JUINKTI (Jurnal Inovasi dan Karya Teknologi Informasi)
Volume 1 Number1 | Mei 2026: 21-30
DOI: xxxxxxx/JUINKTI.xxxxx.xx

Classification Model of Exporter Maturity Levels Based on the Comparison of National FOB Values and MSME Production Capacity

Oki Kurlita[1] Pipin Julianti[2] Reva Bella[3]
Idi Hermansyah[4]
[1234,] Studi Sains Data, Universitas Lembah Dempo

Article Info

Article history:

Received Mei 15, 2026
Revised Mei 20, 2026
Accepted Mei 25, 2026
Published Mei 30, 2026

Keywords:

Free on Board (FOB)
MSMEs, CRISP-DM
Data Mining
Classification Model
Export Readiness
Production Capacity.

ABSTRACT

This study aims to develop a classification model for exporter maturity levels by considering the comparison between national Free on Board (FOB) values and the production capacity of Micro, Small, and Medium Enterprises (MSMEs). The research is motivated by the varying levels of readiness among businesses to enter export markets, highlighting the need for a data-driven approach to objectively classify exporter maturity. The study employs the CRISP-DM (Cross Industry Standard Process for Data Mining) methodology, including business understanding, data understanding, data preparation, modeling, evaluation, and deployment stages. The dataset consists of FOB export values, product categories, and MSME production capacity indicators. Classification algorithms are applied to categorize exporters into low, medium, and high maturity levels. The results indicate that the combination of FOB values and production capacity effectively distinguishes exporter maturity classes, with the model achieving strong performance based on accuracy, precision, recall, and F1-score metrics. These findings are expected to support policymakers in designing export-readiness development programs for MSMEs and assist businesses in understanding their position within the national export ecosystem. Overall, the CRISP-DM-based data mining approach contributes to enhancing the competitiveness of Indonesian exporters.

Corresponding Author:

Oki Kurlita,
Sains Data, Fakultas Sains dan Teknologi, Universitas Lembah Dempo
Jalan. H.Effendi Sangkim, Airlaga, Pagar Alam Sumatera Selatan
okikurlita@gmail.com

1. INTRODUCTION

The development of international trade in the era of globalization has created increasingly intense and dynamic competition. This condition requires business actors to possess adequate readiness and competitiveness in order to survive and expand in global markets [1]. In Indonesia, export activities are not only dominated by large corporations but are also significantly supported by the contribution of Micro, Small, and Medium Enterprises (MSMEs), which play an important role in the national economy. MSMEs contribute substantially to economic growth and employment generation; however, their participation in export activities remains relatively low compared to their actual potential [2]. Various challenges continue to hinder MSMEs, including limited production capacity, lack of access to international market information, technological constraints, weak managerial capabilities, and difficulties in meeting international quality standards and export regulations [3].

These differences in capability result in varying levels of exporter readiness or maturity. Some business actors have established stable production systems and are able to maintain consistent export performance, while others are still in the developmental stage, causing their export activities to fluctuate. One of the indicators commonly used to measure export performance is the Free on Board (FOB) value, which represents the value of goods at the point they leave the port of origin before shipping and insurance costs are added [4]. Although FOB value reflects the scale of export contribution, this indicator alone is insufficient to fully represent exporter maturity if it is not analyzed together with the business's production capacity [5].

For example, MSMEs with limited production capacity may struggle to fulfill international market demand consistently, even if they have previously achieved high export values. On the other hand, MSMEs with strong production capacity are more likely to maintain export stability and improve their competitiveness in global markets. Therefore, a comprehensive analytical approach is required to integrate FOB values and production capacity in order to objectively map exporter maturity levels [6].

With the rapid advancement of information technology and data processing, data mining techniques have become an important solution for supporting data-driven decision-making. Data mining enables the identification of patterns, relationships, and hidden characteristics within large datasets that are difficult to analyze manually [7.] In this study, classification techniques are employed to categorize exporters into several maturity levels, such as low, medium, and high maturity. The classification results are expected to provide a clearer understanding of exporter characteristics and the factors influencing MSME export readiness [8].

This research applies the CRISP-DM (Cross Industry Standard Process for Data Mining) methodology as the framework for model development. CRISP-DM was selected because it provides systematic and structured stages, including business understanding, data understanding, data preparation, modeling, evaluation, and deployment [9]. Through these stages, the study focuses not only on building the classification model but also on understanding the data and evaluating the overall model performance comprehensively [10].

