Penelitian ini bertujuan untuk membuat suatu model sistem prediksi tanaman pangan untuk mencari tahu pengaruh cuaca terhadap produksi hasil panen. Data yang digunakan dalam penelitian ini hanya terbatas pada data tanaman pangan dan data iklim di provinsi Sulawesi Utara saja. Terdapat 4 tahapan yang dilakukan dalam penelitian ini yaitu *Data Collection, Data Transformation, Model development* dan *Final model* seperti yang terlihat pada Gambar 1.



Gambar 1 Bagan Alur Penelitian

### 2.1 Data collection

Dalam tahap ini data diperoleh dari sumber data yakni dari data online dari Badan Pusat Statistik (BPS) dan dari Badan Meteorologi Klimatologi dan Geofisika (BMKG) [11]. Data yang dikumpulkan yaitu: iklim/cuaca dan hasil produksi tanaman pangan. Data yang digunakan dalam penelitian ini adalah 5 tanaman pangan dengan hasil produksi tertinggi di provinsi Sulawesi Utara yakni jagung, kacang tanah, padi, ubi jalar, dan ubi kayu dari tahun 2008 sampai dengan tahun 2015.

### 2.2 Data Preprocessing

Setelah data selesai dikumpulkan, selanjutnya dilakukan tahapan *preprocessing* yang terbagi dalam 2 tahapan yaitu *Cleaning data* dan *Data Transformation*.

- Cleaning data process* mencakup antara lain membuang duplikasi data, memeriksa data yang tidak konsisten, dan memperbaiki kesalahan pada data, seperti kesalahan cetak (tipografi) begitu juga dengan data yang tidak memiliki nilai (*missing value*). Dari data yang didapatkan data tersebut kemudian di *cleaning* dimana *missing value* pada semua data baik data cuaca juga data hasil produksi diganti dengan nilai 0 (nol).
- Data Transformation*, pada bagian ini setelah data sudah dibersihkan, dilakukan proses transformasi dalam hal ini mengubah data numerik pada harga produksi ditransformasi menjadi data nominal. Proses ini dilakukan dengan menggunakan metode binning yakni *Equal Width Binning* dimana metode ini membagi data menjadi  $k$  interval dengan ukuran yang sama. Rumus yang digunakan yakni:

$$w = (max - min)/k$$

dimana  $w$  adalah nilai yang menentukan nilai batas untuk masing-masing interval  $k$ . Adapun batas interval yang diperoleh adalah  $min + w, min + 2w, \dots, min + (k - 1)w$  [12]. Berdasarkan metode ini data hasil produksi dikategorikan menjadi 2 (dua) kategori yaitu *low* dan *high*. Kategori *low* menyatakan bahwa hasil produksi berada di bawah nilai rata-rata, sedangkan kategori *high* menyatakan bahwa hasil

produksi berada di atas nilai rata-rata. Parameter hasil produksi ini yang kemudian akan digunakan sebagai parameter klasifikasi untuk masing-masing jenis tanaman pangan. Hasil akhir dari tahapan ini adalah *training dataset* yang akan digunakan untuk pemodelan sistem. Adapun jumlah dari data akhir yang diperoleh yaitu 120 data untuk masing-masing jenis tanaman pangan.

2.3 Model development

Merupakan tahapan membuat model sistem, dengan menggunakan algoritma *decision tree* dimana diterapkan *10-fold cross validation* yang selanjutnya akan diukur dan dievaluasi kinerja dari model yang telah dibuat sehingga diketahui seberapa akurat model yang dibuat untuk memprediksi tanaman pangan tersebut. Proses ini dilakukan dengan menggunakan WEKA, perangkat lunak untuk proses klasifikasi.

2.4 Final model merupakan hasil akhir dari proses pembuatan dan pelatihan algoritma yang digunakan. Model ini yang kemudian diharapkan dapat menjadi model rekomendasi untuk memprediksi tanaman pangan.

