



Perancangan Sistem Identifikasi Entitas Kritis Pada Label Produk Makanan Berbasis Web Menggunakan Integrasi Named Entity Recognition (NER), Optical Character Recognition (OCR), dan Algoritma K-Nearest Neighbor (KNN)

Miftah Fathu Ramadhan¹, A. Nurul Anwar^{2*}

^{1,2} Fakultas Ilmu Komputer, Teknik Informatika, Universitas Pamulang, Kota Tangerang Selatan, Indonesia
Email: ¹ miftahramadhan78@gmail.com, ² dosen02523@unpam.ac.id

Abstrak—Label komposisi pada produk makanan sering memuat istilah ilmiah atau nama bahan yang kurang dipahami oleh masyarakat sehingga menyulitkan proses identifikasi bahan yang berpotensi memengaruhi status kehalalan suatu produk. Penelitian ini bertujuan mengembangkan sistem berbasis web yang dapat membaca dan menganalisis label komposisi produk makanan secara otomatis serta mengidentifikasi bahan berdasarkan status kehalalannya. Metode pengembangan sistem yang digunakan adalah Prototype. Sistem dibangun dengan mengintegrasikan Optical Character Recognition (OCR) untuk mengekstraksi teks dari gambar label, Named Entity Recognition (NER) untuk mengenali nama bahan, serta K-Nearest Neighbor (KNN) untuk mengklasifikasikan bahan ke dalam kategori halal, haram, dan mushbooh. Hasil penelitian menunjukkan bahwa sistem dapat mengekstraksi informasi dari label produk, mengenali entitas bahan makanan, serta mengklasifikasikan status bahan berdasarkan kemiripan dengan data yang telah dipelajari. Informasi tersebut kemudian digunakan untuk menentukan status produk berdasarkan komposisi yang dianalisis. Dengan demikian, sistem yang dikembangkan dapat membantu pengguna memperoleh informasi mengenai status halal suatu produk secara lebih praktis dan efisien.

Kata Kunci: Optical Character Recognition, Named Entity Recognition, K-Nearest Neighbor, label produk makanan, status halal.

Abstract—Food product labels often contain scientific terms or ingredient names that are difficult for consumers to understand, making it challenging to identify ingredients that may affect the halal status of a product. This study aims to develop a web-based system capable of automatically reading and analyzing food product labels and identifying ingredients based on their halal status. The system was developed using the Prototype method by integrating Optical Character Recognition (OCR) to extract text from product labels, Named Entity Recognition (NER) to identify ingredient names, and the K-Nearest Neighbor (KNN) algorithm to classify ingredients into halal, haram, and mushbooh categories. The results indicate that the proposed system is able to extract information from product labels, identify ingredient entities, and classify ingredient status based on their similarity to the training data. The identified ingredients are then used to determine the overall halal status of the product according to its composition. Therefore, the developed system can assist users in obtaining information about the halal status of food products in a more practical and efficient manner.

Keywords: Optical Character Recognition, Named Entity Recognition, K-Nearest Neighbor, food product labels, halal status.

1. PENDAHULUAN

Perkembangan teknologi informasi dan kecerdasan buatan (Artificial Intelligence/AI) telah mendorong pemanfaatannya di berbagai bidang, termasuk industri pangan. Bagi masyarakat Indonesia yang mayoritas beragama Islam, informasi mengenai kehalalan produk menjadi salah satu pertimbangan penting sebelum mengonsumsi suatu produk. Namun, label komposisi pada kemasan sering memuat istilah ilmiah, nama bahan dalam bahasa asing, atau kode aditif pangan yang sulit dipahami sehingga menyulitkan proses identifikasi status halal suatu produk. Beberapa layanan pengecekan halal telah tersedia, tetapi umumnya masih mengandalkan pencarian berdasarkan nama produk, nomor sertifikat, atau barcode. Pendekatan tersebut belum mampu mengidentifikasi bahan yang tercantum pada label produk secara otomatis sehingga pengguna masih perlu melakukan pengecekan secara manual. Salah satu pendekatan yang dapat digunakan untuk mengatasi permasalahan tersebut adalah mengintegrasikan Optical Character Recognition (OCR) untuk mengekstraksi teks dari label produk, Named Entity Recognition (NER) untuk mengenali nama



JRIIN : Jurnal Riset Informatika dan Inovasi
Volume 4, No. 6 Tahun 2026
ISSN 3025-0919 (media online)
Hal 1649-1656

bahan, serta K-Nearest Neighbor (KNN) untuk mengklasifikasikan bahan ke dalam kategori halal, haram, dan mushbooh. Kombinasi ketiga metode tersebut diharapkan dapat memberikan proses identifikasi yang lebih praktis dan efisien.