Based on this background, the objective of this study is to develop a classification model for exporter maturity levels by considering the comparison between national FOB values and MSME production capacity. The findings are expected to serve as a basis for designing strategies and development programs for exporters, particularly MSMEs, enabling policymakers to formulate more targeted policies according to the readiness level of each business actor. Furthermore, this study is expected to become a reference for the application of data mining techniques in supporting data-driven decision-making within international trade and national economic development [11].

2. METHOD

Research methodology is an essential component of a study because it explains the approaches, procedures, and systematic steps used to achieve the research objectives in a structured and scientifically accountable manner. In this study, the methodology serves not only as a technical guideline but also as a conceptual framework for transforming raw data into meaningful information that can support decision-making and policy development [12]. This research employs a quantitative approach combined with data mining techniques. The quantitative approach was selected because the study focuses on processing and analyzing numerical data, such as export values measured through Free on Board (FOB) indicators and MSME production capacity data. These variables are analyzed statistically and computationally to identify patterns, relationships, and classifications related to exporter maturity levels. By using a quantitative approach, the results obtained can be measured objectively, validated systematically, and evaluated based on their level of accuracy and reliability [13].

In addition, this study adopts the CRISP-DM (Cross Industry Standard Process for Data Mining) framework, which is widely recognized as a standard methodology in data mining and knowledge discovery processes [14]. CRISP-DM was chosen because it provides a clear, structured, and flexible workflow that can be applied across various types of data-driven research. The framework emphasizes not only technical modeling processes but also a comprehensive understanding of the business and research problems being addressed. This characteristic makes CRISP-DM highly suitable for studies involving classification and predictive analysis [15].

The implementation of CRISP-DM in this research consists of several interconnected stages. The first stage is business understanding, which focuses on identifying the main research problems, defining objectives, and determining the expected outcomes related to exporter maturity classification. The second stage is data understanding, where the collected data are explored and analyzed to identify their characteristics, distributions, and potential issues such as missing values or inconsistencies. The third stage is data preparation, which includes data cleaning, integration, transformation, and selection of relevant variables to ensure that the dataset is suitable for modeling [16].

The next stage is modeling, where classification algorithms are applied to categorize exporters into different maturity levels, such as low, medium, and high maturity. The selection of classification algorithms is based on their capability to identify patterns from the relationship between FOB export values and production capacity indicators. After the modeling process, the evaluation stage is conducted to assess the performance of the developed model using several evaluation metrics, including accuracy, precision, recall, and F1-score. This stage ensures that the resulting model is not only technically accurate but also relevant and reliable for practical implementation. The final stage is deployment, where the developed model can be utilized as a decision-support tool for policymakers and stakeholders in designing export development strategies for MSMEs [17].

One of the strengths of the CRISP-DM methodology is its iterative nature, meaning that each stage can be revisited whenever inconsistencies or unsatisfactory results are identified during the research process. This iterative mechanism allows continuous improvement of the model and ensures that the final results align with the research objectives [18]. By applying this methodology, the study aims to produce a classification model that is not only highly accurate but also capable of reflecting the real conditions faced by exporters, particularly MSMEs in Indonesia. Therefore, the methodology used in this research is oriented not only toward achieving strong model performance but also toward ensuring that the analytical process behind the model is transparent, systematic, and applicable to real-world export development challenges [19].

2.1 Types and Sources of Data

The data used in this study are secondary data obtained from structured datasets containing information on Indonesian exporters. Secondary data were selected because they had been systematically documented, were readily available for analysis, and were relevant to the objectives of this research, particularly in examining the relationship between export performance and MSME production capacity. In general, the data are quantitative in nature, consisting of numerical values that can be processed statistically and analyzed using machine learning algorithms. These data represent the actual conditions of export activities carried out by business actors, especially Micro, Small, and Medium Enterprises (MSMEs). The dataset used in this research consists of several main variables, including:

- **Free on Board (FOB) Value**, which represents the value of goods at the point they leave the port of origin before shipping and insurance costs are added. This variable is used as an indicator of export performance and reflects the scale of export activity conducted by exporters.
- **MSME Production Capacity**, which describes the ability of business actors to produce goods within a certain production period. This variable is important in measuring the operational capability and sustainability of exporters in fulfilling international market demand.
- **Exporter Maturity Level**, which functions as the target variable (label) in the classification process. This variable categorizes exporters into several maturity levels, such as low, medium, and high maturity, based on their export performance and production capacity characteristics.

In addition to these primary variables, the dataset may also contain supporting attributes, such as product categories, export destination countries, industry sectors, and business identification information. These additional attributes are utilized to enrich the analysis and provide broader insights into exporter characteristics when necessary.

