



Available online at : <http://bit.ly/InfoTekJar>

InfoTekJar : Jurnal Nasional Informatika dan Teknologi Jaringan

ISSN (Print) 2540-7597 | ISSN (Online) 2540-7600



Deteksi Penyakit Daun pada Tanaman Padi Menggunakan Algoritma Decision Tree, Random Forest, Naïve Bayes, SVM dan KNN

Annida Purnamawati¹, Wawan Nugroho¹, Destiana Putri², Wahyutama Fitri Hidayat²

¹Ilmu Komputer, STMIK Nusa Mandiri, Jl. Kramat Raya No. 18, Senen, Jakarta Pusat

²Ilmu Komputer, Universitas Bina Sarana Informatika, Jl. Kramat Raya No. 98, Senen, Jakarta Pusat

KEYWORDS

Daun Padi, Tanaman, *Decision Tree*, *Random Forest*, *Naïve Bayes*, SVM dan KNN

CORRESPONDENCE

E-mail:

14002403@nusamandiri.ac.id

ABSTRACT

Pertanian sebagai salah satu sektor industri menjadi bagian pekerjaan yang menunjang pemenuhan kebutuhan makanan pokok masyarakat seperti tanaman pangan. Tanaman padi merupakan tanaman pangan yang rentan terserang hama. Pengendalian terhadap jenis hama yang menyerang merupakan langkah awal yang sangat penting untuk menunjang keberhasilan dalam usaha pengendaliannya. Hama tanaman padi tersebut dapat menjadi kendala bagi petani untuk bisa meningkatkan produksi. karena hama tersebut dapat merusak tanaman padi hingga membuat gagal panen. Oleh sebab itu perlu dilakukan deteksi klasifikasi pada hama daun padi untuk mencari akurasi dengan menggunakan perbandingan berbagai macam metode algoritma yaitu dengan *Decision Tree*, *Random Forest*, *Naïve Bayes*, SVM dan KNN. Sehingga diharapkan mampu menangani hama secara tepat, agar tidak terjadi kerusakan dan gagal panen. Dengan menggunakan *dataset Rice Leaf Diseases Detection* untuk Deteksi dan Klasifikasi Penyakit Padi. Dataset ini memiliki tiga kelas/penyakit yang diantaranya yaitu: Bakteri daun busuk, bercak coklat, dan daun api, masing-masing memiliki 40 gambar dengan format gambar jpg. Dari perbandingan ke lima metode algoritma tersebut dapat dihasilkan 3 macam model yaitu Model *Overfit* (*Random Forest*, *Decision Tree* dan *Naïve Bayes*), Model *Underfit* (SVM) dan *Good Models* (KNN). Jadi metode terbaik diantara kelima tersebut yaitu metode KNN dengan nilai akurasi 87%, karna model ini konsisten baik pada kedua evaluasi. KNN tidak terbukti memiliki masalah *overfitting* karena secara konsisten berkinerja baik pada data *train* dan data *test*.

Agriculture as one of the industrial sectors has become a part of the work that supplies people's basic food needs, such as food crops, rice is a food that is susceptible to pests. Identifying the host of pests is a vital first step toward promoting success in its control. The pest of the rice plant can pose a challenge for farmers to increase production. Because such pests can damage the crops to the point of failure. It is therefore necessary to assess the classification of rice leaf pests for accuracy by using a variety of algorithm-based methods of decision tree, random forest, naive bayes, SVM and KNN in the hope that farmers will soon discover the type of rice pests and their ferocity levels. And so it is expected to be able to handle eve properly, lest the damage and failure of the harvest, by using datasets Rice leaf diseases diseases to detect and classification rice diseases. This dataset has three classes The underlying diseases: leaf rot, chocolate patches, and fire leaves produce 40 images each with JPG in their format. In comparison to the five possible methods of the algorithm, the three types of models are the overfit models (Random Forest, Decision Tree dan Naive Bayes), Model Underfit (SVM) dan Good Models (KNN). So the fifth prime method, however, is the KNN method with an accuracy of 87%, because it is consistent with both evaluations. KNN has no evidence of overfitting problems because it consistently performs well on train data and test data.

