



Analisis Pengolahan Citra Produk Pangan Olahan Kemasan Untuk Deteksi Kehalalan Menggunakan Metode *Optical Character Recognition* (Ocr) Dan Klasifikasi *Support Vector Machine* (SVM) Serta *Convolutional Neural Network* (CNN)

*Muhamad Quraisy¹, Agung Budi Susanto², Tukiya³

^{1,2,3}) Teknik Informatika, Pascasarjana, Universitas Pamulang, Kota Tangerang Selatan, Banten

Email: ¹muhammadquraisy.26@gmail.com, ²dosen02680@unpam.ac.id, ³dosen02711@unpam.ac.id

ABSTRACT

This study aims to detect the halal status of packaged processed food products by integrating Optical Character Recognition (OCR) technology with machine learning-based classification methods. The growing number of packaged food products available in the market, together with the difficulty consumers experience in understanding ingredient information printed on product labels, has increased the need for an automated halal identification approach. Previous studies have applied OCR to extract text from food packaging and Named Entity Recognition (NER) to identify non-halal ingredients. However, comparative studies evaluating the performance of Support Vector Machine (SVM) and Convolutional Neural Network (CNN) for classifying halal status based on OCR-extracted text remain limited. In this study, product packaging images collected through direct documentation and official sources were preprocessed using grayscale conversion, Otsu thresholding, and 2× resizing before text extraction with Tesseract OCR. The extracted text was subsequently processed through lowercasing, normalization, stopword removal, and non-halal entity detection using spaCy-based NER. The labelled data were then classified using SVM with TF-IDF representation and CNN with token-sequence representation under 80:20 and 70:30 data split scenarios. The results show that both models achieved excellent classification performance, with SVM consistently providing better stability and predictive accuracy than CNN. The highest performance was obtained by SVM, achieving an accuracy of 99.36% and a macro F1-score of 97.05% in the 80:20 scenario. These findings demonstrate that the proposed approach can support automatic halal identification of packaged processed food products and may assist consumers, researchers, and related institutions in halal product verification.

Keywords: OCR; SVM; CNN; text classification; kehalalan; haram; NER

ABSTRAK

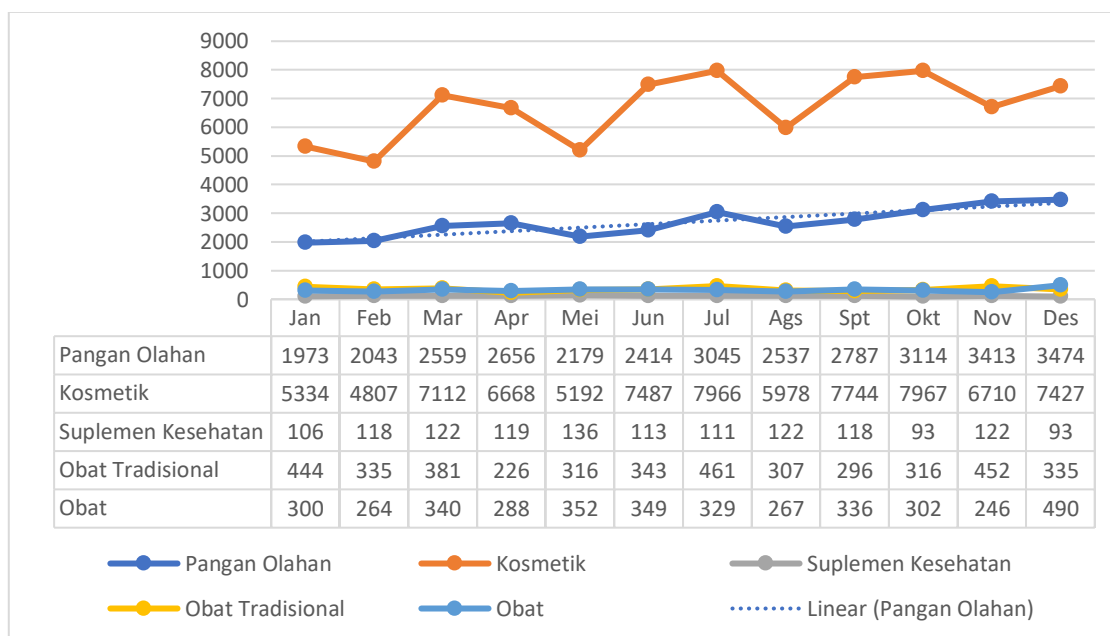
Penelitian ini bertujuan untuk mendeteksi status kehalalan produk pangan olahan kemasan dengan mengintegrasikan teknologi *Optical Character Recognition* (OCR) dan metode klasifikasi berbasis machine learning. Semakin banyaknya produk pangan olahan yang beredar di pasaran serta kesulitan konsumen dalam memahami informasi komposisi yang tercantum pada kemasan mendorong perlunya suatu pendekatan otomatis untuk membantu proses identifikasi kehalalan produk. Penelitian terdahulu telah memanfaatkan OCR untuk mengekstraksi teks dari kemasan pangan dan *Named Entity Recognition* (NER) untuk mendeteksi bahan non-halal. Namun, penelitian yang membandingkan kinerja *Support Vector Machine* (SVM) dan *Convolutional Neural Network* (CNN) dalam klasifikasi status kehalalan berdasarkan teks hasil OCR masih relatif terbatas. Pada penelitian ini, citra kemasan produk yang diperoleh melalui dokumentasi langsung dan sumber resmi diproses menggunakan *grayscale*, *Otsu thresholding*, dan *resize* 2× sebelum dilakukan ekstraksi teks dengan Tesseract OCR. Teks hasil ekstraksi kemudian melalui proses *lowercasing*, normalisasi, *stopword removal*, serta deteksi entitas bahan non-halal menggunakan NER berbasis spaCy. Data yang telah diberi label selanjutnya diklasifikasikan menggunakan SVM dengan representasi TF-IDF dan CNN dengan representasi token-sekuens pada skenario pembagian data 80:20 dan

70:30. Hasil penelitian menunjukkan bahwa kedua model mampu mengklasifikasikan status kehalalan dengan performa yang sangat baik, namun SVM memberikan stabilitas dan akurasi yang lebih tinggi dibandingkan CNN. Performa terbaik diperoleh SVM dengan akurasi sebesar 99,36% dan *macro F1-score* sebesar 97,05% pada skenario 80:20. Temuan ini menunjukkan bahwa pendekatan yang diusulkan berpotensi mendukung proses identifikasi kehalalan produk pangan olahan kemasan secara otomatis serta memberikan manfaat bagi konsumen, peneliti, dan lembaga terkait dalam mendukung proses verifikasi produk halal.