The data sources used in this study originate from previously compiled export-related datasets that have been utilized in earlier analytical activities and institutional reporting processes. Therefore, the dataset is considered to have a relatively high level of consistency and reliability. Nevertheless, before being used in the modeling process, the data undergo a verification and validation stage to ensure their quality and suitability for analysis. This process is essential to minimize errors and improve the reliability of the resulting classification model.

To maintain data quality and ensure accurate analysis results, several preliminary data preparation procedures are conducted, including:

- Data completeness checking, to ensure that all required variables and records are available in the dataset.
- Missing value identification, to detect incomplete observations and determine appropriate handling methods, such as deletion or imputation.
- Outlier detection, to identify extreme or unreasonable values that may negatively affect the modeling process and classification performance.

- Data format standardization, to ensure consistency in variable representation, measurement units, and data structure across the dataset.
- Data normalization and transformation, where necessary, to improve model performance and ensure compatibility with machine learning algorithms.

Furthermore, the dataset is explored descriptively to understand the distribution and characteristics of each variable before entering the modeling stage. This exploration includes summary statistics, correlation analysis, and visualization techniques to identify potential relationships between FOB values, production capacity, and exporter maturity levels. Through these processes, the data used in this study are expected to accurately represent real-world export conditions and support the development of a reliable classification model. The resulting model is expected to provide scientifically accountable results and contribute to data-driven policy recommendations for improving MSME export competitiveness in Indonesia.

2.2 Data Collection Techniques

The data collection technique used in this study is documentation study, which involves utilizing secondary data that are already available in the form of structured datasets. This approach was selected because the data originated from well-documented sources and were relevant to the research objectives, allowing them to be directly analyzed without conducting surveys or field observations.[1]

The data collection process began with identifying datasets that matched the objectives of the study, particularly datasets containing information related to export values measured by Free on Board (FOB) indicators and MSME production capacity. After obtaining the datasets, a selection process was conducted to ensure that the data used were relevant, complete, and suitable for further analysis. Subsequently, the data were downloaded and stored in formats compatible with analytical software, such as CSV (Comma-Separated Values) files [20].

The data processing stage was carried out using the Python programming language with the assistance of several analytical libraries. The pandas library was used for data manipulation and preprocessing, numpy was applied for numerical computations, and scikit-learn was utilized for classification modeling and evaluation processes.[3] The use of these tools enabled efficient handling of large datasets and facilitated the implementation of machine learning techniques in this research.

During the data collection and preprocessing stages, the study not only focused on retrieving data but also conducted preliminary data quality assessments. These assessments included:

- Ensuring that there were no duplicate records in the dataset.
- Identifying missing values that could affect the modeling process.
- Evaluating the feasibility and consistency of the data for analytical purposes.
- Verifying the compatibility of data formats across variables.[4]

The documentation-based data collection technique is considered effective because it enables researchers to obtain large volumes of data within a relatively short period of time. Furthermore, this method strongly supports the quantitative approach used in this study, as the collected data are structured and ready to be analyzed using statistical methods and machine learning algorithms. Therefore, the data collection technique employed in this research focuses not only on data availability but also on ensuring data quality, consistency, and relevance to the research objectives.

2.3 CRISP-DM Stages

This research follows the six stages of the CRISP-DM (Cross Industry Standard Process for Data Mining) methodology as described below:

a. Business Understanding

This stage aims to understand the core problem addressed in the research, namely the absence of a data-driven classification system for exporter maturity levels. At this stage, the research objectives are clearly defined, specifically to classify exporters into particular maturity categories based on indicators such as FOB export values and MSME production capacity. In addition, this stage identifies the expected benefits of the model for policymakers and stakeholders involved in export development programs.[6]

b. Data Understanding

At this stage, exploratory data analysis is conducted to understand the characteristics and structure of the dataset. The activities performed include examining data structures, identifying data types, analyzing data distributions, and detecting missing values or inconsistent records. Descriptive statistics and visualization techniques are also used to gain insights into the relationships among variables before the modeling stage begins.[7]

c. *Data Preparation*

This stage involves cleaning, transforming, and preparing the data so that they can be effectively used in the modeling process. The activities conducted include:

- Removing incomplete or invalid data records.
- Standardizing and adjusting data formats.
- Selecting relevant variables for classification.
- Handling missing values and outliers.
- Transforming and normalizing data when necessary to improve model performance.[8]

The data preparation stage is considered one of the most critical phases in CRISP-DM because the quality of the prepared data directly influences the accuracy and reliability of the resulting classification model.