3. HASIL DAN PEMBAHASAN

Atribut yang digunakan dalam proses klasifikasi meliputi suhu (°C) rata-rata per tahun, kelembaban rata-rata (%), curah hujan (mm), lama penyinaran (jam), dan hasil produksi (ton) sebagai atribut kelas. Tabel 1 menunjukkan hasil akurasi dari model yang dikembangkan. Dapat dilihat bahwa nilai akurasi dapat berbeda bagi tiap jenis tanaman pangan. Hal ini dapat terjadi karena perbedaan jumlah produksi per tahunnya. Adapun hasil akurasi secara keseluruhan dari model ini adalah 69.48%. Berdasarkan hasil ini maka model yang kami kembangkan dapat menjadi bahan pertimbangan bagi para pembuat keputusan sebelum mengambil keputusan.

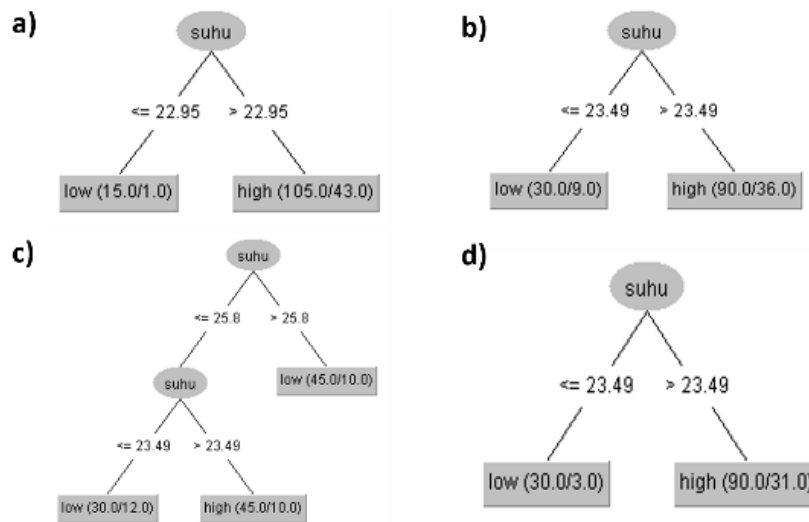
Tabel 1 Hasil Akurasi untuk Setiap Jenis Tanaman

No.	Jenis Tanaman Pangan	Akurasi (Accuracy, %)
1	Jagung	62.5
2	Kacang Tanah	71.6
3	Padi	72.5
4	Ubi Jalar	62.5
5	Ubi Kayu	78.3

Penelitian ini juga memberikan hasil yang menunjukkan atribut yang paling mempengaruhi prediksi masing-masing jenis tanaman pangan. Gambar 2 menunjukkan hasil visualisasi penerapan algoritma *decision tree* untuk tanaman jagung (a). Dapat dilihat bahwa atribut suhu menjadi simpul akar. Hal ini juga menunjukkan bahwa atribut yang paling mempengaruhi tanaman jagung adalah suhu, dimana hasil produksi menjadi lebih tinggi apabila suhu rata-ratanya lebih dari 22.95 °C. Hasil yang sama juga berlaku pada tanaman kacang tanah (b), padi (c), dan ubi jalar (d) dimana suhu menjadi atribut yang paling mempengaruhi hasil produksi tanaman.

Adapun hasil untuk tanaman ubi kayu sedikit berbeda dibandingkan dengan tanaman pangan lainnya. Untuk tanaman ubi kayu atribut kelembaban, curah hujan, dan suhu yang menjadi faktor yang mempengaruhi hasil produksinya. Terdapat 2 (dua) kemungkinan untuk memperoleh produksi ubi kayu yang tinggi yakni pertama dengan memperhatikan kelembaban dan curah hujan, dimana apabila nilai kelembaban lebih kecil atau sama dengan 83.14% dan curah hujan lebih besar dari 7.12 mm, atau apabila nilai curah hujan lebih kecil atau sama

dengan 7.12 mm akan diperhitungkan nilai suhu rata-ratanya yakni lebih dari 23.49 °C maka hasil produksi ubi kayu tergolong tinggi.



