



## **UNDERGRADUATE THESIS**

# **DEVELOPMENT OF RECEIPT DATA EXTRACTION IN FINANCIAL MANAGEMENT APPLICATIONS USING OPTICAL CHARACTER RECOGNITION**

**SYAMMARI RAMA JANITRA**  
NPM 20081010150

## **THESIS ADVISORS**

Dr. Firza Prima Aditiawan, S.Kom., MTI  
Hendra Maulana, S.Kom., M.Kom.

**MINISTRY OF HIGHER EDUCATION, SCIENCE, AND TECHNOLOGY  
UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAWA TIMUR  
FACULTY OF COMPUTER SCIENCE  
INFORMATICS STUDY PROGRAM  
SURABAYA  
2026**

## APPROVAL SHEET

### DEVELOPMENT OF RECEIPT DATA EXTRACTION IN FINANCIAL MANAGEMENT APPLICATIONS USING OPTICAL CHARACTER RECOGNITION

By:  
SYAMMARI RAMA JANITRA  
NPM. 20081010150

Has been defended before, and accepted by, the Board of Assessors of the Thesis Examination of the Informatics Study Program, Faculty of Computer Science, Universitas Pembangunan Nasional Veteran Jawa Timur, on July 20, 2026:

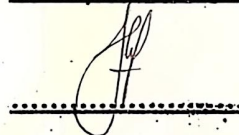
Approved,

Dr. Firza Prima Aditiawan, S.Kom., MTI  
NIP. 19860523 202121 1 003



(Advisor I)

Hendra Maulana, S.Kom., M.Kom.  
NIP. 19831223 202521 1 040



(Advisor II)

Retno Mumpuni, S.Kom., M.Sc  
NIP. 198707162025212045



(Head Assessor)

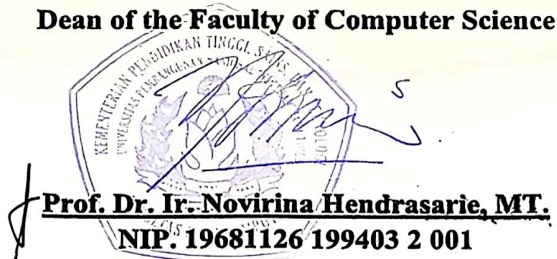
Henni Endah Wahanani, ST. M.Kom.  
NIP. 19780922 202121 2 005



(Assessor I)

Acknowledge by,

Dean of the Faculty of Computer Science



Prof. Dr. Ir. Novirina Hendrasarie, MT.  
NIP. 19681126 199403 2 001

## **APPROVAL SHEET**

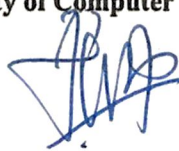
### **DEVELOPMENT OF RECEIPT DATA EXTRACTION IN FINANCIAL MANAGEMENT APPLICATIONS USING OPTICAL CHARACTER RECOGNITION**

**By:**  
**SYAMMARI RAMA JANITRA**  
**NPM. 20081010150**

Approved to proceed to the Thesis Examination

**Approved by,**

**Coordinator of Informatics Study Program**  
**Faculty of Computer Science**



**Dr. Intan Yuniar Purbasari, S.Kom. MSc.**

**NIP. 19800602 202521 2 029**

## STATEMENT OF ORIGINALITY

I am the undersigned:

Student Name : Syammari Rama Janitra  
NPM : 20081010150  
Degree Program : Bachelor (S1)  
Study Program : Informatics  
Faculty : Faculty of Computer Science

Hereby declares that this undergraduate thesis contains no part of any other scientific work that has been submitted to obtain an academic degree at any higher education institution. Furthermore, it does not contain any work or opinions previously written or published by others, except for those which are explicitly cited in this thesis and listed completely in references.

And I declare that this scientific document is free from elements of plagiarism. If in the future indications of plagiarism are found in this Thesis, I am willing to accept sanctions in accordance with the applicable laws and regulations.

Thus, I made this statement without any coercion from anyone and to be used as it should.