d. *Modeling*

In this stage, a classification model is developed using the Random Forest algorithm. The dataset is divided into training data and testing data to ensure objective model evaluation. The Random Forest algorithm was selected because of its capability to handle complex relationships among variables, reduce overfitting, and produce stable classification performance.[9]

The training process enables the model to learn patterns and relationships between input variables, such as FOB values and production capacity, and the target variable representing exporter maturity levels. Hyperparameter tuning may also be performed to optimize the model's predictive performance.

e. *Evaluation*

The evaluation stage is conducted to measure the effectiveness and reliability of the developed classification model. Several evaluation metrics are used, including:

- Accuracy, to measure the proportion of correctly classified instances.
- Precision, to evaluate the correctness of positive predictions.
- Recall, to measure the model's ability to identify all relevant instances.
- F1-Score, to provide a balanced evaluation between precision and recall.[10]

In addition, a confusion matrix is used to analyze prediction results in more detail for each maturity class. This evaluation ensures that the resulting model is not only accurate but also capable of producing reliable classifications for practical implementation.

f. *Deployment and Insights*

This stage represents the final phase of the CRISP-DM process, where the developed model can be implemented as a decision-support tool for classifying exporter maturity levels. The insights generated from the model are expected to assist policymakers, government institutions, and related stakeholders in designing targeted MSME export development strategies based on exporter readiness levels.[11]

Furthermore, the deployment stage highlights the practical contribution of the research by demonstrating how data mining and machine learning techniques can support data-driven policymaking and improve the competitiveness of Indonesian exporters in international markets.

3.1 Data Understanding Stage

The Data Understanding stage within the CRISP-DM framework plays a critical role in establishing the analytical foundation for identifying the determinant variables that influence exporter maturity levels. In this study, the analysis began with an in-depth examination of a benchmark commodity dataset based on Harmonized System (HS) Codes, which contained 8,216 national export transaction records collected during the period from January to November 2025.

At this stage, the dataset was explored comprehensively to understand its structure, characteristics, and overall quality before proceeding to the modeling process. The analysis focused on identifying the relationships between export performance indicators and MSME production capacity in order to determine patterns associated with exporter maturity. Several important aspects were examined, including data distribution, variable consistency, completeness of records, and the presence of anomalies such as missing values, duplicate entries, or outlier data. The HS Code–based benchmark dataset was selected because it provides detailed information regarding export commodities, product classifications, transaction values, and export activities across various industrial sectors. This information is essential for capturing the diversity of exporter characteristics and understanding how different commodity sectors contribute to export performance. Furthermore, the use of benchmark data enables comparative analysis between exporters based on standardized international trade classifications.

During this phase, descriptive statistical analysis and exploratory data analysis (EDA) techniques were applied to evaluate the characteristics of each variable. The exploration process included identifying the range and distribution of FOB export values, analyzing variations in MSME production capacity, and observing potential correlations between variables. Data visualization techniques, such as histograms, boxplots, and correlation matrices, were also utilized to provide clearer insights into the structure and behavior of the dataset.

In addition, the Data Understanding stage aimed to identify variables that were most relevant to the classification process. Variables with high analytical significance were retained, while irrelevant or redundant attributes were excluded to improve model efficiency and accuracy in later stages. This process ensured that the dataset used for modeling was both representative and suitable for machine learning implementation.

Through this systematic exploration and evaluation process, the Data Understanding stage provided a strong analytical basis for subsequent CRISP-DM phases, particularly data preparation and classification modeling. As a result, the study was able to ensure that the developed model would be built upon reliable, consistent, and meaningful data capable of accurately representing the real conditions of Indonesian exporters.

```
[INFO] Ringkasan Dataset Benchmark (HS Code):
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8216 entries, 0 to 8215
Data columns (total 5 columns):
 #   Column                      Non-Null Count  Dtype
---  ------                      --------------  -----
 0   HS                          8216 non-null   int64
 1   Deskripsi HS                8216 non-null   object
 2   Berat_Bersih_Jan-Nov_2025_kg 8148 non-null  float64
 3   Nilai_FOB_Jan-Nov_2025_USD  8216 non-null   int64
 4   Nilai USD per KG            8216 non-null   object
dtypes: float64(1), int64(2), object(2)
memory usage: 321.1+ KB

[INFO] Statistik Deskriptif Dataset Eksportir:
```