Berdasarkan permasalahan tersebut, penelitian ini bertujuan mengembangkan sistem berbasis web yang mampu membaca label komposisi produk makanan, mengidentifikasi bahan yang terkandung di dalamnya, serta menentukan status halal produk berdasarkan hasil analisis komposisi. Sistem yang dikembangkan diharapkan dapat membantu masyarakat memperoleh informasi mengenai status halal produk secara lebih mudah dan cepat.

2. METODE

2.1 Metode Pengumpulan Data

1. Studi Literatur
Studi literatur dilakukan dengan mengkaji berbagai referensi ilmiah yang relevan, seperti jurnal, buku, dan artikel ilmiah yang membahas produk halal, label makanan, serta penelitian-penelitian terkait, guna memperoleh landasan teori yang mendukung penelitian.
2. Observasi
Observasi dilakukan untuk mengumpulkan data berupa gambar label produk pangan olahan. Pengumpulan data dilakukan melalui observasi tidak langsung, yaitu dengan mengambil gambar label produk dari berbagai sumber di internet, serta observasi langsung, yaitu dengan mengambil foto label produk makanan olahan yang tersedia di pasaran. Data yang dikumpulkan berupa gambar label makanan yang memuat daftar bahan dan digunakan sebagai dasar perancangan sistem..

2.2 Metode Pengembangan Sistem.

Metode pengembangan sistem yang digunakan dalam penelitian ini adalah metode prototype. Metode prototype dipilih karena memungkinkan peneliti untuk merancang sistem secara bertahap dan memperoleh gambaran sistem secara cepat sesuai dengan kebutuhan pengguna. Tahapan metode prototype dalam penelitian ini meliputi:

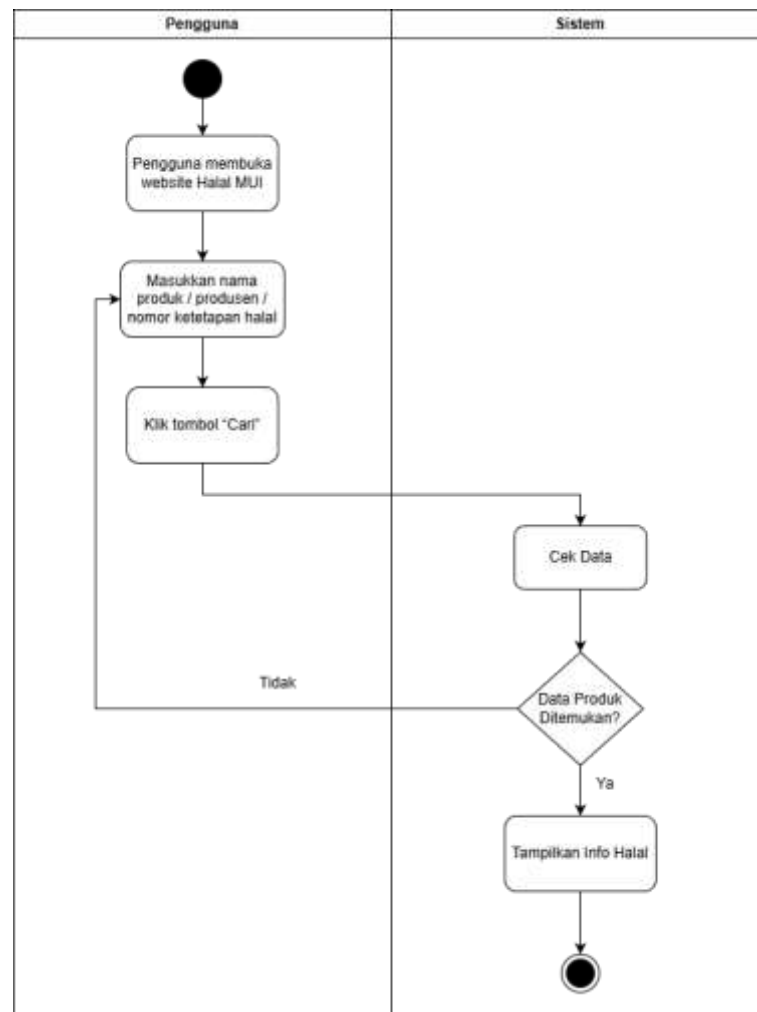
1. Identifikasi kebutuhan sistem, berdasarkan hasil studi literatur dan observasi.
2. Perancangan prototype awal, berupa rancangan alur sistem dan tampilan website.
3. Evaluasi prototype, dilakukan melalui uji coba awal untuk menilai kesesuaian rancangan dengan tujuan penelitian.

