

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Informasi yang tercantum pada label makanan merupakan hal penting yang harus diperhatikan sebelum membeli dan mengonsumsi suatu makanan, terutama terkait dengan konsumen yang memiliki sistem kekebalan tubuh manusia (sistem imun) yang bereaksi terhadap zat tertentu yang seharusnya tidak berbahaya atau yang biasa disebut sebagai alergi (Gangga Febriandi, 2020). Pada label makanan suatu produk, bahan-bahan makanan yang digunakan biasanya dicantumkan secara jelas, namun dengan beragam format dan desain yang membuat pemrosesan informasi ini secara manual menjadi tidak efisien dan rentan terhadap kesalahan manusia (*human error*). Hal ini dapat mempengaruhi ketepatan dan akurasi bagi penderita alergi dalam mendeteksi alergen yang terkandung dalam produk makanan.

Di era teknologi yang berkembang pesat, *Artificial Intelligence* (AI) telah diterapkan dalam berbagai bidang untuk menyelesaikan berbagai masalah dan meningkatkan efisiensi dan akurasi proses bisnis serta memudahkan konsumen, termasuk dalam industry makanan. Salah satu teknologi *Artificial Intelligence* (AI) yang dapat dimanfaatkan adalah *Optical Character Recognition* (OCR). OCR memungkinkan teks pada gambar label makanan untuk dipindai, diekstraksi, dan diolah secara otomatis (Andria & Reza Yogi, 2019). Dengan demikian OCR dapat membantu mengotomatisasi deteksi alergen pada label makanan, mengurangi keterlibatan manual, dan meminimalkan potensi kesalahan manusia.

Namun, tantangan yang dihadapi oleh OCR adalah kemampuannya dalam mengenali format dan karakter yang bervariasi pada label makanan (Iskandar et.al., 2024), terutama dalam mendeteksi alergen. Maka dari itu, untuk mengatasi hal ini diperlukan sistem yang tidak hanya mampu mengekstraksi teks menggunakan OCR, tetapi juga mampu mengklasifikasikan informasi alergen. Sehingga, metode klasifikasi *Gradient Boosting* dipilih dalam penelitian ini. *Gradient Boosting* merupakan metode klasifikasi yang akurat dan tepat untuk menganalisis teks yang telah diekstraksi oleh OCR (Wicaksono et al., 2024). Metode ini dapat mengkategorikan informasi dari label makanan dengan meningkatkan akurasi

melalui kombinasi prediksi dari berbagai model. Dengan kemampuan ini, *Gradient Boosting* dapat membantu mendeteksi produk yang mungkin berisiko bagi konsumen yang memiliki alergi tertentu.

Sistem berbasis *Optical Character Recognition* (OCR) yang digunakan untuk mendeteksi teks pada produk makanan telah dikembangkan pada beberapa penelitian. Riska Tri Mardhatillah pada "Sistem Pendekripsi Alergen pada Produk Kemasan untuk Penderita *Eczema* Menggunakan Metode OCR" mengembangkan sistem deteksi alergen pada produk kemasan menggunakan kombinasi OpenCV dan Tesseract OCR. Sistem ini mencapai rata-rata akurasi deteksi teks sebesar 61,88% dan akurasi deteksi alergen sebesar 83,06%, penelitian ini bertujuan untuk membantu penderita *eczema* dalam mengidentifikasi kandungan alergen pada produk makanan. Selain itu, penelitian oleh Sidiq Maulana pada "Sistem Deteksi Tekst pada Kemasan Produk Makanan Menggunakan *Optical Character Recognition* (OCR)" yang merancang sistem OCR yang bertujuan membantu para tunanetra mengenali nama produk makanan. Sistem ini menggunakan MATLAB untuk memproses gambar dari kamera Raspberry Pi atau gambar dari laptop yang kemudian menghasilkan *output* teks yang dibacakan melalui teknologi *text-to-speech*, sehingga memudahkan tunanetra mengenali produk makanan. Berdasarkan beberapa penelitian sebelumnya, dapat disimpulkan bahwa penelitian mengenai hal ini masih terbatas pada penggunaan OCR untuk deteksi teks umum tanpa fokus pada klasifikasi spesifik terkait alergen atau hanya terfokus untuk satu penderita penyakit. Selain itu, belum ada penelitian yang memanfaatkan metode *Gradient Boosting* untuk meningkatkan akurasi klasifikasi informasi alergen pada label makanan.

