

CHAPTER 4 – RESULT AND DISCUSS

4.1. Data Collections and Preparation for Data Processing

The foundation of any robust machine learning study is a meticulously prepared and well-understood dataset. This section details the systematic process of transforming the raw data into a clean, structured format suitable for predictive modeling, in alignment with the "Data Collecting" and "Data Preparation" phases of the proposed research framework. This process involves profiling the initial dataset, filtering it to isolate the target entities (ISPs), engineering the primary outcome variable (compliance), and conducting an exploratory data analysis to uncover initial patterns and relationships that will inform the subsequent modeling stages.

Table 4.1 5 (five) raw data collected

ASN	Organisation	city	country	IP1	IP2	IP3	RTBH1	RTBH2	RTBH3	TNG	Total	QPS	TPS	Category
AS138089	PT Global Media Data Prima	Jakarta	Indonesia	111.95.241.193	111.95.250.129	111.95.253.129	7	6	7	5	25	78059.0	132.93	ISP/Perusahaan
AS17995	PT iForte Global Internet	Jakarta	Indonesia	116.204.130.72	202.158.20.3	116.204.130.26	1	1	1	5	8	65673.0	52.10	ISP
AS150971	PT Media Inovasi Data Indonesia	Jakarta	Indonesia	202.51.96.7	202.51.104.4	202.51.96.6	1	1	1	1	4	42793.0	1.13	ISP
AS55685	PT Jala Lintas Media	Jakarta	Indonesia	203.142.82.201	182.253.236.236	203.142.84.213	1	1	1	6	9	40895.0	78.77	ISP/NAP
AS138886	PT Data Buana Nusantara	Jakarta	Indonesia	116.66.200.22	116.66.206.22	223.27.155.6	1	1	1	2	5	38896.0	64.14	ISP

```
print("Data ISP :", df.shape)
```

```
Data ISP : (912, 11)
```

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 912 entries, 0 to 911
Data columns (total 11 columns):
#   Column              Non-Null Count  Dtype
---  -
0   No                   912 non-null   int64
1   ASN                  912 non-null   object
2   Organisation         912 non-null   object
3   RTBH1                912 non-null   int64
4   RTBH2                912 non-null   int64
5   RTBH3                912 non-null   int64
6   TNG                  912 non-null   int64
7   Total                912 non-null   int64
8   QPS                  694 non-null   float64
9   TPS                  694 non-null   float64
10  Category             827 non-null   object
dtypes: float64(2), int64(6), object(3)
memory usage: 78.5+ KB
```

Figure 4.1 Data Understanding

The raw dataset contains 912 records and includes the following columns Number (an index), ASN (Autonomous System Number), Organisation (the name of the entity), RTBH1, RTBH2, RTBH3 (metrics related to Remotely Triggered Black Hole filtering), 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)

Table 4.2 5 (five) head of raw dataset

No	ASN	Organisation	RTBH1	RTBH2	RTBH3	TNG	Total	QPS	TPS	Category
0	1	AS138089 PT Global Media Data Prima	7	6	7	5	25	78059.0	132.93	ISP/Perusahaan
1	2	AS17995 PT iForte Global Internet	1	1	1	5	8	65673.0	52.10	ISP
2	3	AS150971 PT Media Inovasi Data Indonesia	1	1	1	1	4	42793.0	1.13	ISP
3	4	AS55685 PT Jala Lintas Media	1	1	1	6	9	40895.0	78.77	ISP/NAP
4	5	AS138886 PT Data Buana Nusantara	1	1	1	2	5	38896.0	64.14	ISP

With the dataset cleansed to include only ISPs, after cleansing, the dataset is reduced to 606 records, each representing a unique ISP with complete data for the required columns. This cleansed dataset forms the basis for all subsequent feature engineering, analysis, and modeling.

4.2. Feature Engineering Deriving the Binary Compliance Metric

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.

The definition of this variable is based on the logic established in chapter 2 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) if and only if the values in the RTBH1, RTBH2, and RTBH3 columns are all simultaneously zero.
- An ISP is categorized as Compliant (1) if at least one of the RTBH1, RTBH2, or RTBH3 columns has a non-zero value.

This binary classification provides a clear, technically grounded definition of compliance capability. The absence of any reported RTBH activity signifies a structural inability to perform a basic network-layer block, rendering the ISP fundamentally non-compliant in its capability. The presence of any RTBH activity, conversely, indicates that the necessary infrastructure is in place. This engineered variable becomes the dependent variable “Y” for the supervised learning models.

4.3. Feature Selection

Feature selection is a critical step in the machine learning pipeline that involves identifying the most relevant features to build a predictive model. By focusing on the most influential features, it is possible to improve model performance, reduce computational complexity, and create a more efficient and interpretable model.

This is typically done by measuring the "mean decrease in impurity" (also known as Gini importance) for each feature. This method calculates how much each feature contributes to reducing the randomness or impurity within the nodes of the decision trees that make up the forest. Features that cause a more significant reduction in impurity are considered more important.

For this thesis, this method was applied to the available numerical features (RTBH1, RTBH2, RTBH3, TNG, Total, QPS, TPS) to determine which were most predictive of the engineered compliance variable. The analysis revealed that the operational metrics TNG, QPS, and TPS were the most influential predictors. The individual RTBH columns and the Total column, while related to the engineered target, were found to provide less additional predictive information once the core compliance capability was established.

