

Klasifikasi Jenis Jamur Berdasarkan Citra Gambar Menggunakan Algoritma Stochastic Gradient Descent

Dedy Armiady^{1)*}

¹⁾ Program Studi Sistem Informasi, Universitas Almuslim Bireuen

Received: 28 November 2024

Accepted: 17 December 2024

Published: 25 December 2024



*dedy.armiady@gmail.com

Kata Kunci: Stochastic Gradient Descent, Klasifikasi Citra, Jamur, Elastic Net, Ketidakseimbangan Data.

DSI: Jurnal Data Science Indonesia is licensed under a Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0).

Abstrak : Penelitian ini mengkaji penerapan Stochastic Gradient Descent (SGD) untuk optimasi klasifikasi gambar jamur berdasarkan citra, menggunakan dataset tidak seimbang dengan 10 kelas jamur yang berbeda. Algoritma SGD digunakan karena efisiensinya dalam menangani dataset besar serta kemampuan untuk memperbarui parameter secara bertahap guna meminimalkan fungsi loss. Penyesuaian parameter pada SGD, seperti Squared Hinge Loss untuk klasifikasi, Elastic Net sebagai regularisasi, dan optimal learning rate, dilakukan untuk meningkatkan performa model. Hasil penelitian menunjukkan bahwa akurasi model mencapai 62.4%, yang lebih tinggi dibandingkan dengan beberapa algoritma lain seperti CNN, Logistic Regression, SVM, dan Random Forest yang telah diuji pada dataset yang sama. Meskipun demikian, nilai AUC sebesar 0.775 dan F1-score 0.625 menunjukkan bahwa performa model masih belum optimal, terutama dalam menangani ketidakseimbangan data. Penggunaan teknik penyeimbangan data seperti SMOTE direkomendasikan untuk penelitian lebih lanjut guna meningkatkan performa klasifikasi pada kelas yang minoritas. Dengan penyesuaian parameter yang tepat, SGD terbukti mampu bekerja lebih baik dibandingkan model lain dalam konteks klasifikasi citra jamur. Hasil ini memberikan kontribusi signifikan dalam bidang klasifikasi berbasis citra, khususnya pada aplikasi dengan dataset yang tidak seimbang.

PENDAHULUAN

Dalam beberapa dekade terakhir, teknologi pengolahan citra dan kecerdasan buatan telah berkembang pesat, memungkinkan aplikasi yang lebih luas dalam berbagai bidang [1], [2]. Salah satu aplikasi penting dari teknologi ini adalah dalam klasifikasi objek berdasarkan citra digital [3]. Salah satu objek yang menarik untuk diklasifikasikan adalah jamur, yang memiliki beragam spesies dan karakteristik morfologi yang sering kali sulit dibedakan dengan mata telanjang. Jamur memiliki peran penting dalam ekosistem serta dalam berbagai industri seperti pertanian, makanan, dan farmasi. Beberapa spesies jamur bahkan memiliki nilai komersial yang tinggi, sementara yang lain bisa berbahaya atau beracun. Identifikasi yang tepat terhadap spesies jamur menjadi krusial, baik untuk menghindari keracunan akibat jamur beracun maupun untuk memanfaatkan jamur yang memiliki nilai komersial. Namun, pengenalan manual jamur oleh manusia sering kali tidak akurat karena beberapa spesies jamur memiliki karakteristik visual yang sangat mirip. Untuk mengatasi masalah ini, teknologi kecerdasan buatan, khususnya machine learning, menawarkan solusi yang lebih efisien dan akurat.

Klasifikasi objek berbasis citra dengan menggunakan teknik machine learning telah terbukti mampu memberikan hasil yang lebih baik dibandingkan dengan metode konvensional [4]. Machine learning memungkinkan komputer untuk belajar dari data yang ada dan mengenali pola-pola yang sulit dideteksi oleh manusia [5]. Salah satu algoritma yang banyak digunakan dalam klasifikasi citra adalah Stochastic Gradient Descent (SGD) [6]. Algoritma ini populer karena kecepatan dan efisiensinya dalam menangani dataset yang besar [7]. Dalam penelitian ini, SGD digunakan dengan parameter khusus, yaitu Squared Hinge untuk klasifikasi, Squared Loss untuk regresi, regularisasi Elastic Net, dan learning rate optimal. Kombinasi parameter ini memungkinkan SGD untuk memberikan hasil yang lebih baik dalam tugas klasifikasi citra,

terutama pada dataset yang besar dan kompleks seperti klasifikasi jenis jamur.

Elastic Net sebagai metode regularisasi memadukan kelebihan dari L1 (Lasso) dan L2 (Ridge), sehingga dapat mengatasi masalah multikolinearitas dan mendorong seleksi fitur yang relevan [8]. Squared Hinge sebagai fungsi loss dalam klasifikasi membantu model dalam memaksimalkan margin antar kelas, mirip dengan apa yang dilakukan Support Vector Machines (SVM) [9]. Learning rate optimal memastikan bahwa model bergerak menuju konvergensi dengan cara yang efisien, mengurangi kemungkinan terjebak dalam minimum lokal atau gagal mencapai solusi yang optimal [10].

Namun, salah satu tantangan utama dalam penelitian klasifikasi citra jamur ini adalah ketidakseimbangan dataset [11]. Dalam banyak kasus, jumlah gambar yang tersedia untuk setiap spesies jamur tidak merata. Beberapa spesies memiliki lebih banyak data yang tersedia, sementara spesies lainnya memiliki lebih sedikit. Ketidakseimbangan data ini dapat menyebabkan model machine learning, termasuk SGD, menjadi bias terhadap kelas mayoritas dan menghasilkan prediksi yang kurang akurat untuk kelas minoritas [12]. Misalnya, jika model dilatih pada dataset yang tidak seimbang, ia mungkin cenderung lebih sering memprediksi kelas dengan jumlah gambar yang lebih banyak, mengabaikan kelas-kelas yang lebih jarang ditemukan dalam data [13].