INTRODUCTION

Pertanian merupakan kegiatan dengan memanfaatkan sumber daya hayati yang dilakukan oleh kebanyakan manusia dengan bertujuan untuk menghasilkan suatu bahan pangan. Kegiatan seperti ini merupakan sektor industri dan juga sumber energi

untuk dapat mengelola lingkungan hidup. Dalam budidaya tanaman padi tidak akan terlepas dari ancaman penyakit yang sering menyerang. Penyakit yang menyerang tanaman padi ini tanpa disadari oleh petani dan sering kali para petani kurang ngerti mengenai penyakit yang menyerang tanaman padi sehingga terlambat dalam penanganan untuk didiagnosa gejala membuat produksi tanaman padi menurun[1].

Oleh karena itu perlunya pengetahuan petani untuk mengenali jenis-jenis penyakit pada tanaman padi agar petani dapat mengidentifikasi dan bisa menerapkan pengendalian secara tepat, cepat dan akurat. Dengan adanya identifikasi diharapkan para petani dapat mengendalikan serangan penyakit tanaman padi dengan maksud dan tujuan dari budidaya tanaman padi dapat tercapai. Penyakit pada tanaman padi ini telah di deteksi dan di klasifikasi menjadi tiga jenis atau tiga kelas diantaranya sebagai berikut; Bacterial Leaf Blight, Brown Spot dan Leaf Smut.

Dengan semakin berkembangnya teknologi beserta segala alternatif yang telah di tawarkan, peneliti akan melakukan *Machine Learning Classifier*. Untuk dapat melihat klasifikasi dari jenis penyakit ini maka penulis akan mencari akurasi dengan menggunakan perbandingan berbagai macam metode algoritma yaitu dengan Decision Tree, Random Forest, Naive Bayes, SVM dan KNN. Penulis akan melakukan perbandingan dengan kelima metode tersebut menggunakan lima metode algoritma.

METHOD

Teknik Klasifikasi adalah sebuah model dalam *data mining* dimana classifier dikonstruksi untuk memprediksi *categorical* label. Kategori tersebut dapat direpresentasikan dengan nilai yang sesuai kebutuhannya. Klasifikasi merupakan tugas yang sama dengan data mining, dimana tujuan utama dari klasifikasi adalah prediksi label kelas.

Setiap Teknik Klasifikasi yang menggunakan suatu algoritma pembelajaran untuk mendapatkan suatu model yang paling memenuhi kebutuhan antara himpunan atribut dan label kelas dalam data masukan. Data masukan dari model klasifikasi adalah sekumpulan *record (training set)*. Tiap *record* adalah himpunan dari *attributes* yang merupakan *class*. Model dalam atribut kelas yaitu suatu fungsi dari nilai atribut lainnya. Suatu data *test* digunakan untuk menentukan keakuratan model yang biasanya dibagi menjadi Data *training* dan Data *test*, dimana Data *training* digunakan untuk membangun model dan Data *Test* adalah untuk memvalidasi [2]. Berikut ini algoritma klasifikasi yang akan digunakan dalam penelitian ini adalah sebagai berikut:

1. Naive Bayes Classifier

Naive Bayes Classifier merupakan pengklasifikasi probabilitas sederhana berdasarkan pada teorema Bayes. Teorema Bayes dikombinasikan dengan “Naive” yang berarti setiap atribut/variabel bersifat bebas (*independent*). Naive Bayes Classifier dapat dilatih dengan efisien dalam pembelajaran terawasi (*supervised learning*). Keuntungan dari klasifikasi adalah bahwa ia hanya membutuhkan sejumlah kecil data pelatihan untuk memperkirakan parameter (sarana dan varians dari variabel) yang diperlukan untuk klasifikasi. Karena variabel independen diasumsikan,

hanya variasi dari variabel untuk masing-masing kelas harus ditentukan, bukan seluruh matriks kovarians [3].