Kata kunci: OCR; SVM; CNN; klasifikasi teks; kehalalan; haram; NER

1. PENDAHULUAN

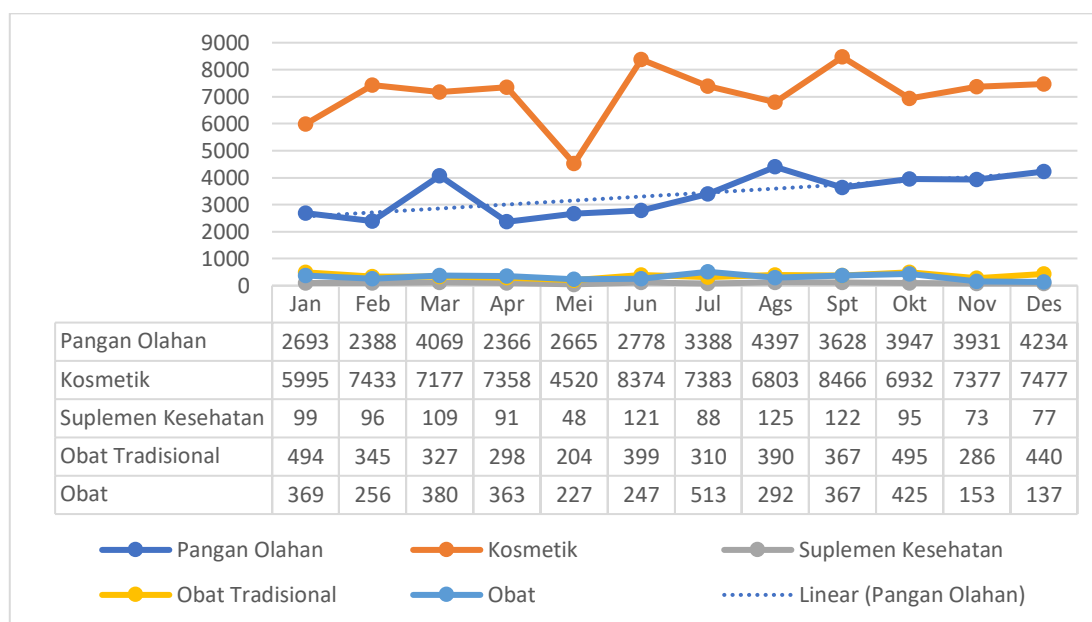
Perkembangan teknologi digital telah membawa dampak besar dalam berbagai sektor termasuk dalam industri pangan olahan. Produk pangan olahan kemasan dari berbagai negara masuk dengan mudah ke pasar Indonesia dan memberikan tantangan tersendiri bagi umat Muslim untuk memastikan produk yang dikonsumsi sesuai prinsip kehalalan. Konsep halal mengacu pada produk yang sesuai dengan hukum Syariah, yang berarti produk tersebut diperbolehkan untuk dikonsumsi oleh umat Muslim.



Grafik 1. Statistik Produk Teregistrasi (2021)

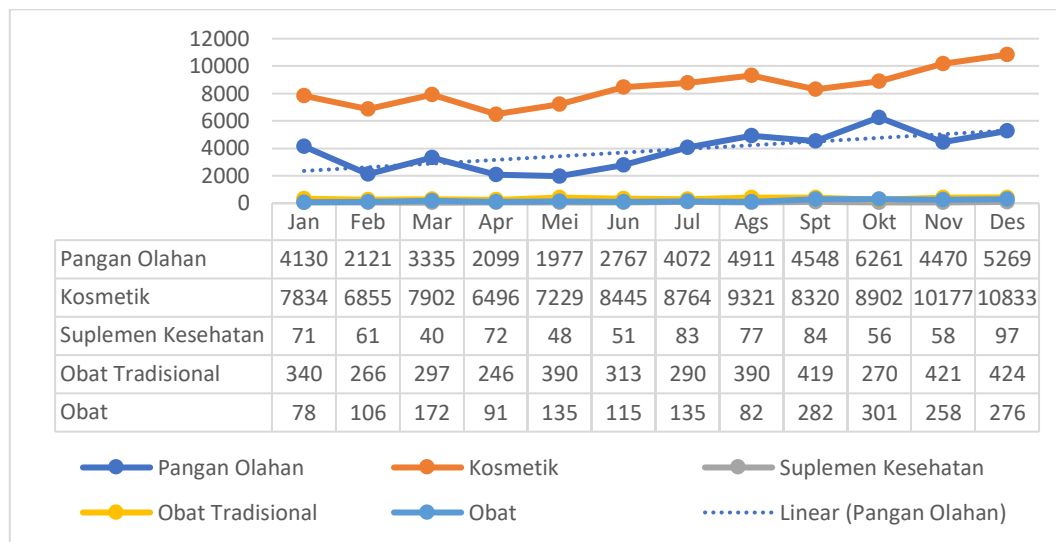
Grafik 1 menunjukkan jumlah produk yang teregistrasi di BPOM sepanjang tahun 2021 pada lima kategori, yaitu pangan olahan, kosmetik, suplemen kesehatan, obat tradisional, dan obat. Terlihat bahwa kategori kosmetik memiliki jumlah registrasi tertinggi hampir di setiap bulan, sedangkan pangan olahan menempati urutan berikutnya dengan kecenderungan meningkat pada paruh akhir tahun. Sementara itu, jumlah registrasi pada kategori suplemen kesehatan, obat tradisional, dan obat relatif lebih rendah

serta cenderung stabil dibandingkan dua kategori lainnya. Kondisi ini menunjukkan bahwa produk pangan olahan merupakan salah satu kelompok produk yang terus mengalami pertumbuhan registrasi dan menjadi objek yang relevan dalam penelitian mengenai identifikasi status kehalalan produk.



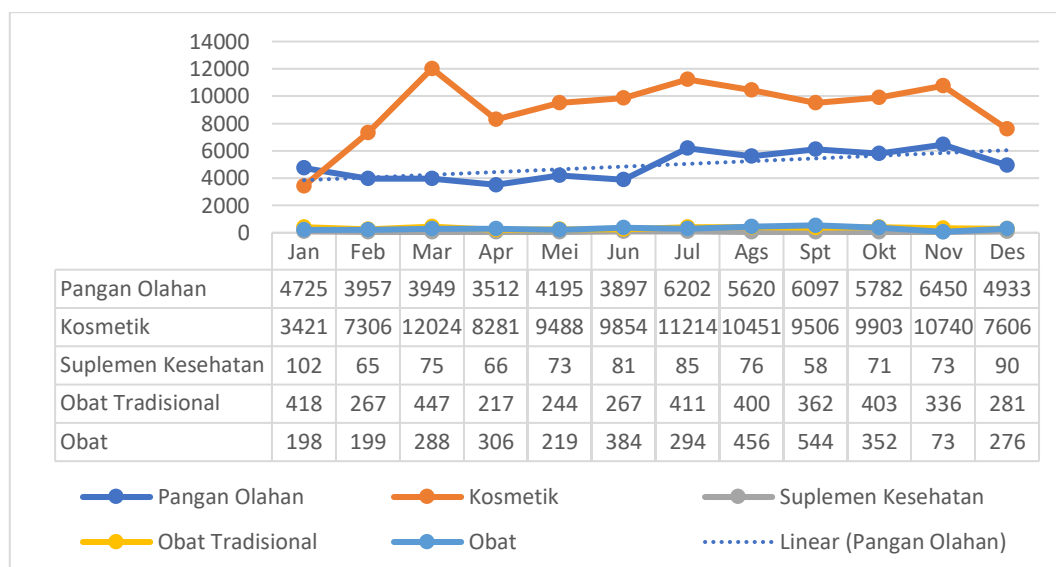
Grafik 2. Statistik Produk Teregistrasi (2022)

Grafik 2 memperlihatkan perkembangan jumlah produk teregistrasi sepanjang tahun 2022. Dibandingkan tahun sebelumnya, jumlah registrasi pada kategori pangan olahan menunjukkan kecenderungan meningkat, meskipun masih mengalami fluktuasi pada beberapa bulan. Kategori kosmetik tetap mendominasi jumlah registrasi, sedangkan kategori suplemen kesehatan, obat tradisional, dan obat relatif tidak mengalami perubahan yang signifikan. Peningkatan jumlah produk pangan olahan tersebut mengindikasikan bahwa informasi komposisi pada kemasan menjadi semakin penting untuk membantu konsumen dalam mengenali kandungan produk yang dikonsumsi.