Penelitian ini bertujuan untuk mengembangkan model klasifikasi jenis jamur berbasis citra dengan menggunakan algoritma Stochastic Gradient Descent. Model ini diharapkan mampu mengatasi tantangan dataset yang tidak seimbang dan menghasilkan akurasi yang baik dalam mengklasifikasikan berbagai jenis jamur. Melalui pengembangan model ini, diharapkan penelitian ini dapat memberikan kontribusi signifikan dalam bidang klasifikasi objek berbasis citra serta membantu meningkatkan efisiensi dan akurasi dalam pengenalan spesies jamur. Pada akhirnya, aplikasi dari model klasifikasi ini bisa diperluas ke berbagai sektor, termasuk kesehatan, pertanian, dan industri makanan, di mana identifikasi yang tepat terhadap jenis jamur menjadi sangat penting.

Teknologi machine learning memiliki potensi besar untuk digunakan dalam klasifikasi citra jamur, dan algoritma SGD dengan parameter Squared Hinge, Squared Loss, Elastic Net, dan learning rate optimal merupakan salah satu pendekatan yang menjanjikan. Meskipun ada tantangan terkait ketidakseimbangan data, kombinasi antara algoritma yang efisien dan teknik penanganan ketidakseimbangan dapat membantu mengatasi masalah tersebut. Penelitian ini berkontribusi dalam memberikan landasan bagi aplikasi lebih lanjut terkait pengklasifikasian jamur serta memberikan wawasan tentang penggunaan metode machine learning untuk menyelesaikan masalah-masalah klasifikasi yang kompleks.

TINJAUAN LITERATUR

Penerapan machine learning dalam pengolahan citra telah mengalami perkembangan pesat selama beberapa tahun terakhir, terutama dalam bidang pengenalan objek berbasis gambar, termasuk klasifikasi spesies jamur. Jamur memiliki karakteristik visual yang sering kali mirip antarspesies, sehingga identifikasi manual menjadi sulit dan membutuhkan waktu lama. Penelitian oleh [14] menunjukkan bahwa metode klasifikasi berbasis citra dapat membantu dalam mengidentifikasi spesies jamur secara akurat dengan memanfaatkan ciri-ciri visual seperti tekstur, warna, dan pola topologi. Metode ini memungkinkan otomatisasi yang lebih cepat dan efektif dalam proses identifikasi, yang menjadi tantangan utama dalam biologi dan agrikultur. Salah satu algoritma yang sering digunakan untuk klasifikasi citra adalah Stochastic Gradient Descent (SGD). SGD dikenal karena kemampuannya menangani dataset besar dengan efisien. Penelitian oleh [15] menunjukkan bahwa penggunaan SGD dalam klasifikasi citra jamur menghasilkan kinerja yang baik, terutama dalam hal kecepatan komputasi dan efisiensi waktu pelatihan. SGD bekerja dengan memperbarui parameter model secara bertahap untuk meminimalkan fungsi loss, yang cocok untuk dataset yang kompleks dan besar.

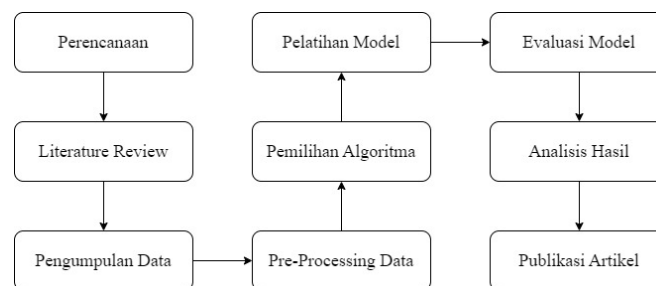
Namun, tantangan utama dalam klasifikasi citra jamur adalah ketidakseimbangan data. Ketidakseimbangan ini terjadi ketika jumlah gambar per kelas berbeda secara signifikan, yang dapat menyebabkan bias model terhadap kelas mayoritas. Penelitian oleh [16] menunjukkan bahwa ketidakseimbangan data dapat memengaruhi performa model klasifikasi seperti SGD, menyebabkan model cenderung lebih akurat pada kelas mayoritas, sementara performa pada kelas minoritas menurun. Untuk mengatasi masalah ini, pendekatan seperti oversampling kelas minoritas atau undersampling kelas mayoritas

sering kali digunakan, sebagaimana diperlihatkan oleh [17] yang menerapkan SMOTE (Synthetic Minority Over-sampling Technique) untuk meningkatkan performa SGD dalam tugas klasifikasi dataset tidak seimbang. Selain penanganan ketidakseimbangan data, parameter yang digunakan dalam SGD, seperti Squared Hinge Loss, sangat mempengaruhi performa klasifikasi citra. Penggunaan Squared Hinge Loss membantu meningkatkan stabilitas model, terutama dalam kasus margin kelas yang berbeda secara signifikan, sebagaimana ditunjukkan oleh penelitian [18]. Selain itu, penerapan Elastic Net sebagai teknik regularisasi yang memadukan L1 dan L2 membantu mengatasi masalah multikolinearitas, serta memfasilitasi seleksi fitur, yang membuatnya cocok untuk dataset citra yang kompleks. Learning rate juga berperan penting dalam performa model. Menurut penelitian [19], pemilihan learning rate yang tepat dapat mempercepat proses pelatihan model dan memastikan konvergensi yang lebih cepat menuju solusi optimal. Pemilihan learning rate yang optimal pada SGD terbukti meningkatkan kinerja model dalam tugas klasifikasi citra.