2. Naive Bayes Classifier

Naive Bayes Classifier merupakan pengklasifikasi probabilitas sederhana berdasarkan pada teorema Bayes. Teorema Bayes dikombinasikan dengan “Naive” yang berarti setiap atribut/variabel bersifat bebas (*independent*). Naive Bayes Classifier dapat dilatih dengan efisien dalam pembelajaran terawasi (*supervised learning*). Keuntungan dari klasifikasi adalah bahwa ia hanya membutuhkan sejumlah kecil data pelatihan untuk memperkirakan parameter (sarana dan varians dari variabel) yang diperlukan untuk klasifikasi. Karena variabel independen diasumsikan, hanya variasi dari variabel untuk masing-masing kelas harus ditentukan, bukan seluruh matriks kovarians [4].

3. Random forest

Random forest merupakan metode yang populer digunakan dalam klasifikasi dan clustering berdasarkan ensemble dari decision tree (pohon keputusan). Keputusan klasifikasi diambil dari suara terbanyak diantara semua pohon. Random forest ini merupakan pengembangan dari metode CART, yaitu dengan menerapkan metode bagging dan random feature selection [5].

4. Decision Tree

Decision tree Merupakan algoritma yang digunakan untuk proses klasifikasi maupun prediksi melalui tahap pembentukan pohon keputusan. Pohon keputusan digunakan untuk prosedur penalaran untuk mendapatkan jawaban dari sebuah masalah [6]. Sedangkan menurut [7] Decision tree merupakan metode klasifikasi dan prediksi yang sangat kuat dan terkenal. Metode decision tree mengubah fakta yang sangat besar menjadi pohon keputusan yang mempresentasikan aturan.

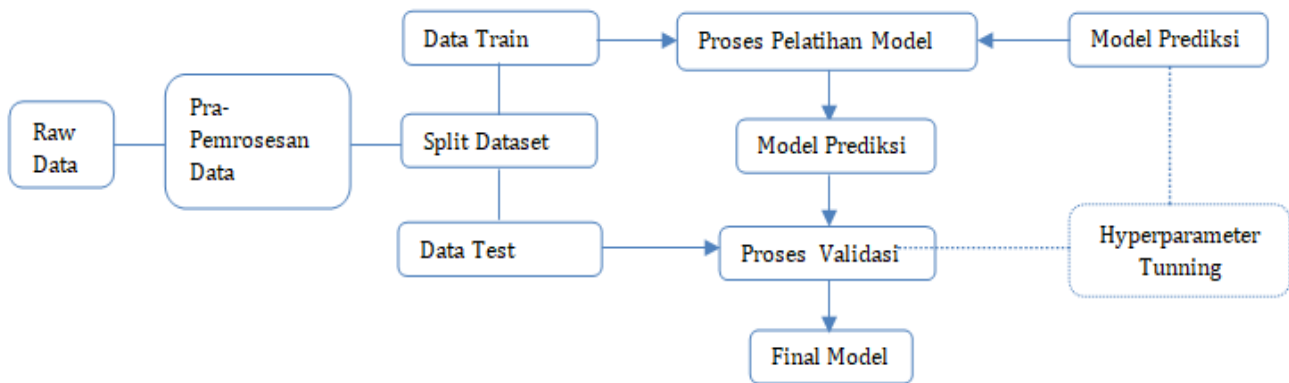
5. K-Nearest Neighbor (KNN)

Metode K-Nearest Neighbor (KNN) merupakan metode klasifikasi terhadap obyek berdasarkan data pembelajaran dengan jarak yang terdekat. Metode KNN diterapkan untuk mengatasi permasalahan identifikasi yang diukur secara kualitatif maupun kuantitatif. Sebelum mencari jarak data ke tetangga adalah menentukan nilai K tetangga (*neighbor*) [8].

