

PREDICTING INDONESIAN ISP COMPLIANCE WITH NEGATIVE CONTENT
DATA SYNCHRONIZATION: A SUPERVISED LEARNING APPROACH

By

Dika Aghniyasyak Priyo Faqih
22251022

MASTER'S DEGREE
in

MASTER OF INFORMATION TECHNOLOGY
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 August 13, 2025

STATEMENT BY THE AUTHOR

I hereby declare that the work I am submitting is entirely my own. I confirm that it does not contain any material published or written by others, nor does it include content that has been significantly submitted for a degree or diploma at any educational institution, unless proper credit has been given within this thesis.


Dika Aghniyasyak Priyo Faqih

Student

27 August 2025

Date

Approved by:


Dr. Ir. Eka Budiarto, S.T., M.Sc.

Thesis Advisor

 27. 8. 2025

Date


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

Thesis Co-Advisor

27.8.2025

Date


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

Dean

27.8.2025

Date

Dika Aghniyasyak Priyo Faqih

ABSTRACT

PREDICTING INDONESIAN ISP COMPLIANCE WITH NEGATIVE CONTENT DATA SYNCHRONIZATION: A SUPERVISED LEARNING APPROACH

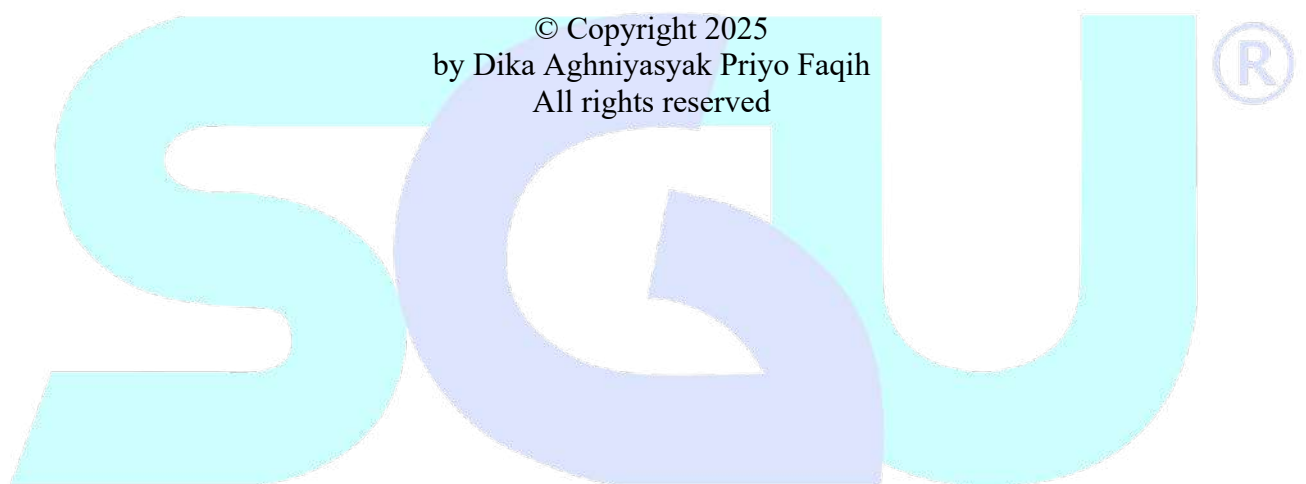
By

Dika Aghniyasyak Priyo Faqih
Dr. Ir. Eka Budiarto, S.T., M.Sc, Advisor
Dr. Ir. Maulahikmah Galinium, S.Kom., M.Sc, Co-Advisor

SWISS GERMAN UNIVERSITY

The Indonesian government, through the Ministry of Communication and Information Technology (Kominfo), mandates that all Internet Service Providers (ISPs) actively filter negative content, creating a significant regulatory and technical challenge. This study presents a data-driven approach to predict ISP compliance with these mandates using supervised machine learning. Leveraging a dataset of operational metrics from 606 Indonesian ISPs for April 2025, this research defines compliance based on an ISP's foundational capability to perform network-level blocking, proxied by the implementation of Remotely Triggered Black Hole (RTBH) filtering. Key performance indicators, including DNS Queries Per Second (QPS), a DNS trustworthiness score (TNG), and Transactions Per Second (TPS), were used as predictive features. Five machine learning algorithms Logistic Regression, Support Vector Machine (SVM), Decision Tree, Naive Bayes, and Random Forest were trained and evaluated. The results demonstrate high predictive accuracy, with the Decision Tree and Random Forest models achieving 97% accuracy. Feature importance analysis, using the Gini impurity method, revealed that QPS was the most significant predictor, serving as a latent indicator of network

Keyword : Internet Service Provider (ISP), Regulatory Compliance, Supervised Machine Learning, Predictive Modeling, Content Filtering Indonesia.