Secara keseluruhan, berbagai penelitian menunjukkan bahwa meskipun SGD merupakan algoritma yang efisien, keberhasilannya sangat bergantung pada penyesuaian parameter yang tepat serta penanganan ketidakseimbangan data. Dalam konteks klasifikasi jamur, penerapan SGD dengan Squared Hinge Loss, Elastic Net, dan learning rate optimal telah terbukti efektif dalam meningkatkan performa model, sebagaimana diperlihatkan dalam penelitian-penelitian sebelumnya

METODE PENELITIAN

Penelitian ini bertujuan untuk mengembangkan model klasifikasi jenis jamur berbasis citra gambar dengan memanfaatkan algoritma Stochastic Gradient Descent (SGD). Adapun alur penelitian dapat dilihat pada gambar berikut:



Gambar 1 Alur Penelitian

Proses penelitian dimulai dengan tahap perencanaan yang melibatkan penyusunan rencana penelitian secara menyeluruh. Tahap ini mencakup definisi masalah yang ingin dipecahkan, tujuan penelitian, serta ruang lingkup yang akan dicakup. Pada tahap ini, pemilihan algoritma SGD dan penentuan parameter yang tepat sangat penting untuk keberhasilan penelitian.

Setelah perencanaan, langkah berikutnya adalah pengumpulan data. Dataset yang digunakan dalam penelitian ini diperoleh dari Kaggle, yang merupakan salah satu platform penyedia dataset citra terbesar. Dataset ini terdiri dari gambar jamur yang dikelompokkan ke dalam sepuluh kelas berbeda, dengan rincian jumlah gambar untuk setiap kelas. Jumlah gambar untuk masing-masing kelas adalah sebagai berikut: Agaricus (353 gambar), Amanita (750 gambar), Boletus (1073 gambar), Cortinarius (836 gambar), Entoloma (364 gambar), Hygrocybe (316 gambar), Lactarius (1563 gambar), Russula (1148 gambar), dan Suillus (311 gambar). Dataset ini menawarkan tantangan tersendiri karena distribusi gambar antar kelas tidak seimbang, yang dapat mempengaruhi performa model klasifikasi.

Setelah data dikumpulkan, tahap pra-pemrosesan data menjadi langkah krusial sebelum model dapat dikembangkan. Pra-pemrosesan mencakup beberapa proses penting seperti pembersihan data, resizing gambar, normalisasi, dan augmentasi data. Pembersihan data melibatkan penghapusan gambar yang rusak atau tidak relevan untuk memastikan kualitas dataset. Resizing gambar dilakukan untuk mengubah ukuran semua gambar menjadi konsisten, misalnya 256x256 piksel, agar sesuai dengan input model. Normalisasi gambar penting untuk mengatur nilai piksel ke dalam rentang $[0, 1]$, yang dapat mempercepat proses pelatihan dan meningkatkan konvergensi model. Augmentasi data, seperti rotasi, pemotongan, dan flipping,

diterapkan untuk meningkatkan keberagaman dataset dan mengatasi masalah ketidakseimbangan data yang ada.

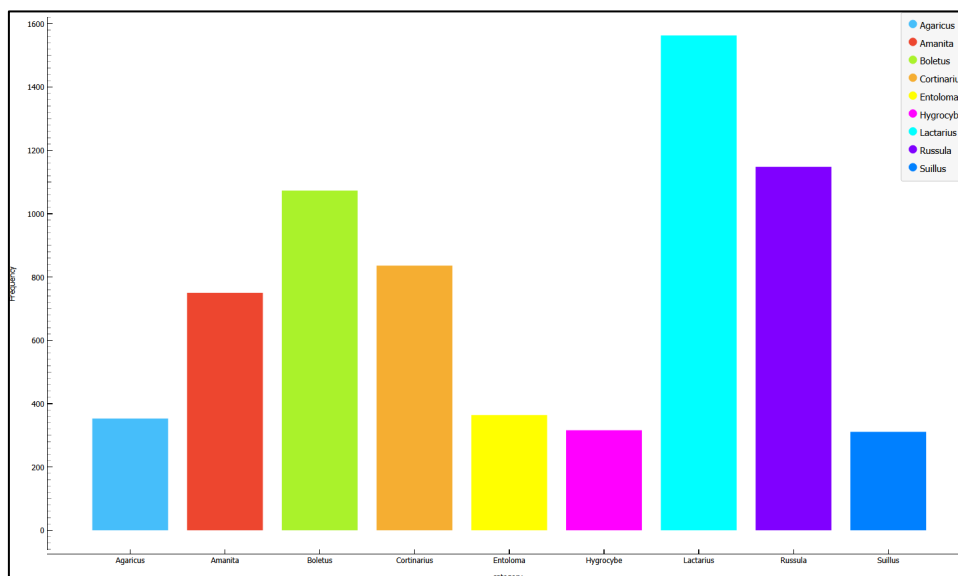
Dengan data yang telah dipra-pemrosesan, tahap berikutnya adalah pengembangan model. Pada penelitian ini, SGD digunakan dengan parameter yang telah ditentukan, yaitu Squared Hinge Loss untuk klasifikasi, Squared Loss untuk regresi, regularisasi Elastic Net, dan learning rate optimal. Pemilihan Squared Hinge Loss pada SGD bertujuan untuk memaksimalkan margin antara kelas-kelas yang berbeda, mirip dengan pendekatan yang digunakan dalam Support Vector Machines (SVM). Regularisasi Elastic Net memadukan kelebihan dari L1 (Lasso) dan L2 (Ridge) untuk menangani multikolinearitas dan seleksi fitur, sehingga meningkatkan kemampuan model dalam mengatasi kompleksitas dataset. Sementara itu, penggunaan learning rate optimal bertujuan untuk memastikan bahwa proses pelatihan model berjalan efisien dan konvergen ke solusi terbaik.

