

**Implementasi *Class Weight*, *Data Augmentation* dan *Resnet101* Untuk
Mengatasi *Overfitting* dan *Imbalanced Data* Pada Klasifikasi Kacang
*Pistachio***

SKRIPSI

Diajukan Untuk Memenuhi Sebagian Syarat Guna
Memperoleh Gelar Sarjana Komputer (S.Kom.)
Pada Program Studi Sistem Informasi



OLEH :

Moch. Raffi Dwi Saktya Rahman

NPM: 2213030128

FAKULTAS TEKNIK DAN ILMU KOMPUTER (FTIK)
UNIVERSITAS NUSANTARA PERSATUAN GURU REPUBLIK INDONESIA
UN PGRI KEDIRI
2026

Skripsi oleh :

Moch. Raffi Dwi Saktya Rahman

NPM : 2213030128

Judul

**Implementasi *Class Weight*, *Data Augmentation* dan *Resnet101* Untuk
Mengatasi *Overfitting* dan *Imbalanced Data* Pada Klasifikasi Kacang
*Pistachio***

Telah disetujui untuk diajukan Kepada

Panitia Ujian/Sidang Skripsi Program Studi Sistem Informasi

FTIK UN PGRI Kediri

Tanggal : 11 Juni 2026

Pembimbing I



Arie Nugroho, S.Kom, M.M, M.Kom

NIDN. 0712108103

Pembimbing II



Aidina Ristyawan, S.Kom, M.Kom

NIDN. 0721018801

Skripsi oleh :

Moch. Raffi Dwi Saktya Rahman

NPM : 2213030128

Judul

**Implementasi *Class Weight*, *Data Augmentation* dan *Resnet101* Untuk
Mengatasi *Overfitting* dan *Imbalanced Data* Pada Klasifikasi Kacang
*Pistachio***

Telah dipertahankan di depan Panitia Ujian/Sidang Skripsi

Program Studi Sistem Informasi FTIK UN PGRI Kediri

Tanggal: 25 Juni 2026

Dan Dinyatakan telah Memenuhi Persyaratan

Panitia Penguji :

1. Ketua : Arie Nugroho, S.Kom, M.M, M.Kom [..........]

2. Penguji 1 : Anita Sari Wardani, S.Kom, M.Kom [..........]

3. Penguji 2 : Aidina Ristyawan, S.Kom, M.Kom [..........]

Mengetahui,

Dekan Fakultas Teknik dan Ilmu Komputer



Dr. Sulistiono, M.Si.

NIDN. 0007076801

PERNYATAAN

Yang bertanda tangan di bawah ini saya,

Nama : Moch. Raffi Dwi Saktya Rahman
Jenis Kelamin : Laki-laki
Tempat/tgl. lahir : Kediri/1 Oktober 2003
NPM : 2213030128
Fak/Prodi. : FTIK/ S1-Sistem Informasi

menyatakan dengan sebenarnya, bahwa dalam Skripsi ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu perguruan tinggi, dan sepanjang pengetahuan saya tidak terdapat karya tulis atau pendapat yang pernah diterbitkan oleh orang lain, kecuali yang secara sengaja dan tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Kediri, 18 Juni 2026

Yang Menyatakan



Moch. Raffi Dwi Saktya Rahman
NPM: 2213030128

MOTTO

“Bukan kesulitan yang membuat kita takut, tapi ketakutan yang membuat kita sulit.”

Ali bin Abi Thalib

“Tugas manusia hanya berjuang, bukan memaksakan hasil. Kita punya kendala tapi Allah punya kendali. Yakinlah jika Allah sudah ikut andil, maka tidak ada kata mustahil.”

Ustadz Hanan Attaki

“Keberhasilan bukanlah milik orang pintar. Melainkan milik mereka yang senantiasa berusaha.”

B.J Habibie

ABSTRAK

Moch. Raffi Dwi Saktya Rahman: Implementasi *Class Weight*, *Data Augmentation* dan *Resnet101* Untuk Mengatasi *Overfitting* dan *Imbalanced Data* Pada Klasifikasi Kacang *Pistachio*, Skripsi, Sistem Informasi, Fakultas Teknik dan Ilmu Komputer UN PGRI Kediri, 2025.

Kacang *Pistachio* merupakan komoditas pertanian bernilai ekonomi tinggi yang proses klasifikasi pasca panennya masih dilakukan secara manual, sehingga rentan terhadap kesalahan akibat kemiripan morfologi antar varietas. Penelitian sebelumnya menggunakan *Resnet50* menunjukkan *overfitting* parah dengan *gap training-validation accuracy* mencapai 78%, serta belum menangani ketidakseimbangan distribusi data antara kelas *Kirmizi* (1232 sampel) dan *Siirt* (916 sampel) secara eksplisit. Penelitian ini bertujuan meningkatkan akurasi klasifikasi sekaligus mengatasi *overfitting* dan *imbalanced data* melalui implementasi *Resnet101* yang dikombinasikan dengan *data augmentation* dan *class weighting* menggunakan kerangka *CRISP-DM*, dengan dataset 2148 citra yang dibagi rasio 80:10:10. Hasil penelitian menunjukkan bahwa model *Resnet101* dengan *class weighting* tanpa *augmentation* menghasilkan performa terbaik dengan *testing accuracy*, *precision*, *recall*, dan *F1-score* masing-masing sebesar 98,31%, melampaui *baseline* Inception-V3 (96%) sebesar 2,31 poin persentase dan *Resnet50* (86%) sebesar 12,31 poin, dengan selisih *F1-score* antar kelas hanya 0,51 poin. Sebaliknya, *data augmentation* justru menurunkan akurasi karena karakteristik latar belakang citra yang sudah sangat bersih dan homogen. Penelitian ini memiliki keterbatasan berupa dataset yang hanya mencakup dua varietas dengan kondisi citra terkontrol, sehingga penelitian selanjutnya disarankan untuk mengeksplorasi arsitektur yang lebih modern, memperluas dataset dengan kondisi pengambilan citra yang lebih bervariasi.