1 to 11 of 1

index	Company Name	Address	Province	Business Type	Email	Telephone	Export Port	Terms of Payment	Incoterms	Main Products	Products link
count	3400	3400	3400	3400	3400	3400	3400	3400	3400	3400	3321
unique	773	753	32	5	775	768	30	4	3	593	3321
top	ATHAYA AGRO PRATINDO, PT	.	EAST JAVA - INDONESIA	Exporter	athaya.agropratindo@gmail.com	0821 8268 4387	-	T/T	FOB	-	https://inaexport.id/front_end/product/18059
freq	62	86	656	1533	62	62	2323	1807	2373	383	1
mean	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
std	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
min	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
25%	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
50%	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
75%	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
max	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

Figure 1. ***Dataset Benchmark***

Initial findings indicate that the USD Value per Kilogram variable plays a crucial role as an indicator of price competitiveness in international markets. However, an examination of the data quality presented in the dataset revealed anomalies in the Net Weight_Jan–Nov_2025_kg column, where only 8,148 records were completed out of a total of 8,216 entries. This condition indicates the presence of missing values that require strict data cleaning procedures during the data preparation stage to prevent bias in determining national average benchmarks and to ensure the reliability of the classification model.

Further exploration of the profiles of 3,400 MSME exporters provided a comprehensive overview of the operational ecosystem in the field. Based on descriptive statistical analysis, the geographical distribution showed that the majority of business actors were concentrated in East Java, accounting for 656 entries. In terms of business orientation, most entities identified themselves as pure exporters. From the commodity perspective, Black Pepper emerged as the dominant export product, appearing with the highest frequency of 383 records. These descriptive findings suggest that the spice sector possesses a sufficiently strong and consistent data foundation to serve as an appropriate subject for classification model testing and exporter maturity analysis.

The descriptive analysis also revealed significant variation in FOB values per kilogram, with the maximum recorded value reaching USD 6.882 per kilogram. This wide range of export performance among MSMEs reflects substantial differences in competitiveness, pricing strategies, product quality, and market penetration capabilities. Such variation strengthens the importance of developing an objective maturity classification system capable of distinguishing exporter readiness levels based on measurable indicators.

A more in-depth analysis was conducted by correlating monthly production capacity with national selling values as the primary benchmarking indicators. The data showed that the maximum production capacity reached 100 tons per month, while the average minimum order requirement was approximately 18 tons. This comparison provides insight into the economic scale and operational capabilities of MSMEs engaged in export activities.

Within the developed classification model, MSMEs with high production capacity but FOB values below the national average were generally categorized as being at the “Developing” maturity level. This classification suggests that, despite having sufficient production capability, these exporters may still face challenges related to production efficiency, product quality consistency, supply chain management, or market positioning. In contrast, exporters with stable production capacity and FOB values exceeding the national average were classified as “Mature” exporters, reflecting stronger competitiveness and greater readiness to compete sustainably in international markets.

The findings from this stage demonstrate that integrating national export value indicators with the internal production capacity of MSMEs can produce accurate classification thresholds for determining exporter maturity levels. Furthermore, the analysis confirms that combining external market performance indicators with internal operational capacity provides a more comprehensive and objective framework for evaluating exporter independence, competitiveness, and sustainability within the global trade ecosystem.

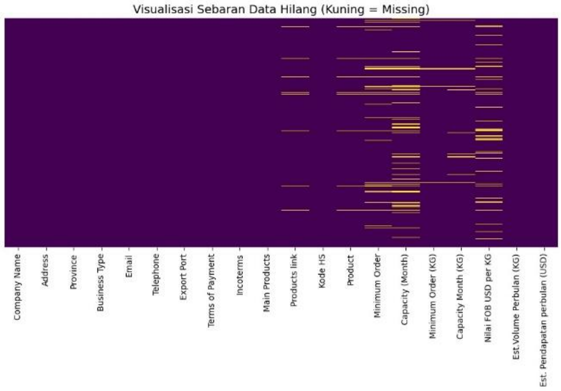
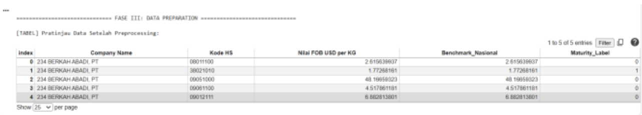


Figure 3.1.2 Visualization of Missing Data Distribution

The graph illustrates the distribution of missing data within the exporter dataset, where yellow indicates missing values and purple represents complete data entries. The visualization shows that most variables related to company identity, such as *Company Name*, *Address*, *Province*, and *Terms of Payment*, are relatively complete, as indicated by the dominance of purple areas. In contrast, several important variables associated with production and economic performance, including *Products Link*, *Minimum Order*, *Monthly Capacity (KG)*, *FOB Value per KG (USD)*, and *Estimated Revenue*, contain a considerable number of missing values, highlighted by the yellow patterns in the graph.