Surabaya, July 20 2026  
Declarant,



SYAMMARI RAMA J.  
NPM. 20081010150

## ABSTRACT

Student Name / NPM : Syammari Rama Janitra/20081010150  
Thesis Title : Development of Receipt Data Extraction in  
Financial Management Applications Using Optical  
Character Recognition  
Advisor : 1. Dr. Firza Prima Aditiawan, S.Kom., MTI  
2. Hendra Maulana, S.Kom, M.Kom

Personal financial management often faces obstacles in terms of consistency in recording transactions due to manual data entry processes that are time-consuming and prone to human error. This research aims to develop a web-based financial management application called Pockane that integrates Optical Character Recognition (OCR) technology for automated data extraction from shopping receipts. The system is built using a decoupled architecture, utilizing the Laravel framework for the user interface and Flask deployed on DigitalOcean infrastructure as the OCR processing engine using Tesseract. The research method includes the implementation of various image preprocessing techniques such as grayscaling, denoising, and binarization to improve character recognition accuracy. The test results show that the combination of image preprocessing significantly reduces the Character Error Rate (CER) in various receipt conditions. System quality evaluation conducted by a few respondents using the System Usability Scale (SUS) method resulted in a score of 75, categorizing the Pockane application in the "Good" predicate. Thus, the integration of OCR in the Pockane application is proven effective in assisting users to perform financial recording automatically, accurately, and efficiently.

**Keywords:** Financial Management, OCR, Tesseract, Laravel, Flask, Preprocessing, System Usability Scale.

## ACKNOWLEDGEMENTS

Praise and gratitude be to Allah SWT for all His blessings, guidance, and grace, which have enabled the author to successfully complete this undergraduate thesis entitled **“Development of Receipt Data Extraction in Financial Management Applications Using Optical Character Recognition”**.

The author wishes to express the sincerest gratitude to Dr. Firza Prima Aditiawan, S.Kom., MTI as the primary advisor, and Mr. Hendra Maulana, S.Kom, M.Kom as the co-advisor, for generously dedicating their time to provide invaluable guidance, advice, and motivation throughout this research. The author has also received immense support—morally, spiritually, and materially—from various parties. Therefore, the author would like to extend deep appreciation to:

1. Prof. Dr. Ir. Novirina Hendrasarie, MT., as the Dean of the Faculty of Computer Science, Universitas Pembangunan Nasional “Veteran” Jawa Timur.
2. Dr. Intan Yuniar Purbasari, S.Kom., M.Sc., as the Head of the Informatics Study Program, Faculty of Computer Science, Universitas Pembangunan Nasional “Veteran” Jawa Timur.
3. Mr. Andreas Nugroho Sihananto, S.Kom., M.Kom., as the Undergraduate Thesis Coordinator of the Informatics Study Program, Faculty of Computer Science, Universitas Pembangunan Nasional “Veteran” Jawa Timur.
4. Mrs. Retno Mumpuni, S.Kom., M.Sc., as a lecturer in the Informatics Study Program, Faculty of Computer Science, Universitas Pembangunan Nasional “Veteran” Jawa Timur, and the first examiner, for evaluating and approving this research.
5. Mrs. Henni Endah Wahanani, ST., M.Kom., as a lecturer in the Informatics Study Program, Faculty of Computer Science, Universitas Pembangunan Nasional “Veteran” Jawa Timur, and the second examiner, for evaluating and approving this research.

6. All lecturers of the Informatics Study Program, Faculty of Computer Science, Universitas Pembangunan Nasional “Veteran” Jawa Timur, who have imparted valuable and beneficial knowledge during the author’s academic years.
7. The author’s beloved mother and father, whose endless support and prayers enabled the completion of this research.
8. The author’s sibling and sibling-in-law, who have accompanied and guided the author through difficult times, providing both mental and financial support, and acting as steadfast guardians. Thank you so much.

The author recognizes that this undergraduate thesis is far from perfect and may contain shortcomings. Therefore, constructive criticism and suggestions from all readers are highly welcomed for the improvement of this work. Finally, despite its limitations, the author sincerely hopes that this report will be beneficial to all readers in general and to the author in particular.