DEDICATION

My family, comprising my wife and my children, has been a constant source of support and encouragement, always believing in my abilities. They serve as my greatest strength and inspiration throughout this journey.

My friends, who have supported me through both my struggles and successes, offering unwavering encouragement and empathy.

I would like to express my gratitude to my colleagues and workplace for creating an environment that prioritizes growth and learning, allowing me to reach this milestone while managing my professional responsibilities.



ACKNOWLEDGEMENT

I am sincerely grateful to Dr. Ir. Eka Budiarto, S.T., M.Sc, and Mr. Dr. Ir. Maulahikmah Galinium, S.Kom., M.Sc for their invaluable support, guidance, and advice as my Thesis advisors throughout this research project and the preparation of this thesis report. Their contributions have been instrumental in bringing this thesis and the entire research project to fruition.

I want to express my heartfelt gratitude to my family for the immense moral support they've given me throughout my journey of knowledge at SGU. I truly appreciate all the time you've dedicated to helping me complete this thesis. Thank you so much!

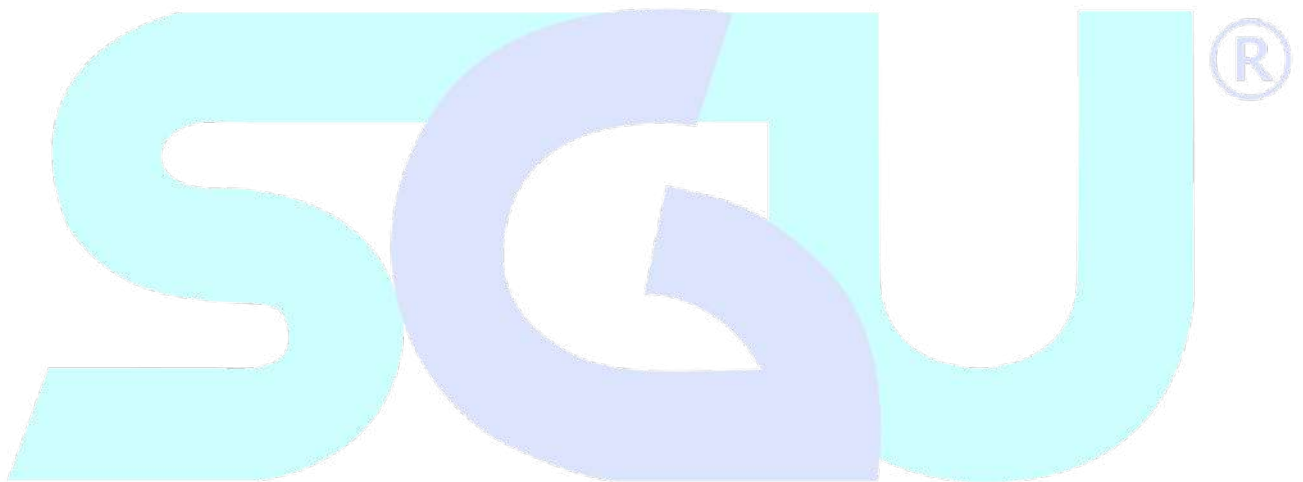
I want to express my gratitude to my direct supervisor, my work team, and my friends, particularly Batch 31 MIT SGU, for their support and companionship. I also extend my thanks to the many individuals who contributed to this research project, whether directly or indirectly.

TABLE OF CONTENTS

	Page
STATEMENT BY THE AUTHOR	2
ABSTRACT	3
DEDICATION	5
ACKNOWLEDGEMENT	6
TABLE OF CONTENTS	7
LIST OF FIGURES	10
LIST OF TABLES	11
CHAPTER 1 – INTRODUCTION	12
1.1. Background	12
1.2. Research Problem	16
1.3. Research Objective	17
1.4. Research Question	17
1.5. Hypothesis	18
1.6. Research Scope and Limitation	18
1.7. Significance of Study	19
1.8. Thesis Structure	20
CHAPTER 2 - LITERATURE REVIEW	23
2.1. Ministry of Communication Informatic	23
2.2. Existing Regulations Related to Domain and IP Blocking in Indonesia	24
2.3. APJII	25
2.4. ISP	26
2.5. Domain Name Server (DNS)	27
2.6. Response Policy Zone (RPZ)	28
2.7. DNS Trust Next Generation (TNG)	29
2.8. Remotely Triggered Black Hole (RTBH) Filtering	30
2.9. Machine Learning	31
2.10. Supervised Machine Learning	32
2.11. Random Forest	33
2.12. Logistic Regression	34
2.13. Support Vector Machines (SVM)	34

