

Implementasi Korelasi untuk Seleksi Fitur pada Klasifikasi Jamur Beracun Menggunakan Jaringan Syaraf Tiruan

Arief Hermawan¹, Adityo Permana Wibowo^{2*}

¹Program Studi Magister Teknologi Informasi

²Program Studi Sistem Informasi Program Diploma Tiga

^{1,2}Universitas Teknologi Yogyakarta, Yogyakarta, 55285, Indonesia

ariefdb@uty.ac.id, adityopw@uty.ac.id*

Abstrak

Jamur merupakan salah satu jenis flora yang tumbuh subur di negara tropis, salah satunya Indonesia. Jamur terdiri atas 2 jenis, yaitu jamur yang bisa dimakan (menyehatkan) dan jamur yang tidak bisa dimakan (jamur beracun). Tujuan yang ingin dicapai dalam penelitian ini adalah meningkatkan keakuratan klasifikasi jamur dengan pendekatan analisis korelasi. Peningkatannya akurasi dilakukan dengan menghilangkan atribut yang berkorelasi di atas 0,8. Dataset yang digunakan diambil dari Kaggle.com yang terdiri atas 22 variabel input, dan 2 variabel output. Setelah dilakukan simulasi dengan menggunakan jaringan syaraf tiruan perambatan galat mundur, diperoleh hasil dari 22 variabel yang dipergunakan untuk mengklasifikasi jamur beracun atau tidak beracun, ditemukan ada 2 variabel yang secara bersama-sama tidak berkontribusi terhadap hasil klasifikasi dan cenderung menjadi variabel pengganggu. Dua variabel tersebut adalah variabel Bruises dan Ringtype. Dengan dihilangkannya 2 variabel tersebut, terjadi peningkatan keakuratan klasifikasi dari 97,97%, menjadi 99,02%.

Kata kunci: Klasifikasi, Korelasi atribut, Jaringan syaraf tiruan.

Abstract

Mushrooms are one type of flora that thrives in tropical countries, such as Indonesia. Mushrooms consist of 2 types, namely edible mushrooms (healthy) and inedible mushrooms (poisonous mushrooms). The aim of this research is to increase the accuracy of classification performance by using a correlation analysis approach. Improvement accuracy is done by removing the correlated attributes above 0.8. The dataset used is taken from Kaggle.com which consists of 22 input variables and 2 output variables. After performing a simulation using backpropagation neural network, the results obtained that from 22 variables used to classify poisonous or non-toxic mushrooms, it was found that there were 2 variables which together did not contribute to the classification results and tended to be confounding variables. The two variables are Bruises and Ringtype variables. By eliminating these 2 variables, there was an increase in the accuracy of the classification from 97,97% to 99,02%.

Keywords: Classification, Attribute Correlation, Neural Network.

1. PENDAHULUAN

Indonesia merupakan salah satu negara beriklim tropis, salah satu tanaman yang mudah tumbuh di negara beriklim tropis adalah jamur. Saat ini lebih dari 200.00 spesies jamur tumbuh di Indonesia (Annisia et al., 2017). Beberapa spesies jamur diketahui dapat dimakan bahkan mengandung zat yang bisa digunakan sebagai obat-obatan (Putra, 2020b), (Permana Putra, 2020). Namun banyak juga jamur yang tidak dapat dimakan karena mengandung zat beracun (Putra, 2020a), (Hidayati & Ika, 2016).