Penyempurnaan prototype, berdasarkan hasil evaluasi hingga diperoleh rancangan sistem yang sesuai.

3. HASIL DAN PEMBAHASAN

3.1 Analisa Sistem Saat Ini

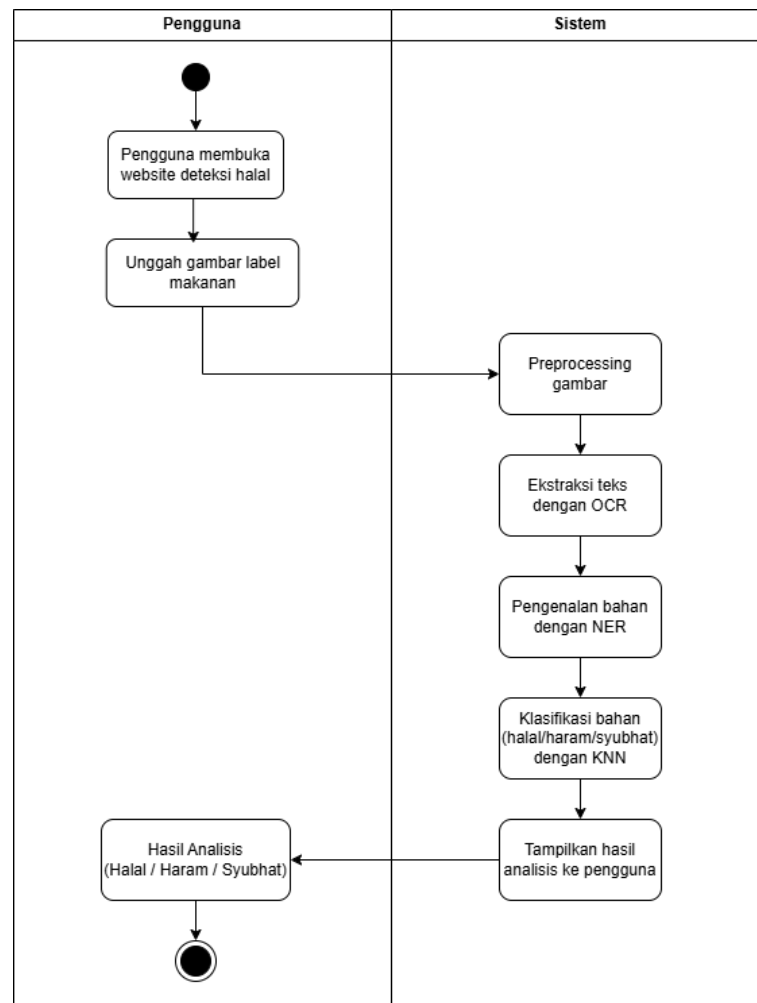
Saat ini, pengecekan status halal produk dapat dilakukan melalui website resmi Halal MUI. Namun, sistem tersebut masih mengandalkan pencarian berdasarkan informasi produk, seperti nama produk atau nomor sertifikat halal, sehingga belum dapat mengidentifikasi status halal secara otomatis dari label atau komposisi produk.



Gambar 1. Activity Diagram Sistem Berjalan

3.2 Analisa Sistem Usulan

Penelitian ini mengusulkan sistem berbasis web yang mengintegrasikan teknologi kecerdasan buatan untuk membaca label komposisi produk makanan, mengenali nama bahan, dan mengklasifikasikan statusnya ke dalam kategori halal, haram, atau mushbooh. Berbeda dengan sistem yang masih mengandalkan pencarian berdasarkan informasi produk, sistem yang dikembangkan memungkinkan pengguna cukup mengunggah gambar label makanan. Selanjutnya, sistem akan mengekstraksi teks dari gambar, mengidentifikasi bahan yang terkandung, mengklasifikasikan status setiap bahan, dan menentukan status produk berdasarkan hasil analisis komposisi.



