

CHAPTER 1 – INTRODUCTION

1.1. Background

Within a decade, the growth of internet users in Indonesia has shown a very rapid increase. Based on data in Figure 1.1 from We Are Social, the number of internet users in Indonesia rose from 106 million users in January 2018 to 185.3 million users in December 2024. This growth represents an almost increase in the period of years, placing Indonesia as one of the one country with the fastest internet development in the world. This phenomenon shows that internet access has become an inseparable part of Indonesian people's lives, both in terms of communication, entertainment, and economic and social activities.

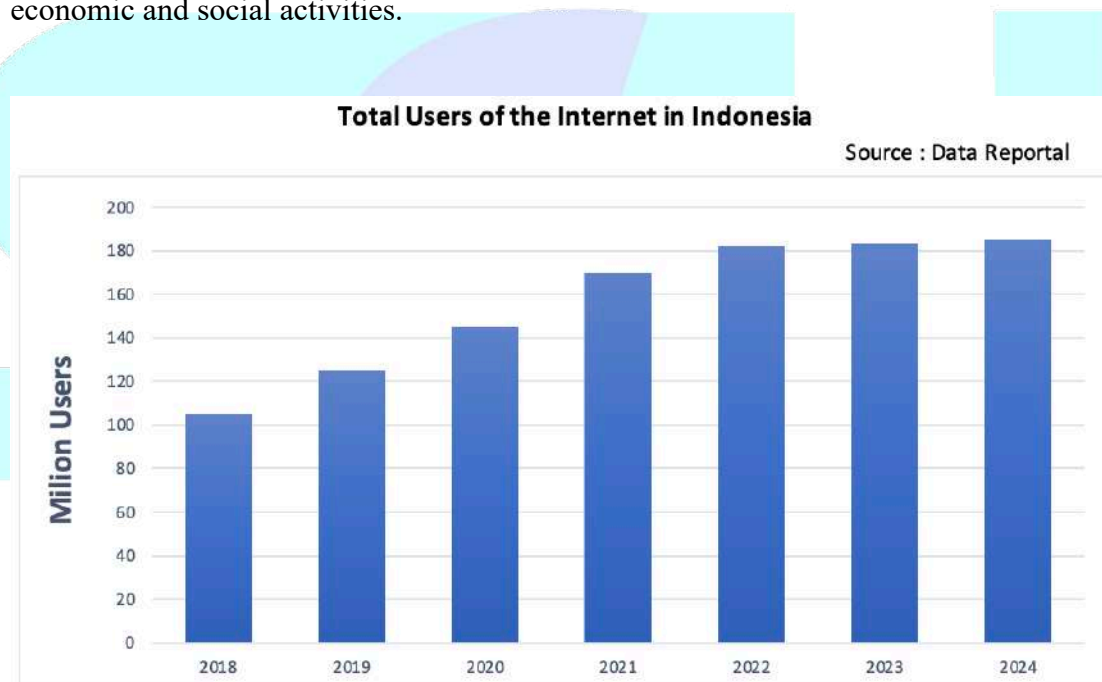


Figure 1.1 Statistics of Internet Users in Indonesia

However, along with the increase in the number of internet users, there has also been a growth in negative content circulating in cyberspace. Content such as gambling and pornography is still a rampant problem and difficult to overcome. This creates challenges for the government and internet service providers in maintaining a healthy and safe digital ecosystem. Based on statistics published by the Ministry of Communication and Information (Kominfo), more than 1.6 million gambling sites and

2.8 million pornographic sites have been blocked by 2022, a corresponding increase in the number of internet users each year contributes to the persistent growth of this problem (Direktorat Jenderal Aplikasi Informatika, 2021). Therefore, effective and innovative steps are needed in filtering and blocking negative content in order to create a safer and more productive internet environment for Indonesian people.

The internet has become an integral part of people's lives today as a medium of information. It is not only needed for individuals, but also for companies and offices. However, the cost of providing internet is very expensive, but the demand for internet networks is always high. This condition provides a business opportunity that can be exploited by creating an Internet Service Provider (ISP).