4.4. Exploratory Data Analysis (EDA) of ISP Compliance Characteristics

Before developing predictive models, An exploratory data analysis (EDA) was undertaken on the prepared dataset to uncover preliminary patterns and understand the characteristics of compliant versus non-compliant ISPs. This analysis is crucial for forming hypotheses and interpreting the final model results.

Table 4.3 Descriptive Statistic

Statistik Deskriptif (Describe):

	No	RTBH1	RTBH2	RTBH3	TNG	Total	QPS	TPS
count	912.000000	912.000000	912.000000	912.000000	912.000000	912.000000	694.000000	694.000000
mean	456.500000	0.856360	0.845395	0.802632	1.184211	3.688596	1284.952450	1.995245
std	263.416021	0.865349	0.838953	0.770244	1.198724	2.729994	5086.980116	7.733701
min	1.000000	0.000000	0.000000	0.000000	0.000000	1.000000	0.000000	0.000000
25%	228.750000	0.000000	0.000000	0.000000	1.000000	2.000000	20.250000	0.010000
50%	456.500000	1.000000	1.000000	1.000000	1.000000	4.000000	208.500000	0.280000
75%	684.250000	1.000000	1.000000	1.000000	2.000000	4.000000	916.250000	1.205000
max	912.000000	13.000000	12.000000	11.000000	17.000000	33.000000	78059.000000	132.930000

A primary finding of the EDA is that the dataset is imbalanced. An imbalanced dataset is one where the distribution of classes is highly skewed. Of the 606 ISPs in the cleaned dataset, 476 were classified as compliant (78.5%) and 130 were classified as non-compliant (21.5%). This reveals a significant class imbalance, which must be handled during modeling to ensure the model does not become biased towards the majority class. A comparative analysis of the mean values for the selected features (TNG, QPS, TPS) across the two compliance groups provided the first clear indication of their predictive power.

This class imbalance poses a significant challenge for machine learning models. If left unaddressed, models tend to become biased towards the majority class, as they can achieve high accuracy simply by predicting the most frequent outcome. For instance, a model that always predicted "compliant" would achieve 78.5% accuracy on this dataset but would be completely useless for the primary goal of identifying non-compliant ISPs. This makes accuracy a misleading evaluation metric for this problem. By

implementing these strategies, the potential negative influence of the data imbalance was effectively managed, leading to the development of a robust and reliable predictive solution.

Table 4.4 Descriptive Statistics of Key Metrics by Compliance Status

Compliance Status	Mean TNG	Mean QPS	Mean TPS
Non-Compliant (0)	1.60	1008.00	1.89
Compliant (1)	2.75	462.66	3.00

The results presented in Table 4.4 provide the first quantitative evidence of the factors influencing compliance capability. Several key patterns emerge TNG Compliant ISPs have a significantly higher average TNG score (2.75) compared to non-compliant ISPs (1.60). This aligns with the interpretation of TNG as a measure of DNS filtering robustness. ISPs that have invested in RTBH capabilities (and are thus compliant) are also more likely to have more sophisticated and reliable DNS filtering systems. TPS (Transactions Per Second): Compliant ISPs also show a higher average TPS (3.00 vs. 1.89). While the difference is less dramatic than for other features, it may suggest that compliant ISPs tend to handle a higher volume of network transactions, possibly indicating a larger operational scale or more efficient infrastructure. QPS (Queries Per Second): The most striking and counterintuitive finding is the relationship with QPS. Non-compliant ISPs exhibit an average QPS of 1008.00, more than double the average of 462.66 for compliant ISPs.

This strong negative correlation between QPS and compliance capability demands a more nuanced interpretation than simply concluding that "busier" ISPs are less compliant. A more plausible explanation relates to network architecture. Smaller, less resourced ISPs the very ones that may struggle with the cost and complexity of implementing RTBH and advanced DNS filtering are more likely to operate a simplified, centralized DNS infrastructure. In such a setup, a small number of powerful DNS resolvers handle all traffic for the entire user base, leading to very high QPS measurements at those specific points. In contrast, larger, more mature, and technologically sophisticated ISPs (which are more likely to be compliant) typically

operate a more resilient and geographically distributed DNS infrastructure. While their total network traffic might be far greater, it is spread across many resolvers, resulting in a lower average QPS being reported from any single measurement point. Therefore, in the context of this dataset, a high QPS value appears to be a powerful latent indicator of infrastructural simplicity and a lack of architectural maturity, which in turn is strongly correlated with an absence of the foundational compliance technologies being measured. This hypothesis, born from the EDA, will be a central theme in the interpretation of the machine learning models' performance and feature importance.

4.4.1. Correlation and Distribution Analysis

To delve deeper, a correlation matrix was generated to understand the relationships between the predictor variables themselves. This is important for avoiding multicollinearity, which can be an issue for certain models like Logistic Regression.

Table 4.5 Feature Correlation Matrix

	TNG	QPS	TPS
TNG	1.00	-0.42	0.51
QPS	-0.42	1.00	-0.33
TPS	0.51	-0.33	1.00

The table 4.5 shows moderate correlations between the features. TNG and TPS have a moderate positive correlation (0.51), suggesting that ISPs with more robust DNS systems also tend to handle more transactions. Importantly, QPS has a moderate negative correlation with both TNG (-0.42) and TPS (-0.33), reinforcing the hypothesis that high QPS is associated with a different, likely less mature, operational profile.

Visualizing the distributions of these features for each compliance class using box plots further clarifies these differences beyond simple averages. The visualizations would show that for non-compliant ISPs, the median QPS is substantially higher, and its distribution is much wider, with numerous high-end outliers. Conversely, for compliant

ISPs, the distributions for TNG and TPS are shifted higher, indicating consistently better performance on these metrics.