Gambar 2. Activity Diagram Sistem Usulan

3.3 Perancangan Modul OCR

Modul Optical Character Recognition (OCR) dibangun menggunakan pustaka EasyOCR dengan dukungan bahasa Indonesia dan bahasa Inggris. Gambar label yang diunggah pengguna diproses untuk mengekstraksi teks komposisi produk. Selanjutnya, hasil ekstraksi diperbaiki menggunakan modul OCR Corrector yang menerapkan pendekatan rule-based melalui regular expression dan fuzzy string matching. Proses ini bertujuan meningkatkan kualitas hasil OCR dengan mencocokkan teks terhadap kamus bahan makanan sehingga kesalahan pembacaan dapat diminimalkan sebelum memasuki tahap identifikasi entitas.

3.4 Perancangan Modul NER

Modul Named Entity Recognition (NER) digunakan untuk mengidentifikasi nama-nama bahan makanan yang terdapat pada teks hasil preprocessing. Proses identifikasi dilakukan untuk mengenali entitas bahan dan mengelompokkannya ke dalam kategori HALAL, HARAM, atau MUSHBOOH.

Selain itu, sistem juga menangani kode aditif pangan (E-number) dengan mengubahnya menjadi nama bahan yang sesuai sehingga statusnya dapat dikenali. Apabila suatu bahan belum dapat diidentifikasi secara langsung, proses klasifikasi dilanjutkan menggunakan algoritma K-Nearest Neighbor (KNN) berdasarkan tingkat kemiripan dengan data pada basis data bahan makanan. Hasil identifikasi tersebut kemudian digunakan untuk menentukan status produk secara keseluruhan.

3.5 Perancangan Modul NER

Modul Named Entity Recognition (NER) digunakan untuk mengidentifikasi nama-nama bahan makanan dari teks hasil preprocessing. Selanjutnya, setiap bahan dikenali dan dikelompokkan ke dalam kategori HALAL, HARAM, atau MUSHBOOH. Sistem juga mendukung identifikasi kode aditif pangan (E-number) dengan mengonversinya menjadi nama bahan yang sesuai sehingga dapat diproses pada tahap klasifikasi.

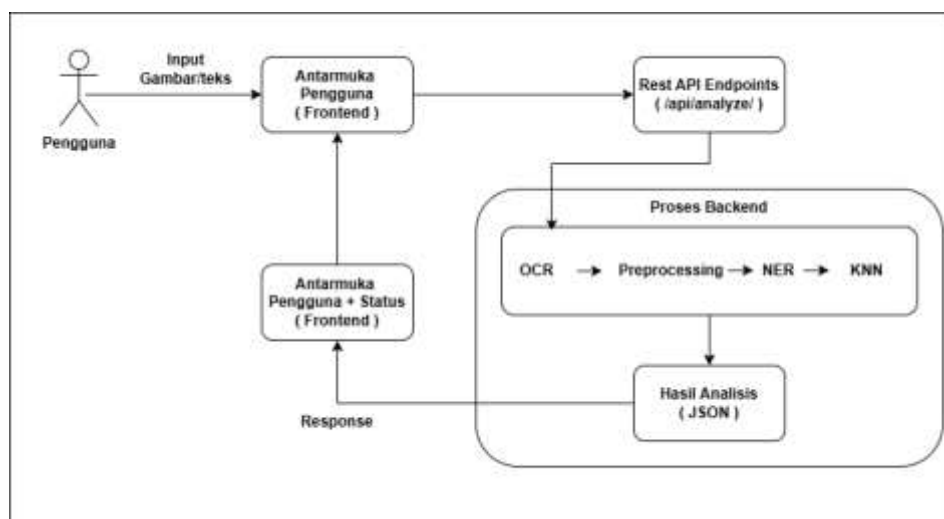
Proses klasifikasi dilakukan menggunakan algoritma K-Nearest Neighbor (KNN) dengan representasi fitur TF-IDF Character N-Gram. Pendekatan ini dipilih karena mampu menangani variasi penulisan nama bahan sehingga proses pencocokan menjadi lebih baik. Pengukuran kemiripan dilakukan menggunakan Cosine Similarity, sedangkan metode Distance-Weighted KNN digunakan untuk memberikan bobot yang lebih besar pada tetangga terdekat dalam menentukan hasil klasifikasi. Basis data yang digunakan sebagai data pelatihan terdiri atas 110 bahan makanan yang dikelompokkan ke dalam tiga kategori, yaitu 74 bahan halal, 19 bahan mushbooh, dan 17 bahan haram. Ringkasan data pelatihan dapat dilihat pada Tabel 1.