Surabaya, July 20 2026

Author

## TABLE OF CONTENTS

|                                                     |             |
|-----------------------------------------------------|-------------|
| <b>APPROVAL SHEET .....</b>                         | <b>iii</b>  |
| <b>APPROVAL SHEET .....</b>                         | <b>v</b>    |
| <b>STATEMENT OF ORIGINALITY .....</b>               | <b>vii</b>  |
| <b>ABSTRACT .....</b>                               | <b>ix</b>   |
| <b>ACKNOWLEDGEMENTS.....</b>                        | <b>i</b>    |
| <b>TABLE OF CONTENTS.....</b>                       | <b>iv</b>   |
| <b>IMAGE LIST.....</b>                              | <b>viii</b> |
| <b>TABLE LIST .....</b>                             | <b>xii</b>  |
| <b>CHAPTER I INTRODUCTION.....</b>                  | <b>1</b>    |
| 1.1 Background .....                                | 1           |
| 1.2 Problem Formulation.....                        | 2           |
| 1.3 Objectives of the Research.....                 | 2           |
| 1.4 Benefits of the Research.....                   | 3           |
| 1.5 Scope and Limitations.....                      | 3           |
| <b>CHAPTER II LITERATURE REVIEW .....</b>           | <b>5</b>    |
| 2.1 Related Works .....                             | 5           |
| 2.2 Theoretical Background .....                    | 7           |
| 2.2.1 Software Engineering .....                    | 7           |
| 2.2.2 Optical Character Recognition .....           | 7           |
| 2.2.3 Preprocessing .....                           | 8           |
| 2.2.4 Pattern Recognition .....                     | 9           |
| 2.2.5 Feature Extraction.....                       | 10          |
| 2.2.6 Postprocessing .....                          | 11          |
| 2.2.7 Regular Expression Algorithm .....            | 11          |
| 2.2.8 Levenshtein Distance.....                     | 12          |
| 2.2.9 Sequence Matching Similarity Evaluation ..... | 12          |
| 2.2.10 Tesseract OCR.....                           | 13          |
| 2.2.11 PaddleOCR .....                              | 13          |
| 2.2.12 EasyOCR .....                                | 14          |
| 2.2.13 Model.....                                   | 14          |

|                                               |           |
|-----------------------------------------------|-----------|
| 2.2.14 Training Data .....                    | 15        |
| 2.2.15 Financial Reports .....                | 15        |
| 2.2.16 Charts .....                           | 15        |
| 2.2.17 Receipts .....                         | 18        |
| 2.2.18 Website .....                          | 19        |
| 2.2.19 Database .....                         | 19        |
| 2.2.20 Web Server .....                       | 19        |
| 2.2.21 Framework .....                        | 20        |
| 2.2.22 Laravel .....                          | 20        |
| 2.2.23 Use Case Diagram .....                 | 20        |
| 2.2.24 Activity Diagram .....                 | 21        |
| 2.2.25 Conceptual Data Model .....            | 21        |
| 2.2.26 Physical Data Model .....              | 21        |
| 2.2.27 Class Diagram .....                    | 22        |
| 2.2.28 System Usability Scale .....           | 22        |
| <b>CHAPTER III RESEARCH METHODOLOGY .....</b> | <b>23</b> |
| 3.1 Research Workflow .....                   | 23        |
| 3.2 Literature Study .....                    | 24        |
| 3.3 Context and User Analysis .....           | 24        |
| 3.3.1 Context .....                           | 24        |
| 3.3.2 User .....                              | 26        |
| 3.4 Application Concept .....                 | 26        |
| 3.5 Model Development and Testing .....       | 27        |
| 3.5.1 Dataset .....                           | 27        |
| 3.5.2 Preprocessing .....                     | 29        |
| 3.5.3 Model Testing .....                     | 33        |
| 3.6 System Design .....                       | 33        |
| 3.6.1 Use Case User .....                     | 34        |
| 3.6.2 User Activity Diagram .....             | 41        |
| 3.6.3 Conceptual Data Model .....             | 49        |
| 3.6.4 Physical Data Model .....               | 50        |
| 3.6.5 Class Diagram .....                     | 51        |

