

CHAPTER 3 – RESEARCH METHODS

3.1. Research Framework

For this thesis, we introduce a framework proposed to Monitoring and Data Synchronization in Optimizing DNS Filtering for Negative Content Blocking. The research framework proposed can see in Figure 3.1.

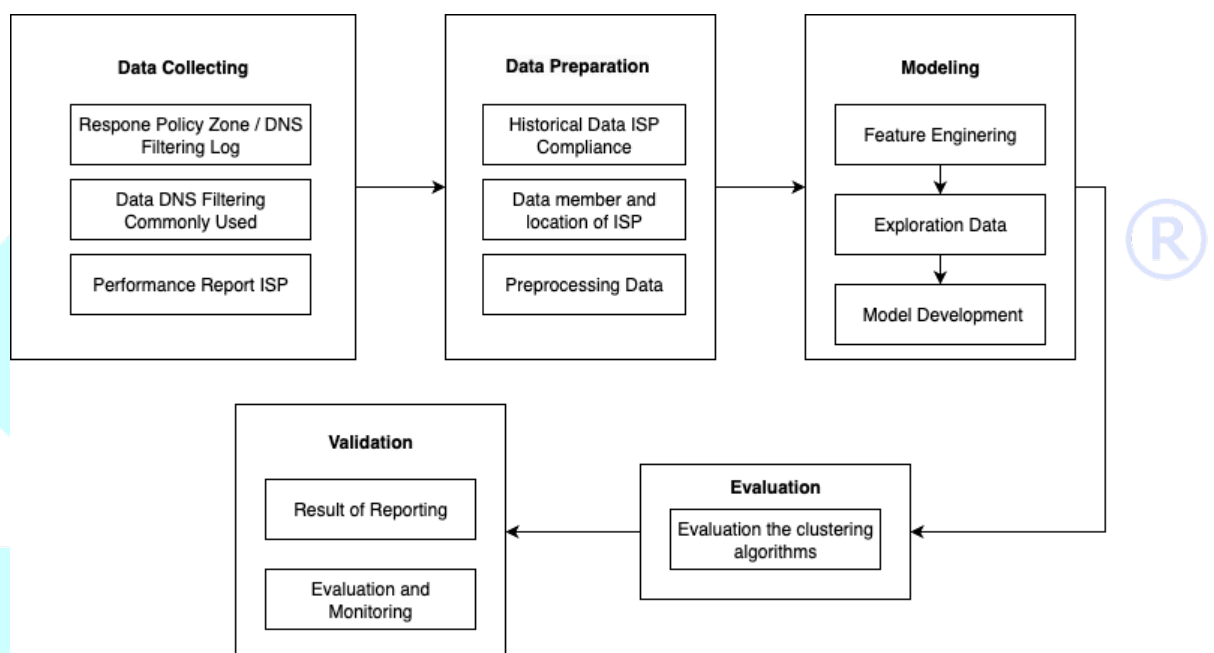


Figure 3.1 Research Framework

3.2. Data Collection

This phase involves gathering and organizing data to build both the ML model and the KG. The sources of data include :

1. This dataset is presented as a snapshot of performance metrics for telecommunications service operators in Indonesia, collected in April 2025, and is attributed to Kominfo's Directorate General of Posts and Informatics Operations (PPI), the body responsible for overseeing internet services

regulation. Feature Attributes Technical specifications and synchronize data from databases defining the characteristics of each ISPs.

2. Data DNS Filtering

The DNS filtering process to sync database negative content in Indonesia involves ISPs utilizing Kominfo's systems. When a user tries to access a website, their device sends a DNS query to their ISP's DNS server. The ISP's DNS server, acting as a slave to Kominfo's RPZ (Response Policy Zone) or employing DNS Trust-NG, processes this query.