Beberapa penelitian yang melakukan klasifikasi jamur menggunakan beberapa metode antara lain, aplikasi android untuk klasifikasi jamur menggunakan Metode KNN (Darsilowati et al., 2021). Selain itu ada juga penelitian yang melakukan klasifikasi jamur berdasarkan pengolahan citra genusnya menggunakan metode CNN. Penelitian tersebut menghasilkan akurasi sebesar 62% dengan menggunakan optimizer adam dan epoch 100, perbandingan data train validation 80:20, ukuran kernel 3x3 dan learning rule 0,001 (Putri, 2020). Penelitian lain yang menggunakan CNN juga pernah dilakukan

dengan modifikasi arsitektur VGG 16 dengan 10 layer yang terdiri dari 7 layer convolution dan 3 layer fully connected, menghasilkan nilai akurasi sebesar 81% (Tanuwijaya & Roseanne, 2021). Penelitian yang menggunakan pengujian try-error pada arsitektur Jaringan Syaraf Tiruan juga pernah dilakukan. Penelitian tersebut menggunakan 5 model, dan menghasilkan nilai akurasi terbaik pada model 4-8-5-1 dengan nilai akurasi 71% (Hayadi et al., 2021). Kemudian penelitian lain yang menggunakan pengolahan citra untuk mendeteksi jamur juga pernah dilakukan menggunakan Backpropagation Neural Network. Penelitian tersebut mengolah jamur berdasarkan segmentasi warna, deteksi tepi, dan kontur, dan menghasilkan akurasi sebesar 80% (Enggari et al., 2022). Penelitian yang melakukan klasifikasi jamur beracun dan tidak beracun menggunakan Jaringan Syaraf Tiruan Backpropagation dengan hasil terbaik menghasilkan akurasi sebesar 99.25% (Alkronz et al., 2019). Penelitian tersebut menggunakan 22 variabel input.

Tujuan yang ingin dicapai dalam penelitian ini adalah mengimplementasikan korelasi untuk seleksi fitur pada klasifikasi jamur.

Korelasi adalah salah satu metode seleksi fitur dengan melihat kedekatan antar variabel, semakin tinggi nilai korelasi antar variabel, maka semakin identik variabel tersebut. Nilai korelasi memiliki rentang -1 sd 1, semakin mendekati -1 atau 1 maka variabel tersebut semakin identik, Jika nilainya positif maka keidentikannya searah jika negatif maka berlawanan arah berlawanan arah (Puteri et al., 2021). Korelasi telah berhasil digunakan sebagai salah satu metode seleksi fitur. Sehingga dengan adanya seleksi fitur ini diharapkan terjadi kenaikan akurasi dengan beban komputasi yang lebih ringan (Blessie & Karthikeyan, 2012).

2. METODE

Untuk mencapai tujuan penelitian yang telah ditetapkan, dilakukan Langkah-langkah sebagai berikut:

2.2. Dataset

Data diambil dari situs <https://www.kaggle.com/uciml/mushroom-classification>. Data tersebut terdiri dari 22 variabel input yang merupakan ciri-ciri jamur

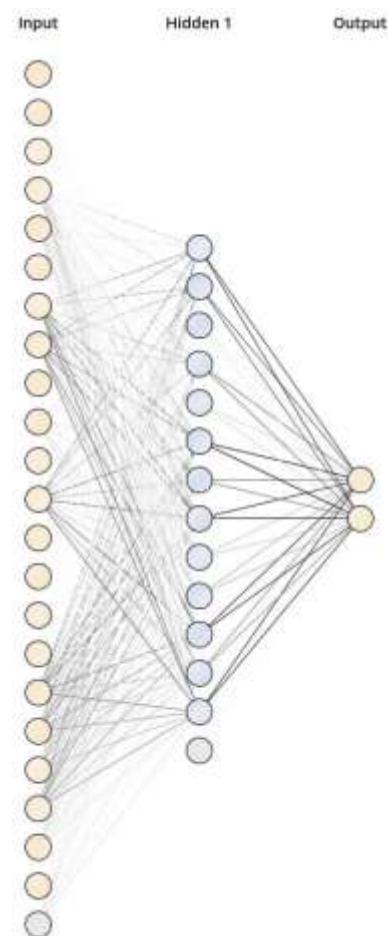
dan 2 variabel output yang merupakan kelas jenis jamur beracun atau tidak beracun.