2.14. Linear Regression	35
2.15. KNN	35
2.16. Decision Trees	35
2.17. Naive Bayes	36
2.18. Random Forest	36
2.19. Successful Metrics	37
2.19.1 Quality of Metrics	38
2.20. Related Works	40
2.21. Evaluation Methods	47
CHAPTER 3 – RESEARCH METHODS	50
3.1. Research Framework	50
3.2. Data Collection	50
3.3. Data Preprocessing	52
3.4. Modelling	53
3.4.1. Feature Engineering	53
3.4.2. Exploration Data	54
3.4.3. Model Development	54
3.5. Evaluation	55
3.6. Validation	57
CHAPTER 4 – RESULT AND DISCUSS	60
4.1. Data Collections and Preparation for Data Processing	60
4.2. Feature Engineering Deriving the Binary Compliance Metric	61
4.3. Feature Selection	62
4.4. Exploratory Data Analysis (EDA) of ISP Compliance Characteristics	63
4.4.1. Correlation and Distribution Analysis	65
4.5. Feature Importance Analysis with Gini Impurity	66
4.6. Supervised Learning Models for Compliance Prediction	67
4.6.1. Model Selection and Research Setup	67
4.6.2. Logistic Regression	69
4.6.3. Support Vector Machine (SVM)	70
4.6.4. Decision Tree	71
4.6.5. Naive Bayes	73
4.6.6. Random Forest	73
4.7. Comparative Analysis and Identification of Compliance	75
4.7.1. Confusion Matrix Analysis	75
4.8. Cross-Model Performance Evaluation	78

4.9. Key Factors Influencing Compliance	79
4.10. Justification and Implications of Feature Importance	82
4.11. Research Validation via Focus Group Discussion (FGD)	83
CHAPTER 5 – CONCLUSION AND RECOMENDATION	87
5.1. Conclusion	87
5.2. Recomendation	87
5.2.1. Actionable Recommendations For Regulators (Kominfo)	88
5.2.2. Actionable Recommendations For Non-Compliant and At-Risk ISPs	88
5.3. Future Work	89
REFERENCES	90
GLOSSARY	94
Appendix	100



LIST OF FIGURES

Figures	Page
Figure 1.1 Statistics of Internet Users in Indonesia	12
Figure 1.2 Site Bloking Mechanism	14
Figure 1.3 Total Internet Service Providers in Indonesia 2023	15
Figure 1.4 Fishbone Diagram about ISP compliance	16
Figure 2.1 Model Tree of Domain	28
Figure 2.2 Overview of DNS RPZ	29
Figure 3.1 Research Framework	50
Figure 3.2 Arcitecture DNS Negative Content Filtering to ISPs comply	51
Figure 3.3 Data Collection Process	52
Figure 3.4 Flow Data Cleansing part of Data Processing	53
Figure 4.1 Data Understanding	60
Figure 4.2 Permutation Importance	79
Figure 4.3 SHAP Interaction Plot between TPS and QPS	80

LIST OF TABLES

Table	Page
Table 2.1 Number of Members and Internet Penetration	27
Table 2.2 Comparison table of unsupervised learning and supervised learning	32
Table 2.3 S.M.A.R.T Goals	37
Table 2.4 Comparison of Related Works in ISP Compliance and Negative Content Data Synchronization Metrics	41
Table 4.1 5 (five) raw data collected	60
Table 4.2 5 (five) head of raw dataset	61
Table 4.3 Descriptive Statistic	63
Table 4.4 Descriptive Statistics of Key Metrics by Compliance Status	64
Table 4.5 Feature Correlation Matrix	65
Table 4.6 Logistic Regression Performance Metrics	69
Table 4.7 Support Vector Machine Performance Metrics	71
Table 4.8 Decision Tree Performance Metrics	72
Table 4.9 Naive Bayes Performance Metrics	73
Table 4.10 Random Forest Performance Metrics	74
Table 4.11 Logistic Regression Confusion Matrix	76
Table 4.12 Support Vector Machine (SVM) Confusion Matrix	76
Table 4.13 Decision Tree Confusion Matrix	77
Table 4.14 Naive Bayes Confusion Matrix	77
Table 4.15 Random Forest Confusion Matrix	77
Table 4.16 Consolidated Performance Metrics Across All Models	78
Table 4.17 Feedback and perspectives from practicality in FGD	84