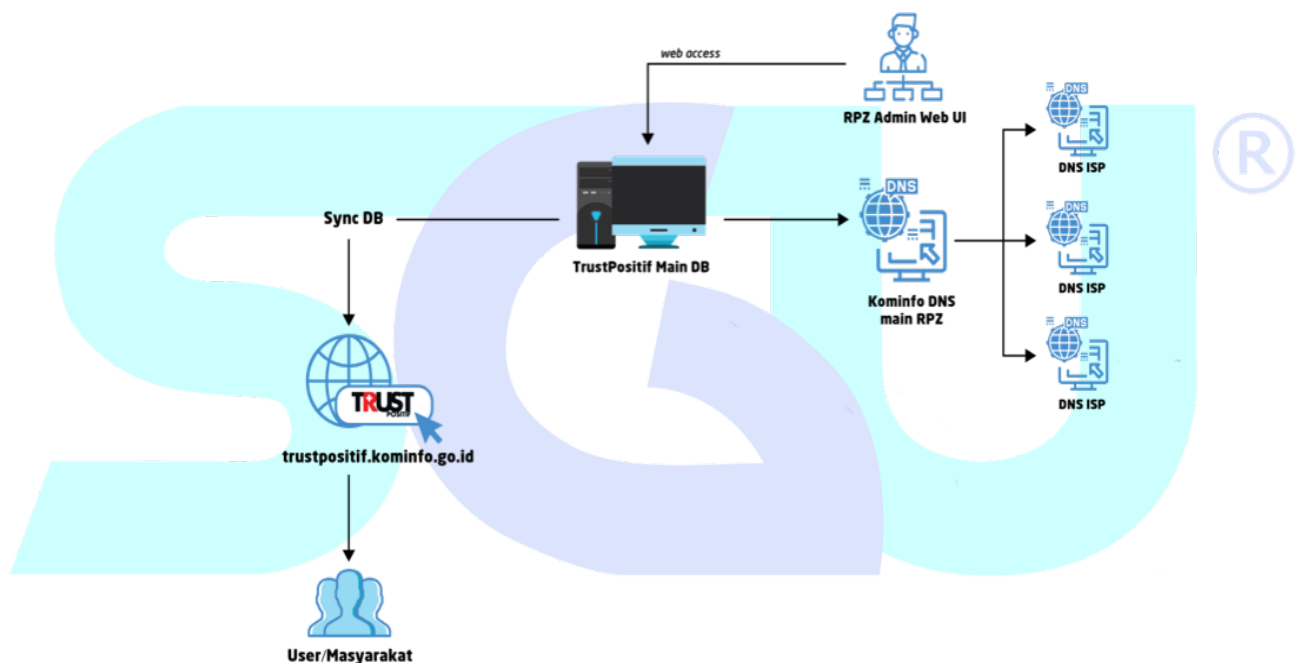


Figure 3.2 Arcitecture DNS Negative Content Filtering to ISPs comply

This figure 3.2 is arcitecture ensures that ISPs comply with their obligation to block content that violates Indonesian laws and regulations, as mandated by Kominfo. For RPZ, the ISP's BIND server synchronizes with Kominfo's "trustpositifkominfo" master zone, which lists domains containing negative content. If the requested domain is found in either the RPZ or DNS Trust-NG database, the ISP's DNS server will block access, either by redirecting the user

to a landing page or returning an response. If the content is not classified as negative, the DNS query proceeds through the standard DNS hierarchy to resolve the legitimate IP address, allowing the user to access the content. This system ensures that ISPs comply with their regulatory obligations to negative content filtering.

3. One of the primary sources is the AXFR logs, which stand for Authoritative Zone Transfer. This protocol is essential for transferring complete DNS zone data from a primary DNS server to a secondary DNS server. AXFR is particularly useful for maintaining consistency between DNS servers, as it ensures that all records are up-to-date and synchronized, There are three logs taken from the Trustpositive server, namely: `xfrouit-feed1.log`, `/var/log/bind9/xfrouit-feed2.log`, and `/var/log/bind9/xfrouit-feed3.log`. Each log has a different server location, so in this thesis it is assumed to be the RTBH1, RTBH2, and RTBH3 attributes. which can be leveraged for various analytical purposes can see in Figure 3.2.

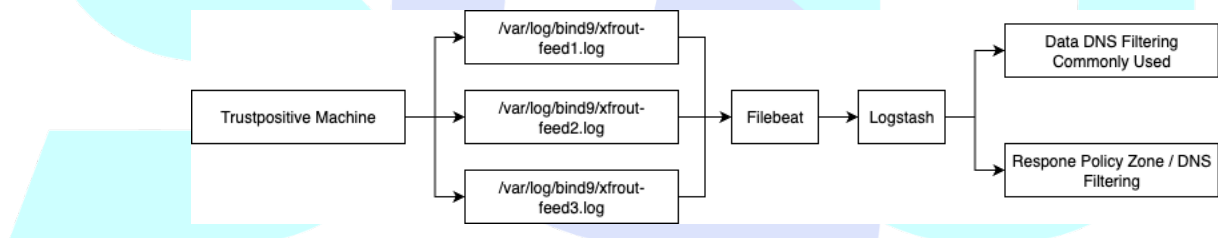


Figure 3.3 Data Collection Process

3.3. Data Preprocessing

Preprocessing steps data cleaning is a removing inconsistencies, errors, and handling missing data to ensure quality. Pandas provides two methods, `.isna()` and `.isnull()`, which both do the same thing for identify missing data. These methods will return another DataFrame with boolean values (True or False) indicating whether a particular element is missing or not. The raw dataset contains 912 records and includes the following columns: No (an index), ASN (Autonomous System Number), Organisation (the name of the entity), RTBH1, RTBH2, RTBH3 (metrics related to source log), TNG (a metric interpreted as DNS filtering trustworthiness), Total (a sum of the RTBH and

TNG metrics), QPS (Queries Per Second), TPS (Transactions Per Second), and Category. The Category column is particularly important as it identifies the type of organization, with values including 'ISP', 'NAP' (Network Access Point), 'Perusahaan' (Company), 'Universitas' (University), and 'Dinas' (Government Agency).