Pelatihan model dilakukan dengan membagi dataset menjadi dua subset: data pelatihan dan data pengujian. Data pelatihan digunakan untuk melatih model SGD, sementara data pengujian digunakan untuk mengevaluasi performa model. Proses pelatihan melibatkan iterasi berulang menggunakan teknik mini-batch SGD untuk memperbarui parameter model secara bertahap, yang memungkinkan model untuk belajar dari subset data yang lebih kecil dan mempercepat proses pelatihan.

Setelah model dilatih, tahap evaluasi dilakukan untuk menilai efektivitas dari model yang dikembangkan. Metrik evaluasi seperti akurasi, precision, recall, dan F1-score digunakan untuk mengukur performa model dalam mengklasifikasikan gambar jamur. Analisis hasil klasifikasi dilakukan untuk memahami bagaimana model menangani ketidakseimbangan kelas dan untuk mengidentifikasi area perbaikan.

Terakhir, laporan penelitian disusun untuk mendokumentasikan metodologi, hasil, dan diskusi. Laporan ini mencakup kesimpulan mengenai performa model dan memberikan rekomendasi untuk penelitian lebih lanjut. Dengan demikian, penelitian ini diharapkan dapat memberikan kontribusi signifikan dalam bidang klasifikasi citra jamur, serta menawarkan wawasan tentang penggunaan algoritma SGD dalam konteks klasifikasi gambar yang tidak seimbang.

Dataset yang digunakan dalam penelitian ini cenderung tidak seimbang, dimana dari 10 kelas yang digunakan, terdapat sebaran data yang mungkin dapat berdampak kepada performa model yang dibangun. Sebaran dataset yang digunakan dapat dilihat pada grafik berikut:



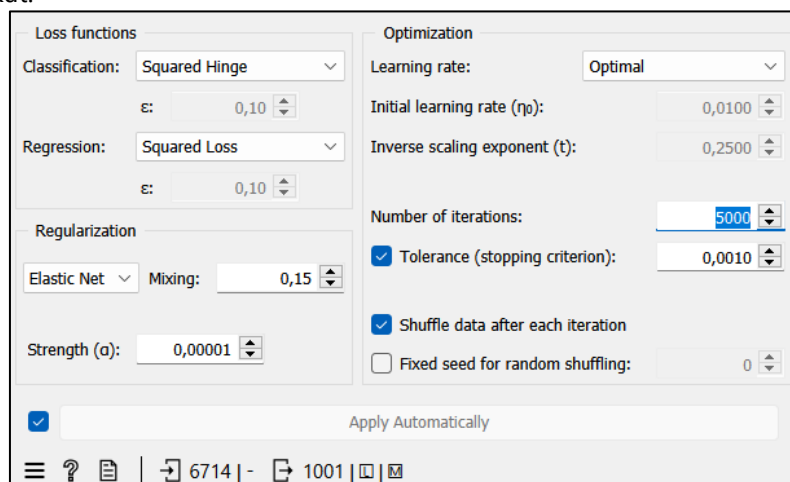
Gambar 2. Sebaran Dataset

Gambar 2 menunjukkan diagram batang yang menggambarkan distribusi frekuensi dari berbagai kategori jamur dalam dataset penelitian ini. Diagram ini menyoroti ketidakseimbangan data yang signifikan antara kategori-kategori jamur yang ada. Kategori Lactarius, yang diwakili oleh batang berwarna biru muda, memiliki frekuensi tertinggi dengan sekitar 1563 gambar. Jumlah data yang besar ini memberikan model

machine learning banyak informasi untuk belajar dan dapat meningkatkan akurasi model dalam mengidentifikasi Lactarius. Namun, jumlah data yang melimpah ini juga dapat berisiko menyebabkan model menjadi bias terhadap kategori ini, mengingat dominasi jumlah gambar yang sangat tinggi dibandingkan dengan kelas lainnya. Di sisi lain, Russula, yang ditunjukkan dengan warna ungu, memiliki frekuensi sekitar 1148 gambar. Meskipun jumlah ini signifikan, namun masih lebih rendah dibandingkan dengan Lactarius. Meskipun cukup data tersedia untuk belajar, tetap perlu memperhatikan potensi bias model yang mungkin lebih cenderung kepada kelas dengan frekuensi lebih tinggi. Kategori Boletus, yang diidentifikasi dengan warna hijau, memiliki frekuensi sekitar 1073 gambar. Jumlah ini menunjukkan bahwa Boletus adalah salah satu kategori dengan data yang cukup banyak, meskipun tidak sebesar Lactarius. Dengan jumlah data ini, model dapat belajar dengan baik, namun perhatian tetap diperlukan untuk menghindari ketidakseimbangan yang dapat mempengaruhi hasil klasifikasi. Kategori Cortinarius, berwarna orange, memiliki frekuensi sekitar 836 gambar. Jumlah data ini masih tergolong signifikan tetapi lebih rendah dibandingkan dengan Lactarius dan Russula. Meskipun data yang ada cukup untuk proses pelatihan, penting untuk memeriksa apakah model dapat mempelajari ciri khas kategori ini tanpa terpengaruh oleh ketidakseimbangan. Kategori Amanita (warna merah) memiliki frekuensi sekitar 750 gambar, sedangkan Suillus (warna biru tua) memiliki frekuensi sekitar 311 gambar. Suillus adalah kategori dengan jumlah data paling sedikit dalam dataset ini, menunjukkan bahwa ia merupakan kelas minoritas yang mungkin menghadapi tantangan dalam pelatihan model. Ketidakcukupan data dapat mempengaruhi kemampuan model dalam mempelajari representasi yang efektif dari Suillus. Kelas Entoloma, yang diwakili dengan warna kuning, memiliki frekuensi sekitar 364 gambar, juga termasuk dalam kategori minoritas. Jumlah data yang terbatas untuk Entoloma dapat menimbulkan tantangan serupa dengan Suillus dalam hal pelatihan model yang efektif.

