

## ABSTRAK

Proses jual beli telah berubah seiring berjalannya waktu mengikuti perkembangan zaman. Proses transaksi dapat dilakukan secara langsung melalui toko fisik atau dengan sistem pemesanan yang diikuti dengan dokumen berupa pesanan pembelian. Melalui pesanan pembelian, pengusaha memperoleh banyak manfaat, seperti menghindari kesalahan pengadaan barang, menentukan harga pasar, sebagai bukti pemesanan dari pembeli, dan lain-lain. Akan tetapi, banyak dari pengusaha retail masih belum menyimpan data tersebut secara terstruktur, sehingga sulit untuk melakukan analisis dan pelacakan. Meskipun banyak dari pesanan pembelian telah berbentuk digital seperti PDF, pencatatan terstruktur bersumber dari dokumen digital masih memerlukan upaya lebih dalam waktu dan usaha, serta rawan kesalahan jika dilakukan secara manual oleh manusia. Berangkat dari permasalahan tersebut, tugas akhir ini bertujuan untuk membuat model *information extraction* dari pesanan pembelian berbentuk PDF. Alur kerja dari tugas akhir ini dimulai dengan pengumpulan data, data *pre-processing*, *information extraction*, evaluasi, dan penyimpanan ke dalam *database*. Data yang digunakan pada tugas akhir ini adalah pesanan pembelian dari “Toko Maju” yang berbentuk PDF. Kemudian, pesanan pembelian akan dirubah ke format JPEG, sebelum dilakukan proses pelabelan dan pembentukan *bounding boxes*. Proses *information extraction* meliputi proses *object detection* dan *Optical Character Recognition (OCR)*. *Object detection* akan menggunakan model Keras RetinaNet. Setelah letak daerah ekstraksi ditemukan, maka akan dilakukan deteksi karakter atau OCR dengan menggunakan *library* Tesseract. Informasi hasil ekstraksi kemudian akan disimpan ke database MySQL dengan menggunakan *library* pymysql. Model *information extraction* memperoleh nilai *confidence* sebesar 95.6% dan nilai *accuracy* sebesar 95.5%. Dengan menggunakan model *information extraction* ini, diharapkan pengusaha retail dapat melakukan pencatatan terstruktur dengan lebih mudah.

**Kata Kunci:** *Information Extraction, RetinaNet, Tesseract, Object Detection*

## ABSTRACT

*The buying and selling process has changed over time following the development of the era. Transactions can be done directly through physical stores or with an ordering system followed by purchasing orders. Through purchasing orders, entrepreneurs can benefit greatly by avoiding procurement errors, determining market prices, providing evidence of orders from buyers, and others. However, many retail entrepreneurs still need to store this data in a structured way, making it difficult to analyse and track. Although many purchasing orders have become digital in the form of PDFs, structured recording sourced from digital documents still requires more time and effort and is prone to error if done manually by humans. Starting from these problems, this final project aims to create an information extraction model from PDF purchasing orders. The workflow of this final project starts with data collection, data pre-processing, information extraction, evaluation, and storage in a database. The data used in this final project are purchasing orders from "Toko Maju" in PDF format. Then, the purchase order will be converted to JPEG format before the labelling process and bounding boxes are created. The information extraction process includes object detection and Optical Character Recognition (OCR). Object detection will use the Keras RetinaNet model. After the extraction area location is found, character detection or OCR will be performed using the Tesseract library. The extracted information will then be saved to a MySQL database using the pymysql library. The information extraction model obtained a confidence value of 95.6% and an accuracy value of 95.5%. This information extraction model allows retail entrepreneurs to make structured recordings more easily.*

**Keywords:** *Information Extraction, RetinaNet, Tesseract, Object Detection.*