IFC-Bank of Italy Workshop on "Data science in central banking"

18-20 February 2025

Enhancing external debt nowcasting: a data-driven approach using advanced analytics[1]

Candra Puspita Dewi, Aditya Wisnugraha Sugiyarto
and Sony Panji Wicaksono,
Bank Indonesia

[1] This contribution was prepared for the workshop. The views expressed in this publication are those of the authors and do not necessarily represent the official views of the Committee, its members, or the BIS.

Enhancing external debt nowcasting: a data-driven approach using advanced analytics

Candra Puspita Dewi[1], Aditya Wisnugraha Sugiyarto[2], Sony Panji Wicaksono[3]

Abstract

External debt is a critical component of a country's Balance of Payments (BoP), reflecting the financial transactions between residents and non-residents that are vital for economic activities and policymaking. In emerging economies like Indonesia, external debt plays a crucial role in supplementing domestic resources for investment and growth. Therefore, accurate forecasting of debt disbursement and repayment is essential for effective economic planning and management. However, current methods face challenges due to insufficient data sources and the unreliability of projections for future transactions. This paper introduces an advanced analytics approach to nowcasting external debt by integrating time series analysis with machine learning techniques. By leveraging historical debt transaction data, a granular nowcast model is developed, which identifies patterns in disbursement and repayment across various debt types, creditors, and corporate sectors. Time series analysis allows for precise future predictions, while machine learning groups transactions based on their historical patterns and characteristics. This combined approach significantly improves the accuracy of short- and medium-term external debt forecasts, offering a more reliable alternative to conventional methods. Our initial findings show that this approach can effectively project patterns in external debt transactions, especially in trade and other loan. The enhanced nowcasting model contributes to more informed decision-making in external sector management and strengthens the accuracy of Indonesia's BoP analysis.

Keywords: balance of payments, time series analysis, external debt, nowcasting, machine learning

JEL classification: C82, E58, F14, O19

[1] Department of Statistics, Bank Indonesia, email: candra_pd@bi.go.id

[2] Department of Statistics, Bank Indonesia, email: aditya_wisnugraha@bi.go.id

[3] Department of Statistics, Bank Indonesia, email: sony_pw@bi.go.id

Contents

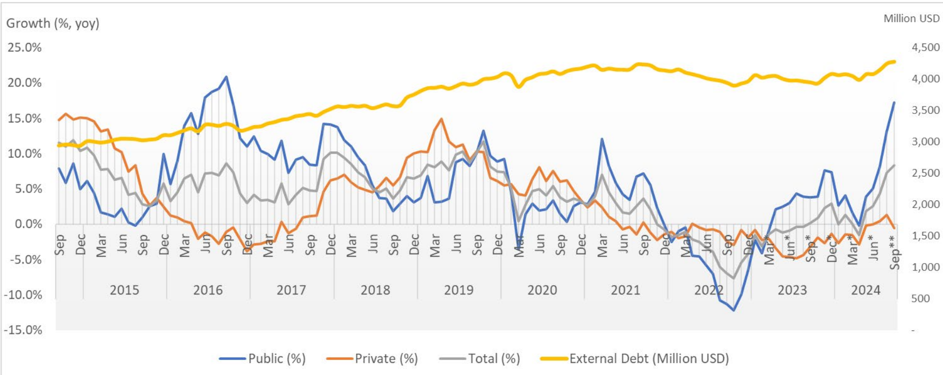
1. Background

External debt (ED) is one of the primary sources of financing frequently used by emerging countries. Naturally, emerging countries prioritize economic growth, which requires substantial capital investments. However, emerging countries often face challenges such as current account deficits, insufficient domestic savings to meet investment needs (saving-investment gap), and fiscal imbalances. These conditions contribute to the increase of external debt as a vital financing option (Dawood, 2021).

Indonesia, as an emerging country, experiences similar conditions. The increased of external debt over the past decade prove this condition. Indonesia's external debt grew from USD 293.8 billion at the end of 2014 to USD 427.8 billion by the end of third quarter 2024 (Fig. 1). This growth stems from increases in both public and private external debt, with an average annual growth rate of 4.0% (*y-o-y*). After a contraction period during the COVID-19 (2020-2022), Indonesia’s external debt has resumed growth post-pandemic. In terms of proportion, the proportions of public and private external debt remain relatively balanced.

Indonesian External Debt Growth (yoy)

Figure 1



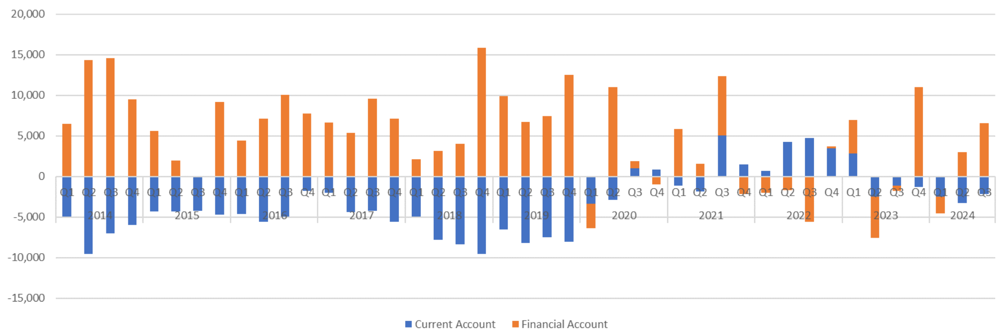
Source: Indonesian External Debt Publication (September 2024)

External debt also plays an essential role in the Balance of Payments (BoP). External debt transactions are part of the financial account liabilities, besides equity. These transactions appear in the form of direct investments, portfolio investments, or other investment. In the BoP framework, external debt disbursement often offset current account deficits, thereby maintaining the overall balance of payments stability (Bulut, 2011).

Indonesia’s BoP have similar patterns. From 2014 to 2020, Indonesia's current account deficit was offset by financial account surplus (Fig. 2). This trend reversed slightly during the pandemic, as the current account recorded a surplus driven by reduced imports due to the decline in economic activity. Additionally, the rise in commodity prices bolstered export values, further contributing to the current account surplus. Conversely, the financial account experienced a deficit because of capital outflows in line with the global economics uncertainties. This anomaly ended post-pandemic, with current account returning to deficit, balanced by surplus in financial account. In addition, external debt forms 55% of financial account liabilities, surpassing equity (Fig.3).

Indonesian Balance of Payment

Figure 2

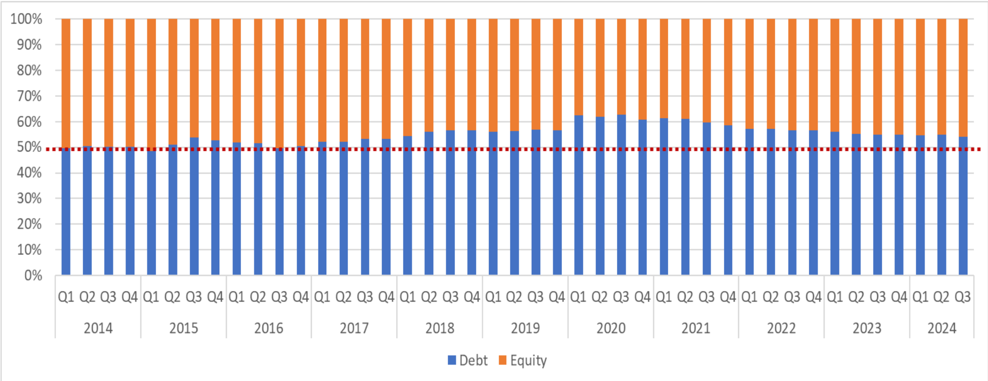


Source: Indonesian Balance of Payment (Q3-2024)

The substantial share of external debt in financing strengthens the needs of debt sustainability monitoring. This not only monitoring external debt transaction realization but also projecting future disbursements and repayments. Such projections are crucial for support preventive policies to ensure external debt sustainability and for estimating foreign exchange demands.

Indonesian Liability on International Investment Position

Figure 3



Source: International Investment Position Publication (Q3-2024)

Forecasting foreign exchange demands aligns with the role of Bank Indonesia as the central bank, whose primary objective is to achieve and maintain Rupiah stability. To achieve this objective, Bank Indonesia conduct various strategies that require reliable information, including projections of foreign exchange demands.

Bank Indonesia projections for external debt disbursements and repayments is a part of foreign exchange demands forecasting. For public external debt, the method is by doing coordination with the Ministry of Finance to estimate disbursements and repayments for government loans and bonds. For private external debt, projections rely on reports submitted by corporations. While projections for public external debt are considered reliable due to the dependable data source from Ministry of Finance, private external debt projections are considered less reliable because of dependency on corporate reports. Therefore, there is a need to enhance methodologies for projecting private external debt disbursements and repayments using alternative approaches in addition to corporate reports.

