

**MAPPING REGIONAL DETERMINANTS USING MACHINE LEARNING  
FOR PREDICTING SATUSEHAT ADOPTION ACROSS JAVA, INDONESIA**

By

Priskila Ranindhita Hapsari  
22451064

MASTER'S DEGREE  
in

DATA SCIENCE BUSINESS INFORMATICS  
FACULTY OF ENGINEERING AND INFORMATION TECHNOLOGY



SWISS GERMAN UNIVERSITY  
The Prominence Tower  
Jalan Jalur Sutera Barat No. 15, Alam Sutera  
Tangerang, Banten 15143 - Indonesia

Revision after the Thesis Defense on 5 February 2026

### STATEMENT BY THE AUTHOR

I hereby declare that this submission is my own work and to the best of my knowledge, it contains no material previously published or written by another person, nor material which to a substantial extent has been accepted for the award of any other degree or diploma at any educational institution, except where due acknowledgement is made in the thesis.



Priskila Ranindhita Hapsari

Student

18 FEB 2026

Date

Approved by:



Dr. Ir. Eka Budiarto ST M.Sc

Thesis Advisor

Date

Dr. Ir. Heru Purnomo Indung M. Eng

Thesis Co-Advisor

18 FEB 2026

Date

Dr. Ir. Maulahikmah Galinium, S. Kom., M.Sc., Ph.D.

Dean

Date

## ABSTRACT

### MAPPING REGIONAL DETERMINANTS USING MACHINE LEARNING FOR PREDICTING SATUSEHAT ADOPTION ACROSS JAVA, INDONESIA

By

Priskila Ranindhita Hapsari  
Dr. Ir. Eka Budiarto ST M.Sc  
Dr. Ir. Heru Purnomo Ipung M. Eng

SWISS GERMAN UNIVERSITY

This study examines SatuSehat adoption patterns across 119 kabupaten/kota in Java Island, Indonesia, utilizing data from 68,330 healthcare facilities (clinics and independent practices) observed across three time points (December 2024, July 2025, November 2025). Two research questions are addressed through quantitative analysis: (1) how regional characteristics correlate with adoption rates, and (2) whether machine learning models can predict facility-level adoption likelihood. Pearson correlation analysis reveals that educational attainment and population density demonstrate the strongest positive associations with adoption rates, while poverty shows moderate negative correlation. Provincial analysis identifies substantial variation, with DKI Jakarta achieving higher adoption compared to East Java, reflecting underlying socioeconomic disparities. Facility-level predictive modeling using Logistic Regression and Random Forest achieves 72-74% classification accuracy, with facility type and BPJS partnership emerging as strongest predictors. Findings indicate that SatuSehat adoption is determined by multi-level factors spanning regional development conditions and organizational characteristics. Policy implications suggest differentiated implementation strategies are required, with comprehensive support targeting low-development regions and capacity-building interventions for independent practices to achieve equitable national health information exchange integration.

---

*Keywords: SatuSehat, Health Information Exchange, Technology Adoption, Machine Learning, Regional Disparities.*





© Copyright 2026  
by Priskila Ranindhita Hapsari  
All rights reserved

## DEDICATION

I dedicate this thesis to Indonesia and to everyone committed to advancing better data and facilities in health for all.



## ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my family for their constant support, encouragement, and understanding throughout this journey. Their belief in me has been an invaluable source of strength.



## TABLE OF CONTENTS

|                                                                             | Page      |
|-----------------------------------------------------------------------------|-----------|
| <b>CHAPTER 1 - INTRODUCTION.....</b>                                        | <b>14</b> |
| 1.1 Background .....                                                        | 14        |
| 1.1.2. Problem Statement .....                                              | 16        |
| 1.2 Research Questions .....                                                | 17        |
| 1.3 Research Objectives.....                                                | 18        |
| 1.5. Hypothesis.....                                                        | 18        |
| 1.6. Scope and Limitations.....                                             | 19        |
| 1.7. Thesis Structure .....                                                 | 21        |
| <b>CHAPTER 2 - LITERATURE REVIEW.....</b>                                   | <b>22</b> |
| 2.1 . Digital Health Transformation Challenges in Developing Countries..... | 22        |
| 2.2. Health Information Exchange Implementation Complexity.....             | 22        |
| 2.3. Indonesian Healthcare System Context .....                             | 23        |
| 2.4. Technology Adoption Theoretical Frameworks .....                       | 23        |
| 2.4.1. Technology Acceptance Model (TAM) and Extended Models.....           | 23        |
| 2.4.2. Diffusion of Innovation Theory .....                                 | 24        |
| 2.4.3. Implementation Science Frameworks.....                               | 24        |
| 2.4.4. Multi-Level Analysis in Technology Adoption.....                     | 25        |
| 2.5. Predictive Modeling in Healthcare Technology Adoption .....            | 25        |
| 2.6. Machine Learning Methodologies for Healthcare Adoption Research .....  | 26        |
| 2.6.1. Supervised Learning Approaches .....                                 | 26        |
| 2.6.2. Logistic Regression Analysis.....                                    | 27        |
| 2.6.3. Decision Tree Methods .....                                          | 27        |
| 2.6.4. Unsupervised Learning Approaches .....                               | 27        |
| 2.6.5. Clustering Analysis .....                                            | 27        |
| 2.6.6. Temporal Cross-Validation for Healthcare Data .....                  | 28        |
| 2.7. Feature Engineering Principles for Healthcare Adoption Studies .....   | 28        |
| 2.7.1. Categorical Variable Encoding .....                                  | 28        |
| 2.7.2. Geographic Feature Engineering .....                                 | 29        |
| 2.7.3. Temporal Feature Engineering .....                                   | 29        |
| 2.7.4. Organizational Feature Engineering .....                             | 29        |
| 2.8. Statistical Concepts and Evaluation Metrics.....                       | 30        |
| 2.8.1 P-value and Statistical Significance.....                             | 30        |