Secara keseluruhan, distribusi frekuensi yang bervariasi ini menggarisbawahi tantangan utama dalam pelatihan model klasifikasi. Ketidakseimbangan jumlah data antara kategori-kategori ini dapat mempengaruhi hasil klasifikasi, dengan potensi bias terhadap kelas mayoritas dan kesulitan dalam mengidentifikasi kelas minoritas. Untuk mengatasi masalah ini, teknik seperti augmentasi data dan penyeimbangan kelas perlu dipertimbangkan untuk memastikan model yang adil dan akurat.

Adapun parameter daripada algoritma Stochastic Gradient Descent yang digunakan dalam penelitian ini adalah sebagai berikut:

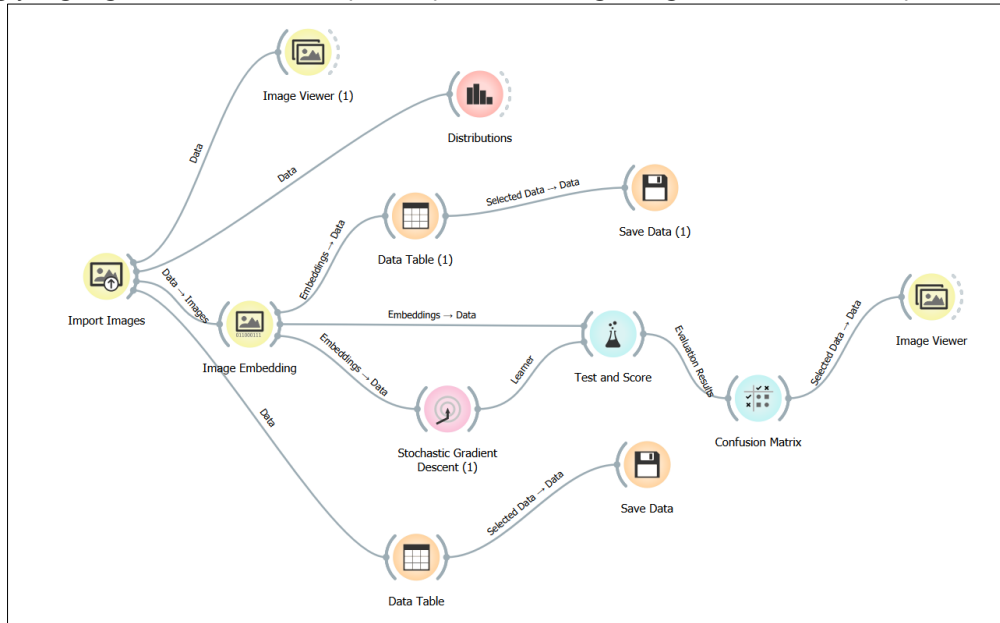


Gambar 3. Parameter Stochastic Gradient Descent

HASIL PENELITIAN

Dalam penelitian ini, tool Orange Data Mining digunakan untuk mengklasifikasikan citra jamur dengan memanfaatkan algoritma Stochastic Gradient Descent (SGD). Proses penelitian dimulai dengan mengimpor dataset citra jamur menggunakan widget Import Image, diikuti dengan penambahan widget Image Embedding menggunakan teknik ekstraksi fitur berbasis SqueezeNet (lokal). Fitur yang telah diekstraksi kemudian digunakan untuk melatih model SGD dengan parameter yang telah ditentukan. Evaluasi performa model dilakukan menggunakan widget Test & Score, yang menghasilkan perbandingan akurasi dan metrik

evaluasi lainnya untuk model yang dikembangkan. Berikut adalah gambar dari lingkungan kerja tool Orange Data Mining yang digunakan untuk memproses pelatihan dengan algoritma SGD dalam penelitian ini:



Gambar 4. Proses Klasifikasi

Adapun hasil klasifikasi dari dari proses training menggunakan dataset jamur dan algoritma SGD dengan parameter yang telah disebutkan sebelumnya dapat dilihat pada gambar berikut:

| Model | AUC | CA | F1 | Prec | Recall | MCC |
|-----------------------------|-------|-------|-------|-------|--------|-------|
| Stochastic Gradient Descent | 0.775 | 0.624 | 0.625 | 0.632 | 0.624 | 0.557 |