In accordance with Law No. 19 of 2016 concerning Amendments to Law Number 11 of 2008 concerning Electronic Information and Transactions, especially as written in Article 40 paragraph 2, the government is obliged to protect the public interest from all types of interference as a result of misuse of electronic information and transactions. electronic devices that disturb public order, in accordance with the provisions of laws and regulations. These conditions and legal foundations are the basis for handling content and terminating PSE blocking access.

To carry out this task, the Directorate of Information Application Control has an AIS team as a verification and internet content blocking team by blocking domain addresses from web pages that violate regulations and laws. Internet content control activities cannot be separated from public reports, results from internet search engines and through a process of verifying truth and accuracy in accordance with statutory and statutory regulations.



Figure 1.2 Site Blocking Mechanism

Technically in site blocking mechanism at figure 1.2 from , blocking of web pages currently uses the Response Policy Zone (RPZ) machine which synchronizes with the Internet Access Service Provider's Domain Name System (DNS) server. The public can view and check the blacklist of web pages blocked by Kominfo which is published on the trustpositive.kominfo.go.id site. Meanwhile, blocking data access from users/public is still carried out by Internet Access Service Providers (ISPs) by synchronizing with the database. Then, in accordance with the mandate of the ITE Law number 19 of 2016 Article 40 paragraph 2 and the Minister of Communication and Information Regulation Number 10 of 2021 concerning Amendments to the Minister of Communication and Information Regulation Number 5 of 2020 concerning Private Electronic System Operators (Menteri Komunikasi Dan Informatika Republik Indonesia, 2020), the Government of the Republic of Indonesia is obliged to supervise the PSE that does not comply with statutory regulations.

One of the main challenges in synchronizing the Response Policy Zone (RPZ) system is that not all ISPs in Indonesia are connected to this system. In addition, there are differences in the time required by each ISP to synchronize data, causing discrepancies between blocking times and policy implementation on various networks. This is exacerbated by the RPZ architectural model which is distributed or not centralized, allowing some negative content to escape filtering. Apart from that, the RPZ system is

currently not capable of directly blocking IP addresses, so that content accessed via certain IP addresses can still pass through existing filters, and this is a serious obstacle in maintaining network security from negative content. Current conditions in 2023-2024 at figure 1.3, there are around 44 (forty four) Internet Access Gateway Providers (Network Access Providers/NAP), 1011 (one thousand eleven) Internet Access Service Providers and 6 (seven) Cellular Operators.