This paper attempts to adopt a data-driven approach to forecast private external debt disbursements and repayments. Using historical data, we categorize corporations based on their behavior. Clustering is also conducted by grouping corporate external debt types (e.g., loans, debt securities, trade payables, and other debts). Subsequently, time series modeling is applied to each cluster to forecast future external debt disbursements and repayments, a process referred to as a cluster-based hybrid model. This clustering-based approach ensures that similar patterns within each group are captured effectively, enhancing the model's predictive accuracy.

The study focuses on Trade Credits (TC) and Other Loan, as these loan types exhibit clearer and more consistent patterns compared to other categories like loans and debt securities, which are often influenced by irregularities or market volatility. TC and Other Loan are typically tied to operational and trade cycles, making their behaviors more predictable and suitable for hybrid modeling. Concentrating on these loans enhances model accuracy and provides actionable insights for decision-making.

Additionally, when data does not meet the required granularity for clustering, the hybrid framework is adapted to include time series decomposition. This method breaks down the data into Trend, Seasonality, and Residual components, enabling a component-wise analysis to further refine predictions. In cases where decomposition or clustering is not feasible due to data limitations, the analysis is performed at an aggregate level to maintain forecast reliability. Hybrid modeling, including the cluster-based approach and decomposition techniques, which has been extensively studied and shown to outperform classical methods, forms the backbone of this study, ensuring robust and flexible predictions across varying data conditions.

This paper consists of five sections. In the next section, we review relevant literature on data-driven approaches, clustering, and time series analysis for projection. The third section elaborates on the methodology, data sources, and data processing techniques. The results are presented in section four, while the final section provides the conclusions drawn from this study.

2. Literature Review

2.1. External Debt Concept

Gross external debt is outstanding amount of those actual current account, and not contingent, liabilities that require payment(s) of principal and/or interest by debtor at some point(s) in the future and that are owed to non-residents by resident of an economy. Principle amount is the amount that debtor owed to creditor, while principal amount payment is all payment by debtor to creditor that reduce the principal amount outstanding. To be categorised as external debt, the debtor must be a residence and the creditor must be a non-resident. Residence is an economic unit, either a Person or an Entity domiciled or plans to reside in a country for at least 1 (one) year, whose main centre of economic activity is in the country of domicile (IMF, 2014).

External debt statistics presentation is based on the institutional sector. The institutional sector presentations are consistent with the Systems of National Account (SNA) 2008 and Balance of Payments and International Investment Position Manual (BPM) 6^{th} Edition. There are 4 institutional sectors:

1. General government

The general government units of a country consist of those authorities and their agencies that are units established by political processes and exercise legislative, judicial, and executive authority over institutional units within a given territorial area.

2. Central bank

The central bank is a financial institution (or institutions) that exercise control over key aspects of financial system.

3. Deposit taking corporation except central bank

This includes all resident units engaging in financial intermediation as principal activity.

4. Other sectors

This last category consists of other financial corporation (other from deposit taking corporation), non-financial corporation, households, and nonprofit institution serving households.

Other statistics presentation is based on financial instruments. There are 5 financial instruments categories, which is debt securities, loans, currency and deposits, trade credit and advances, and special drawing rights (SDRs).

2.2. External Debt Statistics in Indonesia

Indonesian External Debt Statistics is a joint publication between Bank Indonesia (BI) as Indonesian Central Bank and Ministry of Finance (MoF). While MoF responsible for government external debt, BI responsible for central bank and private sector external debt. For institutional presentation, Indonesia External Debt Statistics consists of government, central bank and private sector external debt.

Government external debt refers to the debt held by the central government, consisting of bilateral/multilateral debt, export credit facilities, commercial debt, and leasing, including government securities (SBN) (issued in Indonesia and abroad) held by non-residents. SBNs include Government Debt Securities (SUN) and Government Sharia Debt Securities (SBSN). SUN consists of government bonds lasting more than 12 months and Treasury Bills (SPN) with up to 12 months of tenures. SBSN consists of long-term SBSN (Ijarah Fixed Rate – IFR) and Global Sukuk.

Central Bank external debt refers to the obligations held by Bank Indonesia to support the balance of payments and reserve assets. The debt includes liabilities in the form of securities issued by Bank Indonesia and held by non-residents and non-residents deposits at Bank Indonesia.

Private external debt refers to the obligations of residents (excluding the government and central bank) to non-residents in a foreign currency and/or the rupiah based on loan agreements or other agreements, deposits, and other liabilities. Private external debt includes liabilities in the form of debt securities issued by residents and held by non-residents. The private sector comprises banks and non-banks. The private non-bank consists of Non-Bank Financial Institutions and Non-Financial Corporations (NFC).

Subsequently, Indonesia employs a comprehensive approach to categorizing its external debt by instruments, which includes debt securities, loans, trade credit, and other loans. This classification system allows for a nuanced analysis of the country's financial obligations, facilitating more effective risk management and policy development. By breaking down external debt into these specific categories, policymakers can better assess the nature and structure of the country's liabilities, enabling them to devise targeted strategies to manage and mitigate associated risks.

Furthermore, Indonesia has disclosed Special Drawing Rights (SDRs) as part of the central bank's external debt. SDRs are international reserve assets created by the IMF and allocated to members to supplement existing official reserves. SDR allocations are considered long-term liabilities, with repayment occurring if a member country decides to withdraw from the IMF or if the IMF's SDR Department is liquidated.

2.3. Forecasting of External Debt

Forecasting external debt is a critical component of economic planning and financial stability for any country. An accurate forecasting involves analysing historical data, current economic conditions, and potential future scenarios. The primary goal of forecasting external debt is to predict the future levels of debt and assess the sustainability of current borrowing practices. This process helps policymakers make informed decisions about fiscal policies, debt management strategies, and economic reforms to ensure long-term financial stability (Cuddington, 1997; Reinhart & Rogoff, 2010).

One key methodology in forecasting external debt is econometric modelling, which uses indicators like GDP growth rates, interest rates, exchange rates, and trade balances. Advanced techniques such as vector autoregression (VAR) and structural equation modelling (SEM) capture dynamic relationships between these variables. Scenario analysis further evaluates the impact of different economic conditions on external debt, helping policymakers understand potential risks and develop contingency plans (Chowdhury, 2001).

Nowcasting enhances real-time assessment by using high-frequency data and advanced statistical methods to provide near real-time estimates of economic indicators. This approach allows for timely adjustments to debt management strategies in response to rapidly changing conditions. By integrating nowcasting with traditional forecasting models, central banks can better predict and manage external debt (Banbura et al., 2013). The nowcasting process involves the following steps:

1. Data Collection: Gather high-frequency data from various sources, such as financial markets, trade statistics, and other economic indicators.
2. Modelling: Use statistical models, such as dynamic factor models (DFMs) or Bayesian vector autoregressions (BVARs), to process and analyse the data.
3. Estimation: Apply these models to estimate current economic conditions and predict near-term future trends.

4. Updating: Continuously update the models with new data as it becomes available to refine and improve the accuracy of the forecasts.

By integrating nowcasting techniques with traditional forecasting methods, policymakers can make more informed decisions and respond more swiftly to changing economic conditions.

2.4. Clustering and Time Series

Clustering is a fundamental technique in data analysis, designed to group similar objects into homogeneous clusters without the need for prior labels. Clustering in the realm of time series analysis, is employed to organize complex temporal data, enabling the identification and utilization of hidden patterns. Time series data possess unique characteristics, such as temporal order, high dimensionality, and varying lengths, which pose significant challenges for conventional analytical methods (Chen, 2020; Aghabozorgi et al., 2015; Jorge & Ruben, 2024).

Additionally, clustering serves as an invaluable tool to simplify the complexity when dealing with granular data where each observation exhibits its own patterns and uniqueness, and where it is difficult to discern aggregate time series patterns. This process facilitates the grouping of similar observations, providing an initial structure that is advantageous for subsequent analyses, including the development of predictive models and classification tasks.

2.4.1. Clustering in Time Series Data

Time series clustering can be classified in various approaches:

1. Shape-Based Clustering

This approach directly compares the shapes of time series without additional modifications. It relies on similarity metrics such as:

- Dynamic Time Warping (DTW), which is highly effective in handling differences in speed and temporal deformations between time series, although it requires high computational costs (Javed et al., 2020; Jorge & Ruben, 2024).
- Euclidean Distance, used in methods like k-Means for data that do not have significant time scale differences (Huang, 2016)

Shape-based clustering is often used in applications with simple time series patterns, such as seasonal patterns in stock prices or climate data.