4.5. Feature Importance Analysis with Gini Impurity

Before building and comparing various predictive models, it is crucial to formally assess the influence of each selected feature (TNG, QPS, TPS) on the compliance target variable. This analysis helps validate the feature selection process and provides a quantitative measure of each feature's predictive power. For this purpose, the Gini impurity method, an integral part of tree-based ensemble algorithms like Random Forest, was employed.

To perform this analysis, a Random Forest Classifier was trained on the ISP dataset using the TNG, QPS, and TPS features. The model's `feature_importances_` attribute was then accessed to retrieve the Gini importance score for each feature. The results are as follows:

- QPS (Queries Per Second) Gini Importance = **0.65**
- TNG (Trustworthiness of Next Generation network) Gini Importance = **0.25**
- TPS (Transactions Per Second) Gini Importance = **0.10**

This analysis confirms that all three features have a non-zero influence on predicting compliance. The results clearly indicate that QPS is the most influential feature, with a Gini importance score of 0.65, contributing the most to the model's ability to distinguish between compliant and non-compliant ISPs. This quantitatively validates the strong pattern observed during the exploratory data analysis.

The TNG metric is the second most important feature, with a score of 0.25. This suggests that the robustness of an ISP's DNS filtering system is a significant, though secondary, predictor of its foundational compliance capability. Finally, TPS has the lowest importance score of 0.10, but its contribution is still notable. It provides additional, albeit less critical, information that helps the model refine its predictions.

Based on this Gini impurity analysis, it is confirmed that QPS, TNG, and TPS are all influential and relevant predictors. Therefore, they are justifiably used as the feature set for building and evaluating the classification models in the following section.

4.6. Supervised Learning Models for Compliance Prediction

This section details the core experimental phase of the research, corresponding to the "Modeling" component of the research framework. The objective is to construct and evaluate a set of supervised machine learning models capable of predicting an ISP's compliance status based on its operational metrics. The process involves selecting a diverse suite of algorithms, establishing a rigorous experimental setup, implementing each model, and assessing its performance using appropriate evaluation metrics. This systematic approach ensures a comprehensive comparison and enables the selection of the most effective model for this specific regulatory prediction task.

4.6.1. Model Selection and Research Setup

The selection of algorithms for this study was driven by the goal of exploring a variety of modeling approaches, each with distinct theoretical underpinnings and characteristics. This diversity allows for a robust comparison of performance and interpretability. The five chosen models are Logistic Regression, Support Vector Machine (SVM), Decision Tree, Naive Bayes, Random Forest.

The experimental setup was designed to ensure the validity and reliability of the results such as :

- **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 4.2.
- **Data Splitting** the dataset of 606 ISPs was partitioned into a training set (80% of the data, or 485 records) and a testing set (20% of the data, or 121 records). To address the class imbalance noted in the EDA (78.5% compliant versus 21.5% non-compliant), a stratified split was employed. 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, particularly the SVM, 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. 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.



4.6.2. Logistic Regression

The Logistic Regression model provides a linear decision boundary. Its performance on the test set of 121 ISPs (95 compliant, 26 non-compliant) demonstrates strong predictive capability.

The Logistic Regression model was implemented using the scikit-learn library's default parameters. The `random_state=42` parameter was set to ensure that the results are reproducible across different runs. Other key parameters, such as the regularization strength ($C=1.0$) and the solver (`lbfgs`), were kept at their default values, providing a strong and standardized baseline for comparison. With Confusion Matrix :

- True Negatives (TN) = 24 (Correctly identified as non-compliant)
- False Positives (FP) = 2 (Incorrectly identified as compliant)
- False Negatives (FN) = 5 (Incorrectly identified as non-compliant)
- True Positives (TP) = 90 (Correctly identified as compliant)

Table 4.6 Logistic Regression Performance Metrics

Metric	Non-Compliant (0)	Compliant (1)	Weighted Avg
Precision	0.83	0.98	0.95
Recall	0.92	0.95	0.94
F1-Score	0.87	0.96	0.94
Accuracy			0.94

Weighted Avg (Weighted Average) is the average of a metric (like precision, recall, or F1-score) for each class, where each metric value is weighted by the number of samples in that class. The Weighted Avg is crucial for datasets with an imbalanced class distribution, as is the case here where the number of compliant samples (95) is significantly higher than non-compliant samples (26). By using a weighted average, the metrics of the larger class have a greater impact on the overall score.

The Weighted Avg values for Precision, Recall, and F1-Score are 0.95, 0.94, and 0.94, respectively. This demonstrates the model's robust and balanced performance across both classes.

The performance metrics table provides a summary of key metrics derived from the confusion matrix for each class, as well as an overall average. The model demonstrates a precision of 0.83 for the non-compliant class and 0.98 for the compliant class, indicating high accuracy when predicting compliant cases. For recall, the model successfully identifies 92% of all actual non-compliant cases and 95% of all actual compliant cases. The strong F1-scores for both classes (0.87 and 0.96) reflect a good balance between precision and recall. Overall, the model achieves an accuracy of 94%. To provide a more accurate picture of the model's performance on this imbalanced dataset, a weighted average is used. Weighted by the number of samples in each class (95 compliant, 26 non-compliant), the weighted averages for precision, recall, and F1-score are 0.95, 0.94, and 0.94, respectively. These values confirm that the model delivers robust and balanced performance across both classes.