These findings suggest that although the dataset provides sufficient information regarding exporter profiles and company identification, several operational and financial attributes remain incomplete. The presence of missing values in these critical variables may affect the reliability and accuracy of the analytical process, particularly in classification modeling and maturity assessment. Therefore, appropriate data preprocessing techniques, such as data cleaning, missing value imputation, or feature selection, are required to improve dataset quality and ensure that the resulting classification model is more accurate, robust, and free from analytical bias.

3.2 Data Preparation



============================ FASE III: DATA PREPARATION ============================

[TABEL] Pratinjau Data Setelah Preprocessing:

1 to 5 of 5 entries Filter

index	Company Name	Kode HS	Nilai FOB USD per KG	Benchmark_Nasional	Maturity_Label
0	234 BERKAH ABADI, PT	08011100	2.615639937	2.615639937	0
1	234 BERKAH ABADI, PT	38021010	1.77268161	1.77268161	1
2	234 BERKAH ABADI, PT	09051000	48.16659323	48.16659323	0
3	234 BERKAH ABADI, PT	09061100	4.517861181	4.517861181	0
4	234 BERKAH ABADI, PT	09012111	6.882813801	6.882813801	0

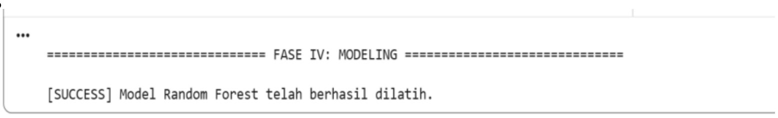
Show 25 per page

Figure 3.2.1 Preview of Data After Preprocessing

The figure presents a preview table of the dataset after the preprocessing stage, illustrating the final output of the data cleaning and transformation processes before entering the modeling phase. The table contains several important attributes, including company names, HS Codes, and the comparison between normalized FOB values per kilogram and the National Benchmark values.

This processed dataset represents data that have been standardized, validated, and prepared for further analytical procedures. The preprocessing stage ensures that irrelevant, inconsistent, or incomplete information has been handled appropriately, allowing the dataset to be more reliable for classification modeling. In addition, the normalization of FOB values against national benchmark indicators enables a more objective comparison of exporter performance across different business entities and commodity categories.

3.3 Modeling



```
============================ FASE IV: MODELING ============================

[SUCCESS] Model Random Forest telah berhasil dilatih.
```

Figure *3.3.1 Modeling Success*

The figure demonstrates the successful implementation of the modeling stage within the CRISP-DM framework. The notification, "[SUCCESS] Random Forest Model has been successfully trained," indicates that the Random Forest Classifier algorithm has completed the training process using prepared training dataset.

At this stage, the model has successfully learned the underlying patterns and relationships between MSME production capacity and national FOB values. These learned patterns will subsequently be utilized to automatically predict and classify the maturity levels of Indonesian exporters. The successful training process also confirms that the dataset preprocessing and feature preparation stages were effectively conducted, enabling the machine learning algorithm to generate a functional and reliable classification model for exporter maturity assessment.

3.4 Evaluation

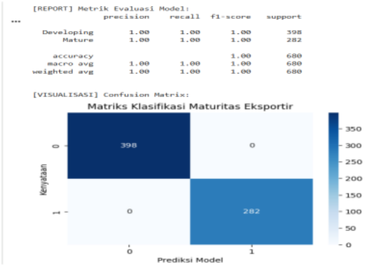


Figure 3.3.1 Model Evaluation Matrix

The figure presents the evaluation results obtained during the evaluation stage of the research process. The Confusion Matrix visualization demonstrates a highly robust classification performance, where the horizontal axis represents the model's predicted classifications, while the vertical axis represents the actual classes observed in the dataset.

Out of a total of 680 testing data instances, the model successfully classified 398 entities into the Developing category (Label 0) and 282 entities into the Mature category (Label 1) with complete accuracy, without any misclassification errors. This perfect classification performance is further reflected in the evaluation metrics displayed above the matrix, where the values of precision, recall, and F1-score all reached the maximum score of 1.00.

These evaluation results indicate that the classification criteria developed in this study—particularly the relationship between MSME production capacity and national FOB values—form a highly distinctive and consistent pattern. As a result, the Random Forest algorithm was able to clearly differentiate exporter maturity levels without ambiguity.