|                                                                                          |           |
|------------------------------------------------------------------------------------------|-----------|
| 2.8.2 R <sup>2</sup> (Coefficient of Determination) .....                                | 32        |
| 2.8.3. Machine Learning Model Performance Metrics.....                                   | 34        |
| 2.9. Review of Previous Studies .....                                                    | 39        |
| 2.9.1. Health Information System Adoption in Developing Countries.....                   | 39        |
| 2.9.2. Geographic and Regional Determinants of Healthcare Technology<br>Adoption .....   | 40        |
| 2.9.3. Machine Learning Applications in Healthcare Technology Adoption<br>Research ..... | 41        |
| 2.9.4. Limited-Attribute Predictive Modeling Challenges.....                             | 44        |
| 2.9.5. Cross-Sectional vs. Longitudinal Predication Approaches.....                      | 45        |
| 2.9.6. Indonesian Healthcare System and Digital Health Context.....                      | 46        |
| 2.9.7. Vendor Ecosystem and Market Dynamics in Health IT .....                           | 48        |
| 2.10. Research Gap Analysis .....                                                        | 49        |
| 2.10.1. Contribution to Knowledge.....                                                   | 51        |
| Theoretical Contributions .....                                                          | 51        |
| Methodological Contributions .....                                                       | 51        |
| Practical Contributions.....                                                             | 52        |
| Contextual Contributions .....                                                           | 52        |
| 2.10. Integrated Conceptual Framework.....                                               | 52        |
| <b>CHAPTER 3 – RESEARCH METHODS.....</b>                                                 | <b>55</b> |
| 3.1. Research Design.....                                                                | 55        |
| 3.1.1. CRISP-DM Framework Application .....                                              | 56        |
| 3.2. Data Collection Methods .....                                                       | 57        |
| 3.2.1. Primary Data Collection: SatuSehat Connectivity Status .....                      | 57        |
| 3.2.2. Secondary Data Collection: Regional Characteristics.....                          | 61        |
| 3.2.3. Data Integration and Merging .....                                                | 63        |
| 3.3. Success Criteria and Measurement Framework.....                                     | 66        |
| 3.3.1. Definition of Success Criteria .....                                              | 66        |
| Primary Success Measure .....                                                            | 66        |
| 3.3.2. Success Criteria Benchmarks.....                                                  | 67        |
| 3.4. Data Analysis Techniques.....                                                       | 68        |
| 3.4.1. Descriptive Analysis .....                                                        | 68        |
| 3.4.1.1. Temporal Trend Analysis .....                                                   | 68        |
| 3.4.1.2. Statistical Summaries.....                                                      | 68        |
| 3.4.2. Correlation Analysis .....                                                        | 68        |
| 3.4.2.1. Regional Characteristics Analysis .....                                         | 68        |
| 3.4.3. Facility-Level Predictive Modeling .....                                          | 68        |
| 3.4.3.1. Classification Models.....                                                      | 68        |
| 3.4.3.2. Predictor Variables.....                                                        | 69        |
| 3.4.3.3. Model Selection .....                                                           | 69        |
| 3.4.3.4. Validation Strategy .....                                                       | 70        |