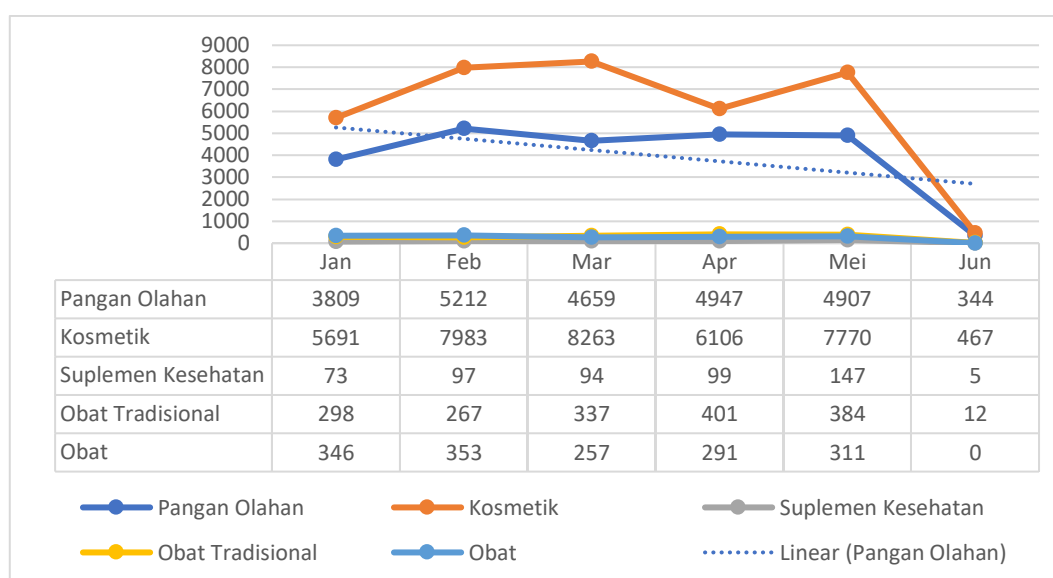
Grafik 3. Statistik Produk Teregistrasi (2023)

Grafik 3 menunjukkan bahwa jumlah registrasi produk pangan olahan pada tahun 2023 masih mempertahankan tren yang relatif meningkat dengan beberapa kenaikan pada pertengahan hingga akhir tahun. Pada periode yang sama, kategori kosmetik kembali mencatat jumlah registrasi tertinggi dibandingkan kategori lainnya. Meskipun jumlah registrasi kategori lain relatif stabil, bertambahnya produk pangan olahan yang beredar memperlihatkan semakin besarnya kebutuhan terhadap metode yang mampu membantu proses identifikasi informasi komposisi bahan secara cepat dan akurat.



Grafik 4. Statistik Produk Teregistrasi (2024)

Grafik 4 memperlihatkan perkembangan registrasi produk selama tahun 2024. Jumlah registrasi pangan olahan masih menunjukkan pola yang cukup konsisten, dengan peningkatan pada beberapa bulan tertentu. Kategori kosmetik tetap menjadi kategori dengan jumlah registrasi terbesar, sedangkan kategori suplemen kesehatan, obat tradisional, dan obat menunjukkan variasi yang relatif kecil. Data tersebut menunjukkan bahwa produk pangan olahan terus mengalami pertumbuhan sehingga kebutuhan akan proses identifikasi kehalalan berbasis teknologi menjadi semakin relevan.



Grafik 5. Statistik Produk Juni Teregistrasi (2025)

Grafik 5 menyajikan data registrasi produk hingga Juni 2025. Data pada grafik ini hanya mencakup periode Januari sampai Juni sehingga tidak dapat dibandingkan secara langsung dengan data tahunan sebelumnya. Meskipun demikian, jumlah registrasi produk pangan olahan masih berada pada tingkat yang cukup tinggi dibandingkan kategori suplemen kesehatan, obat tradisional, dan obat. Temuan ini memperlihatkan bahwa produk pangan olahan tetap menjadi salah satu kategori dengan jumlah registrasi yang besar, sehingga diperlukan pendekatan yang mampu membantu proses identifikasi status kehalalan berdasarkan informasi komposisi yang tercantum pada kemasan produk.

Berdasarkan data BPOM hingga Juni 2025, tercatat sebanyak 280.767 produk pangan olahan telah beredar di Indonesia. Di sisi lain, data BPJPH per 5 November 2024 menunjukkan bahwa 5.492.791 produk telah memperoleh sertifikat halal. Meskipun

jumlah produk bersertifikat halal terus meningkat, konsumen masih menghadapi berbagai kendala dalam memahami informasi komposisi pada kemasan. Permasalahan tersebut antara lain disebabkan oleh kualitas cetak yang kurang baik, kontras warna teks yang rendah, ukuran huruf yang kecil, maupun penggunaan istilah asing yang sulit dipahami. Kondisi ini menunjukkan bahwa proses identifikasi status kehalalan secara otomatis masih memiliki peran penting dalam membantu konsumen memperoleh informasi yang lebih mudah diakses.

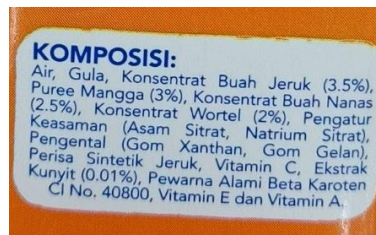
Penelitian sebelumnya telah memanfaatkan OCR untuk mengekstraksi teks komposisi bahan dan NER untuk mendeteksi kandungan haram. Namun, sangat sedikit studi yang secara langsung membandingkan performa *Support Vector Machine* (SVM) dan *Convolutional Neural Network* (CNN) dalam klasifikasi status halal-haram berbasis teks hasil OCR dari kemasan produk. Penelitian ini mengisi celah tersebut dengan mengembangkan sistem klasifikasi status kehalalan produk pangan olahan berbasis teks hasil OCR, sekaligus membandingkan kinerja SVM dan CNN. Proses klasifikasi dilakukan dengan dukungan *preprocessing* citra (*grayscale*, *Otsu thresholding*, *resize*) dan teks (*normalisasi*, *stopword removal*) serta analisis entitas non-halal menggunakan spaCy. Studi ini diharapkan dapat menghasilkan model klasifikasi yang lebih akurat dan konsisten, sekaligus memberikan kontribusi nyata dalam mendukung literasi halal bagi konsumen dan lembaga sertifikasi.

2. METODE

2.1. Data dan Pengumpulan Data

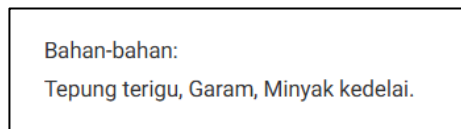
Pengumpulan data dilakukan melalui dua sumber yaitu observasi langsung (primer) dan studi literatur (sekunder). Data utama yang digunakan dalam penelitian ini berupa:

- a. Data primer sebanyak 509 citra produk pangan olahan kemasan diperoleh melalui pemotretan langsung di *supermarket*, toko agen. Foto difokuskan pada bagian komposisi bahan, diambil dengan ponsel kamera resolusi tinggi (HD) dengan pencahayaan yang cukup. Contoh gambar data primer di bawah ini.



Gambar 1. Contoh Foto Langsung Produk Pangan Olahan Kemasan

- b. Data sekunder sebanyak 1.060 citra dikumpulkan dari situs resmi produsen dan platform dagang elektronik. Pemilihan gambar mempertimbangkan resolusi dan keterbacaan informasi komposisi bahan.

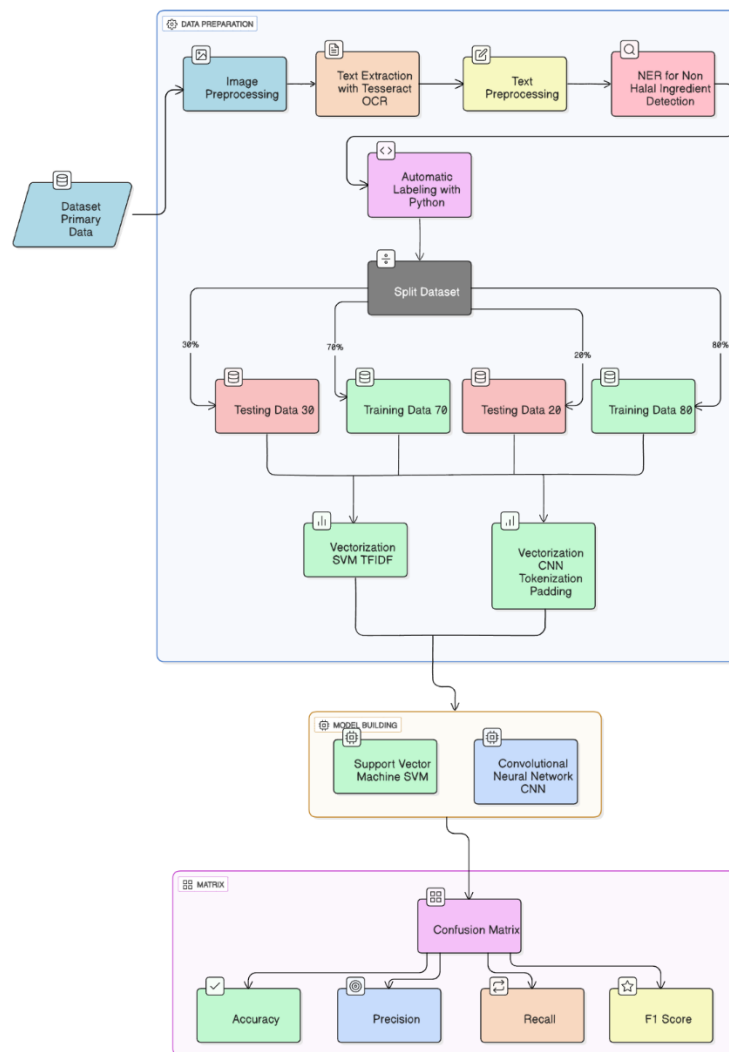


Gambar 2. Contoh Citra Produk Pangan Olahan Kemasan (Sekunder)

Total dataset yang diperoleh sebanyak 1.569 citra. Setelah dilakukan preprocessing dan augmentasi, data menjadi 4.707 citra siap proses OCR. File Excel pendukung berisi nama produk, kategori, dan daftar bahan digunakan untuk mempermudah proses pelabelan.