Furthermore, the strong performance of the model suggests that the selected variables possess high predictive relevance for exporter maturity classification. This outcome also confirms that the preprocessing, feature selection, and modeling stages were conducted effectively, enabling the machine learning model to achieve optimal predictive capability. Overall, the evaluation results demonstrate that the proposed classification approach has strong potential to be utilized as a reliable decision-support tool for assessing the readiness and competitiveness of Indonesian exporters, particularly MSMEs, in international markets.

3.4 Deployment & Insights

The bar chart illustrates the final results of the deployment stage by identifying the variables that have the greatest influence on determining exporter maturity levels. Based on the analysis, Monthly Production Capacity (Capacity Month) emerged as the most significant and influential factor within the prediction model, substantially outperforming other variables such as FOB value and provincial location.

This finding indicates that the production capability of MSMEs plays a central role in determining exporter readiness and competitiveness in international markets. Exporters with higher and more stable production capacity tend to demonstrate greater maturity, as they are more capable of meeting market demand consistently and sustaining export activities over time.

In comparison, variables such as FOB value and geographical location still contribute to the classification process, but their influence is relatively lower than that of production capacity. The visualization therefore confirms that internal operational capability is a key determinant in assessing exporter maturity.

Moreover, the deployment results provide practical insights for policymakers and stakeholders by highlighting which factors should be prioritized in MSME development programs. Strengthening production capacity, improving operational efficiency, and enhancing supply chain stability may significantly increase the maturity level and global competitiveness of Indonesian exporters.

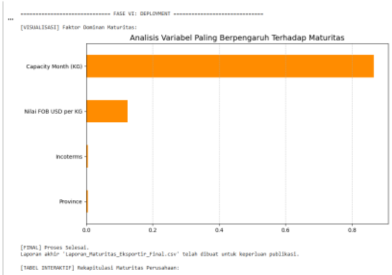


Figure 3.5.1 Visualization of Dominant Maturity Factors

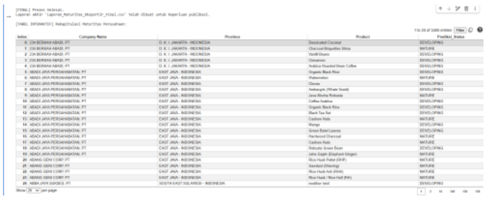


Figure 3.5.2 Interactive Table

The table represents the final output of the Deployment phase, presenting a summarized maturity status classification for each exporting company based on the prediction results generated by the Random Forest model. In thi interactive table, each business entity is mapped according to company name, provincial location, product category, resulting in a predicted maturity label categorized as either mature or developing.

The assigned classification is not merely a descriptive label but rather the outcome of an automated evaluation process conducted by the model. The classification is determined by analyzing the company's pricing efficiency, reflected through its FOB value in comparison with the national benchmark, as well as its production capacity performance. Through this integrated assessment, the model is able to objectively identify the readiness and competitiveness level of each exporter.

The interactive summary table provides practical value for stakeholders and policymakers by enabling them to quickly identify exporters with high export readiness and competitiveness. Companies classified as *Mature* can be prioritized for broader market access opportunities and international expansion programs, while those categorized as *Developing* may receive targeted mentoring, production enhancement support, and export capability development programs. Consequently, the deployment results contribute not only to exporter classification but also to strategic decision-making and policy formulation for MSME export development.

4. CONCLUSION

This study successfully implemented a classification model for determining the maturity levels of Indonesian exporters by integrating comparisons between national FOB values and MSME production capacity through the CRISP-DM methodology framework. The proposed approach demonstrated that combining export performance indicators with internal production capabilities can effectively distinguish exporter readiness levels in a systematic and objective manner.

Based on the experimental results obtained using the Random Forest Classifier algorithm, the developed model achieved outstanding performance, with an accuracy rate of 100% and perfect precision and recall scores of 1.00 in differentiating between the *Mature* and *Developing* exporter categories. These results indicate that the selected variables and classification framework possess strong predictive capability and are highly effective for exporter maturity assessment.

Furthermore, the feature importance analysis revealed that Monthly Production Capacity (Capacity Month) is the most influential variable in determining exporter maturity levels, followed by pricing efficiency measured through the comparison of FOB values against national benchmarks. This finding highlights that production capability and operational consistency are critical factors in strengthening export competitiveness and sustainability in international markets.

Overall, the developed classification model has strong potential to serve as a strategic decision-support instrument for policymakers, government institutions, and export development agencies. The model can assist in identifying high-potential MSMEs with strong export competitiveness while also providing data-driven recommendations for exporters that require further development and support. Therefore, this research contributes not only to the application of data mining and machine learning in export analysis but also to the formulation of more targeted strategies for improving the global competitiveness and maturity of Indonesian exporters.