2. Feature-Based Clustering

This approach transforms time series data into a feature space with more concise and relevant representations, such as:

- Autocorrelation and Spectral Transformation, to identify periodic patterns.
- Wavelet Transformation, which captures high and low-frequency details.

The extracted features are then used by conventional clustering algorithms like K-Means or Hierarchical Clustering. This approach is more robust at handling noise and effective for large-scale datasets.

3. Model-Based Clustering

This model relies on parametric approaches to represent time series, such as Hidden Markov Models (HMM) and ARIMA, which are suitable for data with complex statistical patterns. These parametric models allow for trend-based analysis and long-term predictions, although they have limitations in handling high-dimensional or noisy data (Jorge & Ruben, 2024).

4. Hybrid Approaches

This approach combines shape-based, feature-based, and model-based techniques to leverage the strengths of each method. An example is the R-Clustering algorithm, which uses random convolutional kernels to improve efficiency in handling high-dimensional time series (Jorge & Ruben, 2024).

2.4.2. Challenges in Time Series Clustering

Despite its potential, time series clustering faces the following challenges:

1. High Dimensionality and Large Data Volume. Time series data often consist of hundreds of thousands of data points, requiring representation approaches such as dimensionality reduction to speed up the process (Aghabozorgi et al., 2015; Huang et al., 2016).
2. Length Variability. Differences in the length of time series necessitate approaches like normalization or the use of robust distance metrics such as DTW (Javed et al., 2020).
3. Noise and Outliers. Data imperfections require preprocessing techniques such as smoothing or robust feature extraction (Aghabozorgi et al., 2015).

2.4.3. Applications and Developments in Time Series Clustering

Time series clustering has wide applications across various domains, such as:

1. Finance: Analysis of seasonal patterns, volatility, and stock price predictions (Huang et al., 2016; Jorge & Ruben, 2023).
2. Healthcare: Identification of patterns in ECG signals or breathing patterns (Chen, 2020).
3. Environmental Science: Analysis of weather pattern changes or pollutant concentrations (Aghabozorgi et al., 2015).
4. Technology: Anomaly detection in internet traffic or user behaviour (Jorge & Ruben, 2024).

Clustering serves as an essential preliminary step before modelling, especially when datasets have granular characteristics that make aggregate pattern analysis difficult. Through clustering, the initial structure of the data can be identified to produce more targeted information.

2.5. Hybrid Model of Time Series

The hybrid model in time series analysis is an increasingly popular approach to overcoming the limitations of single models. By combining the strengths of various models, the hybrid approach can enhance the accuracy and robustness of forecasting, especially for data with complex, non-linear, or non-stationary patterns (Xu, et al., 2019).

Hybrid models integrate two or more approaches to address the shortcomings of single models. According to Xu (2019), hybrid models can be divided into three main structures:

1. Parallel Structure: Data is processed simultaneously through multiple models, and the results are combined using specific methods. This structure is suitable for data with many parallel patterns.
2. Series Structure: The first model produces an output that becomes the input for the next model, allowing for sequential processing of linear and non-linear patterns. For example, ARIMA is used to capture linear patterns, while its residuals are processed by a neural network for non-linear patterns.
3. Parallel-Series Structure: A combination of parallel and series approaches to leverage the advantages of both methods.

2.5.1. Advantages and Challenges of Hybrid Models

Hybrid models have proven to be superior to single models in various aspects:

- Higher Accuracy: By combining complementary methods, hybrid models can capture complex patterns, including long-term trends and short-term fluctuations.
- Adaptation to Non-Stationary Data: Models such as Empirical Mode Decomposition (EMD) and Recursive Empirical Mode Decomposition (REMD) are often used to decompose data into simpler components, each of which can be modelled more effectively.
- Multidomain Flexibility: Hybrid models have been used in various fields, including finance, energy, health, and meteorology, demonstrating their adaptability to different types of data and scenarios.

However, despite these advantages, hybrid models also face challenges, including design complexity. It's because the models combining various models requires a complex architectural design and a deep understanding of data characteristics.

2.5.2. Applications of Hybrid Models

These hybrid models have been applied in various fields, including:

- Financial Sector: Stock market and currency price predictions often use a combination of ARIMA with deep learning methods such as LSTM or Deep Belief Network (DBN) to capture complex patterns, including market volatility (Rahman & Pujiati, 2021; Xu et al., 2019), and public sector external debt predictions (Razinkova et al., 2023).

- Energy and Environmental Sector: Forecasting electricity load and renewable energy using hybrid models that combine statistical techniques with neural networks or ensemble methods to capture seasonal fluctuations and long-term trends (Hajirahimi & Khashei, 2019).
- Health Sector: Hybrid models are applied to forecast disease spread, such as COVID-19, using ARIMA for linear patterns and LSTM for non-linear patterns. This integration has helped in better health policy planning (Kontopoulou et al., 2023).
- Meteorological Sector: In weather prediction, hybrid models utilize decomposition methods such as wavelet transform combined with LSTM to capture complex patterns of wind, rain, or temperature (Yang et al., 2021).

3. Methodology

3.1. Data Source

Indonesian External Debt Statistics use various data source, 1) government external debt is compiled using MoF internal data, which consist of outstanding, disbursement, repayment and projection/scheduled external debt; 2) central bank external debt is compiled using BI internal data; 3) private external debt is compiled mainly using data form reporting system which is External Debt Reporting System (EDIS), in addition with data from bank report, data provider such as C-Bonds, and other anecdotal information. In accordance with Bank Indonesia's regulations on the requirement to report external debt, all corporations that have external debt must report their debt to Bank Indonesia through EDIS.

While government and central bank external debt projection data is more reliable, as the result of using institutional data, private external debt projection is more difficult. Currently, EDIS as private external debt main data sources have function to differentiated between scheduled and actual debt repayment. However, EDIS purposes is for data collection, so it doesn't have estimating function. In addition, disbursements and repayments of debt is very related with the company's financial condition, so there can be difference between scheduled and actual debt repayments. This presents a challenge for Bank Indonesia to ensure the accuracy and completeness of the data needed for effective analysis and policy making. In line with that condition, this paper is mainly focus is on the projection of private external debt.

For compiling private external debt, data from EDIS is transform into database named the "realization cube" and the "projection cube". Realization cube contains detailed information on the position values, as well as the realization of principal and interest payments, which are updated monthly. The dimensions of this database are quite complex, including the individual names of corporations, names of creditors, maturity periods, types of debt, ownership status, economic sectors, and original currencies. This comprehensive data structure allows for precise tracking and analysis of private external debt, facilitating effective monitoring and management.

Meanwhile, a projection cube accommodates planned values for disbursements and repayments, as well as the position of debt until maturity. The "projection cube" dimensions are similar with the "realization cube," ensuring consistency and comparability between planned and actual data. This dual-database approach enhances our ability to forecast and manage external debt, supporting informed decision-making and strategic planning to maintain financial stability. However as mentioned above, the accuracy for planned disbursements and repayments data is still not reliable.

In this paper, we use several fields or variables, namely:

- Company name, as a standard name use by each corporation.
- Loan type, consist of loan agreement, debt securities, trade credit and other loan.
- Outstanding, is the principal debt amount.
- Disbursement, is the amount that corporation withdraw in certain period.
- Repayment, is the amount of payment in certain period.
- Investment status, which categorize into direct investment and non-direct investment.

Furthermore, in this paper we utilize data starting from January 2022 until September 2024. We divide this data into 2 categories, first is data for develop the model or data training. For these purposes we use data from January 2022 until March 2024. Meanwhile for testing or out of sample data, we use data from April 2024 until September 2024.

3.2. Data Processing

This research uses a hybrid model. To perform this data processing, we undertake four stages, that are data analysis, data preprocessing, data modelling, and evaluation.

a) Data Analysis

This step includes exploring the collected external debt data to identify patterns and relationships that can be used in the development of the nowcasting model. For example, this analysis includes examining historical data regarding debt disbursements and repayments based on loan type, investment status and issuer. The primary objective is to understand the fundamental behaviour of external debt transactions. This includes identifying any monthly patterns in disbursements and repayments within each company, as well as determining which entities exhibit similar repayment behaviours to their disbursements within the same month.

b) Data Preprocessing

This stage focuses on preparing raw data before modelling. In the context of this paper, these steps include cleaning up external debt data, handling missing or inconsistent values, and normalizing the dataset.

Additionally, this stage involves selecting key entities based on the Pareto 80/20 principle. Only corporations contributing to 80% of the total external debt (as determined from historical data) are prioritized for further analysis. This approach not only reduces computational complexity but also optimizes the results by concentrating on corporations with the most significant impact on the overall external debt data. By focusing on these key contributors, the analysis remains both efficient and impactful.