|                                                   |                                                                 |           |
|---------------------------------------------------|-----------------------------------------------------------------|-----------|
| 3.7                                               | Model Implementation and Application Testing .....              | 51        |
| 3.7.1                                             | Model Implementation within the Application .....               | 52        |
| 3.7.2                                             | Application Testing .....                                       | 52        |
| 3.8                                               | Evaluation Design .....                                         | 52        |
| <b>CHAPTER IV IMPLEMENTATION AND RESULTS.....</b> |                                                                 | <b>55</b> |
| 4.1                                               | Preprocessing and OCR Implementation .....                      | 55        |
| 4.1.1                                             | Combination Testing Scenarios.....                              | 55        |
| 4.1.2                                             | Character and Text Accuracy Analysis .....                      | 56        |
| 4.1.3                                             | Analysis and Testing of Five Preprocessing Combinations .....   | 57        |
| 4.1.4                                             | Testing the Best Preprocessing with Different OCR Engines ..... | 65        |
| 4.1.5                                             | OCR Model Development .....                                     | 68        |
| 4.1.6                                             | Integrating the Model with the Web.....                         | 69        |
| 4.2                                               | System Implementation.....                                      | 69        |
| 4.2.1                                             | Homepage .....                                                  | 69        |
| 4.2.2                                             | Registration.....                                               | 70        |
| 4.2.3                                             | Login.....                                                      | 71        |
| 4.2.4                                             | Forgot Password .....                                           | 72        |
| 4.2.5                                             | Dashboard .....                                                 | 73        |
| 4.2.6                                             | Profile .....                                                   | 74        |
| 4.2.7                                             | Settings .....                                                  | 75        |
| 4.2.8                                             | Adding Expense Transactions .....                               | 76        |
| 4.2.9                                             | Adding Income Transactions .....                                | 77        |
| 4.2.10                                            | Adding Transfer Transactions .....                              | 78        |
| 4.2.11                                            | Adding Expense Transactions Using OCR.....                      | 78        |
| 4.2.12                                            | User Guide for Uploading Receipts.....                          | 79        |
| 4.2.13                                            | Alert for Reviewing OCR Results .....                           | 81        |
| 4.2.14                                            | User Guide for Weight Conversion .....                          | 82        |
| 4.2.15                                            | Transaction History .....                                       | 83        |
| 4.2.16                                            | Planned Payments .....                                          | 83        |
| 4.2.17                                            | Budget Planning.....                                            | 84        |
| 4.2.18                                            | Total Balance .....                                             | 85        |
| 4.2.19                                            | Spending Analysis .....                                         | 86        |

|                                   |                                                                  |            |
|-----------------------------------|------------------------------------------------------------------|------------|
| 4.2.20                            | Cash Flow .....                                                  | 86         |
| 4.2.21                            | Balance Forecasts and Reminders .....                            | 87         |
| 4.3                               | Financial Management Application Feature Testing .....           | 88         |
| 4.3.1                             | Functionality Testing (Black Box Testing) .....                  | 88         |
| 4.3.2                             | Financial Chart Visualization Validation .....                   | 91         |
| 4.4                               | System Quality Analysis .....                                    | 93         |
| 4.4.1                             | Evaluation Procedure with Respondents .....                      | 93         |
| 4.4.2                             | Results of the SUS Questionnaire Assessment by Respondents ..... | 94         |
| 4.4.3                             | Analysis of Respondent Feedback and Findings .....               | 103        |
| <b>CHAPTER V CONCLUSION .....</b> |                                                                  | <b>105</b> |
| 5.1                               | Conclusions .....                                                | 105        |
| 5.2                               | Recommendations for Future Work .....                            | 106        |
| <b>REFERENCES.....</b>            |                                                                  | <b>109</b> |
| <b>APPENDICES .....</b>           |                                                                  | <b>113</b> |