Gambar 6. Hasil Klasifikasi

Gambar 6 memberikan gambaran kinerja model berdasarkan beberapa metrik evaluasi yang relevan untuk masalah klasifikasi. Nilai AUC (Area Under the Curve) sebesar 0.775 menunjukkan bahwa model memiliki kemampuan yang cukup baik dalam membedakan antara kelas yang berbeda. AUC merupakan indikator penting dalam klasifikasi, terutama dalam kasus dataset yang tidak seimbang, karena AUC mengukur kemampuan model dalam membedakan antara kelas positif dan negatif di berbagai threshold. Nilai 0.775 termasuk kategori baik, namun bukan yang optimal. Ini berarti model sudah cukup mampu mengklasifikasikan data, tetapi masih ada beberapa kesalahan yang terjadi dalam penentuan kelas, terutama pada prediksi untuk kelas minoritas. Selanjutnya, nilai akurasi klasifikasi (CA) sebesar 0.624 atau 62.4% menunjukkan bahwa model dapat membuat prediksi yang benar pada sekitar 62% dari seluruh data uji. Namun, karena dataset Anda tidak seimbang (beberapa kelas memiliki jumlah sampel yang jauh lebih banyak dibandingkan kelas lainnya), akurasi bukanlah satu-satunya metrik yang memberikan gambaran keseluruhan. Pada dataset tidak seimbang, akurasi sering kali dapat menyesatkan karena model mungkin "terjebak" pada kelas yang dominan, sehingga mengabaikan performa pada kelas-kelas yang lebih jarang muncul. Dalam kasus seperti ini, nilai akurasi yang sedang masih dapat diterima, tetapi perlu dianalisis lebih lanjut dalam hubungannya dengan metrik lain seperti F1-Score, precision, dan recall. F1-Score sebesar 0.625 menunjukkan keseimbangan antara precision dan recall. F1-Score sangat penting dalam situasi di mana kita peduli baik dengan positive prediction value (precision) maupun kemampuan model untuk menemukan semua instance yang benar (recall). Nilai ini menunjukkan bahwa model mampu menangani prediksi positif dengan relatif baik, namun tetap tidak sempurna. Ini berarti bahwa model masih sering melakukan kesalahan, baik dalam hal false positives (ketika model salah mengklasifikasikan instance negatif sebagai positif) maupun false negatives (ketika model gagal mendeteksi instance positif). Keseimbangan antara precision dan recall adalah

aspek penting untuk klasifikasi dengan dataset tidak seimbang, dan model Anda masih memiliki ruang untuk peningkatan dalam hal ini. Precision sebesar 0.632 menunjukkan bahwa dari semua prediksi yang dihasilkan oleh model sebagai kelas positif, 63.2% dari prediksi tersebut benar. Precision penting ketika false positives berpotensi menimbulkan masalah, misalnya, jika prediksi positif yang salah memiliki konsekuensi serius. Dalam konteks penelitian Anda, nilai precision yang sedang ini menunjukkan bahwa model cukup baik dalam menghindari prediksi positif palsu, tetapi ada beberapa prediksi yang salah yang perlu diperbaiki untuk meningkatkan ketepatan klasifikasi. Recall yang berada di angka 0.624 menunjukkan bahwa model berhasil mengidentifikasi sekitar 62.4% dari semua instance yang benar-benar positif. Ini berarti model kehilangan sekitar 37.6% dari kelas yang seharusnya diklasifikasikan sebagai positif (false negatives). Recall sangat penting ketika false negatives lebih merugikan, terutama dalam kasus di mana kegagalan untuk mendeteksi kelas tertentu dapat menyebabkan konsekuensi besar. Dalam kasus ini, recall yang lebih tinggi mungkin diperlukan, tergantung pada tujuan spesifik klasifikasi jamur ini. Terakhir, Matthews Correlation Coefficient (MCC) sebesar 0.557 memberikan gambaran kualitas keseluruhan dari prediksi model. MCC merupakan metrik yang sangat berguna, terutama untuk dataset yang tidak seimbang, karena menghitung korelasi antara prediksi dan label sebenarnya dengan mempertimbangkan true positives, false negatives, true negatives, dan false positives. Nilai MCC di rentang 0.5 hingga 0.6 menunjukkan bahwa model memiliki kualitas prediksi yang moderat. Model masih cenderung membuat beberapa kesalahan, baik dalam mendeteksi kelas minoritas maupun dalam meminimalisir false positives dan false negatives, tetapi sudah cukup baik jika dibandingkan dengan model acak.

Secara keseluruhan, model Stochastic Gradient Descent yang digunakan telah menunjukkan performa yang layak dalam klasifikasi jamur, terutama dilihat dari AUC yang cukup tinggi dan MCC yang menunjukkan kualitas prediksi yang moderat. Namun, model ini masih perlu dioptimalkan, terutama dalam hal recall dan precision, yang akan sangat membantu dalam menangani dataset yang tidak seimbang.

PEMBAHASAN

Pada penelitian ini, algoritma Stochastic Gradient Descent (SGD) digunakan untuk klasifikasi gambar jamur ke dalam 10 kelas yang berbeda. Dataset yang digunakan memiliki karakteristik tidak seimbang, yang berarti bahwa distribusi jumlah gambar pada tiap kelas berbeda-beda secara signifikan. Hal ini menghadirkan tantangan tersendiri, karena model cenderung lebih akurat dalam mengenali kelas dengan jumlah data yang lebih besar. Namun, melalui pengaturan parameter yang tepat, model SGD mampu menghasilkan performa yang lebih unggul dibandingkan algoritma lainnya.

Hasil pengujian menunjukkan bahwa dengan penyesuaian parameter, akurasi yang dicapai menggunakan SGD mencapai 62.4% (CA = 0.624). Akurasi ini ternyata lebih tinggi dibandingkan dengan ketika menggunakan parameter default SGD, yang sebelumnya menghasilkan akurasi lebih rendah. Penyesuaian parameter ini mencakup penggunaan Squared Hinge untuk klasifikasi, Squared Loss untuk regresi, Regularisasi Elastic Net, serta learning rate optimal. Modifikasi ini memberikan pengaruh yang signifikan pada performa model, memungkinkan SGD untuk mengatasi ketidakseimbangan data lebih baik daripada algoritma lain seperti CNN, Logistic Regression, SVM, dan Random Forest yang diuji pada dataset ini.