Kata Kunci: *Resnet101*, *Transfer Learning*, *Data Augmentation*, *Class Weight*, Klasifikasi *Pistachio*.

KATA PENGANTAR

Puji Syukur Kami panjat kan kehadiran Allah Tuhan Yang Maha Kuasa, karena hanya atas perkenan- Nya tugas penyusunan proposal ini dapat diselesaikan. Penyusunan proposal ini merupakan bagian dari rencana penelitian guna penyusunan skripsi sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer pada Program Studi Sistem Informasi.

Pada kesempatan ini diucapkan terimakasih dan penghargaan yang setulus-tulusnya kepada:

1. Dr. Zainal Afandi, M.Pd. selaku Rektor Universitas Nusantara PGRI Kediri.
2. Dr. Sulistiono, M.Si. selaku Dekan Fakultas Teknik dan Ilmu Komputer Universitas Nusantara PGRI Kediri.
3. Dr. Sucipto, M.Kom, selaku Ketua Program Studi Sistem Informasi Universitas Nusantara PGRI Kediri.
4. Arie Nugroho, M.Kom, selaku Dosen Pembimbing 1 yang telah memberikan bimbingan, arahan, motivasi, dan koreksi sehingga proposal penelitian ini dapat diselesaikan dengan baik.
5. Aidina Ristyawan, S.Kom, M.Kom, selaku Dosen Pembimbing 2 yang telah memberikan bimbingan dan saran kepada penulis.
6. Seluruh dosen dan staf Program Studi Sistem Informasi yang telah memberikan ilmu pengetahuan dan pelayanan selama masa perkuliahan.
7. Teruntuk kedua orang tuaku tercinta, Ayah Rahmani Eko Handriono dan Ibu Umi Maslikah, terimakasih tiada terhingga penulis sampaikan atas segala cinta kasih, arahan, dukungan dan apapun yang telah diberikan.
8. Teruntuk kakak ku tercinta, Sema Arga Tama, terimakasih atas segala cinta kasih, dukungan dan doa-doa terbaik yang selalu dipanjatkan dan terimakasih kepada istri dari kakak ku Clara Desintya Dhea serta dua keponakan ku yang tingkah nya selalu menghibur penulis, Terimakasih semuanya sudah ikut andil dalam setiap proses yang dijalani penulis.
9. Teman-teman seperjuangan Program Studi Sistem Informasi angkatan 2022, yang telah memberikan semangat, motivasi, dan bantuan selama penyusunan proposal ini.

10. Ucapan terima kasih juga disampaikan kepada pihak-pihak lain yang tidak dapat disebutkan satu persatu, yang telah banyak membantu menyelesaikan Skripsi ini.

Penulis menyadari penuh bahwa proposal penelitian ini masih jauh dari sempurna. Oleh karena itu, kritik dan saran yang membangun dari semua pihak sangat penulis harapkan demi perbaikan dan kesempurnaan proposal ini. Semoga proposal penelitian ini dapat memberikan manfaat bagi pengembangan ilmu pengetahuan terutama dalam bidang *Computer Vision*.

Kediri, 18 Juni 2026



Moch. Raffi Dwi Saktya Rahman
NPM: 2213030128

DAFTAR ISI

HALAMAN JUDUL	ii
HALAMAN PERSETUJUAN.....	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERNYATAAN.....	iv
HALAMAN MOTTO	v
ABSTRAK	vi
KATA PENGANTAR.....	vii
DAFTAR ISI.....	ix
DAFTAR GAMBAR	xii
DAFTAR TABEL	xiii
DAFTAR LAMPIRAN	xiii
BAB I PENDAHULUAN.....	1
A. Latar Belakang	1
B. Batasan Masalah	6
C. Rumusan Masalah	7
D. Tujuan Penelitian	7
E. Manfaat Penelitian	8
BAB II KAJIAN PUSTAKA	10
A. Kajian Teori.....	10
1. <i>Pistachio</i> (Kacang <i>Pistachio</i>).....	10
2. <i>Convolutional Neural Network (CNN)</i>	11
3. <i>Transfer Learning</i>	13
4. <i>Resnet101</i>	13
5. <i>Augmentasi Data</i>	14
6. <i>Class weight</i>	14
7. Metrik <i>Evaluasi</i> Klasifikasi	15
B. Kajian Hasil Penelitian Terdahulu	16
1. Penelitian menggunakan <i>CNN</i>	16
2. Penelitian tentang Klasifikasi <i>Pistachio</i>	16
3. Perbedaan penelitian ini dengan penelitian sebelumnya	18
C. Kerangka Berpikir.....	18
BAB III METODE PENELITIAN.....	20
A. <i>Business Understanding</i>	20
1. Latar Belakang Masalah Bisnis.....	20
2. Tujuan Bisnis	21

3.	Kriteria Keberhasilan Proyek	21
B.	<i>Data Understanding</i>	22
1.	Pengumpulan Data	22
2.	Deskripsi Data	22
3.	Eksplorasi Data Awal	22
4.	Verifikasi Kualitas Data	23
C.	<i>Data Preparation</i>	23
1.	Seleksi Data	23
2.	Pembersihan Data	23
3.	Transformasi Data	23
4.	Pembuatan Dataset <i>Modeling</i>	24
D.	<i>Modelling</i>	25
1.	Pemilihan Teknik Model	25
2.	Desain dan Pengujian Model	25
3.	Evaluasi Parameter Model	26
4.	Evaluasi Parameter Model	26
E.	<i>Evaluation</i>	26
1.	Evaluasi Model Terbaik	26
2.	Validasi Model terhadap Tujuan Bisnis	27
3.	Tinjauan Ulang Proses CRISP-DM	27
F.	<i>Deployment</i>	27
1.	Rencana Penerapan Hasil	27
2.	Implementasi Model	28
3.	Dokumentasi dan Pelaporan	28
BAB IV HASIL DAN PEMBAHASAN		29
A.	Data Hasil Penelitian	29
1.	Hasil Eksplorasi Data	29
2.	Proses Transformasi Data	30
3.	Hasil Modeling	32
4.	Perbandingan beberapa model	39
B.	Pembahasan	40
1.	Model yang Paling Optimal	40
2.	Evaluasi Hasil <i>Modeling</i> terhadap Tujuan Bisnis	41
3.	Kesesuaian Hasil dengan <i>Business Understanding</i>	42
4.	Validitas dan Generalisasi Hasil	42
5.	Rekomendasi Penerapan Model	43