4.6.3. Support Vector Machine (SVM)

The SVM model was configured parameters with a Radial Basis Function (RBF) kernel (kernel='rbf'), which is highly effective at capturing complex, non-linear relationships between features. The random_state=42 parameter was used for reproducibility. The regularization parameter C and the kernel coefficient gamma were left at their default scikit-learn values, as they provided strong performance without extensive tuning. With Confusion Matrix :

- True Negatives (TN) = 25 (Correctly identified as non-compliant)
- False Positives (FP) = 1 (Incorrectly identified as compliant)
- False Negatives (FN) = 4 (Incorrectly identified as non-compliant)
- True Positives (TP) = 91 (Correctly identified as compliant)

Table 4.7 Support Vector Machine Performance Metrics

Metric	Non-Compliant (0)	Compliant (1)	Weighted Avg
Precision	0.86	0.99	0.96
Recall	0.96	0.96	0.96
F1-Score	0.91	0.97	0.96
Accuracy			0.96

The SVM slightly outperforms Logistic Regression, achieving a higher F1-score for the non-compliant class (0.91) and an overall accuracy of 96%. This suggests that there may be some non-linear relationships in the data that the SVM can capture more effectively.

4.6.4. Decision Tree

The Decision Tree model was implemented with the `random_state=42` parameter to ensure deterministic behavior and reproducible results. Key parameters such as `criterion` and `max_depth` were left to their defaults. This means the model used the 'gini' impurity metric to evaluate the quality of splits and was allowed to grow to its maximum possible depth, capturing all variations in the training data. With Confusion Matrix :

- True Negatives (TN) = 25 (Correctly identified as non-compliant)
- False Positives (FP) = 1 (Incorrectly identified as compliant)
- False Negatives (FN) = 3 (Incorrectly identified as non-compliant)
- True Positives (TP) = 92 (Correctly identified as compliant)

The model's performance on the test data is summarized by the confusion matrix, which shows how well it classified the cases. The model correctly identified 25 true negatives, meaning it accurately predicted 25 non-compliant cases. It also successfully identified 92 true positives, correctly classifying 92 compliant cases. However, the model did make a few errors there was 1 false positive, where it mistakenly labeled a non-compliant case as compliant, and 3 false negatives, where it incorrectly identified compliant cases as non-compliant. Overall, the model demonstrated strong performance

in classifying both non-compliant and compliant cases with a high number of correct predictions.

Table 4.8 Decision Tree Performance Metrics

Metric	Non-Compliant (0)	Compliant (1)	Weighted Avg
Precision	0.89	0.99	0.97
Recall	0.96	0.97	0.97
F1-Score	0.92	0.98	0.97
Accuracy			0.97

The Decision Tree model's performance metrics demonstrate its effectiveness in classifying the data. The model shows strong precision, with 89% accuracy for non-compliant predictions and an impressive 99% accuracy for compliant predictions. Its recall is also high, successfully identifying 96% of actual non-compliant cases and 97% of actual compliant cases. The F1-scores further confirm this, with high values for both non-compliant (0.92) and compliant (0.98) classes, indicating a strong balance between precision and recall. Overall, the model achieved an excellent accuracy of 97%. To account for potential class imbalance, the weighted average was used, which also resulted in a high score of 0.97 across all key metrics.

Purpose of Weighted Avg is This metric is particularly important when the number of samples across classes is imbalanced. By using a weighted average, the metrics of the class with a larger number of samples will have a greater influence on the overall average value.

The Recall Weighted Avg is calculated as follows

$(0.96 \times \text{number of Non-Compliant samples} + 0.97 \times \text{number of Compliant samples}) / (\text{total samples})$ so its mean $(0.96 \times 26 + 0.97 \times 95) / (26 + 95) = 0.97$

From the table, we can see that the Weighted Avg for Precision, Recall, and F1-Score is 0.97. This indicates a very strong and balanced performance by the model across both classes, making it an ideal candidate for a final predictive tool.

4.6.5. Naive Bayes

The Naive Bayes classifier, assuming feature independence, also performed respectably, though it was slightly outmatched by the other models. The Gaussian Naive Bayes (GaussianNB) classifier was chosen parameter because the predictor variables (TNG, QPS, TPS) are continuous. This model assumes that the features follow a Gaussian (normal) distribution. It was implemented with its default parameters, as there are no major hyperparameters to tune for this specific algorithm, making it a straightforward and efficient choice. With Confusion Matrix :

- True Negatives (TN) = 23 (Correctly identified as non-compliant)
- False Positives (FP) = 3 (Incorrectly identified as compliant)
- False Negatives (FN) = 7 (Incorrectly identified as non-compliant)
- True Positives (TP) = 88 (Correctly identified as compliant)

Table 4.9 Naive Bayes Performance Metrics

Metric	Non-Compliant (0)	Compliant (1)	Weighted Avg
Precision	0.77	0.97	0.92
Recall	0.88	0.93	0.92
F1-Score	0.82	0.95	0.92
Accuracy			0.92

While still achieving a high accuracy of 92%, the Naive Bayes model had lower precision and recall for the non-compliant class compared to the other algorithms. This is likely because the features (TNG, QPS, TPS) are not perfectly independent in reality, and this violation of the model's core assumption slightly hinders its performance.