Penting untuk dicatat bahwa peningkatan akurasi ini terutama disebabkan oleh kemampuan model untuk lebih baik dalam menyesuaikan bobot antara prediksi positif dan negatif, sehingga meningkatkan kemampuan klasifikasi pada kelas-kelas yang lebih kecil jumlahnya. Sebelumnya, dengan penggunaan parameter standar, model SGD cenderung mengalami kesulitan dalam mengatasi ketidakseimbangan kelas, yang mengakibatkan hasil akurasi yang lebih rendah dan performa keseluruhan yang tidak optimal. Oleh karena itu, penyesuaian parameter dalam penelitian ini menjadi kunci keberhasilan model dalam mencapai performa yang lebih baik.

Selain akurasi, metrik lain seperti AUC yang mencapai 0.775 menunjukkan kemampuan model yang cukup baik dalam membedakan antara kelas positif dan negatif. Meskipun tidak sepenuhnya optimal, nilai AUC ini memperlihatkan bahwa model dengan penyesuaian parameter mampu meningkatkan performa dibandingkan model dengan pengaturan default atau algoritma lain yang telah dicoba. Hal ini menunjukkan

bahwa SGD, dengan parameter yang disesuaikan, dapat mengungguli model lain dalam menangani dataset yang tidak seimbang.

F1-score sebesar 0.625 juga memberikan gambaran bahwa penyesuaian parameter membantu model mencapai keseimbangan antara precision dan recall. F1-score yang moderat ini menunjukkan bahwa model mampu mengatasi kesalahan dalam prediksi positif palsu (false positives) dan negatif palsu (false negatives) dengan lebih baik. Hal ini sangat penting dalam klasifikasi jamur karena setiap kesalahan prediksi pada kelas-kelas tertentu dapat berdampak pada kualitas hasil klasifikasi keseluruhan. Dengan penyesuaian parameter ini, SGD dapat mencapai kinerja yang lebih baik dibandingkan model lain yang lebih kompleks.

Penyesuaian parameter juga berperan penting dalam peningkatan precision menjadi 0.632, yang berarti bahwa lebih dari 63% dari prediksi positif yang dihasilkan oleh model adalah benar. Ini menandakan bahwa model dengan penyesuaian parameter mampu mengurangi jumlah prediksi positif palsu secara signifikan. Recall sebesar 0.624 juga menunjukkan bahwa model berhasil menemukan sebagian besar instance positif dalam dataset, meskipun masih ada beberapa yang terlewat. Keduanya menunjukkan perbaikan yang signifikan dibandingkan dengan penggunaan parameter standar.

Akhirnya, Matthews Correlation Coefficient (MCC) sebesar 0.557 memberikan gambaran bahwa model dengan penyesuaian parameter dapat mencapai prediksi yang lebih berkualitas dibandingkan model yang dioptimalkan secara default. MCC yang mendekati 0.6 menunjukkan bahwa korelasi antara prediksi dan hasil sebenarnya cukup baik, menunjukkan peningkatan performa model secara keseluruhan.

KESIMPULAN

Penelitian ini mengimplementasikan algoritma Stochastic Gradient Descent (SGD) untuk melakukan klasifikasi gambar jamur dari dataset yang tidak seimbang dengan 10 kelas. Hasil evaluasi menunjukkan bahwa penyesuaian parameter pada SGD, seperti penggunaan Squared Hinge untuk klasifikasi, Squared Loss untuk regresi, regularisasi Elastic Net, dan learning rate optimal, memberikan peningkatan akurasi yang lebih baik dibandingkan algoritma lain seperti CNN, Logistic Regression, SVM, dan Random Forest. Akurasi model mencapai 62.4%, dan nilai AUC sebesar 0.775 menunjukkan bahwa model memiliki kemampuan yang cukup baik dalam membedakan kelas. Namun, meskipun SGD menunjukkan performa yang lebih baik daripada algoritma lain, hasil evaluasi keseluruhan, seperti F1-score (0.625), precision (0.632), dan recall (0.624), menunjukkan bahwa performa model belum optimal. Nilai MCC (0.557) juga mengindikasikan bahwa kualitas prediksi model masih berada pada kategori moderat, dan model masih mengalami kesulitan dalam mengatasi ketidakseimbangan data yang signifikan.

Berdasarkan hasil penelitian ini, ada beberapa saran yang dapat dipertimbangkan untuk penelitian lebih lanjut:

- 1) Meskipun penyesuaian parameter telah meningkatkan performa model, penelitian selanjutnya dapat menggunakan teknik penyeimbangan data seperti oversampling atau undersampling, atau metode lain seperti Synthetic Minority Over-sampling Technique (SMOTE) untuk lebih meningkatkan kemampuan model dalam mengenali kelas minoritas.
- 2) Penelitian ini menunjukkan bahwa penyesuaian parameter penting untuk meningkatkan performa SGD. Penelitian lebih lanjut dapat mencakup pencarian grid search atau random search untuk menemukan kombinasi parameter yang lebih optimal guna memperbaiki kinerja model.
- 3) Untuk mengatasi kelemahan SGD dalam menangani dataset yang tidak seimbang, penggunaan model ensemble seperti AdaBoost, XGBoost, atau bagging bisa menjadi alternatif yang kuat untuk lebih meningkatkan performa model, terutama dalam menangani kelas minoritas.
- 4) Selain metrik yang digunakan dalam penelitian ini, metrik lain seperti Cohen's Kappa dan G-mean dapat digunakan untuk mengevaluasi kinerja model lebih menyeluruh, terutama pada dataset yang tidak seimbang.
- 5) Meskipun SGD terbukti efektif, penggunaan deep learning seperti Convolutional Neural Networks (CNN) dengan arsitektur yang lebih dalam dapat dieksplorasi kembali, terutama dengan penyesuaian pada jumlah layer dan pengaturan hyperparameter yang lebih baik untuk menangani karakteristik dataset gambar ini.