2.2. Tahapan Penelitian

Penelitian ini menggunakan tahapan sistematis dari tahap awal pengumpulan data hingga klasifikasi.



- Tahap pertama adalah pengumpulan data primer berupa citra produk pangan olahan kemasan yang diperoleh melalui dokumentasi langsung menggunakan kamera ponsel beresolusi HD di supermarket dan toko agen. Data primer digunakan untuk memperoleh citra produk dengan kondisi nyata sehingga dapat merepresentasikan variasi kemasan yang ditemui oleh konsumen.
- Image Processing*, setelah data dikumpulkan, seluruh citra diproses menggunakan teknik *grayscale*, *Otsu thresholding*, dan *resize 2×*. Tahap ini bertujuan meningkatkan kualitas visual citra sehingga tulisan pada komposisi bahan menjadi lebih jelas dan mudah dibaca oleh sistem OCR.
- Text Extraction with Tesseract OCR*, citra hasil *preprocessing* kemudian diproses menggunakan Tesseract OCR untuk mengubah informasi berupa gambar menjadi

data teks. Tahap ini dilakukan agar komposisi bahan yang terdapat pada kemasan dapat dianalisis pada proses berikutnya.

- d. *Text Preprocessing*, teks hasil OCR masih mengandung berbagai karakter yang tidak seragam sehingga dilakukan proses *lowercasing*, normalisasi, dan *stopword removal*. Tujuannya adalah menghasilkan data teks yang lebih bersih dan konsisten sebelum dilakukan proses analisis.
- e. *NER for Non-Halal Ingredient Detection*, teks yang telah diproses kemudian dianalisis menggunakan *Named Entity Recognition* (NER) berbasis spaCy untuk mendeteksi keberadaan bahan yang termasuk dalam daftar bahan non-halal. Tahap ini menjadi dasar dalam proses pelabelan data.
- f. *Automatic Labeling with Python*, hasil deteksi entitas digunakan untuk memberikan label secara otomatis menggunakan Python. Produk yang mengandung entitas non-halal diberi label 1 (haram), sedangkan produk yang tidak mengandung entitas tersebut diberi label 0 (halal).
- g. *Split Dataset*, dataset yang telah diberi label kemudian dibagi menjadi dua skenario pengujian, yaitu 80:20 dan 70:30. Pembagian ini bertujuan membandingkan pengaruh proporsi data latih dan data uji terhadap performa model.
- h. *Training Data 70*, sebanyak 70% data digunakan sebagai data pelatihan pada skenario pertama. Data ini digunakan untuk membangun model klasifikasi.
- i. *Testing Data 30*, sebanyak 30% data digunakan sebagai data pengujian pada skenario pertama untuk mengevaluasi kemampuan model terhadap data yang belum pernah dipelajari.
- j. *Training Data 80*, pada skenario kedua, 80% data digunakan sebagai data pelatihan agar model memperoleh informasi yang lebih banyak selama proses pembelajaran.
- k. *Testing Data 20*, sebanyak 20% data digunakan sebagai data pengujian untuk mengukur performa model pada skenario kedua.
- l. *Vectorization SVM TF-IDF*, data teks pada jalur SVM diubah menjadi representasi numerik menggunakan *TF-IDF Vectorizer* sehingga setiap kata memiliki bobot berdasarkan tingkat kemunculannya pada dokumen.

- m. *Vectorization CNN (Tokenization & Padding)*, pada jalur CNN, data teks diproses menggunakan *Tokenization* dan *Padding* agar setiap dokumen memiliki panjang sekuens yang seragam sebelum dimasukkan ke jaringan CNN.
- n. *Support Vector Machine (SVM)*, data hasil vektorisasi TF-IDF digunakan untuk membangun model klasifikasi menggunakan algoritma *Support Vector Machine (SVM)*.
- o. *Convolutional Neural Network (CNN)*, data hasil tokenisasi digunakan untuk membangun model *Convolutional Neural Network (CNN)* yang terdiri atas *Embedding Layer*, *Conv1D*, *Global Max Pooling*, dan *Dense Layer*.
- p. *Confusion Matrix*, hasil prediksi kedua model dianalisis menggunakan confusion matrix untuk mengetahui jumlah prediksi benar dan salah pada masing-masing kelas.
- q. *Accuracy* digunakan untuk mengukur proporsi keseluruhan prediksi yang benar terhadap seluruh data pengujian.
- r. *Precision* digunakan untuk mengetahui ketepatan model dalam memprediksi kelas haram.
- s. *Recall* digunakan untuk mengukur kemampuan model dalam mendeteksi seluruh data yang benar-benar termasuk kelas haram.
- t. *F1-Score* digunakan sebagai ukuran keseimbangan antara nilai precision dan recall sehingga memberikan gambaran performa model secara lebih menyeluruh.

2.3. Teknik Analisis

Penelitian ini menggunakan pendekatan machine learning untuk mengklasifikasikan status kehalalan produk pangan olahan kemasan berdasarkan daftar komposisi bahan. Analisis dilakukan melalui empat tahap utama:

- a. *Data Preparation*: Dari 1.569 citra (509 primer, 1.060 sekunder) dilakukan preprocessing citra (*grayscale*, *Otsu thresholding*, *resize 2x*, augmentasi) sehingga menghasilkan 4.707 citra. Teks komposisi bahan diekstrak menggunakan *Tesseract OCR*, dibersihkan (*lowercasing*, normalisasi, *stopword removal*), kemudian dilakukan deteksi bahan haram dengan *spaCy NER*. Data

dilabeli otomatis menjadi halal (0) atau haram (1) dan dibagi dalam skenario 80:20 dan 70:30.

- b. Analisis Model SVM: Data teks hasil *preprocessing* direpresentasikan menggunakan TF-IDF. Model *LinearSVC* digunakan dengan parameter $C=1.0$, $class_weight='balanced'$, dan $max_iter=5000$. Model dilatih dan dievaluasi pada kedua skenario pembagian data dengan metrik akurasi, *precision*, *recall*, *F1-score*, serta *confusion matrix*.
- c. Analisis Model CNN: Data teks dikonversi dengan *Tokenizer* dan *Padding* (panjang sekuens 100 kata). Arsitektur CNN terdiri dari *Embedding layer* (dimensi 100), *Conv1D* (128 *filter*, *kernel size* 5), *Global Max Pooling*, *Dropout* (0.5 dan 0.3), dan *Dense layer* dengan *sigmoid activation* untuk klasifikasi biner. Model dilatih dengan *binary_crossentropy loss*, *optimizer Adam*, serta *EarlyStopping*.