Gambar 2 Hasil Visualisasi Tanaman Pangan dengan algoritma *decision tree*

#### 4. KESIMPULAN

Dari hasil eksperimen yang dilakukan dengan data yang ada menggunakan algoritma *decision tree J48* dengan *10-folds cross validation* dimana atribut yaitu curah hujan, suhu rata-rata, kelembaban rata-rata, dan lama penyinaran serta *class* yaitu hasil produksi diperoleh nilai akurasi rata-rata yaitu 69.48%. Dengan kata lain, model yang kami kembangkan dapat digunakan untuk membantu para petani dalam menentukan jenis tanaman yang akan ditanam dengan mempertimbangkan keadaan cuaca, dengan harapan dapat terjadi peningkatan produksi tanaman pangan. Selain itu, model ini juga dapat membantu para pengambil keputusan dalam industri yang mengelola hasil tanaman pangan.

#### 5. SARAN

Investigasi selanjutnya dapat dilakukan lagi agar model dibuat bisa memiliki performa yang lebih baik lagi. Investigasi dapat dilakukan dengan mengumpulkan data yang lebih banyak lagi, atau dapat menggunakan parameter lain seperti *Food Price Index*, data konsumsi, ataupun data pertanian lainnya seperti kondisi tanah. Dapat pula menggunakan algoritma klasifikasi lainnya seperti K-Nearest, Random Forest, NaïveBayes dengan menganalisis dan membandingkan kinerja setiap algoritma untuk hasil akurasi yang lebih baik. Selanjutnya dapat pula dikembangkan suatu perangkat lunak ataupun sistem berbasis *web* dari model prediksi yang dihasilkan penelitian ini.

#### UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Fakultas Ilmu Komputer Universitas Klabat yang telah memberi dukungan finansial terhadap penelitian ini.

## DAFTAR PUSTAKA

- [1] Tunjung, Artist, *Analisis Efisiensi Pengelolaan Persediaan Bahan Baku Kedelai Pada Perusahaan Kecap PT. Lombok Gandaria Food Industry Palur Karanganyar*. [Art]. Universitas Sebelas Maret, 2010.
- [2] J. Umboh, Artist, *Pengaruh Perubahan Sosial Terhadap Lahan Tidur di Desa Tombasian Atas Kecamatan Kawangkoan Barat*. [Art]. Universitas Sam Ratulangi Manado, 2014.
- [3] H. Warsani, Artist, *Kajian Pemanfaatan Lahan Sawah di Kecamatan Kuantan Tengah Kabupaten Kuantan Singingi*. [Art]. Universitas Pendidikan Indonesia, 2013.
- [4] "Harga Bahan Pokok Tahun 2015 dan Tahun 2016," Dinas Perindustrian dan Perdagangan Sulawesi Utara.
- [5] D. H. Sunaryono and Rismunandar, *Pengantar Pengetahuan Dasar Hortikultura*, Bandung: Sinar Baru Bandung.
- [6] "Seputar Pertanian," [Online]. Available: <http://www.seputarpertanian.com/2016/03/macam-macam-tanaman-pangan.html>. [Accessed 17 March 2017].
- [7] Suhardi, *Dasar-dasar Bercocok Tanam*, Yogyakarta: Kanisius.
- [8] V. Sellam and E. Poovammal, "Prediction of Crop Yield using Regression Analysis," *Indian Journal of Science and Technology*, vol. 9, no. 38, p. 5, 2016.
- [9] S. S. Dahikar and D. V. Rode, "Agricultural Crop Yield Prediction Using Artificial Neural Network Approach," *INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH IN ELECTRICAL, ELECTRONICS, INSTRUMENTATION AND CONTROL ENGINEERING*, vol. 2, no. 1, p. 4, 2014.
- [10] S. Veenadhari, "Machine Learning Approach for Forecasting Crop Yield based on Climatic Parameters," in *International Conference on Computer Communication and Informatics*, 2014.
- [11] "Data Online Pusat Database - BMKG," [Online]. Available: <http://dataonline.bmkg.go.id/home>. [Accessed June 2017].
- [12] R. K. M. S. J. Dougherty, "Supervised and unsupervised discretization of continuous features," in *Machine Learning Proceedings*, 1995.