|                                                                        |           |
|------------------------------------------------------------------------|-----------|
| 3.4.3.5. Addressing Limited Facility Attributes .....                  | 70        |
| 3.4.4. Model Validation and Performance Assessment .....               | 70        |
| 3.4.4.1. Spatial Cross-Validation for Facility-Level Models.....       | 70        |
| 3.4.4.2. Performance Metrics .....                                     | 71        |
| 3.5. Unit of Analysis and Data Structure .....                         | 71        |
| 3.5.1. Multi-Level Units of Analysis .....                             | 71        |
| 3.5.1.1. Regional-Level Analysis.....                                  | 71        |
| 3.5.1.2. Facility-Level Analysis .....                                 | 71        |
| 3.5.2. Expected Data Structure by Analytical Level.....                | 72        |
| 3.5.2.1. Regional-Level Dataset Structure .....                        | 72        |
| 3.5.2.2. Facility-Level Dataset Structure .....                        | 72        |
| <b>CHAPTER 4 – RESULTS AND DISCUSSIONS.....</b>                        | <b>74</b> |
| 4.1. Initial Evaluation.....                                           | 74        |
| 4.1.1. Data Collection Summary .....                                   | 74        |
| Primary Data Collection: SatuSehat Connectivity Database .....         | 74        |
| Secondary Data Collection: Master Sarana Index .....                   | 76        |
| Secondary Data Collection: Regional Characteristics .....              | 78        |
| Variables Excluded Due to Data Availability Constraints.....           | 79        |
| 4.1.2. Data Quality Assessment .....                                   | 80        |
| Primary Data Quality: SatuSehat Connectivity Database.....             | 80        |
| Secondary Data Quality: Regional Characteristics .....                 | 84        |
| Excluded Variables and Justification .....                             | 88        |
| 4.2 Data Analysis .....                                                | 93        |
| 4.2.1 Regional Characteristics and Adoption Patterns (RQ1) .....       | 94        |
| 4.2.1.1 Correlation Analysis Results.....                              | 94        |
| 4.2.1.2 Scatter Plot Analysis .....                                    | 96        |
| 4.2.1.4 Temporal Adoption Trends.....                                  | 99        |
| 4.2.1.5 Synthesis of RQ1 Findings .....                                | 100       |
| 4.2.2 Facility-Level Predictive Modeling (RQ2) .....                   | 101       |
| 4.2.2.1 Modeling Approach and Dataset Configuration .....              | 101       |
| 4.2.2.2 Model Performance Evaluation .....                             | 103       |
| 4.2.2.3 Confusion Matrix Analysis .....                                | 104       |
| 4.2.2.4 Feature Importance Analysis .....                              | 106       |
| 4.2.2.5 Model Limitations and Data Constraints .....                   | 108       |
| 4.2.2.6 Implications for Targeted Implementation .....                 | 109       |
| 4.2.2.7 Synthesis of RQ2 Findings .....                                | 110       |
| 4.3. Barriers to Adoption: Insights from Field Experts.....            | 111       |
| 4.3.1. Identified Barriers to SatuSehat Adoption .....                 | 111       |
| 4.3.2. Implications for Differentiated Implementation Strategies ..... | 115       |
| 4.3.3. Limitations of Expert Consultation Approach .....               | 115       |

|                                                                  |            |
|------------------------------------------------------------------|------------|
| <b>CHAPTER 5 – CONCLUSIONS AND RECOMMENDATIONS.....</b>          | <b>117</b> |
| 5.1 Conclusions.....                                             | 117        |
| 5.1.1 Regional Characteristics and Adoption Patterns (RQ1) ..... | 117        |
| 5.1.2 Facility-Level Predictive Modeling (RQ2) .....             | 118        |
| 5.1.3 Integration of Multi-Level Findings .....                  | 119        |
| 5.2 Recommendations.....                                         | 119        |
| 5.2.1 Policy Recommendations for Ministry of Health .....        | 119        |
| 5.2.2 Recommendations for Future Research .....                  | 126        |
| 5.3 Research Limitations .....                                   | 129        |
| 5.4 Closing Remarks .....                                        | 131        |
| <b>REFERENCES .....</b>                                          | <b>136</b> |
| <b>EXPERIENCE .....</b>                                          | <b>144</b> |
| <b>EDUCATION .....</b>                                           | <b>144</b> |



---

## LIST OF FIGURES

### Figures

- 1.1 Spatiotemporal distribution pattern of RDI of Java
- 1.2 Fishbone Diagram
- 2.1 Conceptual framework (Multi-level nested structure)
- 3.1 Research Framework Diagram
- 3.2 CRISP-DM Research Process Foundation



---

## LIST OF TABLES

### Table

- 2.1: Confusion Matrix
- 2.2: Comparison of Machine Learning Classification Methods
- 2.3: Research Gap Analysis and Study Contributions
- 2.4: Analytical Alignment with Framework Domains
- 4.1: Primary Dataset Variables
- 4.2: Master Sarana Index Summary
- 4.3: Regional Variables Collected
- 4.4: Primary Dataset Summary by Time Point
- 4.5: Regional Variables Data Completeness
- 4.6: Excluded Regional Variables and Rationale
- 4.7: Variable Set Comparison with Prior Health IT Adoption Studies
- 4.8: Correlation Between Regional Characteristics and Adoption Rates
- 4.9: Provincial-Level Summary Statistics
- 4.10: Classification Model Performance Comparison
- 4.11: Confusion Matrix - Logistic Regression
- 4.12: Confusion Matrix - Random Forest
- 4.13: Feature Importance Rankings
- 5.1: Mapping of Research Insights to Policy Recommendations