Berikut merupakan metodologi dari rancangan penelitian;

A. Rancangan Penelitian

Dalam rancangan penelitian ini, penulis mengusulkan perbandingan metode klasifikasi dalam dataset Rice Leaf Diseases, adapun metodologi yang digunakan dalam penelitian ini yaitu:



Gambar 1. Alur Metodologi

Dari gambar 1 dapat di jelaskan mengenai alur dari penelitian ini yaitu dengan melakukan preprocessing dari data mentah agar data dapat digunakan oleh model data mining setelah itu dibagi menjadi 2 yaitu data train dan data test. Data train adalah data pelatihan yang digunakan untuk membuat model kemudian data test adalah data yang digunakan untuk memvalidasi model apakah sudah baik hasil prediksinya. Setelah itu masuk ke dalam proses pelatihan model dimana disana data akan dilatih dengan mengetahui atau mengenali bagaimana pola dari data tersebut. Kemudian data tersebut di prediksi dengan model yang di terapkan penulis, setelah itu keluar hasil model prediksi maka akan di lakukan proses validasi dengan data test yang sudah di split dari awal. Setelah itu dilakukan pengukuran apakah hasilnya sudah bagus atau belum. Jika belum maka diatur kembali dengan parameter sampai mendapatkan performa yang baik.

B. Sumber Data dan Teknik Pengumpulan Data

Peneliti menggunakan sumber daya yang telah jadi yaitu dari UCI *Machine Learning Repository* dengan dataset Rice Leaf Diseases dengan tiga kelas yaitu *Bacterial leaf blight*, *Brown spot* dan *Leaf smut*.

C. Analisis Data

Untuk mencapai tujuan dari analisis data ini, peneliti menggunakan analisis dengan data kualitatif dimana merupakan suatu analisa yang digunakan apabila angka-angka dan perhitungan menggunakan rumus yang ada hubungannya dengan penulisan ini. Dalam hal ini penulis menggunakan kelima metode yaitu dengan Decision Tree, Random Forest, Naïve Bayes, SVM dan KNN.

RESULTS AND DISCUSSION

Pada tahap ini penulis akan menjelaskan mengenai pembahasan sampai ke hasil dari penelitian. Penelitian ini mengambil data dari UCI *Machine Learning Repository* dengan dataset Rice Leaf Diseases yang akan diuji dengan lima algoritma yaitu *Decision Tree*, *Random Forest*, *Naïve Bayes*, *SVM* dan *KNN*. Hasil dan pembahasan data ini yaitu sebagai berikut:

Data yang digunakan oleh penulis berupa data gambar, maka penulis perlu melakukan Feature Exreaxtion. Featur Extraxtion merupakan suatu “encoding” dari sebuah image menjadi features yang berupa angka-angka yang

merepresentasikan image tersebut. Pada kasus ini feature extraxtion menggunakan model resnet 50 atau biasa disebut pretrained model. Pretrained model merupakan model yang sudah di training sebelumnya dan kita tinggal menggunakan saja. Kenapa penulis tidak melakukan training sendiri karena mengindari error pada saat training lebih baik menggunakan data yag sudah ada untuk menghemat waktu dan bisa diatur sesuka hati.



Gambar 2. Feature Extractor (Resnet50)

Untuk klasifikasi penulis menggunakan machine learning models data yang dihasilkan atau yang diklasifikasikan harus berupa data diskrit untuk itu perlu dilakukan transformasi data gambar menjadi data tabular (angka). Dengan menggunakan resnet50 penulis mentransformasikan 120 gambar ke dalam 120 embedding yang setiap embeddingnya memiliki 2048 features.