BAB V PENUTUP.....	47
A. Kesimpulan	47
B. Saran	48
DAFTAR PUSTAKA	50
LAMPIRAN	50

DAFTAR GAMBAR

Gambar 2.1. <i>Pistachio Kirmizi</i>	10
Gambar 2.2. <i>Pistachio Siirt</i>	11
Gambar 2.3. Arsitektur <i>CNN</i>	12
Gambar 2.4. Arsitektur <i>Resnet-101</i>	14
Gambar 2.5. Alur Kerangka Berpikir.....	19
Gambar 4.1. Distribusi Kelas Dataset <i>Pistachio</i>	29
Gambar 4.2. Sampel Citra <i>Kirmizi</i> dan <i>Siirt</i> 4 Skenario	32
Gambar 4.3. Kurva <i>Training Accuracy</i> Keempat Skenario	36
Gambar 4.4. Kurva <i>Training Loss</i> Keempat Skenario.....	37
Gambar 4.5. <i>Confusion Matrix</i> Keempat Skenario	38
Gambar 4.6. Tampilan Halaman <i>Home</i>	44
Gambar 4.7. Tampilan Halaman <i>About</i>	44
Gambar 4.8. Tampilan Halaman Deteksi	45
Gambar 4.9. Tampilan Halaman Hasil Prediksi.....	46

DAFTAR TABEL

Tabel 3.1. Pembagian Dataset	23
Tabel 3.2. Parameter <i>Data Augmentation</i>	24
Tabel 3.3. Nilai <i>Class Weight</i>	24
Tabel 3.4. Arsitektur <i>Classification Head Resnet101</i>	25
Tabel 3.5. Konfigurasi <i>Callbacks</i>	26
Tabel 3.6. Hasil Evaluasi Model <i>Resnet101</i>	27
Tabel 3.7. Validasi Model terhadap Kriteria Keberhasilan.....	27
Tabel 3.8. Spesifikasi Laptop.....	28
Tabel 4.1. Statistik Deskriptif Dataset <i>Pistachio</i>	29
Tabel 4.2. Deskripsi Keempat Skenario Eksperimen.....	30
Tabel 4.3. Pembagian Dataset.....	31
Tabel 4.4. Ringkasan Hasil Keempat Eksperimen <i>Resnet101</i>	33
Tabel 4.5. <i>Classification Report</i> Skenario 1	34
Tabel 4.6. <i>Classification Report</i> Skenario 2	34
Tabel 4.7. <i>Classification Report</i> Skenario 3	35
Tabel 4.8. <i>Classification Report</i> Skenario 4.....	35
Tabel 4.9. Hasil <i>F1-Score Class Weighting</i>	39
Tabel 4.10. Perbandingan dengan Penelitian Sebelumnya	40

DAFTAR LAMPIRAN

Lampiran 1 Kartu Bimbingan Skripsi	53
Lampiran 2 Surat Keterangan Similarity	54
Lampiran 3 Hasil Cek Similarity	55
Lampiran 4 Bukti Loa Artikel.....	56
Lampiran 5 Lembar Berita Acara Ujian.....	57
Lampiran 6 Lembar Revisi Ujian.....	58

BAB I

PENDAHULUAN

A. Latar Belakang

Pada abad ini, sektor pertanian menjadi aspek yang paling signifikan dalam pembangunan suatu negara dalam hal dukungan ekonomi. Seiring berjalannya waktu, persaingan di sektor pertanian akan semakin intens, yang ditandai dengan produksi yang lebih besar, kualitas yang lebih baik, dan efisiensi bisnis yang lebih tinggi. Inovasi pertanian baru semakin dibutuhkan untuk membantu proses produksi sehingga kualitas dan kuantitas yang dihasilkan dapat dimaksimalkan dan bersaing dengan yang lain (Lisda et al., 2023) (Trihardianingsih L. & Permatasari H., 2025).

Pistachio merupakan salah satu komoditas pertanian yang memerlukan inovasi dalam proses pengelolaannya. *Pistachio* adalah jenis tanaman penghasil biji-bijian yang umumnya dikenal sebagai kacang *Pistachio*, yang tumbuh dengan baik di iklim kering dengan tanah yang kurang subur, seperti Iran, Turkmenistan, dan Azerbaijan bagian barat. Biji berwarna hijau cerah dari tanaman *Pistachio* biasanya digunakan untuk menghias hidangan atau dalam berbagai olahan makanan lezat seperti es krim. *Pistachio* adalah jenis kacang-kacangan yang memiliki nutrisi tinggi karena mengandung protein, serat, asam lemak tak jenuh tunggal, mineral, vitamin, karotenoid, asam fenolik, flavonoid, dan antosianin (Lisda et al., 2023) (Trihardianingsih L. & Permatasari H., 2025).