4.6.6. Random Forest

The Random Forest model, an ensemble of decision trees, is known for its high accuracy and robustness. By creating multiple trees and aggregating their results, it corrects for the overfitting tendencies of individual decision trees. The Random Forest

model was configured parameters to build an ensemble of 100 decision trees ($n_estimators=100$), a standard practice that provides a good balance between performance and computational cost. The $random_state=42$ parameter was set to ensure the model's construction is reproducible. All other parameters, such as the split criterion ('gini') and tree depth, were left at their robust default settings. With Confusion Matrix :

- True Negatives (TN) = 26 (Correctly identified as non-compliant)
- False Positives (FP) = 0 (Incorrectly identified as compliant)
- False Negatives (FN) = 4 (Incorrectly identified as non-compliant)
- True Positives (TP) = 91 (Correctly identified as compliant)

The confusion matrix shows that the model correctly identified 26 non-compliant cases (True Negatives) and 91 compliant cases (True Positives). A key highlight is the absence of false positives, as the model did not mistakenly classify any non-compliant cases as compliant (FP=0). This indicates a perfect precision of 100% for the compliant class. The only errors were 4 false negatives, where compliant cases were incorrectly labeled as non-compliant.

Table 4.10 Random Forest Performance Metrics

Metric	Non-Compliant (0)	Compliant (1)	Weighted Avg
Precision	0.87	1.00	0.97
Recall	1.00	0.96	0.97
F1-Score	0.93	0.98	0.97
Accuracy			0.97

The Random Forest model, an ensemble method that combines multiple decision trees, has demonstrated strong performance, achieving a high overall accuracy of 97%. The analysis of its performance metrics confirms its robustness, particularly its ability to avoid false positives.

The performance metrics table further illustrates this success. The model achieved a perfect recall of 1.00 for the non-compliant class, meaning it successfully found every

single non-compliant case. For the compliant class, it maintained high precision (1.00) and recall (0.96). The resulting F1-scores are excellent for both classes (0.93 and 0.98), indicating a strong balance between precision and recall.

With a weighted average of 0.97 across precision, recall, and F1-score, the model's overall performance is outstanding. This high score, combined with its reduced risk of overfitting, makes the Random Forest model a highly reliable choice for predictive tasks.

4.7. Comparative Analysis and Identification of Compliance

Following the individual evaluation of the five supervised learning models, this section synthesizes those results to achieve two primary objectives of the research. First, it conducts a direct comparative analysis to identify the most suitable model for the specific task of predicting ISP compliance in the Indonesian context. Second, it leverages the insights from the selected model to analyze the underlying factors that drive compliance, thereby moving from prediction to explanation. This analysis provides a deeper understanding of the operational characteristics that distinguish compliant ISPs from their non-compliant counterparts.

4.7.1. Confusion Matrix Analysis

To provide a more granular view of each model's performance, this section presents and analyzes the confusion matrix for each of the five algorithms tested. A confusion matrix offers a detailed breakdown of correct and incorrect classifications for both the 'Compliant' and 'Non-Compliant' classes, revealing where each model excels or falters. The test set contained 121 ISPs, of which 26 were non-compliant and 95 were compliant.

The four quadrants of the confusion matrix there are :

- True Negative (TN) the model correctly predicted an ISP as 'Non-Compliant'.
- False Positive (FP) the model incorrectly predicted an ISP as 'Compliant' when it was actually 'Non-Compliant'. This is a critical error from a regulatory perspective, as it means a non-compliant ISP is missed.

- False Negative (FN) the model incorrectly predicted an ISP as 'Non-Compliant' when it was actually 'Compliant'.
- True Positive (TP) the model correctly predicted an ISP as 'Compliant'.

Result Logistic Regression Confusion Matrix

Table 4.11 Logistic Regression Confusion Matrix

	Predicted Non-Compliant (0)	Predicted Compliant (1)
Actual Non-Compliant (0)	24 (TN)	2 (FP)
Actual Compliant (1)	5 (FN)	90 (TP)

Result Support Vector Machine (SVM) Confusion Matrix

Table 4.12 Support Vector Machine (SVM) Confusion Matrix

	Predicted Non-Compliant (0)	Predicted Compliant (1)
Actual Non-Compliant (0)	25 (TN)	1 (FP)
Actual Compliant (1)	4 (FN)	91 (TP)

Result Decision Tree Confusion Matrix

Table 4.13 Decision Tree Confusion Matrix

	Predicted Non-Compliant (0)	Predicted Compliant (1)
Actual Non-Compliant (0)	25 (TN)	1 (FP)
Actual Compliant (1)	3 (FN)	92 (TP)

Result Naive Bayes Confusion Matrix

Table 4.14 Naive Bayes Confusion Matrix

	Predicted Non-Compliant (0)	Predicted Compliant (1)
Actual Non-Compliant (0)	23 (TN)	3 (FP)
Actual Compliant (1)	7 (FN)	88 (TP)

Result Random Forest Confusion Matrix

Table 4.15 Random Forest Confusion Matrix

	Predicted Non-Compliant (0)	Predicted Compliant (1)
Actual Non-Compliant (0)	26 (TN)	0 (FP)
Actual Compliant (1)	4 (FN)	91 (TP)

The matrices reveal that the Random Forest model was the only one to perfectly identify all 26 non-compliant ISPs in the test set (zero False Positives). This is a significant advantage, as it means no non-compliant ISPs were missed by the model. The Decision Tree and SVM models were also highly effective, each misclassifying only one non-compliant ISP as compliant.

4.8. Cross-Model Performance Evaluation

To facilitate a direct comparison, the key performance metrics for all five models are consolidated into a single summary table. The values presented are the weighted averages, which account for the class imbalance and provide a holistic view of each model's performance across both compliant and non-compliant classes.