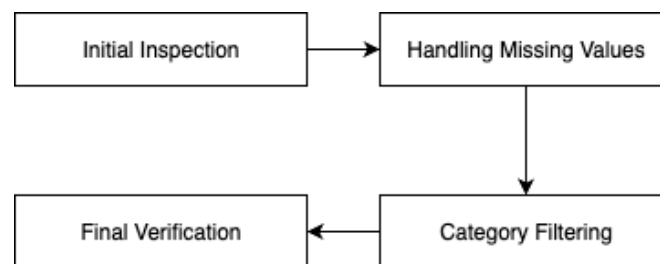


Figure 3.4 Flow Data Cleansing part of Data Processing

In Figure 3.4 The data cleansing process began with an initial inspection to understand the dataset's structure, data types, and identify missing values. This inspection revealed that certain critical columns, specifically QPS and TPS, had missing data in some rows. Given the relatively small number of affected rows and the lack of a reliable method for imputing these performance metrics, the most effective strategy was to remove these incomplete records, thereby ensuring the integrity and validity of the subsequent analysis. Following this, a crucial step was to filter the data by category.

3.4. Modelling

With the dataset cleansed to include only ISPs, the next step is to engineer the target variable—the metric the model will be trained to predict. For this study, a binary compliance variable was created to represent an ISP's foundational capability to adhere to Kominfo's content-blocking mandates.

3.4.1. Feature Engineering

The definition of this variable is based on the logic established in Section 3.3 an ISP's ability to perform network-level blocking via RTBH is a fundamental prerequisite for

compliance. Therefore, the compliance variable was engineered as follows an ISP is categorized as Non-Compliant (0) and ISP is categorized as Compliant (1).

3.4.2. Exploration Data

Before developing predictive models, a comprehensive Exploratory Data Analysis (EDA) was conducted on the prepared dataset. This process is crucial for identifying patterns, validating assumptions, and forming the hypotheses that guide the modeling phase. The analysis included examining the class distribution, comparing descriptive statistics, analyzing feature correlations, and visualizing data distributions.

3.4.3. Model Development

Each of the four models was trained on the scaled training dataset and subsequently evaluated on the unseen, scaled testing dataset. The performance of each model is detailed below, using a confusion matrix to visualize the predictions and a set of key metrics to quantify performance. The metrics are defined as The evaluation of the various models were done on the following measures: Accuracy, Precision, Recall, F1-Score and Training Time. A brief description of these performance measures is given below :

The proportion of total predictions that were correct. Formula (1) :

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}. \quad (1)$$

From all instances predicted as positive (compliant), the proportion that was actually positive. Formula (2) :

$$\text{Precision} = \frac{TP}{TP + FP}. \quad (2)$$

Of all instances that were actually positive (compliant), the proportion that was correctly identified. Formula (3) :

$$\text{Recall} = \frac{TP}{TP + FN} \quad (3)$$

F1-Score the harmonic mean of Precision and Recall, providing a single score that balances both metrics, which is especially valuable for imbalanced datasets.

Formula (4) :

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4)$$

The next step involves developing and training the machine learning model, which will predict ISP Compliance based on part attributes and historical data. Analyse the collected data to understand key trends, correlations, and potential features that impact ISP Compliance classification.

Feature Engineering like Create features from both the raw data and graph-based features extracted from the KG. Model Selection and Training with multiple machine learning models (Logistic Regression , Decision Tree , Naive Bayes, Random Forest, Support Vector Machines) will be evaluated using performance metrics such as accuracy, precision, and recall. Fine-tuning and hyperparameter optimization will be performed to select the best-performing model. And then The model's performance will be tested on unseen data using evaluation metrics. The model that offers the best balance of accuracy and explainability will be selected for integration.

3.5. Evaluation

In this thesis was designed to ensure the validity and reliability of the results such us :

- Feature and Target Selection

The feature matrix, denoted as “X”, was constructed using the three primary operational metrics identified in the Feature Selection process TNG, QPS, and

TPS. The target vector “Y” was the binary compliance variable engineered in Section 3.4.1

- Data Splitting

This technique ensures that the proportion of compliant and non-compliant samples in both the training and testing sets is identical to their proportion in the overall dataset. This prevents the model from being trained or tested on a skewed representation of reality and is crucial for obtaining meaningful performance metrics.