Results on train data

	Model	Accuracy	F1 Score	MCC	Training Time (s)	Predict Time (s)
0	RandomForestClassifier	1.000000	1.000000	1.000000	0.226920	0.009128
1	DecisionTreeClassifier	1.000000	1.000000	1.000000	0.060679	0.000475
2	KNeighborsClassifier	0.843750	0.841354	0.765465	0.007944	0.040643
3	GaussianNB	1.000000	1.000000	1.000000	0.002769	0.005221
4	SVC	0.354167	0.174359	0.000000	0.042485	0.030680

Gambar 3. Result on Train Data

CONCLUSIONS

Dari hasil penelitian yang dibahas mengenai *machine learning clasification* penyakit tanaman, maka penulis dapat menyimpulkan dari hasil penelitian telah diketahui bahwasannya penulis memperoleh 3 jenis hasil; *Overfit Models* yaitu algoritma *Random Forest*, *Decission Tree* dan *Naive Bayes*, *Underfit Models* yaitu algoritma *SVM* dan *Good Models* yaitu algoritma *KNN*. Jadi metode terbaik diantara kelima tersebut yaitu metode *KNN* dengan nilai akurasi 87%, karna model ini konsisten baik pada kedua evaluasi. *KNN* tidak terbukti memiliki masalah *overfitting* karena secura konsisten berkinerja baik pada data *train* dan data *test*. Hal ini memiliki generalisasi lebih dari model lain yang baik mengingat data gambar yang digunakan dalam data produksi mungkin akan berbeda dengan apa yang dimiliki sebagai data train. Model ini berfokus pada

fitur umum untuk memprediksi penyakit daun dari pada merujuk sepenuhnya pada fitur terperinci yang dimiliki oleh data *train*.

REFERENCES

- [1] Fatkhuroji, Santosa, S., & Pramunendar, R. A. (2019). *Prediksi Harga Kedelai Lokal Dan Kedelai Impor Dengan Metode Support Vector Machine Berbasis*. 15, 61–77.
- [2] Hermanto, H., Mustopa, A., & Kuntoro, A. Y. (2020). Algoritma Klasifikasi Naive Bayes Dan Support Vector Machine Dalam Layanan Komplain Mahasiswa. *JITK (Jurnal Ilmu Pengetahuan Dan Teknologi Komputer)*, 5(2), 211–220. <https://doi.org/10.33480/jitk.v5i2.1181>
- [3] Lutfi, M. (2018). *Klasifikasi Kualitas Mutu Jagung dengan Menggunakan Metode Decision Tree Pada Dinas Pertanian Bojonegoro*. 8(1).
- [4] Muzakir, A., & Wulandari, R. A. (2016). Model Data Mining sebagai Prediksi Penyakit Hipertensi Kehamilan dengan Teknik Decision Tree. *Scientific Journal of Informatics*, 3(1), 19–26. <https://doi.org/10.15294/sji.v3i1.4610>
- [5] Sari, S. A. N. P., Ariyanto, Y., & Astuti, E. S. (2016). SISTEM PAKAR DIAGNOSA PENYAKIT PADA TANAMAN PADI MENGGUNAKAN METODE DEMPSTER SHAFER. (*JUSTIN*) *Jurnal Sistem Dan Teknologi Informasi*, 1(1), 1–6. <https://doi.org/10.21776/ub.jeest.2015.002.02.2>
- [6] Syarief, M., Mukminin, A., Prastiti, N., & Setiawan, W. (2017). Penerapan Metode Naïve Bayes Classifier Untuk. *Nero*, 3(1), 61–68.
- [7] Triscowati, D. W., Sartono, B., Kurnia, A., Domiri, D. D., & Wijayanto, A. W. (2019). *Classification of Rice Plant Phase Using Supervised Random Forest Based On Klasifikasi Fase Tanam Padi Menggunakan Supervised Random Forest Pada Data Multitemporal Citra Landsat-8 Classification of Rice Plant Phase Using Supervised Random Forest Based On*. (July).
- [8] Yhurinda, A., Putri, P., Sodik, A., Adhi, I. T., & Suarabaya, T. (2019). Identifikasi Penyakit Tanaman Kopi Arabika dengan Metode K-Nearest Neighbor (K-NN). *Seminar Nasional Sains Dan Teknologi Terapan VII*, 759–764.