2.3. Korelasi

Setelah data didapatkan dilakukan analisis korelasi terhadap variabel input. Analisis korelasi untuk bertujuan untuk melihat kedekatan antar variabel, variabel yang berkorelasi di atas 0.8 dianggap memiliki kedekatan yang tinggi. Variabel-variabel yang memiliki korelasi yang tinggi tersebut dapat diduga menjadi variabel yang dapat dihilangkan (diseleksi) untuk tidak digunakan pada proses klasifikasi.

2.4. Klasifikasi

Melakukan klasifikasi dengan menggunakan jaringan saraf tiruan perambatan galat mundur. Arsitektur jaringan saraf yang digunakan diperlihatkan pada gambar 1. Proses pelatihan-pengujian dilakukan beberapa kali dengan menghilangkan variabel yang memiliki nilai korelasi lebih dari 0.8. Selain itu juga dilakukan proses kombinasi 2 variabel yang berkorelasi lebih dari 0.8.



Gambar 1. Arsitektur JST

Arsitektur jaringan saraf tiruan pada hambar 1 terdiri atas 3 lapisan, lapisan input terdiri atas 22 sel, lapisan tersembunyi terdiri atas 13 sel dan lapisan output terdiri atas 2 sel. Jaringan dilatih dengan parameter 0,01, jumlah iterasi 200, momentum 0,9. Pelatihan dan pengujian menggunakan metode *split validation*, 80% data digunakan untuk pelatihan dan 20% data dilakukan untuk pengujian.

3. Hasil dan Pembahasan

Dataset yang digunakan dalam penelitian ini terdiri atas 22 variabel input dan 2 variabel output. Dua puluh dua variabel input tersebut adalah *lap-0.4hape*, *cap-surface*, *cap-color*, *bruises*, *odor*, *gill-attachment*, *gill-spacing*, *gill-size*, *gill-color*, *stalk-shape*, *stalk-root*, *stalk-surface-above-ring*, *stalk-surface-below-ring*, *stalk-color-above-ring*, *stalk-color-below-ring*, *veil-type*, *veil-color*, *ring-number*, *ring-type*, *spore-print-color*, *population*, *habitat*. Variabel-variabel tersebut akan menentukan 1 variabel output, yaitu sifat jamur tersebut termasuk ke dalam kelas jamur beracun atau jamur tidak beracun.

Setelah data diperoleh dilakukan analisis korelasi. Korelasi antar variabel dihitung, variabel-variabel yang memiliki korelasi lebih dari 0,8 dipisahkan. Pada Tabel 1, ditunjukkan variabel-variabel yang memiliki nilai korelasi lebih besar dari 0,8.

Tabel 1. Hasil Korelasi Atribut lebih dari 0.8

Variabel 1	Variabel 2	Nilai Korelasi
Stalk-Surface-Above-Ring	Ringtype	0,909
Stalk-Surface-Above-Ring	Stalk-Surface-Below-Ring	0,895
Stalk-Surface-Below-Ring	Ringtype	0,878
Bruises	Ringtype	0,830

Berdasarkan tabel 1 terlihat bahwa ada 4 variabel yang memiliki korelasi di atas 0.8, yaitu Bruises, Ringtype, Stalk-Surface-Above-Ring, dan Stalk-Surface-Below-Ring. Korelasi terbesar terjadi antara variabel Stalk-Surface-Above-Ring dan Ringtype, dan korelasi terendah terjadi

antara variabel Bruises dan Ringtype. Variabel Ringtype adalah variabel yang banyak berkorelasi dengan variabel lainnya. Variabel-variabel tersebut akan dihilangkan dan dilihat pengaruhnya terhadap unjuk kerja ketepatan klasifikasinya.

Pada Tabel 2 ditunjukkan hasil akurasi klasifikasi dengan menggunakan berbagai kombinasi input dengan menghilangkan 1 variabel yang memiliki korelasi lebih dari 0.8. Hasil klasifikasi tertinggi terjadi ketika pelatihan dilakukan tanpa variabel Bruises dan Stalk-Surface-Above-Ring, yaitu 97,72%. Berdasarkan hasil tersebut maka variabel Above yang memiliki korelasi tertinggi dengan variabel Ringtype menjadi variabel yang memang layak untuk dihilangkan, sedangkan variabel Ringtype menjadi variabel yang tetap dipertahankan. Variabel Bruises juga menjadi variabel yang layak untuk dihilangkan karena juga menghasilkan nilai akurasi tertinggi jika variabel tersebut dihilangkan.