2.4. Evaluasi Model

Evaluasi dilakukan untuk menilai performa algoritma klasifikasi SVM dan CNN dalam memprediksi status kehalalan produk pangan olahan berdasarkan teks hasil ekstraksi OCR. Data uji berasal dari pembagian dataset ke dalam skenario 80:20 dan 70:30. Lima metrik evaluasi digunakan untuk mengukur kinerja model:

- a. Akurasi (*Accuracy*) mengukur proporsi prediksi benar dari keseluruhan data uji.

$$Akurasi = \left(\frac{TP + TN}{TP + FP + FN + TN} \right) \times 100\% \quad (1)$$

- b. Presisi (*Precision*) menghitung proporsi data yang diprediksi positif (haram) dan benar-benar positif.

$$Precision = \left(\frac{TP}{TP + FP} \right) \times 100\% \quad (2)$$

- c. Recall (*Sensitivity*) mengukur kemampuan model mendeteksi semua data positif (haram).

$$Recall = \left(\frac{TP}{TP + FN} \right) \times 100\% \quad (3)$$

- d. *F1-Score* memberikan keseimbangan antara presisi dan recall melalui rata-rata harmonisasi.

$$F1 - Score = 2x \left(\frac{Recall \times Precision}{Recall + Precision} \right) \quad (4)$$

- e. *Confusion Matrix* digunakan untuk menggambarkan distribusi prediksi model dalam empat kategori: True Positive (TP), False Positive (FP), True Negative (TN), dan False Negative (FN).

Tabel 1. Confusion Matrix

		Prediksi	
		Halal (0)	Haram (1)
Aktual	Halal (0)	TN	FP
	Haram (1)	FN	TP

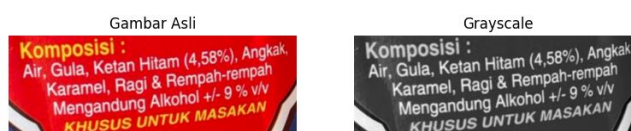
Perhitungan metrik dilakukan menggunakan hasil prediksi model pada data uji. Confusion matrix divisualisasikan dengan Matplotlib dan Seaborn untuk menampilkan perbandingan performa SVM dan CNN secara visual.

3. HASIL DAN PEMBAHASAN

3.1. Proses dan Prosedur

Pada bagian proses dan prosedur terdiri dari empat tahapan, yaitu:

- Akuisisi dan spesifikasi citra produk dengan total 1569 citra dari data primer dan sekunder. Adapun spesifikasi citra yang digunakan menggunakan format file .jpg/.jpeg/.png, rata-rata resolusi gambar yang digunakan yaitu 72dpi dan 32 Bit *depth* dan warna RGB.
- Preprocessing* Citra dilakukan untuk memastikan bahwa kualitas visual citra yang optimal serta dimensi yang seragam sebelum masuk tahap proses ekstraksi teks. Tahapan *preprocessing* citra yang diterapkan adalah *grayscale*, *otsu thresholding*, dan *resize 2x*). Hasil dapat dilihat pada gambar di bawah ini.





Gambar 3. Preprocessing Citra (Grayscale, Otsu Thresholding, Resize 2x)

- c. Ekstraksi teks dengan *Tesseract* OCR difokuskan pada bagian komposisi bahan pada produk pangan olahan kemasan. Hasil ekstraksi teks disimpan dalam format tabular.

Tabel 2. Contoh Hasil Ekstraksi Teks dengan Tesseract OCR

Filename	Ingredients
primer_koepoekoep oehitam3_resized	"KOMPOSISI / INGREDIENTS: Pemanis Alami Sorbitol Sirup, Pengemulsi Gliserin, Pewarna Sintetik (Kuning FCF Cl. 15985, indigotin Cl. 73015, Karmoisin Cl. 14720, Biru Berlian FCF Cl. 42090) - Natural Sweetener Sorbitol Syrup, _ Emulsifier Glycerin, Synthetic Coloring (Sunset Yellow FCF Cl. 15985, Indigotine Cl. 73015, Carmoisine Cl. 14720, Brilliant Blue FCF Cl, 42090) "
sekunder_orangtuab eraskencur_resized	"KOMPOSISI: AIR, HASIL FERMENTASI GULA MERAH, SPIRIT GULA MERAH, GULA, BERAS, KENCUR MADU, KUNING TELUR, AGAR-AGAR, BREM. "

- d. *Preprocessing* Teks terhadap total data ekstraksi dengan *tesseract* OCR sebanyak 4707 baris dari dua atribut tabular dilakukan dalam enam tahapan, sebagaimana disebutkan dalam perancangan penelitian menggunakan Google *Colaboratory*.

Tabel 3. Contoh Hasil *Preprocessing* Teks

Filename	Ingredients	Processed_Ingredients
primer_koep oekoepoehit am3_resized	"KOMPOSISI / INGREDIENTS: Pemanis Alami Sorbitol Sirup, Pengemulsi Gliserin, Pewarna Sintetik (Kuning FCF Cl. 15985, indigotin Cl. 73015, Karmoisin Cl. 14720, Biru Berlian FCF Cl. 42090) - Natural Sweetener Sorbitol Syrup, _ Emulsifier Glycerin, Synthetic Coloring (Sunset Yellow FCF Cl. 15985, Indigotine Cl. 73015, Carmoisine Cl. 14720, Brilliant Blue FCF Cl, 42090) "	komposisi ingredients pemanis alami sorbitol sirup pengemulsi gliserin pewarna sintetik kuning fcf cl 15985 indigotin cl 73015 karmoisin cl 14720 biru berlian fcf cl 42090 natural sweetener sorbitol sirup emulsifier glycerin synthetic coloring sunset yellow fcf cl 15985 indigotine cl 73015 carmoisine cl 14720 brilliant blue fcf cl 42090

sekunder_or	"KOMPOSISI: AIR, HASIL FERMENTASI	komposisi air hasil fermentasi gula merah
angtuaberask	GULA MERAH, SPIRIT GULA MERAH,	spirit gula merah gula beras kencur madu
encur_resize	GULA, BERAS, KENCUR MADU, KUNING	kuning telur agaragar brem
d	TELUR, AGAR-AGAR, BREM. "	

Tabel di atas menunjukkan proses *lowercasing*, normalisasi, dan *stopword removal* yang diterapkan pada data hasil ekstraksi teks sebelum dianalisis lebih lanjut. Pada tabel tersebut menghasilkan teks bersih 100% di bagian *Processed_Ingredients*.

- e. Deteksi Entitas dan Pelabelan Halal-Haram dilakukan dengan pendekatan *Named Entity Recognition* (NER) dengan *EntityRuler* dari spaCy. Daftar entitas haram yang digunakan termasuk “alcohol”, “babi”, dan turunan lainnya dikompilasi secara manual berdasarkan dari studi literatur. Daftar tersebut dijadikan pola pencocokan yang digunakan oleh spaCy untuk menandai entitas haram dan dapat dilihat pada gambar di bawah ini.

```
# Daftar bahan non-halal yang akan dideteksi
bahan_nonhalal = [
    "alkohol", "alcohol", "babi", "bacon", "sake", "brandy", "lard", "porcine", "pork", "daging babi",
    "fermentasi", "fermentasi anggur", "anggur", "wine", "vinegar", "beer", "ethanol", "enzyme babi",
    "darah", "cuka anggur", "whisky", "angkak"
]
```

Gambar 4. EntityRuler

Apabila dalam teks ditemukan entitas dari daftar tersebut, maka sistem akan secara otomatis memberikan label “haram” pada data tersebut. Sebaliknya, jika tidak ditemukan entitas non-halal, data akan dilabeli sebagai “halal”. Label ini bersifat biner, dengan nilai 1 untuk haram dan 0 untuk halal. Hasil nya dapat dilihat pada contoh tabel di bawah ini.