- Feature Scaling

Before training the models, feature scaling was applied. The features have vastly different scales; QPS values can be in the thousands, while TNG values are in the single digits 8. Distance-based algorithms like SVM are sensitive to this, as features with larger scales can disproportionately dominate the model's learning process. To mitigate this, a StandardScaler was applied to the training data, which transforms each feature to have a mean of 0 and a standard deviation of 1. The same scaling transformation was then applied to the test data to ensure consistency.

- Feature Important

Method using SHAP (SHapley Additive exPlanations) is a more advanced and in-depth method. It provides detailed insight into how each feature influences individual predictions. This is extremely valuable when you need to understand why the model made a certain decision, not just which features were the most important. Permutation Importance, on the other hand, is a faster and more direct method. It is an excellent tool for getting a quick, reliable overview of which features contribute most to the model's overall performance.

In practice, a highly effective strategy is to use both methods together. You can start with Permutation Importance to quickly identify the most significant features globally, and then use SHAP to dive deeper and understand the specific interactions and contributions of those features to individual predictions

Feature and Target Selection the feature matrix, denoted as (x) , was constructed using the three primary operational metrics identified in the EDA (TNG, QPS, and TPS). The target vector (y) was the binary compliance variable. **Data Splitting:** The dataset was partitioned into a training set 80% of the data and a testing set 20% of the data.

3.6. Validation

The validation method ensures that the results predictive by the ISP Compliance classification system are reliable and generalizable with following method:

Expert Review and Focus Group Discussion (FGD)

Participant Selection: A Focus Group Discussion (FGD) will be conducted with internal and external operating experts. These participants are chosen based on their experience and expertise in ISP operations, particularly within the context of XYZ.

- Internal Experts the Director at XYZ and a experience in assessing XYZ's regulation and monitoring ISP.
- External Experts from ISP other organizations to provide an industry-standard perspective.

To validate and enrich the findings of this study, Focus Group Discussions (FGDs) were conducted with various stakeholders, involving at least three individuals representing diverse perspectives. The primary objective of these FGDs was to gather feedback on the identified gaps in Indonesia's online gambling regulatory framework and explore potential solutions. The discussions facilitated through the FGDs were designed to foster a collaborative environment where participants could share their insights and expertise on the challenges and opportunities associated ISP compliance with negative content data synchronization. The feedback gathered from these discussions was then analyzed to formulate specific and actionable recommendations for enhancing Indonesia's regulatory approach. The FGDs played a crucial role in determining the most impactful recommendations, focusing on those that could be implemented in the near future to mitigate the adverse effects of ISP compliance with negative content in Indonesia. By engaging with stakeholders directly, the study ensured that the recommendations were not only grounded in research but also aligned with the practical

realities and priorities of those involved in ISP compliance with negative content data synchronization.

Here is a summary of the key points for the FGD validation questions, there are :

1. Challenges in Negative Content Compliance, The summary poin is inquiring about the primary challenges faced by ISPs in adhering to Kominfo negative content filtering regulations from their perspective.
2. Risk-Based Monitoring Approach with Predictive Models, The summary poin is seeking opinions on the recommendation for regulators like Kominfo to use predictive models (as in this research) for risk-based monitoring, focusing audits on high-risk ISPs. Discussing potential benefits and drawbacks from an ISP's viewpoint.
3. Network Architecture Review for High-Risk ISPs, The summary poin is asking about the feasibility of a strategic review of network architecture for ISPs identified as "high-risk", especially for smaller or medium-sized ISPs.
4. Technology Investment and Other Critical Areas, The summary poin is questioning whether investing in foundational technologies like RTBH and enhancing DNS filtering systems (such as DNS Trust-NG or RPZ) are appropriate areas of focus for improving compliance. Discussing other critical areas that might have been overlooked.
5. Role of Data Analytics and AI in Regulatory Compliance, The summary poin is looking ahead, how the role of data analytics and AI is expected to evolve in regulatory compliance for the telecommunications industry, particularly for ISPs, in enhancing compliance with negative content data synchronization.
6. Final Research Feedback, The summary poin is requesting any final thoughts or feedback regarding the research approach and its findings.

The target participants for the FGD will be Directors Regulation and Personal Expert from Indonesian Internet Service Providers (ISPs). The selection criteria will prioritize individuals whose roles and responsibilities are directly related to:

- Network Operations as Individuals responsible for the day-to-day management of network infrastructure, including DNS and routing systems.
- Regulatory Compliance as Managers tasked with ensuring the company adheres to all government regulations, including Komdigi content filtering mandates.
- Strategic Technology Planning as Decision-makers involved in planning and investing in network architecture and compliance technologies.

This specific targeting ensures that the feedback is both technically informed and strategically relevant, providing a comprehensive validation of the research findings.

Evaluation Metrics

Metrics such as average accuracy, precision, recall, and F1 score across all folds will be used to assess the model's robustness and consistency.