Berdasarkan hal tersebut, penelitian ini bertujuan untuk mendekripsi alergen pada label makanan menggunakan ekstraksi teks dengan *Optical Character Recognition* (OCR) dan *Gradient Boosting*. Berbeda dengan penelitian sebelumnya, penelitian ini tidak hanya fokus pada ekstraksi teks tetapi juga pada peningkatan akurasi klasifikasi alergen menggunakan metode *Gradient Boosting*. Solusi ini diharapkan dapat meningkatkan efisiensi dalam mendekripsi alergen pada produk makanan, mengurangi potensi kesalahan manusia, serta membantu produsen dan konsumen dalam memastikan keamanan produk yang dikonsumsi.

## 1.2 Rumusan Masalah

Pada penelitian ini, terdapat beberapa permasalahan yang akan dibahas dan diteliti, antara lain:

1. Bagaimana cara mengekstraksi informasi alergen dan *e-code* dari label makanan secara otomatis menggunakan teknologi *Optical Character Recognition* (OCR) pada sistem berbasis website?
2. Bagaimana penerapan algoritma *Gradient Boosting* dalam mengklasifikasikan teks hasil ekstraksi OCR untuk mendeteksi alergen dan *e-code* pada produk makanan?
3. Seberapa akurat model *Gradient Boosting* dalam mendeteksi alergen berdasarkan teks label makanan yang diekstraksi melalui fitur *upload* gambar atau *live capture*?

## 1.3 Tujuan Penelitian

Tujuan utama dari penelitian ini adalah untuk memberikan solusi atas permasalahan yang telah dirumuskan sebelumnya, dengan rincian sebagai berikut:

1. Mengembangkan sistem yang mampu mendeteksi alergen dan *e-code* pada label makanan menggunakan teknologi *Optical Character Recognition* (OCR).
2. Menerapkan *Gradient Boosting* untuk mengidentifikasi dan mengkategorikan bahan-bahan alergen dan *e-code* dari teks yang telah diekstraksi pada label makanan.
3. Menilai efektivitas sistem dalam mendeteksi alergen dan *e-code* berdasarkan teks label makanan yang diekstraksi melalui fitur *upload* gambar dan *live capture* dari perangkat pengguna.

## 1.4 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan beberapa manfaat, baik dalam lingkup akademis maupun bagi masyarakat luas. Manfaat-manfaat tersebut meliputi:

1. Sistem ini memberikan keamanan bagi konsumen dengan alergi melalui pendekslian otomatis alergen dan *e-code* pada produk makanan, membantu

mereka untuk lebih cepat dan mudah mengidentifikasi produk yang aman dikonsumsi.

2. Penelitian ini berkontribusi pada pengembangan aplikasi AI di bidang pemrosesan teks dan pengenalan pola dalam konteks analisis data alergen dan *e-code*, sekaligus memajukan implementasi teknologi OCR dan *Gradient Boosting* dalam sistem berbasis industri.
3. Penelitian ini menyediakan landasan bagi penelitian lebih lanjut dalam pengembangan teknologi pengolahan teks dan klasifikasi otomatis di berbagai sektor.

### **1.5 Ruang Lingkup dan Batasan Penelitian**

Untuk menjaga fokus penelitian dan menghindari ruang lingkup yang terlalu luas, penelitian ini dibatasi pada aspek-aspek berikut:

1. Jenis alergen dan *e-code* yang dideteksi terbatas pada kata-kata yang muncul secara eksplisit dalam teks label makanan.
2. Penelitian ini terbatas pada deteksi alergen yang tercantum secara eksplisit dalam teks label makanan, dengan jenis alergen yang dianalisis meliputi susu, kedelai, gandum, kacang pohon, kacang, telur, ikan, kerang, wijen, sulfit, gluten, dan almond.
3. Penelitian ini hanya mencakup klasifikasi teks yang terdapat dalam bahasa Inggris pada label makanan.
4. Penelitian ini hanya memproses gambar dengan format JPG, JPEG, dan PNG dalam analisis label makanan.
5. *Output* yang dihasilkan terbatas pada nama alergen yang terdeteksi, alternatif pengganti bahan yang aman bagi individu dengan alergi, dan rekomendasi pengelolaan reaksi alergi.
6. Penelitian ini terbatas pada penggunaan *live capture* pada perangkat dengan dukungan kamera yang memadai dan sistem operasi yang kompatibel.