In addition, data normalization is also used to align variable scales to equalize, improve model performance, reduce bias, and facilitate analysis. Min-Max scaling is used as a normalization function in this study with the formula:

$$X_{normalized} = \frac{X - X_{min}}{X_{max} - X_{min}}$$

The final step in this process is clustering. These techniques are used to group companies based on their external debt behaviour patterns, so that the subsequent analysis and modelling becomes more structured. The clustering technique used is based on Dynamic Time Wrapping (DTW) in accordance with what has been discussed by Aghabozorgi, et. al (2015) with code implementation by Tavenard, et. al (2020).

Once cluster memberships are determined, we apply the "window and horizon" technique to model the time series data. This approach divides the time series into overlapping windows of a fixed length, which are used to train the model. The horizon represents the forecasting period beyond the observed window, where the model predicts future values. This method captures both local and global patterns within the time series, enabling it to handle complex and non-linear relationships.

This technique is chosen because, as supported by previous research (e.g. Sugiyarto & Abadi (2019), Chou & Ngo (2016), Zhou et. al (2018)), it has demonstrated superior performance compared to classical time series analysis. By focusing on smaller windows, the model can adapt to localized trends and changes, while the overlapping structure ensures continuity and consistency in the predictions. This combination enhances the model's accuracy and robustness, particularly when dealing with the variability and complexity inherent in external debt data. Through this approach, we aim to achieve more reliable and precise forecasting outcomes.

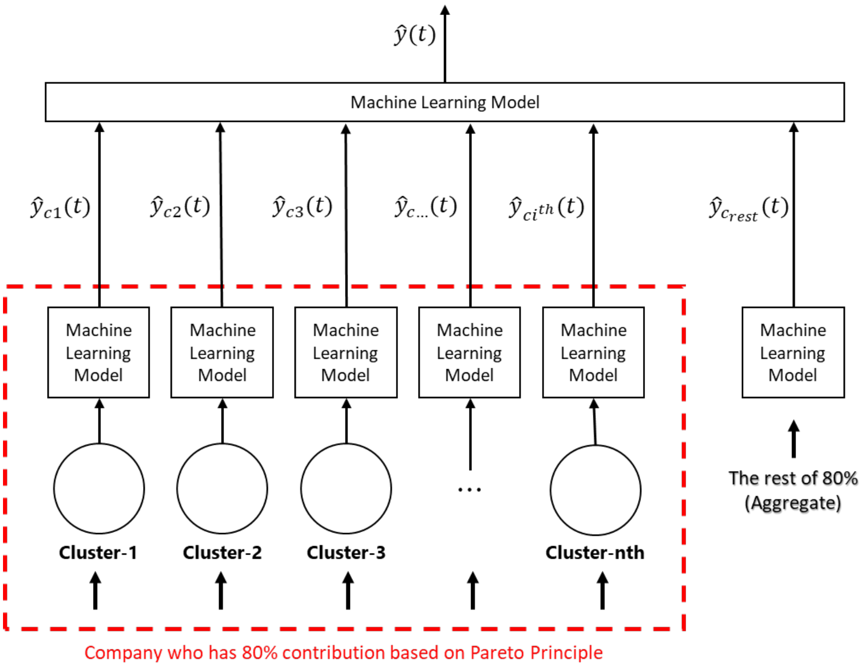
c) Data Modelling

In this step, the clustering method and time series analysis are combined to develop a hybrid nowcasting model. There are 2 (two) types of hybrid model in this paper: 1) Cluster-based hybrid model and 2) Decomposition-based hybrid model.

In cluster-based hybrid model, clustering is used to organize debt data into homogeneous groups, thus allowing the application of a specific time series model for each group. This hybrid approach aims to improve prediction accuracy by leveraging the power of both methods. The scheme of this hybrid model is explained in (Fig. 4):

Clustering-based Hybrid Model

Figure 4

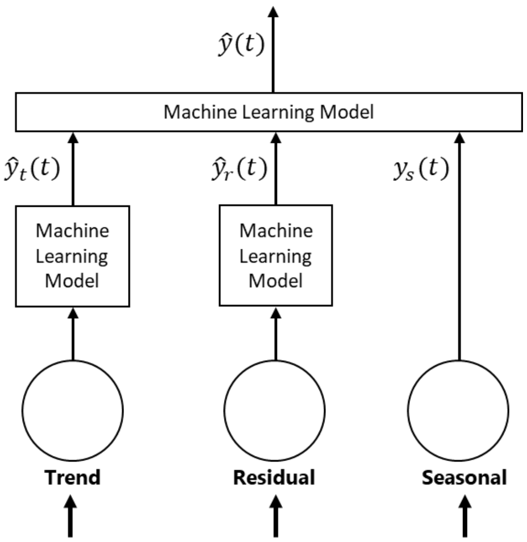


Modelling using machine learning consists of two levels (Fig. 4). At the first level, machine learning is used to model aggregate data from each cluster, which has been grouped based on the characteristics of similar corporation. This ensures that each corporation is modelled specifically according to its unique behaviour. In the second level, machine learning is used to process the prediction results of each first-level model. The goal is to produce more accurate final predictions by reducing errors, which could potentially arise if the results of the first level are simply summed up without further processing.

While the decomposition-based hybrid model schemes presented in (Fig. 5):

Decomposition-based Hybrid Model

Figure 5



In the (Fig. 5) illustrates a hybrid modelling framework using decomposition of time series. The time series is decomposed into 3 (three) components: (i) trend, (ii) seasonality, and (iii) residual. Each component is modelled separately using specialized machine learning models tailored to capture its unique characteristics. The outputs of these component-specific models are then combined by a unified machine learning model in the top layer to produce the final prediction. This modular approach enhances prediction accuracy by addressing the distinct behaviours of each time series component.

Additionally, during the modelling process, we employ Grid Search optimization for hyperparameter tuning to enhance the performance of our models. This approach ensures that the most suitable parameter combinations are identified, optimizing the accuracy of the forecasts. In this study, we utilize a range of algorithms, including Random Forest (RF), Support Vector Regression (SVR), and various linear models such as Lasso Regression, Ridge Regression, and Linear Regression, to forecast and nowcast external debt disbursements and repayments effectively. This diverse selection of algorithms enables the model to capture complex patterns and relationships within the data, ensuring robust and reliable predictions.

d) Evaluation

The last stage is evaluation. This stage assesses the performance of the nowcasting model using metrics such as mean absolute percentage error (MAPE). This evaluation ensures that the model can provide accurate short-term and medium-term predictions for external debt disbursement and repayment. The formula for the metric is as follows:

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{\hat{y}_i - y_i}{y_i}\right| \times 100$$

$\hat{y}_i$: predicted value for i-th data point

y_i : actual value for the i-th data point

n : number of observations

To conclude whether the model is reliable or not, Lewis (1982) divide MAPE criteria as follows (Table 1):

MAPE Criteria for Model Evaluation (Lewis, 1982) Table 1

MAPE	Forecasting Power
<10%	Highly accurate forecasting
10% - 20%	Good forecasting
20% - 50%	Reasonable forecasting
>50%	Weak and inaccurate forecasting

In this paper, we exclusively use MAPE as the evaluation metric because it provides several advantages over other metrics like Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE). Unlike MSE and RMSE, which are sensitive to the range and scale of the data, MAPE expresses the error as a percentage, offering a scale-independent and more interpretable measure. This makes MAPE particularly useful for comparing models across datasets with varying scales and for communicating results to non-technical stakeholders, as it provides a clearer understanding of the model's performance (Myttenaere, B., Grand, & Rossi, 2016).

While MSE and RMSE emphasize larger errors due to their squaring nature, making them more sensitive to outliers, MAPE provides a balanced view by treating all errors proportionally to their actual values. This makes it better suited for cases where understanding relative performance is crucial. Additionally, MAPE is more intuitive, as it reflects the average percentage error, enabling easier interpretation for stakeholders who may not have a technical background.

4. Result and Discussion

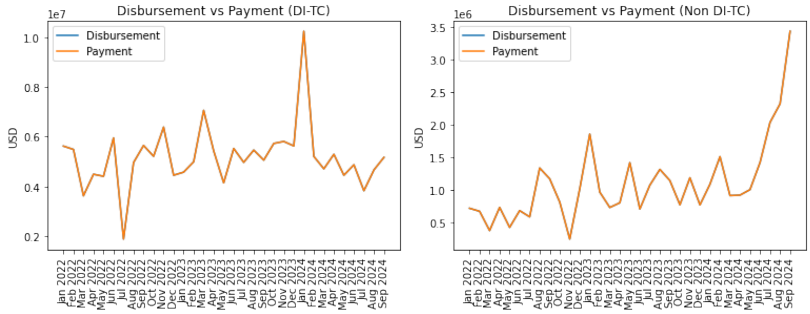
4.1. Data Analysis

In order to enhance external debt nowcasting, the first step involves analysing the values of external debt payments and disbursement to determine how many debtors exhibit disbursement and repayment patterns that align with the "projection cube" dataset. This external debt analysis will be conducted based on different instruments, namely debt securities (DS), loans, trade credit (TC), and other loans.