Pada Tabel 3 diperlihatkan akurasi klasifikasi dengan menghilangkan 2 variabel yang memiliki korelasi terbesar. Pada penghilangan 2 variabel ini di dapat peningkatan yang cukup signifikansi yaitu pada penghilangan variabel Bruises dan Ringtype, diperoleh hasil klasifikasi 99,02%. Hasil ini menunjukkan bahwa 2 variabel tersebut merupakan variabel yang tidak berkontribusi terhadap variabel output dan cenderung menjadi variabel pengganggu

Berdasarkan 2 perlakuan tersebut terlihat bahwa penghilangan 2 variabel secara bersamaan menghasilkan akurasi lebih baik dibandingkan hanya menghilangkan 1 variabel. Dua variabel tersebut adalah Bruises dan Ringtype. Jika dilihat dari hasil analisis korelasinya, terlihat bahwa Ringtype memiliki korelasi yang paling banyak, yaitu dengan 3 variabel lainnya. Hal ini juga menjadi indikator bahwa informasi yang ada dalam Ringtype sudah terepresentasikan di 3 variabel lainnya, sehingga memang layak untuk dihilangkan.

Tabel 2. Hasil Akurasi Klasifikasi Berdasarkan Atribut

Atribut yang digunakan	Full	Tanpa Bruises	Tanpa Ringtype	Tanpa Stalk-Surface-Above-Ring	Tanpa Stalk-Surface-Below-Ring	Tanpa 4 atribut
Split data	80:20	80:20	80:20	80:20	80:20	80:20
Epoch	200	200	200	200	200	200
LR	0.01	0.01	0.01	0.01	0.01	0.01
momentum	0.9	0.9	0.9	0.9	0.9	0.9
Akurasi	97,97	97,72	97,23	97,72	97,05	97,72

Tabel 3. Perbandingan akurasi berdasarkan pengurangan variabel

Variabel 1	Variabel 2	Akurasi
Tanpa Stalk-Surface-Above-Ring	Tanpa Stalk-Surface-Below-Ring	97,23
Tanpa Bruises	Tanpa Tanpa Stalk-Surface-Above-Ring	97,72
Tanpa Bruises	Tanpa Stalk-Surface-Above-Ring	97,97
Tanpa Bruises	Tanpa Ringtype	99,02
Tanpa Stalk-Surface-Above-Ring	Tanpa Ringtype	97,23
Tanpa Tanpa Stalk-Surface-Below-Ring	Tanpa Ringtype	97,23

4. KESIMPULAN

Berdasarkan hasil dan pembahasan dalam penelitian ini dapat disimpulkan bahwa dari 22 variabel yang dipergunakan untuk mengklasifikasi jamur beracun atau tidak ditemukan ada 2 variabel yang secara bersama-sama tidak berkontribusi terhadap hasil klasifikasi dan cenderung menjadi variabel pengganggu. Dua variabel tersebut adalah variabel Bruises dan Ringtype. Dengan dihilangkannya 2 variabel tersebut, terjadi peningkatan keakuratan klasifikasi dari 97,97%, menjadi 99,02%.