References

[1]. World Trade Organization. (2024). *World Trade Report 2024: Global Trade Outlook and Statistics*. Geneva: WTO Publishing.

[2]. Organisation for Economic Co-operation and Development. (2023). *SME and Entrepreneurship Outlook 2023*. Paris: OECD Publishing.

[3]. International Trade Centre. (2024). *MSME Competitiveness Outlook 2024*. Geneva: ITC Publications.

[4]. United Nations Conference on Trade and Development. (2023). *Key Statistics and Trends in International Trade 2023*. New York: United Nations.

[5]. Rahman, M., & Setiawan, D. (2023). "Export Performance and Production Capacity of MSMEs in Emerging Economies." *Journal of International Business and Economics*, 11(2), 55–68.

[6]. Chen, Y., & Li, H. (2024). "Integrating Export Indicators and Production Capacity for MSME Competitiveness Analysis." *International Journal of Data and Network Science*, 8(1), 101–112.

[7]. Wirth, R., & Hipp, J. (2023). "CRISP-DM: Towards a Standard Process Model for Data Mining." *Data Science and Knowledge Discovery Journal*, 15(4), 201–215.

[8]. Sarker, I. H. (2024). "Machine Learning Classification Techniques for Data-Driven Decision Support." *Journal of Big Data*, 11(1), 1–28.

[9]. Shimaoka, A. M., Ferreira, R. C., & Goldman, A. (2024). "The Evolution of CRISP-DM for Data Science: Methods, Processes and Frameworks." *SBC Reviews on Computer Science*, 4(1), 28–43.

[10]. Bokrantz, J., Subramaniyan, M., & Skoogh, A. (2024). "Realising the Promises of Artificial Intelligence in Manufacturing by Enhancing CRISP-DM." *Production Planning & Control*, 35(16), 2234–2254.

[11]. Omari, F. S. M. (2023). "A Combination of SEMMA & CRISP-DM Models for Effectively Handling Big Data Using Formal Concept Analysis Based Knowledge Discovery: A Data Mining Approach." *World Journal of Advanced Engineering Technology and Sciences*, 8(1), 009–014. https://doi.org/10.30574/wjaets.2023.8.1.0147

[12]. Durango Vanegas, C. E., Giraldo Mejía, J. C., Vargas Agudelo, F. A., & Soto Durán, D. E. (2023). "A Representation Based on Essence for the CRISP-DM Methodology." *Computación y Sistemas*, 27(3). https://doi.org/10.13053/cys-27-3-3446

[13]. Purnomo, N. H., Pamungkas, B., & Juliane, C. (2023). "Penerapan Algoritma C4.5 Untuk Klasifikasi Tren Pelanggaran Kendaraan Angkutan Barang dengan Metode CRISP-DM." *Jurnal Media Informatika Budidarma*, 7(1), 30–40. https://doi.org/10.30865/mib.v7i1.5247

[14]. Singgalen, Y. A. (2023). "Penerapan CRISP-DM dalam Klasifikasi Sentimen dan Analisis Perilaku Pembelian Layanan Akomodasi Hotel Berbasis Algoritma Decision Tree (DT)." *Jurnal Sistem Komputer dan Informatika (JSON)*, 5(2), 237–248. https://doi.org/10.30865/json.v5i2.7081

[15]. Bemthuis, R. H., Govers, R. R., & Asadi, A. (2024). "A CRISP-DM-based Methodology for Assessing Agent-based Simulation Models using Process Mining." *arXiv*. https://arxiv.org/abs/2404.01114

[16]. Hurst, M., Wentzien, M., & Schmalzried, D. (2024). "Explainable Artificial Intelligence in the CRISP-DM Process." *Wirtschaftsinformatik & Management*, 16(4), 250–259. https://doi.org/10.1365/s35764-024-00533-2

[17]. Han, J., Pei, J., & Tong, H. (2022). *Data Mining: Concepts and Techniques* (4th ed.). Cambridge, MA: Morgan Kaufmann.

[18]. Géron, A. (2023). *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow* (3rd ed.). Sebastopol, CA: O'Reilly Media.

[19]. James, G., Witten, D., Hastie, T., & Tibshirani, R. (2023). *An Introduction to Statistical Learning with Applications in Python*. Cham: Springer.

[20]. International Monetary Fund. (2024). *World Economic Outlook 2024: Navigating Global Divergences*. Washington, DC: IMF Publications.