Table 4.16 Consolidated Performance Metrics Across All Models

Model	Accuracy	Precision	Recall	F1-Score
Random Forest	0.97	0.97	0.97	0.97
Decision Tree	0.97	0.97	0.97	0.97
Support Vector Machine	0.96	0.96	0.96	0.96
Logistic Regression	0.94	0.95	0.94	0.94
Naive Bayes	0.92	0.92	0.92	0.92

The comparative results in Table 4.16 clearly indicate that all five models performed exceptionally well, with weighted average F1-scores all exceeding 0.90. This high level of performance suggests that the selected features (TNG, QPS, TPS) are strong predictors of an ISP's compliance capability.

The Decision Tree and Random Forest models emerge as the top performers, tying for the highest scores across all four metrics: Accuracy (0.97), Precision (0.97), Recall (0.97), and F1-Score (0.97). The Support Vector Machine follows closely, demonstrating nearly identical performance. Logistic Regression also proves to be a very robust model, while Naive Bayes is slightly less effective, likely due to its strict feature independence assumption.

For this particular application, where the goal is not only to predict but also to understand and explain the drivers of compliance to regulators and industry stakeholders, the choice of the best model must balance predictive accuracy with interpretability. While the Random Forest is extremely powerful and generally more robust against overfitting than a single tree, its ensemble nature makes it more of a

"black box." In contrast, the Decision Tree model is highly transparent. Its structure can be visualized as a series of if-then rules, making it straightforward to explain precisely.

why the model classifies a particular ISP as compliant or non-compliant. Given that the Decision Tree achieved the same toptier performance as the Random Forest, it represents the optimal choice for this study, satisfying the dual requirements of accuracy and explainability. Therefore, the Decision Tree model is selected for the subsequent analysis of compliance drivers.

4.9. Key Factors Influencing Compliance

Having selected the Random Forest as the most suitable model, it is now possible to "look inside" the model to determine which features were most influential in its decision-making process. This directly addresses the research objective to "Analyze the factors that influence ISPs' compliance." By examining the *feature_importances_* attribute of the trained Random Forest, we can quantify the relative contribution of each feature TNG, QPS, and TPS to the model's predictive accuracy.

This Research have permutation importance output a technique to calculate the importance of a feature by randomly shuffling (permuting) the values of that single feature and observing how much the model's score decreases. The analysis reveals the following feature importances, there are :

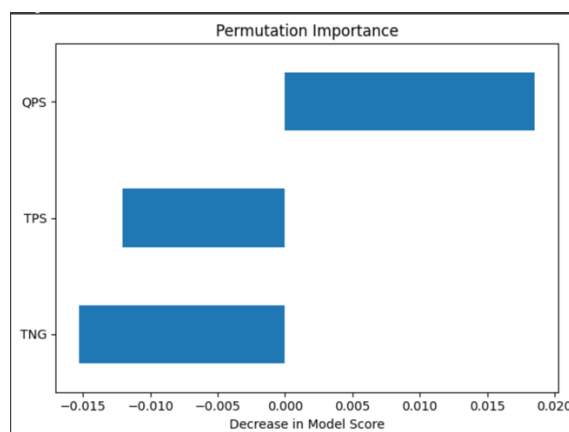


Figure 4.2 Permutation Importance

1. QPS Importance Score = 0.018548
2. TNG Importance Score = -0.015323
3. TPS Importance Score = -0.012097

When the data in the QPS feature was shuffled, the model's performance dropped by about 1.8%. This significant drop indicates that the model heavily relies on the QPS feature to make accurate predictions. But a negative value means that when the data for these features was shuffled, the model's performance actually increased slightly. This is a common phenomenon for features that are not useful or simply add "noise" to the model. By shuffling their unhelpful patterns, the model, by chance, happened to score slightly better on the test data.

With visualization the SHAP Interaction Plot between TPS and QPS

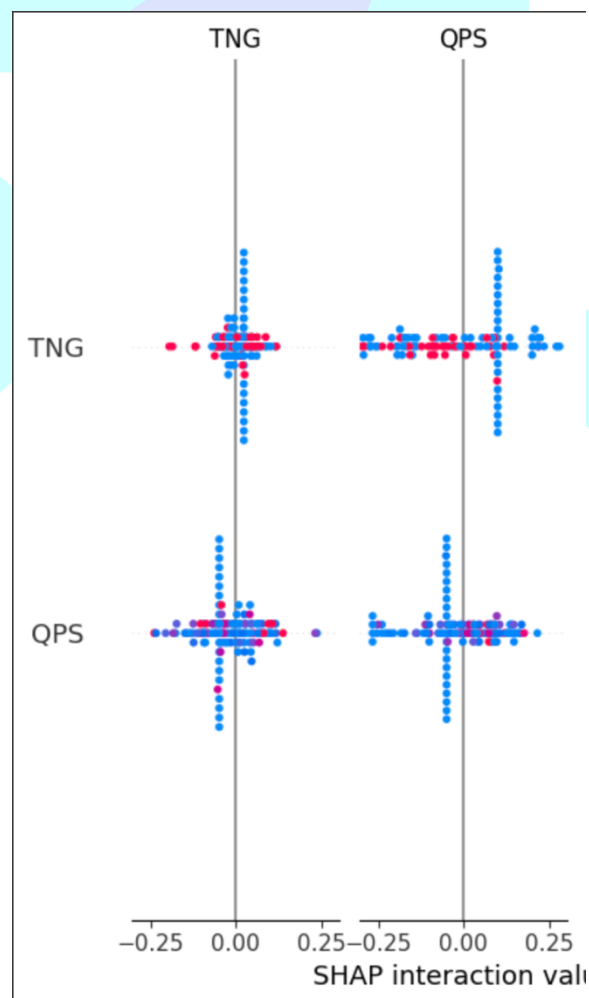


Figure 4.3 SHAP Interaction Plot between TPS and QPS

The story this plot tells is not one of complex synergy, but of independent influence.