Pistachio dipilih sebagai objek penelitian karena merupakan salah satu komoditas pertanian dengan nilai jual tinggi yang permintaannya terus meningkat di pasar global, namun proses klasifikasi pasca panennya masih dilakukan secara manual dan konvensional sehingga rentan terhadap kesalahan Singh et al., (2022) Uzer, (2025). Dibandingkan komoditas pertanian lainnya, *Pistachio* memiliki tantangan klasifikasi yang lebih kompleks karena perbedaan morfologi antar varietasnya sangat halus dan sulit dibedakan secara kasat mata, menjadikannya objek yang relevan dan menantang untuk dikaji menggunakan pendekatan *computer vision* berbasis *deep learning* (Lisda et al., 2023).

Adapun pemilihan varietas *Kirmizi* dan *Siirt* secara spesifik didasarkan pada tiga alasan utama. Pertama, kedua varietas ini merupakan jenis *Pistachio* yang paling banyak diperdagangkan dan memiliki nilai ekonomi tertinggi dibandingkan varietas lainnya, sehingga akurasi klasifikasi keduanya berdampak langsung pada nilai transaksi perdagangan AVUÇLU, (2023) Özkan et al., (2021). Kedua, meskipun ditujukan untuk segmen pasar yang berbeda. *Kirmizi* untuk industri permen dan kue manis, serta *Siirt* untuk camilan, keduanya memiliki kemiripan bentuk fisik yang sangat tinggi sehingga menjadi tantangan klasifikasi yang signifikan Lisda et al., (2023; Trihardianingsih L. & Permatasari H., (2025). Ketiga, tersedianya dataset publik yang telah tervalidasi untuk kedua varietas ini di Kaggle memungkinkan penelitian dilakukan secara terstruktur dan hasil penelitian dapat dibandingkan secara langsung dengan penelitian-penelitian sebelumnya yang menggunakan dataset yang sama (Lisda et al., 2023) (Avuçlu, 2023).

Amerika Serikat, Iran, Turki, dan Suriah sebagai produsen global utama *Pistachio* menyumbang hampir 90% dari total produksi di seluruh dunia (Singh et al., (2022) Avuçlu, (2023) Uzer, (2025). Namun demikian, proses yang digunakan untuk memisahkan kacang *Pistachio* masih dilakukan secara manual dengan pengetahuan dasar. Karena hal ini, terdapat potensi tinggi untuk terjadinya kesalahan dalam proses klasifikasi mengingat setiap varietas kacang *Pistachio* memiliki bentuk yang hampir identik Lisda et al., (2023) Trihardianingsih L. & Permatasari H., (2025). Kesalahan klasifikasi ini dapat menyebabkan kerugian ekonomi yang signifikan karena produk dapat tercampur dan dijual dengan harga yang tidak sesuai dengan jenisnya (Singh et al., 2022) (Uzer, 2025).

Oleh karena itu, inovasi diperlukan untuk mengenali jenis kacang *Pistachio* secara otomatis sehingga proses pengemasan untuk penjualan dapat tepat sasaran Lisda et al., (2023) Trihardianingsih L. & Permatasari H., (2025). Untuk meningkatkan nilai pasar *Pistachio*, diperlukan pemisahan dan klasifikasi otomatis untuk berbagai jenis produk yang berasal dari pemasok yang berbeda AVUÇLU, (2023) Özkan et al., (2021). Oleh karena itu, teknik dan teknologi baru harus digunakan untuk meningkatkan metode yang

digunakan dalam proses mekanisasi pasca panen *Pistachio* dan untuk meningkatkan hasil dari produk yang memiliki nilai ekonomi tinggi (Avuçlu, 2023) (Uzer, 2025).

Perkembangan teknologi *computer vision* dan pemrosesan citra dalam beberapa tahun terakhir telah membuka peluang untuk mengotomatisasi proses klasifikasi produk pertanian. Teknologi berbasis *deep learning*, khususnya *Convolutional Neural Network (CNN)*, telah terbukti mampu mengekstraksi fitur visual secara hierarkis dan otomatis dari citra tanpa memerlukan rekayasa fitur manual. Hal ini menjadikan *CNN* sebagai salah satu pendekatan yang paling menjanjikan untuk diterapkan dalam klasifikasi *Pistachio* yang memiliki kemiripan bentuk visual antar varietas (Lisda et al., 2023) (Trihardianingsih L. & Permatasari H., 2025).

Selain potensi teknologi yang besar, terdapat juga tantangan praktis yang perlu diatasi, yaitu keterbatasan jumlah data citra *Pistachio* yang tersedia untuk melatih model *deep learning* secara optimal. Dataset yang terbatas berpotensi menyebabkan model mengalami *overfitting*, yakni kondisi di mana model hanya mampu menghafal data latih tetapi gagal menggeneralisasi pada data baru. Hal ini memperkuat urgensi penerapan teknik *data augmentation* sebagai solusi untuk memperbanyak variasi data latih tanpa harus mengumpulkan data baru secara fisik (Lisda et al., 2023).

Berbagai penelitian terdahulu telah dilakukan untuk mengklasifikasikan kacang *Pistachio* menggunakan metode *machine learning* dan *deep learning*. Penelitian oleh Farazi et al. menggunakan arsitektur *machine learning* dan *deep learning* untuk mengklasifikasikan *Pistachio* ke dalam 3 kelas yaitu *Pistachio* cangkang terbuka, *Pistachio* cangkang busuk, dan sampah atau serpihan kayu, dengan menggunakan 1000 citra yang diekstraksi menggunakan arsitektur AlexNet dan GoogleNet dan memperoleh akurasi secara berturut-turut sebesar 98% dan 99% Trihardianingsih L. & Permatasari H., (2025). Penelitian lain oleh Omid et al. menggunakan pengolahan citra dengan *machine learning* untuk proses klasifikasi jenis *Pistachio* yang sudah mengalami proses pengupasan, dengan membagi *Pistachio* yang sudah dikupas menjadi 5 kelas menggunakan *Artificial Neural Networks (ANN)* dan *Support Vector Machine*

(SVM), dan hasil klasifikasi dengan menggunakan *ANN* dan *SVM* secara berturut-turut adalah 99,4% dan 99,8% (Trihardianingsih L. & Permatasari H., 2025).