## IMAGE LIST

|                                                                     |    |
|---------------------------------------------------------------------|----|
| Figure 2.1 General workflow of OCR [9] .....                        | 8  |
| Figure 2.2 Grayscale Result [4] .....                               | 9  |
| Figure 2.3 Thresholding Result [4] .....                            | 9  |
| Figure 2.4 Erosion Result [4] .....                                 | 9  |
| Figure 2.5 Illustration of the Pattern Recognition Stage [13] ..... | 10 |
| Figure 2.6 Illustration of the Feature Extraction Stage [22] .....  | 11 |
| Figure 2.7 Column Chart [13] .....                                  | 16 |
| Figure 2.8 Pie Chart [13] .....                                     | 16 |
| Figure 2.9 Stacked Bar Chart [13] .....                             | 17 |
| Figure 2.10 Area Chart [13] .....                                   | 17 |
| Figure 2.11 Supermarket Receipt (Superindo) .....                   | 18 |
| Figure 3.1 Research Methodology Flowchart .....                     | 23 |
| Figure 3.2 Web Application Workflow .....                           | 26 |
| Figure 3.3 Raw Dataset Sample .....                                 | 29 |
| Figure 3.4 Cropped Dataset Sample .....                             | 29 |
| Figure 3.5 Grayscale Conversion Output .....                        | 30 |
| Figure 3.6 Sharpening Output .....                                  | 30 |
| Figure 3.7 Denoising Output .....                                   | 31 |
| Figure 3.8 Thresholding Output .....                                | 32 |
| Figure 3.9 Smoothing Output .....                                   | 32 |
| Figure 3.10 Color Inversion Output .....                            | 33 |
| Figure 3.11 User Use Case Diagram .....                             | 34 |
| Figure 3.12 Register Activity Diagram .....                         | 41 |
| Figure 3.13 Login Activity Diagram .....                            | 42 |
| Figure 3.14 Dashboard Activity Diagram .....                        | 43 |
| Figure 3.15 Settings Activity Diagram .....                         | 44 |
| Figure 3.16 Add Income Report Activity Diagram .....                | 45 |
| Figure 3.17 Add Expense Report Activity Diagram .....               | 46 |
| Figure 3.18 Add Expense Report with OCR Activity Diagram .....      | 47 |
| Figure 3.19 Add Fund Transfer Activity Diagram .....                | 48 |

|                                                                              |    |
|------------------------------------------------------------------------------|----|
| Figure 3.20 View Statistics Activity Diagram .....                           | 49 |
| Figure 3.21 Conceptual Data Model .....                                      | 50 |
| Figure 3.22 Physical Data Model .....                                        | 50 |
| Figure 3.23 Class Diagram.....                                               | 51 |
| Figure 4.1 Preprocessing Pipeline 1 Visual Results.....                      | 58 |
| Figure 4.2 Preprocessing Pipeline 2 Visual Results.....                      | 60 |
| Figure 4.3 Preprocessing Pipeline 3 Visual Results.....                      | 61 |
| Figure 4.4 Preprocessing Pipeline 4 Visual Results.....                      | 63 |
| Figure 4.5 Preprocessing Pipeline 5 Visual Results.....                      | 64 |
| Figure 4. 6 Comparative Accuracy Results Across Preprocessing Scenarios..... | 66 |
| Figure 4. 7 Homepage Interface .....                                         | 70 |
| Figure 4.8 Registration Interface.....                                       | 71 |
| Figure 4.9 Login Interface.....                                              | 72 |
| Figure 4.10 Forgot Password Interface .....                                  | 73 |
| Figure 4.11 Dashboard Interface .....                                        | 74 |
| Figure 4.12 Profile Interface .....                                          | 75 |
| Figure 4.13 Settings Interface .....                                         | 76 |
| Figure 4.14 Manual Expense Entry Interface.....                              | 76 |
| Figure 4.15 Manual Income Entry Interface .....                              | 77 |
| Figure 4.16 Internal Fund Transfer Interface .....                           | 78 |
| Figure 4.17 OCR-Assisted Expense Entry Interface.....                        | 79 |
| Figure 4.18 Receipt Upload User Guidance Interface .....                     | 80 |
| Figure 4.19 OCR Parsing Review Warning Modal.....                            | 81 |
| Figure 4.20 Weight Conversion Educational Banner Interface .....             | 82 |
| Figure 4.21 Weight Conversion Confirmation Modal .....                       | 82 |
| Figure 4.22 Historical Transaction Management Interface.....                 | 83 |
| Figure 4.23 Scheduled Liabilities and Planned Payment Interface.....         | 84 |
| Figure 4.24 Monthly Budget Allocation Interface .....                        | 85 |
| Figure 4.25 Aggregated Net Balance Interface .....                           | 85 |
| Figure 4.26 Categorical Expense Breakdown Interface.....                     | 86 |
| Figure 4.27 Cash Flow Analytics Line Chart Interface .....                   | 87 |
| Figure 4.28 Balance Predictive Analytics and Alert System Interface .....    | 87 |