Tabel 1. Kategori bahan Makana dalam database Model KNN

Kategori	Jml. Bahan	Contoh Bahan	Keterangan
Halal	74	Tepung terigu, minyak sawit, gula pasir, susu sapi, lesitin kedelai, natrium benzoat, asam sitrat	Bahan yang dipastikan halal sesuai syariat Islam
Mushbooh	19	Gelatin (tidak jelas sumber), emulsifier, perisa alami, mono & digliserida asam lemak, gliserin	Bahan yang meragukan; memerlukan konfirmasi sumber/produsen
Haram	17	Gelatin babi, lemak babi (lard), alkohol/etanol, karmin (E120), darah, enzim babi	Bahan yang dipastikan haram berdasarkan dalil syariat

3.6 Perancangan antar muka API

Alur komunikasi antara antarmuka pengguna dan backend menggunakan Flask dan REST API dapat dilihat pada gambar



Gambar 3. Alur Rancangan Antarmuka API

3.7 Implementasi

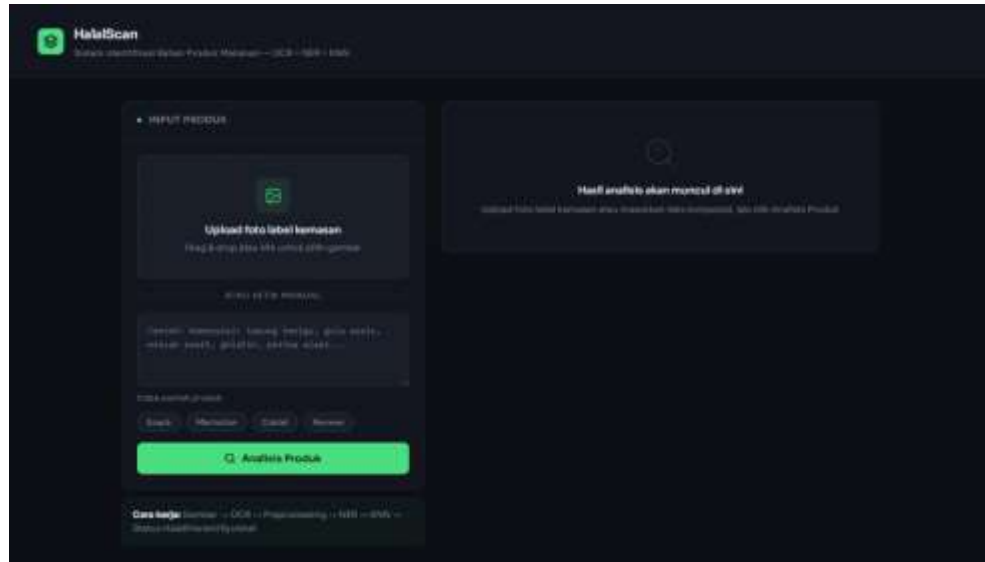
3.7.1 Tampilan Halaman Utama

Halaman utama merupakan halaman pertama yang ditampilkan ketika pengguna mengakses



JRIIN : Jurnal Riset Informatika dan Inovasi
Volume 4, No. 6 Tahun 2026
ISSN 3025-0919 (media online)
Hal 1649-1656