Tabel 4. Contoh Hasil Deteksi Entitas Bahan Haram

Filename	Ingredients	Processed_Ingredients	Detected_NonHalal	Label
	"KOMPOSISI / INGREDIENTS:	komposisi ingredients pemanis	halal	halal
primer_k	Pemanis Alami Sorbitol Sirup,	alami sorbitol sirup pengemulsi		
oepoekoe	Pengemulsi Gliserin, Pewarna	gliserin pewarna sintetis kuning		
poehitam	Sintetik (Kuning FCF Cl. 15985,	fcf cl 15985 indigotin cl 73015		
3_resized	indigotin Cl. 73015, Karmoisin Cl.	karmoisin cl 14720 biru berlian		
	14720, Biru Berlian FCF Cl. 42090)	fcf cl 42090 natural sweetener		

	- Natural Sweetener Sorbitol Syrup,	sorbitol	syrup	emulsifier		
	_ Emulsifier Glycerin, Synthetic	glycerin	synthetic	coloring		
	Coloring (Sunset Yellow FCF Cl.	sunset	yellow	fcf	cl	15985
	15985, Indigotine Cl. 73015,	indigotine	cl	73015	carmoisine	
	Carmoisine Cl. 14720, Brilliant Blue	cl	14720	brilliant	blue	fcf
	FCF Cl, 42090) "	42090				
	"KOMPOSISI: AIR, HASIL	komposisi	air	hasil fermentasi	fermentasi	haram
sekunder_	FERMENTASI GULA MERAH,	gula	merah	spirit	gula	merah
orangtuab	SPIRIT GULA MERAH, GULA,	gula	beras	kencur	madu	kuning
eraskencu	BERAS, KENCUR MADU,	telur	agaragar	brem		
r_resized	KUNING TELUR, AGAR-AGAR,					
	BREM. "					

Setelah proses deteksi entitas bahan non-halal selesai dilakukan, setiap data diberikan label sesuai hasil identifikasi, yaitu halal atau haram. Agar data dapat digunakan pada tahap pelatihan model klasifikasi, label tersebut selanjutnya dikonversi ke dalam bentuk numerik (*binary labeling*), dengan nilai 0 merepresentasikan produk halal dan nilai 1 merepresentasikan produk haram. Proses konversi ini dilakukan secara otomatis menggunakan Python sehingga format label menjadi sesuai dengan kebutuhan algoritma machine learning. Hasil konversi label ditunjukkan pada Gambar 5.

Filename	Ingredients	Processed_Ingredients	Detected_NonHalal	Label	Label_Numeric
sekunder_S-26Promil1Gold_resized	: eel Okt Ey/nKosa (48%), Mica Nabe Bexedelai...	eel okt ey kosa 48 mica nabe bexedelai cleat b...	halal	halal	0
sekunder_Dancow1*ImunutrinVanila_gray	Komposisi/nPodoton Sumy fam Bubvk Skim, Laka...	komposisi podoton sumy fam bubvk skim laka sub...	halal	halal	0
sekunder_Dancow1*ImunutrinVanila_otsu	Komposisi/nPodoton Som) tim Bubok \$k, Lahde 8...	komposisi podoton som tim bubok k lahde 80 we ...	halal	halal	0
sekunder_Dancow1*ImunutrinVanila_resized	Leeda stone =/n/nPodaton Sum fam Bubyk Skim, ...	leeda stone podaton sum fam bubyk skim laktaag...	halal	halal	0
sekunder_orangtuaberaskencur_gray	KOMPOSISI'/n/nAIR, HASIL FERMENTASI GULA MERAH...	komposisi air hasil fermentasi gula merah spir...	fermentasi	haram	1

Gambar 5. Hasil Konversi Label

Berdasarkan Gambar 5, terlihat bahwa setiap data memiliki dua bentuk label, yaitu label kategorikal (*Label*) dan label numerik (*Label_Numeric*). Sebagai contoh, data dengan label halal dikonversi menjadi nilai 0, sedangkan data yang teridentifikasi mengandung bahan non-halal, seperti fermentasi pada contoh data, dikonversi menjadi nilai 1. Representasi numerik tersebut digunakan sebagai target (*target label*) pada proses pelatihan dan pengujian model SVM maupun CNN sehingga kedua algoritma dapat melakukan klasifikasi secara konsisten terhadap status kehalalan produk.

- f. Pembagian Dataset (*Split*), data yang telah diberi label biner berjumlah 4707 dibagi menjadi dua skenario untuk keperluan pelatihan dan pengujian, berikut

adalah skema pembagian data: Skema 1: 80% data latih (3742) dan 20% data uji (936). Skema 2: 70% data latih (3274) dan 30% data uji (1404).

3.2. Analisis Klasifikasi SVM

Pengujian klasifikasi dilakukan pada dua skema pembagian data, yaitu 80:20 dan 70:30, menggunakan algoritma Support Vector Machine (SVM) yang diimplementasikan di Google Colaboratory. Data teks hasil preprocessing diubah menjadi representasi numerik dengan TF-IDF Vectorizer, sehingga setiap kata diberikan bobot berdasarkan frekuensi kemunculannya dalam dokumen dan keseluruhan korpus. Contoh term dengan skor tertinggi seperti “milk” (0.2218) menunjukkan kata unik yang relevan untuk pemisahan kategori. Model SVM dibangun menggunakan LinearSVC dengan parameter utama: $C=1.0$ (mengontrol *margin* dan kesalahan), $max_iter=5000$ (memastikan konvergensi), $class_weight='balanced'$ (mengatasi ketidakseimbangan kelas), $dual=False$ (untuk efisiensi karena fitur lebih banyak dari sampel).

Model *Support Vector Machine* (SVM) dievaluasi menggunakan dua skenario pembagian data, yaitu 80:20 dan 70:30. Evaluasi dilakukan untuk mengetahui kemampuan model dalam mengklasifikasikan status kehalalan produk berdasarkan teks hasil ekstraksi OCR. Parameter yang digunakan meliputi *accuracy*, *precision*, *recall*, *F1-score*, serta nilai *macro average* dan *weighted average* sehingga performa model dapat dianalisis secara menyeluruh. Hasil evaluasi untuk kedua skenario tersebut disajikan pada Tabel 5.

Tabel 5. Metrik Evaluasi Model SVM

Metrik Evaluasi	Nilai 80:20	Nilai 70:30
<i>Accuracy</i>	0.9936	0.9922
<i>Precision</i>	0.9943 (Halal)	0.9917 (Halal)
	0.99808 (Haram)	1.0000 (Haram)
<i>Recall</i>	0.9989 (Halal)	1.0000 (Halal)
	0.9107 (Haram)	0.8690 (Haram)
<i>F1-Score</i>	0.9966 (Halal)	0.9959 (Halal)
	0.9444 (Haram)	0.9299 (Haram)
<i>Macro Avg.</i>	<i>Precision</i> : 0.9876	<i>Precision</i> : 0.9959
	<i>Recall</i> : 0.9548	<i>Recall</i> : 0.9345
	<i>F1-Score</i> : 0.9705	<i>F1-Score</i> : 0.9629

	<i>Precision: 0.9935</i>	<i>Precision: 0.9922</i>
<i>Weighted Avg.</i>	<i>Recall: 0.9936</i>	<i>Recall: 0.9922</i>
	<i>F1-Score: 0.9935</i>	<i>F1-Score: 0.9919</i>

Berdasarkan Tabel 5, model SVM menunjukkan performa yang sangat baik pada kedua skenario pembagian data. Pada skenario 80:20, model memperoleh nilai *accuracy* sebesar 99,36%, sedangkan pada skenario 70:30 nilai *accuracy* mencapai 99,22%. Perbedaan nilai tersebut relatif kecil, yang menunjukkan bahwa model mampu mempertahankan kinerja yang konsisten meskipun proporsi data pelatihan dan data pengujian berbeda.

Nilai *precision*, *recall*, dan *F1-score* pada kelas halal juga menunjukkan hasil yang sangat tinggi pada kedua skenario. Sementara itu, performa pada kelas haram masih berada pada kategori sangat baik meskipun nilai *recall* dan *F1-score* sedikit lebih rendah dibandingkan kelas halal. Kondisi ini menunjukkan bahwa model lebih konsisten dalam mengenali kelas halal, sedangkan pada kelas haram masih terdapat sejumlah kecil data yang belum teridentifikasi secara sempurna.

Selain itu, nilai *macro average* dan *weighted average* yang mendekati 1,00 menunjukkan bahwa model memiliki kemampuan klasifikasi yang baik secara keseluruhan. Hasil ini mengindikasikan bahwa algoritma SVM mampu memanfaatkan representasi fitur TF-IDF secara efektif dalam membedakan produk halal dan haram berdasarkan informasi komposisi yang diperoleh dari hasil ekstraksi OCR.