The results indicate that trade credit and other loans have consistent repayment and disbursement patterns, whereas debt securities and loans show inconsistent patterns. Considering these findings, this paper will focus on analysing the instruments of trade credit and other loans.

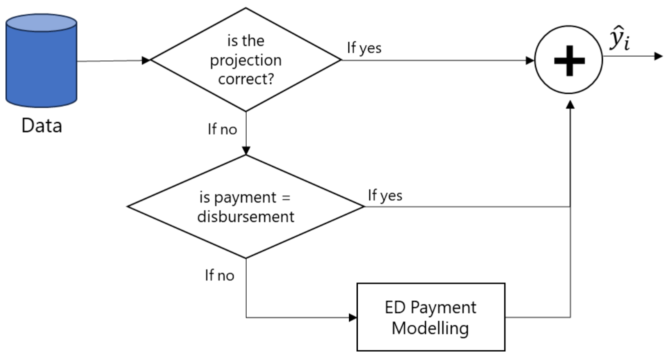
In analysing TC, there are debtors with identical repayment patterns to their disbursement patterns. This pattern suggests several interesting possibilities for further examination. First, the similarity in values may reflect consistency in the repayment patterns of the debtors, indicating financial stability in fulfilling liabilities on schedule. Second, it may also indicate that the transactions are short-term, specifically to meet temporary operational liquidity needs. Additionally, the characteristics of loan types such as TC tend to follow trade cycles. Such patterns have important implications for the development of external debt projection models, particularly in predicting future repayment and disbursement needs. Therefore, identifying entities with similar characteristics can support the accuracy of the nowcasting process and more effective external debt management. In this analysis, there are 4 (four) debtors from the DI-TC group and 9 (nine) debtors from the Non-DI-TC group with such patterns. The examples of these patterns are presented in (Fig. 6):

Examples of Identical Disbursement and Repayment Patterns — Figure 6



Based on the results of the data analysis, the modelling process will be carried out according to the following scheme (Fig. 7):

Modelling scheme (a) Repayments and (b) Disbursement based on data analysis results — Figure 7



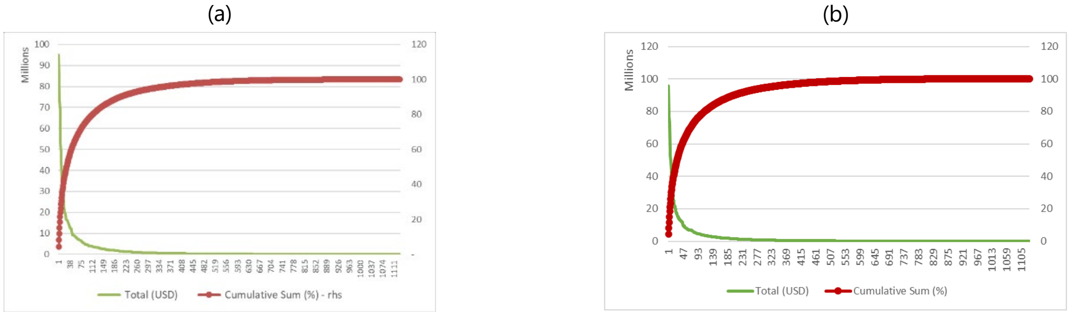
1.2. Data Preprocessing

1.2.1. Data Preprocessing for Clustering-based Hybrid Model

At the data preprocessing stage, an analysis based on Pareto principle was conducted, resulting the following results:

Example of Pareto Calculation Results (a) DI-TC Disbursements (b) DI-TC Repayments

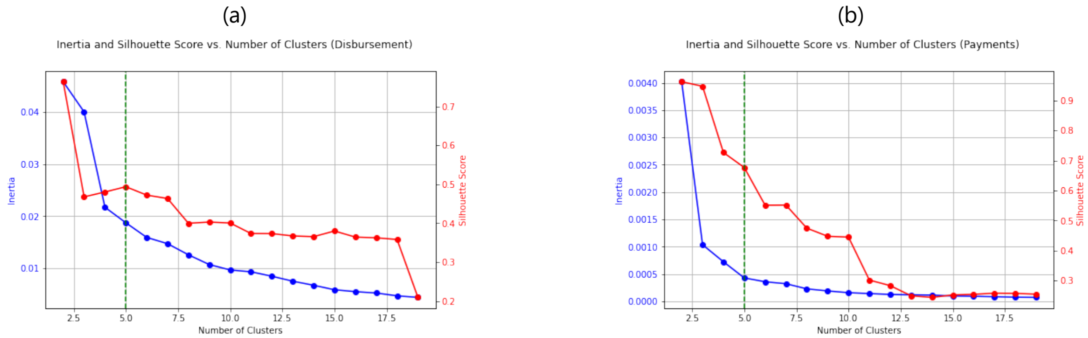
Figure 8



Based on Figure 8, according to the Pareto principle, the data analysis will focus on debtors who contribute 80% of the total external debt. The debtors contributing 80% for the DI-TC loan type (disbursement) are 113 debtors, Non-DI-TC (disbursement) are 115 debtors, DI-TC (payment) are 114 debtors, and Non-DI-TC (payment) are 115 debtors. After that, data normalization is performed to align variable scales to equalize, improve model performance, reduce bias, and facilitate analysis. The next step is the clustering process using DTW, selected based on the Within-Cluster Sum of Squares (WCSS) and Silhouette Score values. The results for the WCSS and Silhouette Score TC are as follows:

Example of Selecting the Optimal Number of Clusters for DI-TC (a) DI-TC Disbursements (b) DI-TC Repayments

Figure 9



In determining the best cluster, we use the WCSS and Silhouette score. The Silhouette score measures how well each point in a cluster fit with its own cluster compared to other clusters, with a value range of [-1,1]. Higher scores indicate better clusters. Meanwhile, the WCSS value measures the total squared distance between points within a cluster and the cluster centre. For selecting the best cluster with WCSS, the Elbow Method is used, where the elbow point on the WCSS graph indicates the optimal number of clusters. Based on data processing, the best clusters for TC payments and disbursements are presented on (Table 2):

Best Cluster Search Results

Table 2

		# Best Clusters	Silhouette Score	WCSS Score
Disbursement	DI-TC	5	0.4936	0.0187
	Non-DI-TC	4	0.7000	0.0007
Repayment	DI-TC	5	0.4206	0.0202
	Non-DI-TC	6	0.4668	0.0137

Based on Table 2, the optimal number of clusters is around 5. Next, aggregate modelling will be performed for each cluster and then aggregated overall.

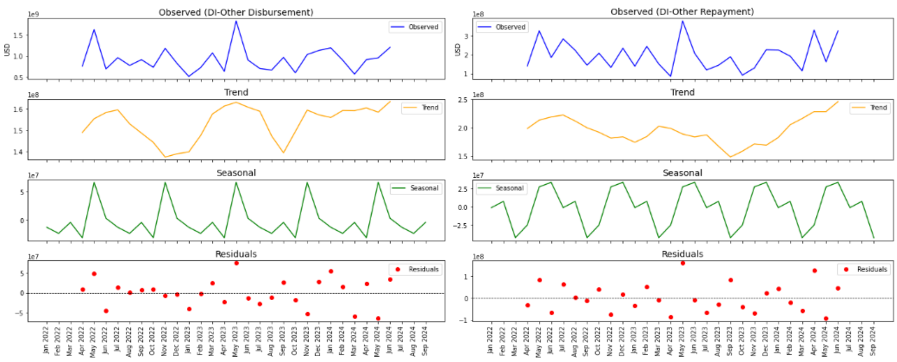
1.2.2. Data Preprocessing for Decomposition-based Hybrid Model

Next, for other types of loans, modelling is conducted using a hybrid model based on time series decomposition. This approach is chosen because, for other loan types, it is not feasible to group debtors into clusters, as indicated by low Silhouette Scores, WCSS values and the MAPE results. During the data preprocessing stage, the raw data is decomposed using the additive method, resulting in three components: Trend, Seasonality, and Residual. The Trend component reflects the overall long-term movement in the data, Seasonality captures recurring patterns or cycles over a fixed period, and Residual represents the random noise or unexplained variability. The results obtained from this decomposition process are presented on (Fig. 10):