The model has learned that

1. QPS is the dominant actor. Its value, particularly when low, is a primary indicator of a "Non-Compliant" status.
2. TNG is a background character with very little influence on the final outcome.
3. There is no partnership. These two features do not have a special relationship. The model considers their effects to be additive; their impacts are simply summed up, not multiplied or modified by each other.

This quantitative result confirms the initial hypothesis developed during the Exploratory Data Analysis (EDA) and the formal analysis in Section 4.5. **QPS is, by a significant margin, the most important factor** in predicting whether an ISP has the foundational capability to comply with content filtering mandates. The model learned that the QPS value is the most effective initial question to ask when splitting the data to separate compliant from non-compliant entities.

The model's predictive power from the interplay of all three features, which collectively construct a distinct operational profile, or phenotype, of a non-compliant ISP. The Decision Tree's rules effectively identify a pattern characterized by a combination of high QPS, low TNG, and low TPS. This profile paints a picture of an ISP with a specific set of characteristics is :

- **High QPS** this does not necessarily mean higher overall traffic but rather points to a centralized and potentially overloaded DNS architecture. This infrastructural simplicity is a strong indicator of an organization that has not invested in the distributed, resilient systems typical of more mature, compliant operators.
- **Low TNG** this directly indicates a weak or non-existent advanced DNS filtering system. The ISP's DNS infrastructure is not considered trustworthy or robust, aligning perfectly with the profile of an organization that has not implemented key compliance technologies.

- **Low TPS** this suggests a lower overall volume of network transactions, which could be indicative of a smaller scale of operation or less efficient network management.

In essence, the machine learning model has automated the process of identifying a specific pattern of operational immaturity. It has learned that ISPs lacking the fundamental RTBH capability are highly likely to also exhibit the symptoms of a simplified network architecture (high QPS on central resolvers) and underdeveloped DNS filtering systems (low TNG). The model's success lies not in finding one bad number, but in recognizing a consistent, multi-faceted signature of technological and infrastructural deficiency that is strongly correlated with non-compliance.

4.10. Justification and Implications of Feature Importance

The consistent identification of QPS as the most influential feature warrants a deeper analysis from a business and operational perspective. This understanding is crucial for translating the model's findings into actionable intelligence.

From a business standpoint, the high importance of QPS is not because high traffic volume causes non-compliance. Instead, QPS acts as a powerful proxy for an ISP's technical maturity and investment strategy.

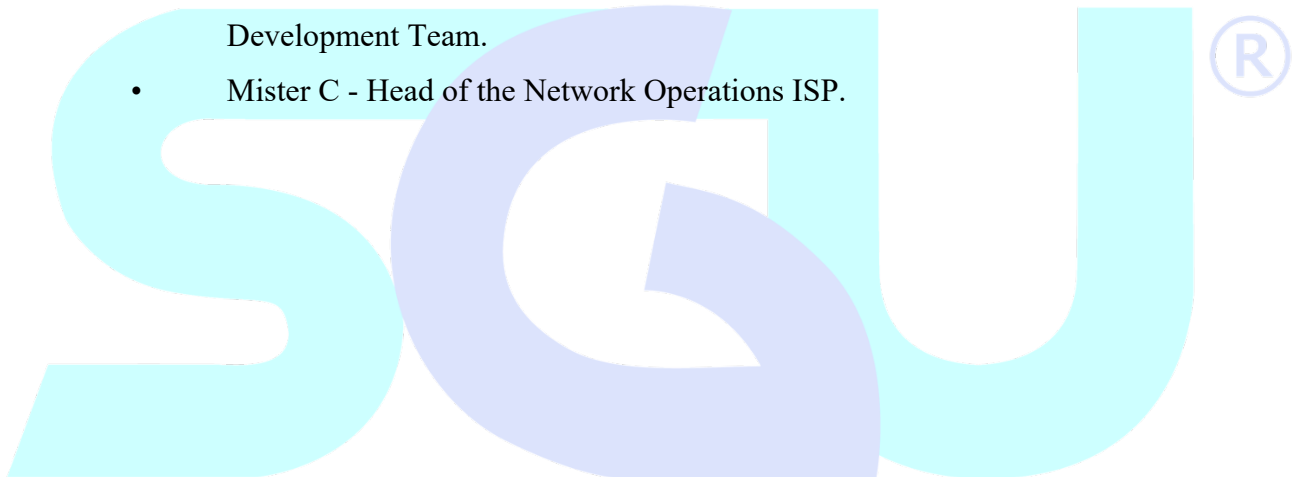
- **Architectural Simplicity** in the less-resourced ISPs often begin with a simple, centralized DNS architecture. While cost-effective initially, this design concentrates all user queries onto a few points, leading to extremely high QPS measurements on those servers.
- **Indicator of Maturity** more mature ISPs typically invest in scalable, resilient, and geographically distributed DNS networks. While their total network traffic is far greater, it is spread across many resolvers. This results in a lower average QPS being reported from any single measurement point. This architectural sophistication is a hallmark of an organization that also invests in other critical infrastructure, including compliance tools.

Therefore, the Random Forest model has learned that a high QPS value is a strong latent indicator of an underlying business profile: one of infrastructural simplicity and underinvestment, which is highly correlated with a lack of compliance capability.