Penelitian yang mengklasifikasikan spesies *Pistachio* menggunakan pendekatan *Transfer Learning* dengan arsitektur *AlexNet*, *VGG16*, dan *VGG19* menggunakan 2148 foto yang terdiri dari 916 jenis *Siirt* dan 1232 jenis *Kirmizi*, dibagi dengan rasio 80:20 untuk data *training* dan *testing*, menghasilkan akurasi masing-masing 94,42%, 98,84%, dan 98,14%. *VGG16* mencapai tingkat keberhasilan klasifikasi terbesar sebesar 98,84% Lisda et al., (2023) Trihardianingsih L. & Permatasari H., (2025). Penelitian lain menggunakan metode *KNN* untuk melakukan klasifikasi *Pistachio* menjadi dua kelas yaitu *Kirmizi* dan *Siirt*, dengan tiga skenario yaitu *KNN*, *Weighted KNN*, dan *PCA based k-NN*, memperoleh hasil secara berturut-turut 83%, 87%, dan 94%, di mana keberhasilan klasifikasi tertinggi dicapai menggunakan *PCA* dengan akurasi 94% (Özkan et al., 2021) (Trihardianingsih L. & Permatasari H., 2025).

Penelitian menggunakan arsitektur *Resnet* untuk klasifikasi *Pistachio* dengan dataset 2148 gambar dan 2 kelas menunjukkan bahwa akurasi tertinggi diperoleh sebesar 88,58% pada *fold-1* dengan rata-rata akurasi klasifikasi 5-*fold* sebesar 86,16% Avuçlu, (2023) Uzer, (2025). Studi lain menggunakan dataset 958 *Pistachio* dengan kondisi cacat dan sempurna untuk klasifikasi, di mana *GoogleNet*, *VGG16*, dan *Resnet models* mencapai akurasi klasifikasi masing-masing sebesar 95,8%, 95,83%, dan 97,2% Uzer, (2025). Studi komparatif tentang evaluasi berbagai jaringan berbasis *CNN* untuk klasifikasi juga menunjukkan bahwa arsitektur *Resnet101* memiliki potensi untuk memberikan hasil yang lebih baik dalam tugas klasifikasi citra dibandingkan *VGG16*, *VGG19*, *GoogleNet*, dan *InceptionV3* ketika diimplementasikan dengan parameter yang sama (Goceri & Karakas, 2020).

Meskipun penelitian sebelumnya telah menunjukkan hasil yang menjanjikan, terdapat beberapa *research gap* yang perlu diperhatikan. Pertama, penelitian-penelitian terdahulu belum mengujicobakan arsitektur *Resnet101* secara spesifik pada dataset *Pistachio* varietas *Kirmizi* dan *Siirt*. Kedua, sebagian besar penelitian tidak menerapkan teknik *data augmentation* untuk

meningkatkan variasi data *training*, padahal jumlah foto dalam dataset dapat mempengaruhi akurasi kategorisasi dari berbagai model. Ketiga, penelitian menggunakan *Resnet* pada dataset yang sama menunjukkan terjadinya *overfitting* parah yang ditandai oleh *gap* besar antara *training accuracy* (98%) dan *validation accuracy* (20%) pada arsitektur *Resnet50*, yang menandakan perlunya strategi penanganan *overfitting* yang lebih baik Lisda et al., (2023). Keempat, ketidakseimbangan distribusi data antara kelas *Kirmizi* (1232 sampel) dan *Siirt* (916 sampel) belum ditangani secara eksplisit melalui teknik *class weighting* dalam penelitian-penelitian sebelumnya.

Selain itu, beberapa studi melaporkan hasil dari *single experimental runs* atau performa terbaik yang dipilih, yang kurang memiliki validasi statistik yang tepat. Praktik ini dapat menyebabkan klaim performa yang terlalu optimis dan reproduktibilitas yang buruk ketika metode diterapkan dalam pengaturan dunia nyata Hoots et al., (2025). Hal ini menunjukkan pentingnya metodologi yang lebih ketat dan transparan dalam penelitian klasifikasi *Pistachio* ke depannya.

Berdasarkan hasil perbandingan berbagai arsitektur *CNN* yang telah dilakukan oleh Goceri & Karakas, (2020) dan saran yang diberikan oleh peneliti Lisda et al., (2023), penelitian ini bertujuan untuk mengimplementasikan arsitektur *Resnet101* yang belum pernah diuji pada dataset *Pistachio* varietas *Kirmizi* dan *Siirt*. Penerapan teknik *data augmentation*, meliputi *rescale*, *rotation*, *horizontal flip*, *width shift range*, *height shift range*, *fill mode*, dan *validation split*, diharapkan dapat meningkatkan variasi data *training* sesuai dengan rekomendasi bahwa jumlah foto dalam dataset dapat mempengaruhi akurasi model (Lisda et al., 2023).

Penerapan *data augmentation* diharapkan dapat mengatasi keterbatasan jumlah dataset dan meningkatkan kemampuan generalisasi model dalam mengklasifikasikan kedua jenis *Pistachio*. Selain itu, penelitian ini juga akan menerapkan *class weighting* untuk mengoptimalkan performa model sekaligus mencegah terjadinya *overfitting* dan bias model terhadap kelas mayoritas. Dengan menggunakan metodologi yang konsisten dengan penelitian sebelumnya dalam hal pembagian dataset, *batch size* 32, dan *epoch* hingga 100 Lisda et al., (2023), serta dengan memperhatikan pentingnya validasi statistik

yang tepat untuk menghindari klaim performa yang terlalu optimis Hoots et al., (2025), penelitian ini diharapkan dapat memberikan kontribusi berupa peningkatan akurasi klasifikasi yang melampaui *baseline* Inception-V3 sebesar 96% (Lisda et al., 2023).