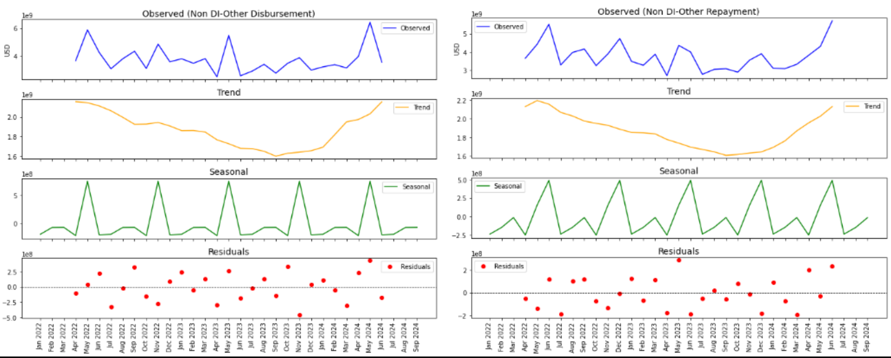
Decomposition Results of (a) DI-Other Loan (b) Non-DI-Other Loan

Figure 10

(a)



(b)



Based on Figure 10, due to the impact of the additive method on the decomposition process, the initial data of 33 data points (Jan 2022 – Sep 2024) was reduced to 27 data points (Apr 2022 – Jun 2024). Each component formed (trend, seasonal, and residual) will undergo different treatments. The trend and residual components will be modelled separately, while the seasonal component will be studied for its seasonal pattern and directly added to the predictions of the trend and residual components. This approach is taken because the seasonal patterns formed through additive decomposition are already very clear, as explained in Section 3.2.c, subsection 2.

1.3. Data Modelling and Evaluation

1.3.1. Clustering-based Hybrid Model on Trade Credit Loans

In this approach, we use a clustering-based hybrid model to forecast Trade Credit (TC) Loans. This method involves grouping time series data using clustering algorithms to form clusters with similar characteristics. The clustering process aims to identify localized patterns within each cluster, enabling a more focused and customized analysis. In this modelling approach, we separate the data based on investment status (Direct Investment (DI) and Non-Direct Investment (Non-DI)) categories for further analysis. The clustering results for the time series disbursement realization data of the DI-TC loan group are presented in (Fig. 11).

Clustering Results from DI-TC Disbursement — Figure 11

Five clusters were generated, each with distinct pattern characteristics:

- Cluster 0, exhibits a simple pattern where most values remain flat with a significant spike toward the end.
- Cluster 1 displays a more complex and fluctuating pattern, with values rising and falling variably over time.
- Cluster 2, shows highly volatile and seemingly random patterns, indicating significant variation within this cluster.
- Cluster 3 stands out with its unique pattern, where values remain stable for most of the period before experiencing a sharp increase at the end.

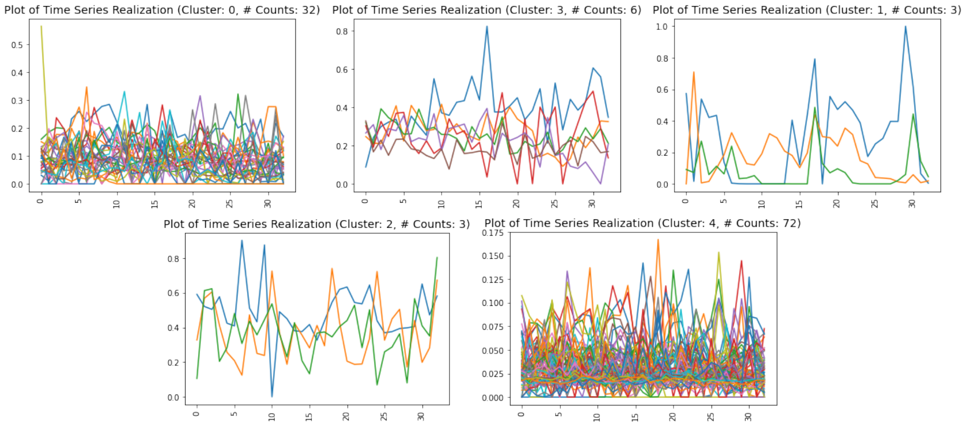
- Cluster 4 demonstrates relatively more organized fluctuations compared to Cluster 2, although variations are still present.

In summary, this clustering result highlights that the time series data for DI-TC Disbursement can be grouped based on similar behavioural patterns. This analysis provides valuable insight into the underlying trends and patterns within the data, which can be further explored for decision-making and interpretation.

Meanwhile, the results for the Non-DI-TC presented in (Fig. 12):

Clustering Results from DI-TC Repayment

Figure 12



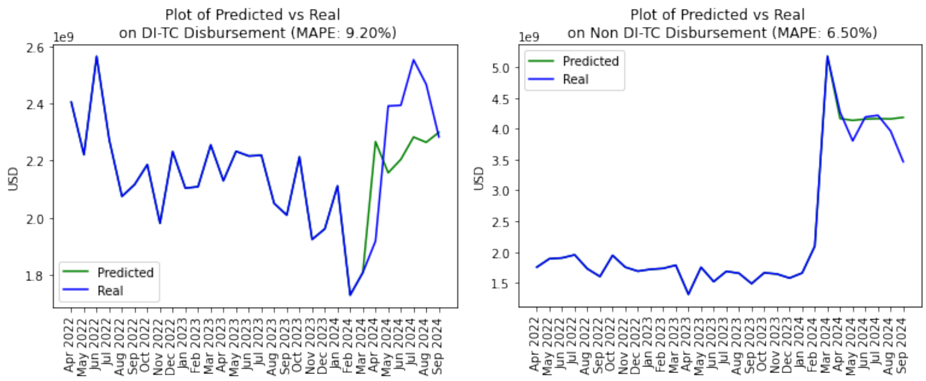
For the Non-DI-TC, five clusters were generated with different characteristics:

- Cluster 0, the lines are densely packed, indicating that the data points share relatively consistent characteristics but with some variations.
- Cluster 1, the patterns are more distinct, suggesting unique characteristics compared to other clusters.
- Cluster 2, the data exhibit different trends, potentially indicating unique underlying behaviours.
- Cluster 3, the patterns show moderate similarity, with slight variations in trends over time.
- Cluster 4, the lines are densely packed and overlap significantly, suggesting that the data points in this cluster share broadly similar trends with minor deviations.

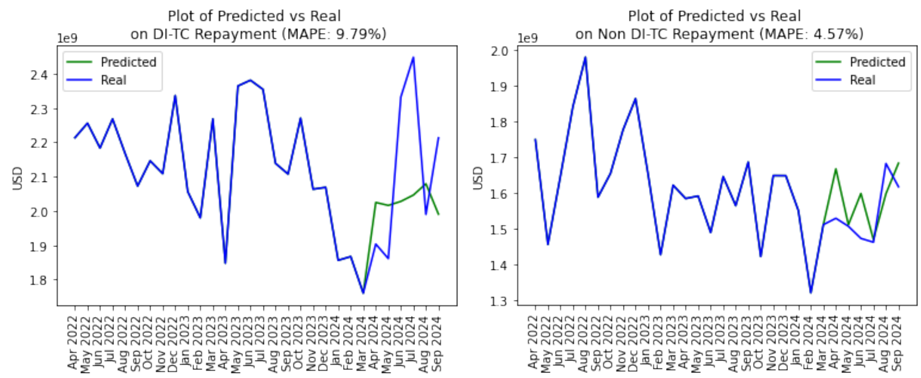
Next, data modelling is performed based on the results from the previous data preprocessing. The following results of the TC modelling are presented in (Fig. 13):

Trade Credit Modelling Results (a) Disbursement (b) Repayment Figure 13

(a)



(b)



Based on Figure 13, the predicted versus actual values of disbursement and repayment for Trade Credit (TC) loans are compared, categorized into Direct Investment (DI) and Non-Direct Investment (Non-DI). In summary, the DI-TC disbursement model achieved a MAPE of 9.2% in medium-term nowcasting (6 months) and 13.9% in short-term nowcasting (2 months), while the repayment model resulted in a MAPE of 9.79% in medium-term and 7.35% in short-term. For Non-DI-TC, the disbursement model demonstrated a MAPE of 6.5% in medium-term and 5.6% in short-term, and the repayment model achieved a MAPE of 4.57% in medium-term and 4.65% in short-term. In this modelling, we perform modelling in each cluster. In Table 3, the details of MAPE and the best model in each cluster are presented.

