

ABSTRAK

Label makanan memiliki peran penting dalam memberikan informasi kepada konsumen, khususnya bagi individu yang memiliki alergi terhadap bahan-bahan tertentu. Namun, format dan tata letak label yang bervariasi seringkali menyulitkan konsumen dalam mengidentifikasi kandungan alergen secara cepat dan akurat. Penelitian ini bertujuan untuk membangun sebuah sistem berbasis *website* yang dapat secara otomatis mengekstraksi informasi alergen dan *e-code* dari label makanan menggunakan teknologi *Optical Character Recognition* (OCR) serta mengklasifikasikannya dengan algoritma *Gradient Boosting*. Sistem ini memungkinkan pengguna untuk mengunggah gambar label atau menggunakan fitur *live capture* dari perangkat yang kompatibel. Teks yang berhasil diekstraksi kemudian diproses dan dianalisis untuk mendeteksi adanya bahan alergen dan *e-code* yang teridentifikasi. *Gradient Boosting* dipilih sebagai metode klasifikasi karena kemampuannya dalam meningkatkan kinerja prediksi melalui pendekatan ansambel. Hasil pengujian menunjukkan bahwa sistem ini mampu mencapai akurasi sebesar 85%, dengan *F1-Score* sebesar 87%, *precision* 90%, dan *recall* 93,3%. Capaian ini menunjukkan bahwa sistem dapat bekerja secara efektif dalam mendeteksi bahan-bahan alergen pada label makanan. Dengan adanya sistem ini, diharapkan konsumen dapat lebih cepat dan aman dalam memilih produk makanan yang sesuai, sekaligus memberikan kontribusi pada pengembangan teknologi kecerdasan buatan di bidang pengolahan teks berbasis citra.

Kata kunci: OCR, *Gradient Boosting*, Deteksi Alergen, *E-code*, Klasifikasi Teks

ABSTRACT

Food labels play a crucial role in providing essential information to consumers, particularly those with allergies to specific ingredients. However, the diverse formats and layouts of food labels often make it challenging for consumers to quickly and accurately identify allergen content. This study aims to develop a web-based system that can automatically extract allergen and e-code information from food labels using Optical Character Recognition (OCR) technology and classify it using the Gradient Boosting algorithm. The system allows users to upload label images or utilize a live capture feature from compatible devices. The extracted text is subsequently processed and analyzed to detect the presence of allergenic substances and identified e-codes. Gradient Boosting was selected as the classification method due to its ability to enhance predictive performance through an ensemble learning approach. The testing results demonstrate that the system achieves an accuracy of 85%, with an F1-Score of 87%, a precision of 90%, and a recall of 93,3%. These results indicate that the system is highly effective in detecting allergenic ingredients in food labels. The proposed system is expected to help consumers make safer and faster food choices, while also contributing to the advancement of artificial intelligence technologies in the field of image-based text processing.

Keywords: *OCR, Gradient Boosting, Allergen Detection, E-code, Text Classification*