|                                                                          |     |
|--------------------------------------------------------------------------|-----|
| Figure 4.29 Electronic Mail Distribution of Evaluation Respondents ..... | 95  |
| Figure 4.30 Respondent Name Identification Directory .....               | 95  |
| Figure 4.31 Score Distribution Response for Item 1.....                  | 96  |
| Figure 4.32 Score Distribution Response for Item 2.....                  | 97  |
| Figure 4.33 Score Distribution Response for Item 3.....                  | 97  |
| Figure 4.34 Score Distribution Response for Item 4.....                  | 98  |
| Figure 4.35 Score Distribution Response for Item 5.....                  | 98  |
| Figure 4.36 Score Distribution Response for Item 6.....                  | 99  |
| Figure 4.37 Score Distribution Response for Item 7.....                  | 100 |
| Figure 4.38 Score Distribution Response for Item 8.....                  | 100 |
| Figure 4.39 Score Distribution Response for Item 9.....                  | 101 |
| Figure 4.40 Score Distribution Response for Item 10.....                 | 101 |
| Figure 4.41 Standardized SUS Score Benchmark Categories.....             | 102 |

## TABLE LIST

|                                                                             |     |
|-----------------------------------------------------------------------------|-----|
| Table 2.1 System Usability Scale Classification [24] .....                  | 22  |
| Table 3.1 Laptop Specifications.....                                        | 25  |
| Table 3.2 Cloud Server (Droplet) Hardware Specifications .....              | 25  |
| Table 3.3 Representative Dataset Samples .....                              | 28  |
| Table 3.4 Average Dataset Accuracy .....                                    | 33  |
| Table 3.5 User Registration Use Case Narrative.....                         | 34  |
| Table 3.6 User Login Use Case Narrative.....                                | 35  |
| Table 3.7 Dashboard Use Case Narrative .....                                | 36  |
| Table 3.8 Settings Use Case Narrative.....                                  | 36  |
| Table 3.9 Add Income Report Use Case Narrative .....                        | 37  |
| Table 3.10 Add Expense Report Use Case Narrative .....                      | 38  |
| Table 3.11 Add Expense Report with OCR Use Case Narrative .....             | 39  |
| Table 3.12 Add Fund Transfer Report Use Case Narrative .....                | 40  |
| Table 3.13 View Statistics Use Case Narrative.....                          | 40  |
| Table 3.14 SUS Questionnaire .....                                          | 52  |
| Table 4.1 OCR Engines Comparison using the Fifth Preprocessing.....         | 66  |
| Table 4.2 Black Box Functionality Test Results.....                         | 88  |
| Table 4.3 Financial Chart Visualization Validation Results .....            | 92  |
| Table 4.4 System Usability Scale (SUS) Evaluation Questionnaire Items ..... | 94  |
| Table 4.5 Individual System Usability Score Calculation Summary.....        | 102 |
| Table 4.6 Respondent Technical Feedback and System Validation .....         | 103 |