Comparison of Medium-Term Nowcasting MAPE in Each Cluster Table 3

<table>
<tr><th colspan="2"></th><th>i-th Clusters</th><th>Best Model[4]</th><th>Best MAPE Cluster (%)</th><th>MAPE Hybrid Model (%)</th></tr>
<tr><td rowspan="9">Disbursement</td><td rowspan="5">DI-TC</td><td>1</td><td>RF</td><td>206.43</td><td rowspan="5">9.2
(RF)</td></tr>
<tr><td>2</td><td>RF</td><td>8.28</td></tr>
<tr><td>3</td><td>Ridge</td><td>9.07</td></tr>
<tr><td>4</td><td>RF</td><td>87.76</td></tr>
<tr><td>5</td><td>Ridge</td><td>8.02</td></tr>
<tr><td rowspan="4">Non-DI-TC</td><td>1</td><td>Linear</td><td>9.76</td><td rowspan="4">6.5
(RF)</td></tr>
<tr><td>2</td><td>RF</td><td>15.86</td></tr>
<tr><td>3</td><td>SVM</td><td>85.36</td></tr>
<tr><td>4</td><td>RF</td><td>22.96</td></tr>
<tr><td rowspan="11">Repayment</td><td rowspan="5">DI-TC</td><td>1</td><td>RF</td><td>37.52</td><td rowspan="5">9.79
(Lasso)</td></tr>
<tr><td>2</td><td>RF</td><td>15.21</td></tr>
<tr><td>3</td><td>SVM</td><td>14.85</td></tr>
<tr><td>4</td><td>RF</td><td>87.73</td></tr>
<tr><td>5</td><td>Linear</td><td>9.04</td></tr>
<tr><td rowspan="6">Non-DI-TC</td><td>1</td><td>RF</td><td>23.49</td><td rowspan="6">4.57
(Lasso)</td></tr>
<tr><td>2</td><td>SVM</td><td>9.86</td></tr>
<tr><td>3</td><td>Linear</td><td>14.88</td></tr>
<tr><td>4</td><td>Lasso</td><td>5.54</td></tr>
<tr><td>5</td><td>Lasso</td><td>8.94</td></tr>
<tr><td>6</td><td>SVM</td><td>39.23</td></tr>
</table>

Based on Table 3, the hybrid modelling approach significantly improves prediction accuracy by leveraging the strengths of different models across clusters, tailoring predictions to the unique characteristics of each group. This flexibility is crucial for handling complex datasets like DI-TC and Non-DI-TC, where variability across clusters can be substantial. For instance, the hybrid model achieves lower overall MAPEs (9.2% for DI-TC disbursement and 6.5% for Non-DI-TC disbursement) compared to individual models. Similarly, for repayment, the hybrid model reduces errors to 9.79% for DI-TC and 4.57% for Non-DI-TC by selecting models like Lasso regression that best fit each cluster. By addressing cluster-specific behaviours, the hybrid approach outperforms single-model strategies, making it a powerful tool for predicting diverse financial data accurately. The DI-TC results show a higher MAPE value compared to Non-DI-TC in terms of both disbursements and repayments. This is consistent with its nature, considering that for external debt, DI-TC is provided by affiliates or shareholders who, at certain periods, will make large debt disbursements or repayments in line with the Company's business needs.

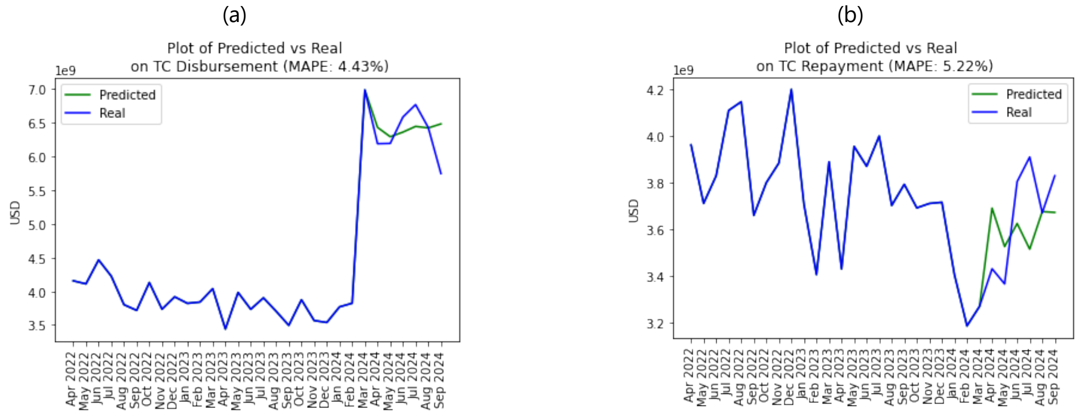
After obtaining the prediction results for DI and Non-DI separately, the results are combined to generate overall predictions for TC Disbursement and TC Repayment. The combined predictions achieve a MAPE of 4.43% (medium-term) and

[4] We utilize Random Forest (RF), Support Vector Regression (SVR), and various linear models such as Lasso Regression, Ridge Regression, and Linear Regression to identify the best model

2.75% (short-term) for TC Disbursement, also for TC Repayment achieve 5.22% (medium-term) and 6.14% (short-term), as illustrated in (Fig. 14):

Results of TC Prediction

Figure 14



1.3.2. Decomposition-based Hybrid Model on Other Loans

For modelling external debt with the "Other" loan type, we applied a time series decomposition-based modelling approach. Following the data preprocessing step, which involved decomposing the time series, each component was subsequently modelled individually. The results of this modelling are presented in Table 4:

Results of Medium-term Nowcasting using Decomposition-based Modelling

Table 4

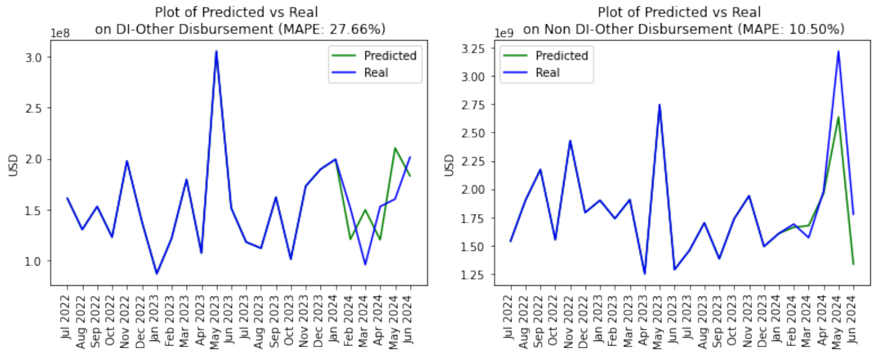
		Time Series Component	Best Model	Best MAPE (%)	MAPE Hybrid Model (%)
Disbursement	DI-Other	Trend	Ridge	2.81	27.66 (RF)
		Residual	Ridge	66.84	
	Non-DI-Other	Trend	Lasso	5.43	10.5 (Ridge)
		Residual	Lasso	88.96	
Repayment	DI-Other	Trend	Ridge	7.39	20.8 (RF)
		Residual	Ridge	67.36	
	Non-DI-Other	Trend	Ridge	4.74	7.78 (Linear)
		Residual	SVR	91.15	

The hybrid model demonstrates strong performance by effectively combining the strengths of different models for the Trend and Residual components, significantly improving prediction accuracy for both Disbursement and Repayment. For non-DI loans, the hybrid model achieves low MAPEs of 7.78% and 10.5%, indicating its robustness in handling predictable trends and variable residuals. While the model also enhances accuracy for DI loans, achieving a MAPE of 27.66% for DI-Other Disbursement, the high MAPE values for Residuals (e.g., 67.36% for DI-Other Repayment) highlight the challenges of modelling unpredictable components. Overall, the hybrid model effectively balances accuracy across components, particularly excelling in non-DI loan predictions, though refinements in handling Residual components could further improve its performance for DI loans. Visualization of the result of decomposition-based hybrid model presented in Figure 15:

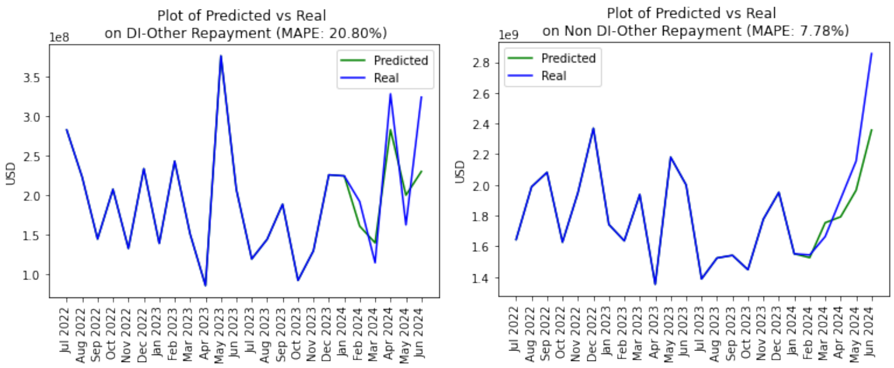
Other Liabilities Modelling Results (a) Disbursement (b) Repayment

Figure 15

(a)



(b)