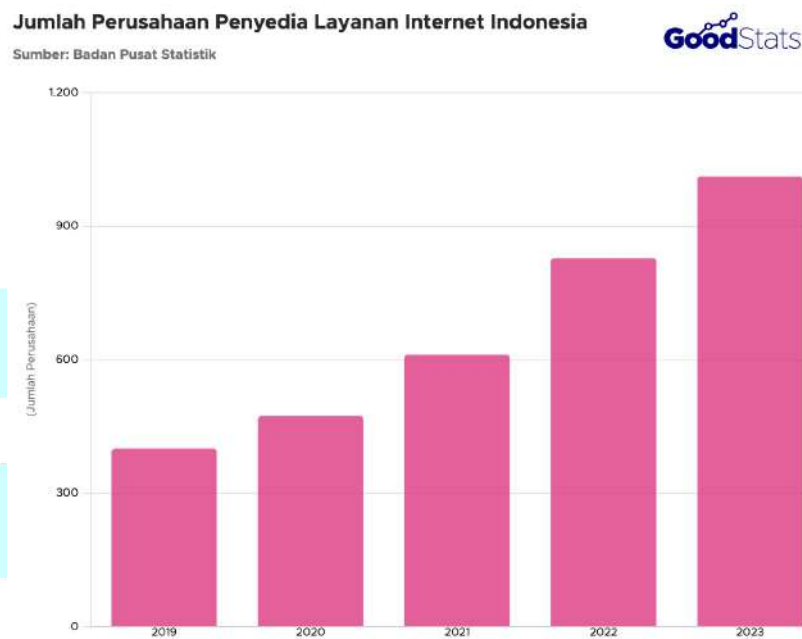


Figure 1.3 Total Internet Service Providers in Indonesia 2023

The very fast growth of Internet content needs to be controlled to create a clean Internet ecosystem, as well as upholding state sovereignty in the cyber area. Blocking by the Ministry of Communication and Information is carried out in 2 ways, namely using the RPZ (DNS Filtering) system, TKPPSE (TCP Reset) system. The DNS Filtering system functions to reach regional ISPs, while the TKPPSE are more for mobile operator Internet Service Providers (ISPs) and Network Access Providers (NAP).

This research uses a directed machine learning approach to predict the compliance of ISPs in Indonesia towards negative content data synchronization. By utilizing the collected data of 521 ISPs combined with data from APJII and additional information collected, a machine learning model will be trained to identify patterns and factors associated with ISP compliance. The results of this study are expected to provide a

better understanding of the factors that influence ISP compliance and assist the government in formulating effective strategies to improve compliance. Given the complexity and large amount of ISP data, a machine learning approach is appropriate for predicting ISP compliance. Machine learning enables the identification of patterns and factors that influence ISP compliance from existing

1.2. Research Problem

Part of the government's effort to control negative content is by requiring ISPs to synchronize negative content data. This synchronization aims to make it easier for the government to identify and block access to prohibited content. However, the effectiveness of this system faces challenges, especially since not all Internet Service Providers (ISPs) and the slowing growth of internet penetration have the potential to affect the effectiveness of content regulation strategies, including the negative content data synchronization policy. In addition, the implementation of this policy faces various challenges, including ISP awareness, infrastructure availability, and other technical factors.

Based on the background in the previous section. I use fishbone diagram to look challenges at four main areas: People, Process, Technology, and Governance in Figure