DAFTAR PUSTAKA

- Alkronz, E. S., Moghayer, K. A., Meimeh, M., Gazzaz, M., Abu-Nasser, B. S., & Abu-Naser, S. S. (2019). Prediction of Whether Mushroom is Edible or Poisonous Using Back-propagation Neural Network. *International Journal of Academic and Applied Research*, 3(2), 1–8. <http://www.ijeais.org/ijaar>
- Annissa, I., Ekamawanti, H. A., & Wahdina. (2017). Keanekaragaman Jenis Jamur Makrokopis Di Arboretum Sylva Universitas Tanjungpura. *Jurnal Hutan Lestari*, 5(4), 969–977.
- Blessie, E. C., & Karthikeyan, E. (2012). Sigmis: A feature selection algorithm using correlation based method. *Journal of Algorithms and Computational Technology*, 6(3), 385–394. <https://doi.org/10.1260/1748-3018.6.3.385>
- Darsilowati, D., Riadi, A. A., & Evanita, E. (2021). Klasifikasi Jenis Jamur Konsumsi Berbasis Android Menggunakan Metode K-Nearest Neighbors (KNN). *Journal of Applied Science and Technology*, 1(02), 23–29. <https://doi.org/10.30659/jast.1.02.22-29>
- Enggari, S., Ramadhanu, A., & Marfalino, H. (2022). Peningkatan Digital Image Processing Dalam Mendeskripsikan Tumbuhan Jamur Dengan Segmentasi Warna, Deteksi Tepi Dan Kontur. *Jurnal Teknologi Dan Sistem Informasi Bisnis*, 4(1), 70–75. <https://doi.org/10.47233/jteksis.v4i1.358>
- Hayadi, B. H., Sudipa, I. G. I., & Windarto, A. P. (2021). Model Peramalan Artificial Neural Network pada Peserta KB Aktif Jalur Pemerintahan menggunakan Artificial Neural Network Back-Propagation. *MATRIK : Jurnal Manajemen, Teknik Informatika Dan Rekayasa Komputer*, 21(1), 11–20. <https://doi.org/10.30812/matrik.v21i1.1273>

- Hidayati, & Ika, P. (2016). Mikrobiologi Dasar. In *Universitas Kanjuruhan Malang*. repository.unikama.ac.id/656/1/BUKU AJAR MIKROBIOLOGI.pdf
- Permana Putra, I. (2020). Studi Taksonomi dan Potensi Beberapa Jamur Liar di Pulau Belitung. *Justek : Jurnal Sains Dan Teknologi*, 3(1), 24. <https://doi.org/10.31764/justek.v3i1.3534>
- Puteri, A. N., Arizal, A., & Achmad, A. D. (2021). Feature Selection Correlation-Based pada Prediksi Nasabah Bank Telemarketing untuk Deposito. *MATRIK : Jurnal Manajemen, Teknik Informatika Dan Rekayasa Komputer*, 20(2), 335–342. <https://doi.org/10.30812/matrik.v20i2.1183>
- Putra, I. P. (2020a). Kasus keracunan *Inocybe* sp. di Indonesia. *Prosiding Seminar Nasional Biologi, September*, 148–153.
- Putra, I. P. (2020b). Note on Macro Fungi on Belitung Island: Description and Potential. *BIOEDUSCIENCE: Jurnal Pendidikan Biologi Dan Sains*, 4(1), 11–20. <https://doi.org/10.29405/j.bes/4111-204416>
- Putri, O. N. (2020). *Implementasi Metode CNN Dalam Klasifikasi Gambar Jamur Pada Analisis Image Processing (Studi Kasus: Gambar Jamur Dengan Genus Agaricus Dan Amanita)*.
- Sibero, M., Putra, I., & Murwani, R. (2021). Deskripsi dan Potensi Jamur Makro Asal Hutan Adat Panembahen, Desa Juhar, Kabupaten Tanah Karo, Sumatera Utara. *Jurnal Mikrobiologi Indonesia*, 5(1), 48–54. <https://doi.org/10.46638/jmi.v5i1.164>. Abstrak
- Tanuwijaya, E., & Roseanne, A. (2021). Modifikasi Arsitektur VGG16 untuk Klasifikasi Citra Digital Rempah-Rempah Indonesia. *MATRIK : Jurnal Manajemen, Teknik Informatika Dan Rekayasa Komputer*, 21(1), 189–196. <https://doi.org/10.30812/matrik.v21i1.1492>