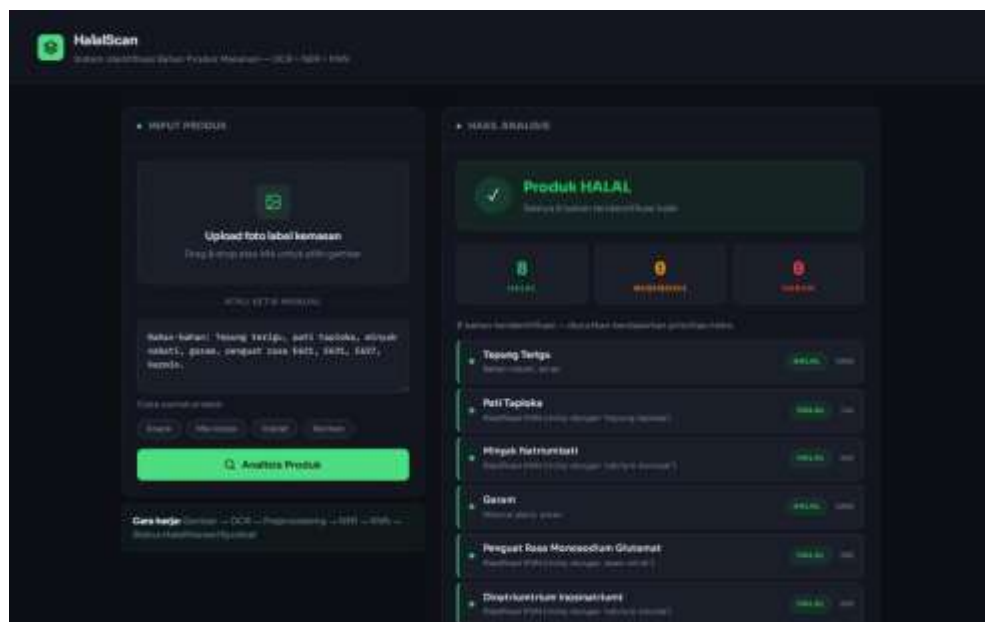
sistem. Halaman ini berfungsi sebagai media interaksi utama antara pengguna dan sistem untuk melakukan analisis status halal produk makanan.



Gambar 4. Tampilan Halaman Utama

3.7.2 Tampilan Halaman Hasil Produk Halal

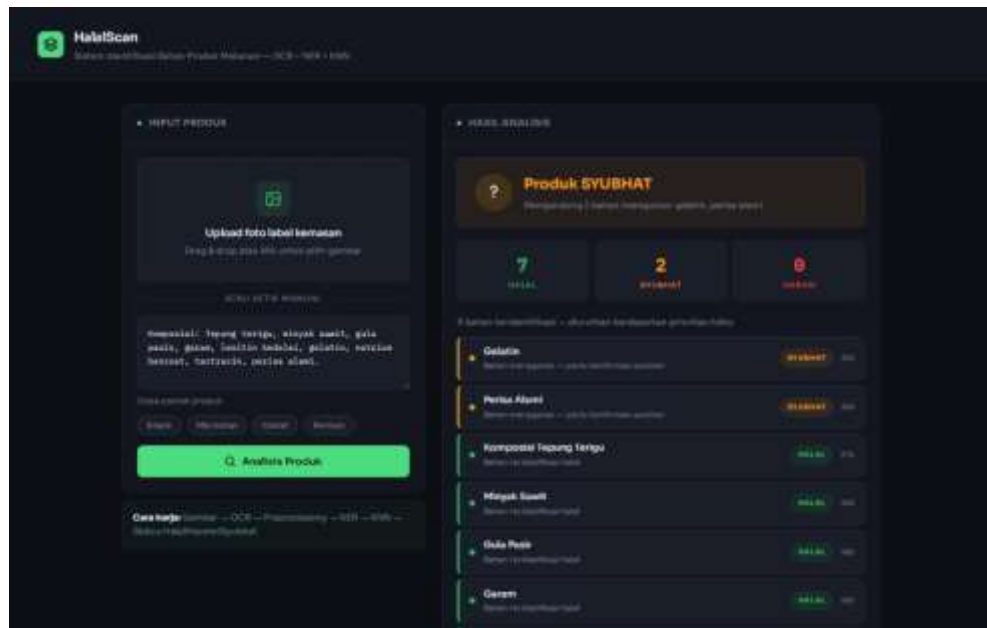
Halaman hasil produk halal ditampilkan ketika seluruh bahan yang teridentifikasi pada komposisi produk dikategorikan sebagai bahan halal. Halaman ini berfungsi untuk memberikan informasi mengenai status produk, jumlah bahan berdasarkan kategori status, serta rincian hasil identifikasi setiap bahan yang terkandung pada produk.