1.4

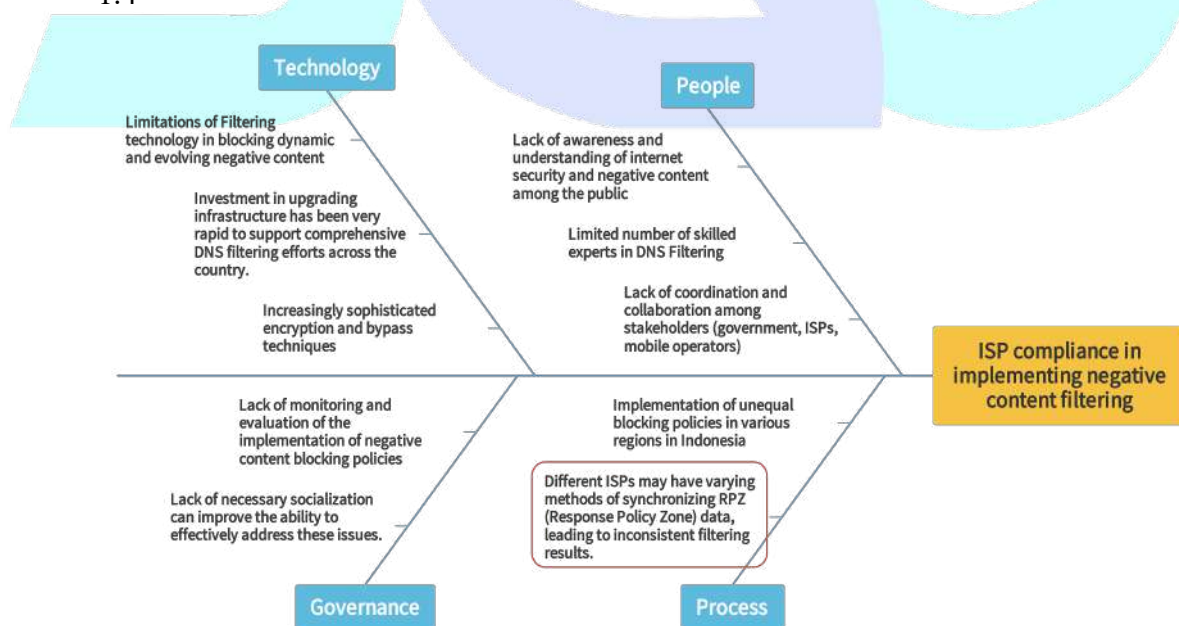


Figure 1.4 Fishbone Diagram about ISP compliance

This research uses a directed machine learning approach to predict the compliance of ISPs in Indonesia towards negative content data synchronization. By utilizing existing data and additional information collected, machine learning models will be trained to identify patterns and factors associated with ISP compliance. The results of this study are expected to provide a better understanding of the factors that influence ISP compliance and assist the government in formulating effective strategies to improve compliance.

1.3. Research Objective

Based on the problem formulation, this study aims to analyze whether there is a significant relationship such as :

1. Analyze the factors that influence ISPs' compliance with negative content data synchronization in Indonesia.
2. Build a machine learning model to predict ISPs' compliance with negative content data synchronization.
3. Identify ISPs that are predicted to be non-compliant and provide recommendations to improve compliance.

1.4. Research Question

This research was conducted to provide solutions to the problems described in the Background which are able to provide an overview of the current conditions along with recommendations by designing a research question, namely:

1. What factors influence ISPs' compliance with negative content data synchronization in Indonesia ?
2. How to build an accurate machine learning model to predict ISPs' compliance with negative content data synchronization ?
3. Which ISPs are predicted to be non-compliant and what are the recommendations to improve their compliance ?

1.5. Hypothesis

The hypothesis for this thesis are as follow :

1. H1 : DNS filtering method has a significant effect on ISP compliance. ISPs that use DNS filtering with manual input on the router are assumed to be less compliant as this method is less efficient and prone to human error.
2. H2 : ISPs located in big cities with more developed telecommunications infrastructure are hypothesized to be more compliant than ISPs located in small towns.
3. H3 : Random Forest will produce a model with the highest prediction accuracy compared to other machine learning algorithms (Logistic Regression, SVM, Decision Trees, Naive Bayes, Random Forest) in predicting ISP compliance with negative content data synchronization. Random Forest is hypothesized to be superior due to its ability to overcome overfitting and handle high data.

1.6. Research Scope and Limitation

The significance study of this research are as follows.

1. The scope of this research focuses on predicting the compliance of ISPs in Indonesia with negative content data synchronization based on data member of ISP from APJII in December 2024 and additional data collected from 912 data.
2. For Limitation
 - i. This research is limited to the available data and may not cover all factors affecting ISP compliance.
 - ii. The accuracy of machine learning models depends on the quality and quantity of data.
 - iii. The results may not be generalizable to all ISPs in Indonesia due to data limitations and the scope of the research area.
 - iv. This study may have limited access to more complete and updated government data related to ISP compliance.

- v. Changes in government regulations and policies related to negative content may affect the results and relevance of this study.

1.7. Significance of Study

This study has important significance in the Indonesian context, given the increasing use of the internet and the threat of spreading negative content. Practically, the results of this study are expected to provide recommendations for Kominfo in optimizing the DNS Filtering system, so that it can block negative content more effectively.

This research contributes to the development of science in the field of machine learning and cybersecurity, especially in the application of supervised learning to predict compliance with cybersecurity regulations. Research on pornography image classification often uses the skin detection method as the main method in the classification process. Research conducted by (Moreira and Fachine, 2018) combines two methods for the process of pornographic image classification. The methods used are the Viola-Jones algorithm and skin detection. The Viola-Jones algorithm is used to detect any pornographic objects contained in an image, examples of objects are breasts, cleavage, and genitals. However, the Viola-Jones algorithm has the disadvantage that it is very sensitive to changes in image orientation of pornographic objects, so it will give a very high false-positive value in the classification process.