Penelitian ini juga akan memberikan analisis komprehensif mengenai performa *Resnet101* pada dataset *Pistachio*, termasuk evaluasi menggunakan berbagai metrik seperti *accuracy*, *precision*, *recall*, *F1-score*, dan *confusion matrix* Lisda et al., (2023) Uzer, (2025), sehingga dapat memberikan gambaran yang lebih lengkap dan valid tentang kemampuan model dalam mengklasifikasikan kedua jenis *Pistachio* tersebut. Hasil penelitian ini diharapkan pula dapat membantu petani dalam meningkatkan efisiensi proses pasca panen dengan mengklasifikasikan berbagai jenis kacang *Pistachio* secara cepat, akurat, dan efisien, serta menangani kebutuhan pasar dengan lebih tepat sasaran (Lisda et al., 2023) (Özkan et al., 2021) (Trihardianingsih L. & Permatasari H., 2025).

B. Batasan Masalah

Agar penelitian ini lebih terarah dan tidak melebar dari tujuan utama, maka ditetapkan batasan masalah sebagai berikut:

1. Penelitian hanya berfokus pada klasifikasi dua varietas kacang *Pistachio*, yaitu *Kirmizi* dan *Siirt*, sebagaimana terdapat pada dataset *Pistachio* publik dari kaggle yang digunakan dalam penelitian sebelumnya.
2. Model deep learning yang dikembangkan hanya menggunakan arsitektur *Resnet101*, sebagai bentuk eksplorasi arsitektur alternatif dari penelitian terdahulu yang menggunakan *Inception-V3*, dan *Resnet50*. Penelitian ini tidak melakukan eksperimen pada arsitektur lain di luar *Resnet101*.
3. Penambahan jumlah data hanya dilakukan melalui teknik data Augmentation, seperti (*rescale*, *rotation*, *horizontal_flip*, *width_shift_range*, *height_shift_range*, *fill_mode*, dan *validation_split*). Penelitian ini tidak menambah data melalui pengambilan gambar baru ataupun sumber dataset eksternal.
4. Penanganan ketidakseimbangan kelas (*imbalance class*) hanya dilakukan menggunakan teknik *class weighting* pada proses *training*. Penelitian ini tidak

membahas metode balancing lain seperti *SMOTE*, *oversampling*, *undersampling*, ataupun *focal loss*.

5. Cakupan evaluasi model terbatas pada metrik klasifikasi standar, yaitu *accuracy*, *precision*, *recall*, *F1-score*, dan *confusion matrix*. Evaluasi yang bersifat deployment juga termasuk dalam ruang lingkup ini.

C. Rumusan Masalah

Berdasarkan identifikasi dan batasan masalah yang telah ditetapkan, maka rumusan masalah dalam penelitian ini adalah sebagai berikut :

1. Seberapa efektif penerapan teknik data Augmentation pada arsitektur *Resnet101* dalam meningkatkan akurasi klasifikasi dan mencegah overfitting pada kacang *Pistachio* varietas *Kirmizi* dan *Siirt* dibandingkan dengan model tanpa data Augmentation?
2. Seberapa akurat tingkat akurasi, *precision*, *recall*, dan *F1-score* yang dicapai oleh model *CNN Resnet101* dengan data *Augmentation* dibandingkan dengan penelitian sebelumnya yang menggunakan baseline Inception-V3 (96%) dan *Resnet50* (86%)?
3. Bagaimana pengaruh penerapan *class weighting* terhadap keseimbangan performa klasifikasi antara kelas *Kirmizi* dan *Siirt* yang memiliki distribusi data tidak seimbang?

D. Tujuan Penelitian

Tujuan dari penelitian ini adalah sebagai berikut :

1. Meningkatkan variasi data *training* dan mencegah overfitting melalui implementasi teknik data Augmentation (*rescale*, *rotation*, *horizontal_flip*, *width_shift_range*, *height_shift_range*, *fill_mode*, dan *validation_split*) pada arsitektur *Resnet101* untuk mengatasi keterbatasan jumlah dataset dan meningkatkan kemampuan generalisasi model, yang ditunjukkan melalui pengurangan gap yang signifikan antara *training* dan *validation accuracy* dibandingkan dengan hasil penelitian sebelumnya pada *Resnet50* yang mengalami *overfitting* parah (*gap* 78%).
2. Meningkatkan akurasi klasifikasi *Pistachio* dengan mengimplementasikan arsitektur *Resnet101* yang lebih dalam dibandingkan *Resnet50*, untuk

mencapai akurasi minimal melampaui baseline Inception-V3 (96%) dan *Resnet50* (86%) dari penelitian sebelumnya.

3. Menyeimbangkan kemampuan model dalam mengenali kedua jenis *Pistachio* (*Kirmizi* dan *Siirt*) dengan menggunakan teknik *class weighting*, sehingga model tidak lebih pintar mengenali jenis yang datanya lebih banyak (*Kirmizi*) dan mengabaikan jenis yang datanya lebih sedikit (*Siirt*).

E. Manfaat Penelitian

1. Manfaat Teoritis

- a. Memberikan kontribusi dalam pengembangan ilmu *computer vision* dan *deep learning* untuk aplikasi klasifikasi produk pertanian dengan karakteristik visual yang serupa.
- b. Menjadi referensi akademis mengenai implementasi arsitektur *Resnet101* pada klasifikasi *Pistachio* yang dapat dijadikan acuan untuk penelitian sejenis di masa mendatang.
- c. Memberikan pemahaman yang lebih mendalam mengenai efektivitas teknik data *Augmentation* dan *class weighting* dalam mengatasi permasalahan dataset terbatas dan tidak seimbang.
- d. Memudahkan peneliti lain untuk mengulangi penelitian ini karena melalui transparansi metodologi yang lengkap, termasuk dokumentasi *hyperparameter* dan strategi *training*.