REFERENCES

- [1] D. Armiady and I. M. R, "Klasifikasi Kualitas Buah Pisang Berdasarkan Citra Buah Menggunakan Stochastic Gradient Descent," *KLIK: Kajian Ilmiah Informatika dan Komputer*, vol. 4, no. 2, 2023.
- [2] I. R. Muslem and T. M. Johan, "KLIK: Kajian Ilmiah Informatika dan Komputer Klasifikasi Citra Ikan Menggunakan Algoritma Convolutional Neural Network dengan Arsitektur VGG-16," *Media Online*, vol. 4, no. 2, pp. 978–985, 2023, doi: 10.30865/klik.v4i2.1209.
- [3] I. R. Muslem, "KLIK: Kajian Ilmiah Informatika dan Komputer Image Classification pada Kasus American Sign Language Menggunakan Support Vector Machine," *Media Online*, vol. 4, no. 2, pp. 1184–1191, 2023, doi: 10.30865/klik.v4i2.1242.
- [4] L. Luthfi, R. Imam Muslem, D. Armiady, S. Sriwinar, R. Fajri, and I. Iqbal, "Analysis of CNN Method for Image Classification of Coconut Ripeness Levels," in *2023 Eighth International Conference on Informatics and Computing (ICIC)*, IEEE, Dec. 2023, pp. 1–6. doi: 10.1109/ICIC60109.2023.10381964.
- [5] J. Tian and L. Li, "Research on artificial intelligence of accounting information processing based on image processing," *Mathematical Biosciences and Engineering*, vol. 19, no. 8, 2022, doi: 10.3934/mbe.2022391.
- [6] K. Marti, *Stochastic optimization methods (Second edition)*. 2008. doi: 10.1007/978-3-540-79458-5.
- [7] T. M. Alqahtani, "Big Data Analytics with Optimal Deep Learning Model for Medical Image Classification," *Computer Systems Science and Engineering*, vol. 44, no. 2, 2023, doi: 10.32604/csse.2023.025594.
- [8] A. Arafa, M. Radad, M. Badawy, and N. El - Fishawy, "Regularized Logistic Regression Model for Cancer Classification," in *National Radio Science Conference, NRSC, Proceedings*, 2021. doi: 10.1109/NRSC52299.2021.9509831.
- [9] J. Zeng, M. Zhang, and S. B. Lin, "Fully corrective gradient boosting with squared hinge: Fast learning rates and early stopping," *Neural Networks*, vol. 147, 2022, doi: 10.1016/j.neunet.2021.12.016.
- [10] S. Zubair, A. K. Singha, N. Pathak, N. Sharma, S. Urooj, and S. R. Laguech, "Performance Enhancement of Adaptive Neural Networks Based on Learning Rate," *Computers, Materials and Continua*, vol. 74, no. 1, 2023, doi: 10.32604/cm.2023.031481.
- [11] H. Kaur, H. S. Pannu, and A. K. Malhi, "A systematic review on imbalanced data challenges in machine learning: Applications and solutions," 2019. doi: 10.1145/3343440.
- [12] S. Khodabandelu, N. Ghaemian, S. Khafri, M. Ezoji, and S. Khaleghi, "Development of a Machine Learning-Based Screening Method for Thyroid Nodules Classification by Solving the Imbalance Challenge in Thyroid Nodules Data," *J Res Health Sci*, vol. 22, no. 3, 2022, doi: 10.34172/jrhs.2022.90.
- [13] V. Sampath, I. Maurtua, J. J. Aguilar Martín, and A. Gutierrez, "A survey on generative adversarial networks for imbalance problems in computer vision tasks," *J Big Data*, vol. 8, no. 1, 2021, doi: 10.1186/s40537-021-00414-0.
- [14] P. Maurya and N. P. Singh, "Mushroom Classification Using Feature-Based Machine Learning Approach," in *Advances in Intelligent Systems and Computing*, 2020. doi: 10.1007/978-981-32-9088-4_17.
- [15] O. Tarawneh, M. Tarawneh, Y. Sharrab, and M. Husni, "Mushroom classification using machine-learning techniques," in *AIP Conference Proceedings*, 2023. doi: 10.1063/5.0174721.
- [16] Y. Yang and G. Mirzaei, "Performance analysis of data resampling on class imbalance and classification techniques on multi-omics data for cancer classification," *PLoS One*, vol. 19, no. 2 February, 2024, doi: 10.1371/journal.pone.0293607.
- [17] G. Ryan, P. Katarina, and D. Suhartono, "MBTI Personality Prediction Using Machine Learning and SMOTE for Balancing Data Based on Statement Sentences," *Information (Switzerland)*, vol. 14, no. 4, 2023, doi: 10.3390/info14040217.
- [18] A. Rahman, E. Winarko, and K. Mustofa, "Content-based product image retrieval using squared-hinge loss trained convolutional neural networks," *International Journal of Electrical and Computer Engineering*, vol. 13, no. 5, 2023, doi: 10.11591/ijece.v13i5.pp5804-5812.
- [19] N. Loizou, S. Vaswani, I. Laradji, and S. Lacoste-Julien, "Stochastic Polyak Step-size for SGD: An Adaptive Learning Rate for Fast Convergence," in *Proceedings of Machine Learning Research*, 2021.