And practically, the results of this research can provide valuable information for the Ministry of Communication and Information Technology (Kominfo) in formulating more effective strategies and policies to improve ISP compliance with negative content data synchronization. Benefits for APJII This research can help APJII understand the factors that influence the compliance of its members and formulate targeted assistance programs to improve compliance. This research can also be a reference for future research related to cybersecurity, internet regulation, and machine learning implementation in Indonesia. Increased ISP compliance with negative content data synchronization will contribute to the creation of a safer and healthier internet environment for Indonesian. By identifying factors that influence ISP compliance and predicting potentially non-compliant ISPs, this research is expected to provide useful

recommendations for stakeholders to improve the effectiveness of negative content regulation enforcement in Indonesia.

This study has crucial social significance, especially in mitigating the negative impact of online gambling which is increasingly prevalent in Indonesia. Online gambling has caused various social and economic problems, such as increased poverty, financial losses, school dropouts, crime, and even suicide. By identifying the factors that influence the effectiveness of DNS Filtering, this study can provide information to the public about how the blocking system works in preventing access to online gambling sites, its limitations, and how to protect themselves and their families from the dangers of online gambling. This knowledge is expected to increase public awareness of the importance of internet safety and wise use of the internet, thereby reducing the number of online gambling victims and creating a healthier internet environment. In addition, this study can also provide encouragement for stakeholders, including the government, ISPs, and the community, to work together to tackle the problem of online gambling through various efforts, ranging from increasing the effectiveness of DNS Filtering to education and digital literacy programs.

1.8. Thesis Structure

The systematics in this writing is organized systematically and divided into chapters in the following order.

Chapter 1 Introduction

This chapter describes the background, problem formulation, research questions, problem boundaries that become the scope of research, objectives, benefits, and writing systematics.

Chapter 2 Literature Review

This chapter will build a theoretical foundation for your research. You will dive into the literature relevant to DNS Filtering and negative content blocking. First, you will define what negative content is, classify it by types such as pornography, gambling, and hate speech, and examine its impact on society. Next, you will cover the concept of DNS Filtering in depth, explain its working mechanism, and discuss the various

techniques used, including RPZ (Response Policy Zone), TCP RST, Trust-Next Generation (TNG), and RTBH (Remote Trigred Black Hole). In addition, you will examine the policies and regulations in Indonesia related to blocking negative content, including regulation of the Ministry of Communication and Information and the applicable laws and regulations. Finally, you will review previous research on the effectiveness of DNS Filtering in blocking negative content, both in Indonesia and in other countries, to gain a comprehensive overview of the developments and challenges in this field.

Chapter 3 Research Methodology

This chapter will detail your research methodology. You will explain the type of research chosen, such as quantitative, qualitative or mixed methods, and why this approach is appropriate to answer the research question. Then, you will describe the data collection methods used, such as literature review, document analysis, observation, or interviews, and explain how the data was collected. The data sources used will also be explained, such as secondary data from the Ministry of Communication and Information, data from ISPs, or data from other open sources. Finally, you will describe the techniques used to analyze the data, such as statistical analysis, content analysis, or descriptive analysis, and explain how these techniques help interpret the data and answer the research questions.

Chapter 4 describes the research results

This chapter will present the research results and discussion in a structured manner. First, you will present the data that has been collected through various methods, either in the form of tables, graphs, or narratives, so that readers can understand the general picture of the data. Next, you will analyze the data using the techniques explained in the previous chapter, answer the research questions, and test the hypotheses that have been proposed. The results of the analysis will then be interpreted and linked to theory and previous research to provide context and depth to the research findings.

Chapter 5 explains the conclusions of the research, recommendations, and future work

This chapter will summarize the entire research and draw comprehensive conclusions. You will start by briefly and concisely summarizing the research objectives, methods and results. Then, you will present the conclusions drawn from the research, answer the research question posed at the beginning, and state whether the hypothesis is accepted or rejected based on the data and analysis. Next, you will mention the limitations that exist in the research, both in terms of data, methods, and scope of research, and provide suggestions for future research in order to complement or develop this research.