2. Manfaat Praktis

- a. Bagi petani dan pelaku usaha *Pistachio*, hasil penelitian dapat menjadi dasar pengembangan sistem klasifikasi otomatis yang meningkatkan efisiensi proses pasca panen, sehingga menghemat biaya tenaga kerja dan meningkatkan keuntungan karena produk dijual sesuai dengan jenisnya.
- b. Bagi industri pengolahan makanan, sistem klasifikasi ini dapat membantu memastikan kualitas bahan baku yang digunakan selalu konsisten dan tidak tercampur antar jenis.
- c. Bagi pengembang teknologi pertanian, penelitian ini dapat menjadi bukti nyata bahwa teknologi klasifikasi otomatis bisa diterapkan untuk pengembangan aplikasi mobile yang dapat diimplementasikan secara *real-time*.

- d. Bagi peneliti selanjutnya, metodologi yang terdokumentasi dengan baik dalam penelitian ini dapat menjadi baseline yang solid untuk penelitian lanjutan dalam klasifikasi *Pistachio* maupun produk pertanian lainnya.

3. Harapan/kontribusi

Secara keseluruhan, penelitian ini diharapkan dapat memberikan kontribusi yang berarti dalam pengembangan model klasifikasi kacang *Pistachio* berbasis arsitektur *CNN Resnet101*, khususnya dalam membuktikan efektivitas kombinasi teknik *data augmentation* dan *class weighting* untuk meningkatkan akurasi dan keseimbangan klasifikasi varietas *Kirmizi* dan *Siirt*. Selain itu, penelitian ini diharapkan dapat menjadi solusi praktis bagi para pelaku industri *Pistachio* dalam meningkatkan efisiensi dan akurasi proses klasifikasi pasca panen secara otomatis, sehingga setiap varietas dapat dipisahkan dan dipasarkan sesuai dengan nilai ekonominya masing-masing.

DAFTAR PUSTAKA

- Adi, T., Yusri, S., Rudhistiar, D., & Ratnasari, A. P. (2025). *KLASIFIKASI VARIETAS KACANG MENGGUNAKAN XGBOOST DENGAN PENYESUAIAN CLASS WEIGHTING* (Vol. 8, Number 1).
- Adzkia, S. A., & Arifin, T. (2025). Perbandingan Arsitektur CNN Berbasis Transfer Learning untuk Klasifikasi pada BreastMNIST. *Jurnal Nasional Teknologi Dan Sistem Informasi*, 11(2), 192–200. <https://doi.org/10.25077/teknosi.v11i2.2025.192-200>
- Akbar, H., & Sanjaya, W. K. (2023). Kajian Performa Metode Class Weight Random Forest pada Klasifikasi Imbalance Data Kelas Curah Hujan. *Jurnal Sains, Nalar, Dan Aplikasi Teknologi Informasi*, 3(1). <https://doi.org/10.20885/snati.v3i1.30>
- Aldiani, D., Dwilestari, G., Susana, H., Hamonangan, R., & Pratama, D. (2024). *IMPLEMENTASI ALGORITMA CNN DALAM SISTEM ABSENSI BERBASIS PENGENALAN WAJAH*. <https://doi.org/10.33795/jip.v10i2.4852>
- Al-Fahrezi, M. A. (2025). Pengaruh Augmentasi Data Terhadap Akurasi Pelatihan Model CNN untuk Klasifikasi Jenis Ikan. *JITSI : Jurnal Ilmiah Teknologi Sistem Informasi*, 6(2), 177–185. <https://doi.org/10.62527/jitsi.6.2.471>
- Ardyansyah, F., Daniati, E., & Ristyawan, A. (2024). Pemanfaatan Data Mining untuk Analisis Keputusan. In *Agustus* (Vol. 8). Online.
- AVUÇLU, E. (2023). Classification of Pistachio Images Using VGG16 and VGG19 Deep Learning Models. *International Scientific and Vocational Studies Journal*, 7(2), 79–86. <https://doi.org/10.47897/bilmes.1328313>
- Avuçlu, E. (2023). Classification Of Pistachio Images With The ResNet Deep Learning Model. *Selcuk Journal of Agricultural and Food Sciences*. <https://doi.org/10.15316/sjafs.2023.029>
- Azmi, K., & Defit, S. (2023). Implementasi Convolutional Neural Network (CNN) Untuk Klasifikasi Batik Tanah Liat Sumatera Barat. *Jurnal Unitek*, 16(1), 2023. <https://doi.org/10.52072/unitek.v16i1.504>
- Bahtiar, A., Ihsan, M., Hutomo, P., Widiyanto, A., & Khomsah, S. (2025). Class Weighting Approach for Handling Imbalanced Data on Forest Fire Classification Using EfficientNet-B1. In *Jurnal Informatika Sunan Kalijaga* (Vol. 10, Number 1). <https://doi.org/10.14421/jiska.2025.10.1.63-73>
- Bhakti, I. N., Sholikhin, A. Z., Lukman, M. A., Daniati, E., & Ristyawan, A. (2024). *Klasifikasi Kategori Berita Menggunakan Naive Bayes*.
- Fathir Rizal Januar, A., Indra, J., Sulistya Kusumaningrum, D., & Faisal, S. (2025). Application of Convolutional Neural Network (CNN) Algorithm with ResNet-101 Architecture for Monkey Pox Detection in Human. In *Journal of Applied Informatics and Computing (JAIC)* (Vol. 9, Number 3). <https://doi.org/10.30871/jaic.v9i3.9621>