3.3. Analisis Klasifikasi CNN

Seperti pada pengujian SVM, model Convolutional Neural Network (CNN) juga dievaluasi menggunakan dua skenario pembagian data, yaitu 80:20 dan 70:30. Sebelum memasuki proses pelatihan, data teks hasil preprocessing direpresentasikan menggunakan Tokenizer Keras dengan maksimum 10.000 kata unik dan padding sepanjang 100 token agar setiap data memiliki panjang sekuens yang seragam. Model CNN dibangun menggunakan Sequential API yang terdiri atas Embedding Layer, Conv1D, Global Max Pooling, Dropout, Dense Layer, dan output layer dengan fungsi aktivasi sigmoid untuk melakukan klasifikasi biner. Evaluasi dilakukan menggunakan metrik *accuracy*, *precision*, *recall*, *F1-score*, serta *macro average* dan *weighted average*. Hasil evaluasi model CNN pada kedua skenario pembagian data disajikan pada Tabel 6.

Tabel 6. Metrik Evaluasi CNN

Metrik Evaluasi	Nilai 80:20	Nilai 70:30
<i>Accuracy</i>	0.9861	0.9808
<i>Precision</i>	0.9921 (Halal)	0.9901 (Halal)
	0.8909 (Haram)	0.8353 (Haram)
<i>Recall</i>	0.9932 (Halal)	0.9894 (Halal)
	0.8750 (Haram)	0.8452 (Haram)
<i>F1-Score</i>	0.9926 (Halal)	0.9898 (Halal)
	0.8829 (Haram)	0.8402 (Haram)
<i>Macro Avg.</i>	<i>Precision</i> : 0.9415	<i>Precision</i> : 0.9127
	<i>Recall</i> : 0.9341	<i>Recall</i> : 0.9173
	<i>F1-Score</i> : 0.9378	<i>F1-Score</i> : 0.9150
<i>Weighted Avg.</i>	<i>Precision</i> : 0.9860	<i>Precision</i> : 0.9809
	<i>Recall</i> : 0.9861	<i>Recall</i> : 0.9808
	<i>F1-Score</i> : 0.9861	<i>F1-Score</i> : 0.9808

Berdasarkan Tabel 6, model CNN menunjukkan performa klasifikasi yang sangat baik pada kedua skenario pembagian data. Pada skenario 80:20, model memperoleh nilai *accuracy* sebesar 98,61%, sedangkan pada skenario 70:30 nilai *accuracy* mencapai 98,08%. Selisih nilai yang relatif kecil menunjukkan bahwa model mampu mempertahankan performa yang stabil meskipun proporsi data pelatihan dan data pengujian berbeda.

Nilai *precision*, *recall*, dan *F1-score* pada kelas halal juga berada pada tingkat yang tinggi pada kedua skenario pengujian. Sementara itu, performa pada kelas haram masih menunjukkan hasil yang baik, meskipun nilainya lebih rendah dibandingkan kelas halal, terutama pada metrik *recall* dan *F1-score*. Hal ini mengindikasikan bahwa model CNN masih mengalami kesulitan dalam mengenali sebagian data pada kelas haram dibandingkan kelas halal.

Selain itu, nilai *macro average* dan *weighted average* yang tinggi menunjukkan bahwa model CNN mampu melakukan klasifikasi secara konsisten pada keseluruhan data pengujian. Meskipun demikian, jika dibandingkan dengan hasil evaluasi model SVM pada penelitian ini, performa CNN masih sedikit lebih rendah pada hampir seluruh metrik evaluasi. Temuan tersebut menunjukkan bahwa pendekatan berbasis TF-IDF pada SVM memberikan hasil yang lebih optimal untuk klasifikasi status kehalalan berdasarkan teks

hasil ekstraksi OCR, sedangkan CNN tetap mampu memberikan performa yang kompetitif dalam membedakan produk halal dan haram.

3.4. Pembahasan Analisis Perbandingan

Hasil evaluasi menunjukkan bahwa *Support Vector Machine* (SVM) maupun *Convolutional Neural Network* (CNN) sama-sama mampu mengklasifikasikan status kehalalan produk pangan olahan berdasarkan teks hasil ekstraksi OCR dengan tingkat performa yang sangat baik pada kedua skenario pembagian data, yaitu 80:20 dan 70:30. Ringkasan hasil evaluasi kedua model disajikan pada Tabel 7, yang memperlihatkan bahwa seluruh metrik evaluasi berada pada nilai yang tinggi. Kondisi tersebut menunjukkan bahwa tahapan *preprocessing* citra, ekstraksi teks menggunakan Tesseract OCR, *text preprocessing*, serta deteksi entitas non-halal berbasis spaCy telah menghasilkan data yang cukup representatif untuk digunakan pada proses klasifikasi.

Tabel 7. Metrik Perbandingan SVM dan CNN

Model	Accuracy	Macro Precision	Macro Recall	Macro F1-Score
SVM (80:20)	0.9936	0.9876	0.9548	0.9705
SVM (70:30)	0.9922	0.9959	0.9345	0.9629
CNN (80:20)	0.9861	0.9415	0.9341	0.9378
CNN (70:30)	0.9808	0.9127	0.9173	0.9150

Berdasarkan Tabel 7, model SVM memperoleh nilai *accuracy* sebesar 99,36% pada skenario pembagian data 80:20 dan 99,22% pada skenario 70:30, sedangkan model CNN memperoleh 98,61% dan 98,08% pada skenario yang sama. Perbedaan tersebut memang tidak terlalu besar, namun memperlihatkan bahwa SVM mampu memberikan hasil yang lebih konsisten pada kedua skenario pengujian. Konsistensi ini juga terlihat pada nilai *macro F1-score*, yaitu 97,05% dan 96,29% untuk SVM, sedangkan CNN memperoleh 93,78% dan 91,50%.

Performa SVM yang lebih baik dapat dipahami dari karakteristik representasi fitur yang digunakan. Pada penelitian ini, teks hasil OCR direpresentasikan menggunakan TF-IDF, sehingga setiap kata memperoleh bobot berdasarkan tingkat kemunculannya dalam dokumen. Representasi tersebut memungkinkan SVM memanfaatkan informasi kata-kata penting yang berkaitan dengan komposisi bahan secara efektif untuk membedakan produk halal dan haram. Mengingat data yang digunakan berupa teks komposisi dengan

struktur relatif pendek dan memiliki kosakata yang cukup spesifik, pendekatan berbasis TF-IDF mampu memberikan representasi yang sesuai untuk proses klasifikasi.

Di sisi lain, model CNN juga menunjukkan performa yang sangat baik meskipun masih berada sedikit di bawah SVM. CNN memiliki kemampuan mempelajari pola hubungan antarkata melalui proses *embedding* dan operasi konvolusi sehingga tidak hanya mempertimbangkan keberadaan suatu kata, tetapi juga pola kemunculannya dalam suatu urutan teks. Kemampuan tersebut menjadi salah satu keunggulan CNN dalam menangkap informasi kontekstual dari data teks. Namun, pada penelitian ini keunggulan tersebut belum sepenuhnya memberikan peningkatan performa dibandingkan SVM. Hal ini dapat disebabkan oleh karakteristik data yang sebagian besar berupa daftar komposisi bahan dengan pola penulisan yang relatif sederhana, sehingga representasi TF-IDF sudah mampu menangkap informasi yang diperlukan untuk proses klasifikasi.

Perbandingan hasil pada dua skenario pembagian data juga menunjukkan bahwa kedua model memiliki tingkat stabilitas yang baik. Penurunan nilai *accuracy* maupun *macro F1-score* antara skenario 80:20 dan 70:30 relatif kecil, baik pada SVM maupun CNN. Kondisi ini menunjukkan bahwa model tidak mengalami perubahan performa yang signifikan ketika proporsi data pelatihan dikurangi. Dengan kata lain, model yang dibangun memiliki kemampuan generalisasi yang cukup baik terhadap data pengujian pada kedua skenario.