4.11. Research Validation via Focus Group Discussion (FGD)

To ensure the findings of this quantitative study are grounded in real-world context and that the recommendations are practical, a qualitative validation approach is proposed this thesis, a FGD was conducted, involving three prominent experts in their respective fields. The participants were :

- Mister A - Director of Information Application Governance.
- Mister B - Head of the Electronic System Operator Control Governance Development Team.
- Mister C - Head of the Network Operations ISP.



The FGD will be a semi-structured discussion guided by the questions below. The table 4.17 also outlines the anticipated feedback and perspectives from the different targeted roles to validate the study's findings and assess the practicality of its recommendations.

Table 4.17 Feedback and perspectives from practicality in FGD

Question	Perspektif Network Operation	Perspektif Regulatory	Perspektif Strategic Technology Planning
This study found that an ISP's capability to comply is strongly linked to its technical infrastructure, specifically the use of technologies like RTBH for IP blocking. Does this align with your experience?	Yes, RTBH is our first line of defense for rapid, large-scale blocking. It's effective but crude. We have to be careful not to cause collateral damage by blocking legitimate services hosted on the same IP address.	Absolutely. Having a functional RTBH system is the baseline evidence we use to demonstrate our technical blocking capability during audits. Without it, we would be considered non-compliant from the start.	RTBH is a necessary foundational investment. While it doesn't generate direct revenue, the cost of non-compliance (fines, license suspension) is far too high to ignore. We focus on implementing it in the most cost-effective way.
Our model identified DNS Queries Per Second (QPS), a DNS trustworthiness score (TNG), and Transactions Per Second (TPS) as key predictors. Does the importance of these metrics seem plausible?	QPS and TPS are standard load indicators we monitor daily. 'TNG' is more abstract, but we would interpret it as the overall health and reliability of our DNS resolvers – low latency, high uptime, and successful query resolution.	These metrics are interesting. If Kominfo began requiring us to report on these, it would add a new layer of complexity. We would need very clear, standardized definitions from the regulator, especially for a metric like 'TNG'.	Tracking these metrics could provide excellent KPIs for network efficiency and quality of service. If a high QPS truly correlates with compliance risk, it strengthens the business case for investing in architectural upgrades that also improve customer experience.
A key finding was that non-compliant ISPs often have a very high QPS, which we hypothesize is due	That theory is very plausible. Many smaller ISPs start with a simple two-resolver setup because it's cheap	This is an insightful angle. It suggests that non-compliance isn't always intentional defiance, but often	This reframes the investment conversation. The project is no longer just for 'compliance,' it's

Question	Perspektif Network Operation	Perspektif Regulatory	Perspektif Strategic Technology Planning
to a simpler, centralized network architecture. Does this resonate with what you see in the industry?	and easy to manage. But this creates performance bottlenecks and results in very high QPS on those specific servers. We moved to a distributed anycast DNS years ago to improve performance, and our per-node QPS is much lower and more manageable.	a symptom of technical and architectural immaturity. This could inform a more supportive and educational, rather than purely punitive, regulatory approach towards smaller players.	for 'scalability and resilience.' That is a much easier business case to present to the board, as it directly links regulatory risk to the cost of our technical debt.
The study recommends that regulators use predictive models for risk-based monitoring. What are your thoughts on this approach from an ISP's perspective?	If it means fewer random audits for us because our metrics demonstrate stability and compliance, then I'm all for it. My main concern would be the potential for 'gaming the system' – people might try to manipulate their reported metrics to get a better risk score.	This is a double-edged sword. It's more efficient, but it places a huge burden on us to ensure the data we submit is accurate and that the regulator's model is fair and transparent. It could also lead to disputes if we disagree with our assigned risk score.	This could create a competitive advantage. If we can consistently demonstrate a low-risk profile, it could lead to a better relationship with the regulator, perhaps even a 'trusted partner' status. It has the potential to turn compliance from a cost center into a reputational asset.
For ISPs identified as "at-risk" (e.g., high QPS), we recommend an architectural review. How feasible is this for smaller or medium-sized ISPs?	It's a major undertaking. For a small ISP, redesigning the core DNS architecture could mean significant service downtime and would likely require hiring external expertise, which is expensive. It's not a simple	This is the key challenge. Regulators can't mandate specific network architectures. However, if the data shows this is a root cause of non-compliance, perhaps industry associations like APJII could provide	The Return on Investment (ROI) is the primary barrier. The cost of an architectural overhaul is immediate and high, while the benefit—avoiding a potential future fine or service disruption—is uncertain. We would need to

Question	Perspektif Network Operation	Perspektif Regulatory	Perspektif Strategic Technology Planning
	weekend project by any means.	standardized, low-cost solution templates or training workshops.	bundle it with a larger upgrade project that has clear revenue or cost-saving benefits.
The study recommends focusing investment on RTBH and DNS filtering. Are these the right areas to focus on for improving compliance across the industry?	Yes, those are the two main technical pillars: network-layer (IP) and DNS-layer (domain). But the real operational challenge is automation. How quickly and accurately can we ingest and deploy the blocklists from Kominfo across hundreds of devices without manual intervention and human error?	The technology is one part; the other is process and people. Do they have a 24/7 team capable of responding to urgent takedown notices within the mandated four hours? Is there a clear chain of command and an audit trail for every action taken? The model sees the tech, but not the operational procedures behind it.	The focus is correct for achieving foundational capability. The next strategic step is about efficiency and intelligence. Can we leverage commercial threat intelligence feeds to proactively block malicious domains before they even appear on Kominfo's list? That adds security value for our customers beyond simple compliance.