- Faturrahman, R., Hariyani, Y. S., & Hadiyoso, S. (2023). Klasifikasi Jajanan Tradisional Indonesia berbasis Deep Learning dan Metode Transfer Learning. *ELKOMIKA: Jurnal Teknik Energi Elektrik, Teknik Telekomunikasi, & Teknik Elektronika*, 11(4), 945. <https://doi.org/10.26760/elkomika.v11i4.945>
- Goceri, E., & Karakas, A. A. (2020). COMPARATIVE EVALUATIONS OF CNN BASED NETWORKS FOR SKIN LESION CLASSIFICATION. *International Conferences Computer Graphics, Visualization, Computer Vision and Image Processing 2020*.
- Harahap, H., Rahman, S., & Zen, M. (2024). Analisis Existing Convolutional Neural Network Untuk Klasifikasi Usia Pengunjung Rumah Sakit: Studi Kasus Pemantauan Anak dan Dewasa. In *Journal of Computer Science and Information Technology E-ISSN* (Vol. 4, Number 1). <https://doi.org/10.47065/explorer.v4i1.881>
- Hisyam Pradhana, A., & Daniati, E. (2025). *Prosiding Seminar Nasional Teknologi dan Sistem Informasi (SITASI) 2025 Surabaya*.
- Hoots, D., Patel, Y., Charan, R., & Harsha Sankranthi, S. (2025). *Statistical Validation of VGG16 Transfer Learning for Pistachio Classification: A Rigorous Baseline Study*. <https://github.com/DavaughnHoots>
- Lisda, L., Kusriani, K., & Ariatmanto, D. (2023). Classification of Pistachio Nut Using Convolutional Neural Network. *Inform : Jurnal Ilmiah Bidang Teknologi Informasi Dan Komunikasi*, 8(1), 71–77. <https://doi.org/10.25139/inform.v8i1.5685>
- Marsheila Islami, B., & Nugroho, A. (2025). Analisis Algoritma KNN dan Penerapan SMOTE dalam Deteksi Dini Kanker Paru-Paru. *Jurnal Qua Teknika*, 15(2), 15. <https://doi.org/10.35457/quateknika.v15i02.4603>
- Nurona Cahya, F., Hardi, N., Riana, D., & Hadiani, S. (2021). *SISTEMASI: Jurnal Sistem Informasi Klasifikasi Penyakit Mata Menggunakan Convolutional Neural Network (CNN)*. <https://doi.org/10.32520/stmsi.v10i3.1248>
- Özkan, I. A., Köklü, M., & Saraçoğlu, R. (2021). Classification of pistachio species using improved k-NN classifier. *Progress in Nutrition*, 23(2). <https://doi.org/10.23751/pn.v23i2.9686>
- Ristyawan, A., & Nugroho, A. (2026). Analysis of Preprocessing Technique Combinations and Hyperparameter Tuning for Building a Reliable Random Forest–Based Stroke Prediction Model. *International Journal of Engineering and Computer Science Applications (IJECSA)*, 5(1), 9–22. <https://doi.org/10.30812/ijeCSA.v5i1.6080>
- Shelin Sahira, M., Dessela Putri, F., Ristyawan, A., & Daniati, E. (2024). Penggunaan Algoritma K-Nearest Neighbor Untuk Klasifikasi Data Diabetes Pada Wanita. *Prosiding Seminar Nasional Inovasi Teknologi*, 8, 2549–7952. <https://doi.org/doi.org/10.29407/inotek.v8i1.4981>

- Singh, D., Taspinar, Y. S., Kursun, R., Cinar, I., Koklu, M., Ozkan, I. A., & Lee, H. N. (2022). Classification and Analysis of Pistachio Species with Pre-Trained Deep Learning Models. *Electronics (Switzerland)*, 11(7). <https://doi.org/10.3390/electronics11070981>
- Solihin, A., Iskandar Mulyana, D., & Yel, M. B. (2022). *Klasifikasi Alat Musik Tradisional Papua menggunakan Metode Transfer Learning Dan Data Augmentasi*. <https://doi.org/10.47970/siskom-kb.v5i2.279>
- Sucipto, A., Nuraji, D., Enrique Lazuardi Ramadany, M., Vitasari, J., & Aji Widarso, R. (2024). *Implementasi Sistem Deteksi Otomatis pada Tanaman Cabai Rawit Menggunakan Convolutional Neural Network (CNN) Berbasis TensorFlow untuk Optimasi Pertanian Modern Implementation of an Automatic Detection System for Chili Plants Using a TensorFlow-Based Convolutional Neural Network (CNN) for Modern*. (1). <https://doi.org/10.25047/nacia.v2i1.254>
- Sucipto, S., Dwi Prasetya, D., & Widiyaningtyas, T. (2024). Educational Data Mining: Multiple Choice Question Classification in Vocational School. *MATRIK : Jurnal Manajemen, Teknik Informatika Dan Rekayasa Komputer*, 23(2), 379–388. <https://doi.org/10.30812/matrik.v23i2.3499>
- Trihardianingsih L., & Permatasari H. (2025). *PENGARUH OPTIMIZER TERHADAP AKURASI KLASIFIKASI PISTACHIO MENGGUNAKAN MOBILENETV2*.
- Triwardani, A. A., Ulfiah, M. H., Ristyawan, A., & Daniati, E. (2024). Prediksi Kematian Akibat Gagal Jantung Menggunakan Algoritma K-Nearest Neighbors. In *Agustus* (Vol. 8). Online.
- Uzer, M. S. (2025). Deep Learning-Based Classification Consisting of Pre-Trained Models and Proposed Model Using K-Fold Cross-Validation for Pistachio Species. *Applied Sciences (Switzerland)*, 15(8). <https://doi.org/10.3390/app15084516>
- Vierisyah, A., & Maulana Fajri, R. (2023). KLASIFIKASI KANKER PARU PARU MENGGUNAKAN CNN DENGAN 5 ARSITEKTUR. *Jini Global : Journal of Intelligent Networks and IoT Global*. <https://doi.org/doi.org/10.36982/jinig.v1i2.3643>