Selain nilai *accuracy*, hasil evaluasi juga memperlihatkan bahwa nilai *precision*, *recall*, dan *F1-score* untuk kelas halal cenderung lebih tinggi dibandingkan kelas haram pada kedua model. Temuan ini mengindikasikan bahwa sebagian besar produk halal dapat dikenali dengan sangat baik oleh sistem, sedangkan pada kelas haram masih terdapat sejumlah kecil data yang belum teridentifikasi secara sempurna. Meskipun demikian, nilai metrik pada kelas haram tetap berada pada kategori yang tinggi sehingga kedua model masih mampu memberikan hasil klasifikasi yang baik.

Secara keseluruhan, hasil penelitian menunjukkan bahwa kombinasi *preprocessing* citra, ekstraksi teks menggunakan Tesseract OCR, *text preprocessing*, deteksi entitas non-halal berbasis spaCy, serta klasifikasi menggunakan algoritma machine learning mampu menghasilkan sistem identifikasi status kehalalan dengan performa yang tinggi. Berdasarkan hasil pengujian pada dua skenario pembagian data,

model SVM memberikan performa terbaik pada penelitian ini, sedangkan CNN tetap menunjukkan kemampuan klasifikasi yang kompetitif. Temuan tersebut menunjukkan bahwa kedua pendekatan dapat digunakan untuk mendukung proses identifikasi status kehalalan produk pangan olahan berdasarkan informasi komposisi yang terdapat pada kemasan, dengan SVM menjadi pilihan yang lebih optimal pada karakteristik data yang digunakan dalam penelitian ini.

4. KESIMPULAN

Berdasarkan hasil implementasi dan evaluasi yang telah dilakukan, penelitian ini berhasil mencapai tujuan untuk mendeteksi status kehalalan produk pangan olahan kemasan melalui analisis teks hasil ekstraksi *Optical Character Recognition* (OCR) dengan memanfaatkan metode klasifikasi *Support Vector Machine* (SVM) dan *Convolutional Neural Network* (CNN). Tahapan penelitian yang meliputi *preprocessing* citra menggunakan *grayscale*, *Otsu thresholding*, dan *resize*, ekstraksi teks menggunakan Tesseract OCR, *text preprocessing*, serta deteksi entitas bahan non-halal menggunakan *Named Entity Recognition* (NER) berbasis spaCy mampu menghasilkan data yang siap digunakan pada proses klasifikasi. Selain itu, proses pelabelan berhasil dikonversi ke dalam bentuk biner, yaitu 0 untuk produk halal dan 1 untuk produk haram sesuai kebutuhan algoritma machine learning.

Hasil pengujian menunjukkan bahwa kedua model memiliki performa yang sangat baik pada skenario pembagian data 80:20 maupun 70:30. Model SVM memberikan hasil terbaik dengan nilai *accuracy* sebesar 99,36% dan *macro F1-score* sebesar 97,05% pada skenario 80:20, serta tetap menunjukkan performa yang konsisten pada skenario 70:30. Sementara itu, model CNN juga menghasilkan performa yang tinggi dengan *accuracy* sebesar 98,61% pada skenario 80:20 dan 98,08% pada skenario 70:30. Hasil tersebut menunjukkan bahwa SVM lebih optimal dalam mengklasifikasikan data teks komposisi produk pada penelitian ini, sedangkan CNN tetap memberikan kemampuan klasifikasi yang kompetitif.

Secara keseluruhan, integrasi OCR, text preprocessing, NER berbasis spaCy, dan metode klasifikasi berbasis machine learning dapat menjadi pendekatan yang efektif untuk mendukung identifikasi status kehalalan produk pangan olahan kemasan secara

otomatis. Penelitian selanjutnya disarankan menggunakan dataset yang lebih beragam, mencakup variasi jenis kemasan, kualitas citra, dan produk impor, serta mengeksplorasi pengembangan metode OCR maupun algoritma klasifikasi lainnya untuk meningkatkan kemampuan sistem dalam mengenali komposisi bahan dan mempertahankan performa klasifikasi pada berbagai kondisi data.

5. DAFTAR PUSTAKA

- [1] Angela, S. M., A. E., & Mauliana, M. I. (2024). Pengembangan Teknologi Optical Character Recognition di Flutter Berupa Deteksi Teks pada Gambar. *Jurnal TEKINKOM*, 7, 17–24.
- [2] Arora, V. S., Sahu, P., Sharma, S., Mishra, S., & Bajpai, P. P. (2024). *Evaluating Packaged Food Nutritional Value and Ingredient Analysis for Informed Choices. International Research Journal of Modernization in Engineering Technology and Science*, 3677–3680.
- [3] Borade, R., Gupta, A., Kathpalia, A., Jain, T., & Chakurkar, P. (2024). *Machine Learning Model for Optical Character Recognition-Based Food Allergen Detection with Recommendation System for Alternative Food. International Journal of Intelligent Systems and Applications in Engineering*, 1869–1875.
- [4] Ghofar, M. A., Luhur, A., & Paryasto, M. W. (2023). *Analysis of Egg Quality and Size Through Digital Image Using Support Vector Machine. e-Proceeding of Engineering*, 10(1), 196–203.
- [5] Jacky, & Ma'muriyah, N. (2021). Perancangan dan Pembuatan Alat untuk Mendeteksi Teks Hangul dan Inggris pada Menu Makanan Menggunakan Metode OCR (*Optical Character Recognition*). *TELCOMATICS*, 6(1), 1–10.
- [6] Khairani, D., Bangkit, D. A., Rozil, N. F., Masruroh, S. U., Oktaviana, S., & Rosyadi, T. (2022). *Named-Entity Recognition and Optical Character Recognition for Detecting Halal Food Ingredients: Indonesian Case Study. 2022 10th International Conference on Cyber and IT Service Management (CITSM)*.
- [7] Ladopurab, Y. U., & Kaesmetan, Y. R. (2025). Analisis Perbandingan *Convolutional Neural Network* dan *Support Vector Machine* pada Analisis Sentimen RUU Pilkada di Aplikasi X. *Journal of Information System, Applied*,

- Management, Accounting and Research, 908–927.
- [8] Mulyanto, A., Susanti, E., Rosi, F., Wajiran, & Borman, R. I. (2021). Penerapan Convolutional Neural Network (CNN) pada Pengenalan Aksara Lampung Berbasis *Optical Character Recognition* (OCR). Jurnal Edukasi dan Penelitian Informatika, 7, 52–57.
 - [9] Nurhaliza, S. S., Subali, M., ETP, L., & Rozi. (2022). Analisis Kinerja *Optical Character Recognition* untuk Membaca Dokumen Secara Otomatis. Seminar Nasional Teknologi Informasi dan Komunikasi STI&K (SeNTIK), 135–140.
 - [10] Rakhmawati, N. A., Utomoa, G. K., Indraswaria, R., & Susetyoa, I. R. (2022). *Halal Food Prediction Using the Similarity Graph Algorithms*. International Journal of Food System Dynamics, 165–173.
 - [11] Rahmawati, A. N., Wibowo, S. A., & Sunarya, U. (2021). *Performance Analysis of Optical Character Recognition*. e-Proceeding of Engineering, 8(5), 4777–4785.
 - [12] Septiani, D., & Isabela, I. (2022). Analisis *Term Frequency Inverse Document Frequency* (TF-IDF) dalam Temu Kembali Informasi pada Dokumen Teks. SINTESIA: Jurnal Sistem dan Teknologi Informasi Indonesia.
 - [13] W, B. K., Prasetyo, H., L, G. A., Wibisono, B. A., & Prasvita, D. S. (2021). Analisis Perbandingan Algoritma SVM dan CNN untuk Klasifikasi Buah. SENAMIKA.
 - [14] Yanti, R. M., Santoso, I., & Suadaa, L. H. (2021). *Application of Named Entity Recognition via Twitter on spaCy in Indonesian (Case Study: Power Failure in the Special Region of Yogyakarta)*. Indonesian Journal of Information Systems (IJIS), 76–86.
 - [15] Yogi, M., Sari, M. I., & Handayani, R. (2023). Sistem Deteksi Teks pada Kemasan Produk Makanan Menggunakan Metode *Deep Learning* Berbasis Raspberry Pi. e-Proceeding of Applied Science, 9, 2198–2205.