Gambar 5. Halaman Hasil Produk Halal

3.7.3 Tampilan Halaman Hasil Produk Syubhat

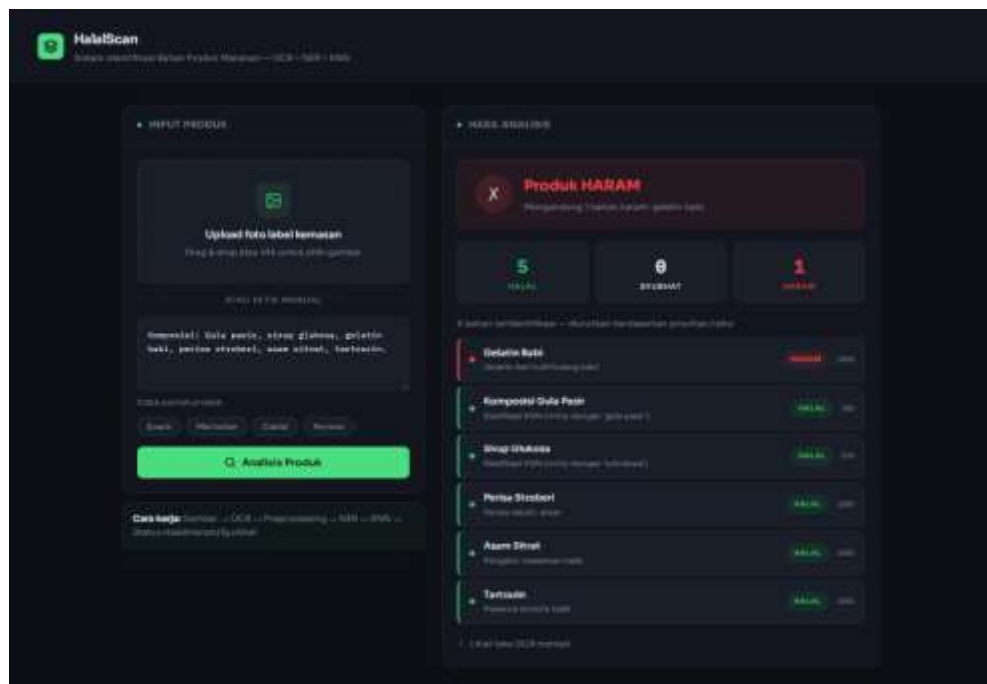
Halaman hasil produk syubhat ditampilkan ketika sistem mengidentifikasi adanya satu atau lebih bahan yang termasuk dalam kategori MUSHBOOH atau meragukan.



Gambar 6. Halaman Hasil Produk Syubhat

3.7.4 Tampilan Halaman Hasil Produk Haram

Halaman hasil produk haram ditampilkan ketika sistem mengidentifikasi adanya satu atau lebih bahan yang termasuk dalam kategori HARAM.



Gambar 7. Halaman Hasil Produk Haram

4. KESIMPULAN

Penelitian ini berhasil mengembangkan sistem berbasis web untuk mengidentifikasi status bahan pada label produk makanan melalui integrasi Optical Character Recognition (OCR), Named



JRIIN : Jurnal Riset Informatika dan Inovasi
Volume 4, No. 6 Tahun 2026
ISSN 3025-0919 (media online)
Hal 1649-1656

Entity Recognition (NER), dan algoritma K-Nearest Neighbor (KNN). Sistem mampu mengekstraksi teks dari label produk, mengenali entitas bahan, mengklasifikasikan bahan ke dalam kategori halal, haram, dan mushbooh, serta menentukan status produk berdasarkan komposisi yang dianalisis. Hasil pengujian menunjukkan bahwa integrasi ketiga metode tersebut dapat mendukung proses identifikasi bahan secara otomatis sehingga memberikan kemudahan bagi pengguna dalam memperoleh informasi mengenai status halal suatu produk.

REFERENCES

- Amien, M. (2023). Sejarah dan Perkembangan Teknik Natural Language Processing (NLP) Bahasa Indonesia: Tinjauan Tentang Sejarah, Perkembangan Teknologi, dan Aplikasi NLP dalam Bahasa Indonesia. *ELANG: Journal of Interdisciplinary Research*, 99-105.
- Angelina, S. J., Negara, A. B., & H. M. (2023). Analisis Pengaruh Penerapan Stopword Removal Pada Performa Klasifikasi Sentimen Tweet Bahasa Indonesia. *JUARA (Jurnal Aplikasi dan Riset Informatika)*, 165-173.
- Hadi, K., & Utami, E. (2025). ANALISIS K-NN DENGAN INTEGRASI BOW, TF-IDF, DAN N-GRAMS UNTUK KLASIFIKASI UJARAN KEBENCIAN PADA TWITTER. *JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, 2971-2983.
- Khairani, D., Bangkit, D. A., & Rozi, N. F. (2022). Named-Entity Recognition and Optical Character Recognition for Detecting Halal Food Ingredients: Indonesian Case Study. *The 10 th International Conference on Cyber and IT Service Management (CITSM 2022)*.
- Marunung, J., Saragih, H., Prabukusumo, M. A., & Firdaus, E. A. (2025). Optimizing the performance of the K-Nearest Neighbors algorithm using grid search and feature scaling to improve data classification accuracy. *Jurnal Mandiri IT Vol. 14 No. 2*, 260-268.
- Maulana, M. R., Nugroho, A. P., Adinata, F. C., Haidar, N. B., & Setiawan, A. (2025). KNN-Based Handwritten Digit Classification with Accuracy Analysis and Visualization. *JURNAL ILMIAH SISTEM INFORMASI (JUSI)*, 142-150.
- Nanda, I. A., Seputra, K. A., & Dermawan, K. T. (2026). Perbandingan Performa RESTful API Menggunakan Framework Python Flask dan Go Gin pada Pengembangan Aplikasi Alumni Circle. *Journal of Artificial Intelligence and Digital Business (RIGGS)*, 12113-12119.
- Rosyadi, A. W., Ma'shummah, S., Zaman, M. Q., & Fajar, M. R. (2024). Ingredients Identification Through Label Scanning Using PaddleOCR and ChatGPT for Information Retrieval. *JURNAL RESTI (Rekayasa Sistem dan Teknologi Informasi)*, 758-767.
- Salsabella, A., Amalia, A. N., & Rahma, D. A. (2024). SYUBHAT IN EDUCATION: BALANCING KNOWLEDGE. *Islamologi : Jurnal Ilmiah Keagamaan*.
- Sari, L. A. (2024). Analysis of Public Sentiment on Google Play Store Tije Application Users Using Naïve Bayes Classifier Method. *Jurnal Teknik Informatika (Jutif)*, 243-251.
- Setiawan, A., & Hasan, F. N. (2025). ANALISIS SENTIMEN TANGGAPAN PENGGUNA APLIKASI BALE BY BTN MENGGUNAKAN METODE SUPPORT VECTOR MACHINE (SVM). *STORAGE – Jurnal Ilmiah Teknik dan Ilmu Komputer*, 315-326.
- Sholeh, M., Aji, W. L., Yusdhistira, R., & Qasthari, B. L. (2022). Pengelolaan Pemesanan Menu Makanan Menggunakan Framework Flask Python. *JATISI (Jurnal Teknik Informatika dan Sistem Informasi)*, 916-929.
- Sholihin, R. (2024). KONSEP HALAL DAN HARAM (PERSPEKTIF HUKUM DAN PENDIDIKAN). *JOURNAL OF ISLAMIC AND LAW STUDIES*, 1-8.
- Syahrina, A., Krismono, B., & Zulfikri, M. (2025). IMPLEMENTASI OPTICAL CHARACTER RECOGNITION DAN ALGORITMA AHO-CORASICK UNTUK DETEKSI BAHAN BERBAHAYA PADA PRODUK SKINCARE. *INDEXIA: Informatic and Computational Intelligent Journal*, 88-96.
- Ummah, A. C., Moh, B., & Hilal, S. (2023). Sertifikasi Halal dan Kesadaran Halal Pada Minat Beli Produk Makanan. *Ekonomi, Keuangan, Investasi dan Syariah (EKUITAS)*, 1113-1119.
- Wijaya, J., Putra, F. S., Irsyad, H., & Rahman, A. (2025). IMPLEMENTASI TF-IDF DAN KNN PADA REKOMENDASI JURNAL OTOMATIS. *Journal of Informatics and Computer Engineering Research (JICER)*, 18-24.
- Wijayanto, C., & Susetyo, Y. A. (2022). IMPLEMENTASI FLASK FRAMEWORK PADA PEMBANGUNAN APLIKASI SISTEM INFORMASI HELPDESK (SIH). *JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, 858-868.