After obtaining the results for each investment status, all results are combined to produce the final predictions for disbursement and repayment of external debt for the “Other” loan type. The combined result gives better evaluation score, since DI investment type give worse result than non-DI. This condition is due to the behaviour of DI company, that have more flexibility on disbursement and repayment from its parent company. The modelling results for the "Other" loan type are illustrated in Figure 16:

Results of Other Loan Prediction

Figure 16

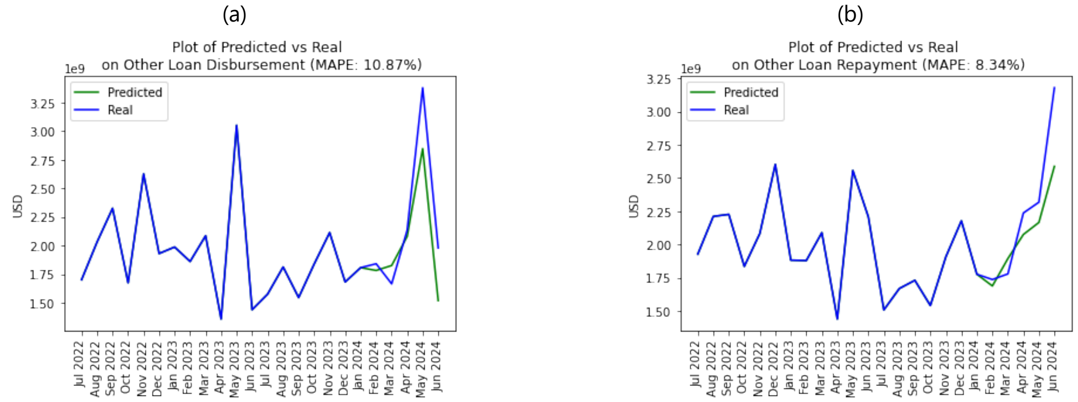


Figure 16 shows the prediction results for disbursement and repayment of other loans, with disbursement MAPE are 10.87% (medium-term) and 6.41% (short-term) also for repayment MAPE are 8.34% (medium-term) and 4.64% (short-term). The disbursement predictions align well with actual values but show slight deviations during periods of sharp spikes, while repayment predictions demonstrate higher accuracy and a closer fit across the time series. These results highlight the hybrid model's effectiveness in capturing overall trends and patterns for other loans, though slight improvements could be made to handle volatility and residual noise more effectively.

1.3.3. Comparison to Existing Method

Currently, in preparing the Balance of Payments for the liabilities section of the Financial Account, the projections/nowcasting of disbursements are solely based on transaction projections submitted by debtors. For repayments, we apply a cleansing technique (e.g., correcting the transaction values based on their behaviours) in the reporting. The comparison of prediction results with the existing method is presented in Table 5:

Comparison of Proposed vs Existing Method

Table 5

MAPE (%)		Short-term (2 months)		Medium-term (6 months)	
		Proposed	Existing	Proposed	Existing
Disbursement	TC	**2.75**	14.33	**4.43**	24.54
	Other	**6.41**	17.89	**10.87**	16.95
Repayment	TC	**6.14**	21.33	**5.22**	11.18
	Other	**4.64**	18.01	**8.34**	20.78

Based on Table 5, the proposed method achieves remarkable accuracy improvements for disbursement. In the case of Trade Credit (TC) disbursement, the proposed method reduces the short-term MAPE to 2.75% compared to the existing method's 14.33%, and in the medium term, the proposed method records a MAPE of

4.43%, far better than 24.54% with the existing method. Similarly, for other loans, the proposed method achieves a MAPE of 6.41% in the short term and 10.87% in the medium term, compared to the consistently high MAPE of 17.89% and 16.95%, respectively, for the existing method.

For repayment, the proposed method also demonstrates significant improvements in accuracy. For TC repayment, the short-term MAPE is reduced to 6.14% compared to 21.33% with the existing method, and in the medium term, the proposed method achieves a MAPE of 5.22%, outperforming the existing method’s 11.18%. Similarly, for other loans, the proposed method achieves a MAPE of 4.64% in the short term and 8.34% in the medium term, significantly better than the existing method's 18.01% and 20.78%, respectively.

Overall, the proposed method exhibits consistent superiority in accuracy, particularly for disbursement predictions, where the existing method struggles with extremely high MAPE values. The improvements in repayment predictions further highlight the robustness and reliability of the proposed approach for both short-term and medium-term forecasting. These results demonstrate that the proposed method is far more effective and reliable for predicting external debt disbursement and repayment compared to the existing method.

2. Conclusion and Future Works

Based on this discussion, several conclusions from this paper are as follows:

a) Advanced analytics, particularly specifically clustering-decomposition and time series methodologies, can be used for projecting private external debt (trade credit and other loan) disbursement and repayment.

b) The implementation of clustering-decomposition and time series methodologies enhances the reliability of projections. This is proven by projection evaluation using MAPE, as shown in the Table 6.

MAPE Score Result – TC

Table 6

<table>
<tr><th colspan="2"></th><th>MAPE Score</th><th>Description</th></tr>
<tr><td rowspan="3">Disbursement</td><td>DI-TC</td><td>9,20%</td><td rowspan="6"><10%, highly accurate forecasting</td></tr>
<tr><td>Non-DI TC</td><td>6,50%</td></tr>
<tr><td>TC Disb. (in total)</td><td>4,43%</td></tr>
<tr><td rowspan="3">Repayment</td><td>DI-TC</td><td>9,76%</td></tr>
<tr><td>Non-DI-TC</td><td>4,57%</td></tr>
<tr><td>TC Repay. (in total)</td><td>5,22%</td></tr>
</table>

MAPE Score Result – Other Liabilities

Table 7

		MAPE Score	Description
Disbursement	DI-Other Liab.	27,66%	20-50%, Reasonable forecasting
	Non-DI Other Liab.	10,5%	10-20%, Good forecasting
	Other Liab Disb. (in total)	**10,87%**	**10-20%, Good forecasting**
Repayment	DI-Other Liab.	20,8%	20-50%, Reasonable forecasting
	Non-DI-Other Liab.	7,78%	<10%, highly accurate forecasting
	Other Liab. Repay. (in total)	**8,34%**	**<10%, highly accurate forecasting**

c) Implementing advanced analytics for projecting private external debt, can enhance work process efficiency. This is achieved by reducing dependencies on corporate reporting data, thereby eliminating coordination processes such as data confirmation.

Nevertheless, in this paper, the advanced analytics methodology (clustering-decomposition and time series) is limited to the projection of trade credit and other loans. The use of this methodology for other types of external debt, such as loan agreements, still requires further research. In the future, BI will develop this methodology and explore other data-driven methodologies for external debt projection, while also considering several external indicators that may influence external debt. Advanced analytics methodologies can also be further explored to project other components in the Balance of Payments, such as investment, trade, or income payment/receipt projections.

References

Aghabozorgi, S., Shirkhorshidi, A. S., & Wah, T. Y. (2015). Time-series clustering – A decade review. *Information Systems, 53*, 16-38.

Awan, R. U., Anjum, A., & Rahim, S. (2015). An econometric analysis of determinants of external debt in Pakistan. *British Journal of Economics, Management & Trade, 5*(4), 382-391.

Banbura, M., Giannone, D., Mondugno, M., & Reichlin, L. (2013). Now-Casting and the Real-Time Data Flow. *Centre for Economic Policy Research*.

Biqiong, Z., Al-Asfour, A., & Nilofar, M. (2020). External Debt and Economic Growth: A Heterogeneous Static Panel Study in Asian Developing and Transition Economies. *Studies of Applied Economics*.

Bulut, L. (2011). External Debts and Current Account Adjustments. *The B.E. Journal of Macroeconomics*.

Chen, M. (2020). Time series clustering and classification. *Journal of the American Statistical Association, 115*(531), 1558.

Chou, J., & Ngo, N. (2016). Time series analytics using sliding window metaheuristic optimization-based machine learning system for identifying building energy consumption patterns. *Applied Energy*, 751-770. doi:10.1016/j.apenergy.2016.05.074

Chowdhury, A. (2001). External Debt and Growth in Developing Countries: A Sensitivity and Causal Analysis. *World Institute for Development Economic Research (UNU-WIDER)*.

Cuddington, J. (1997). Analyzing the Sustainability of Fiscal Deficits in Developing Countries.

Dawood, M., Baidoo, S. T., & Shah, S. M. (2021). An empirical investigation into the determinants of external debt in Asian developing and transitioning economies. *Development Studies Research*.

Hajirahimi, Z., & Khashei, M. (2019). Hybrid structures in time series modeling and forecasting: A review. *Engineering Applications of Artificial Intelligence, 86*, 83-106.

Huang, X. Y. (2016). Time series k-means: A new k-means type smooth subspace clustering for time series data. *Information Sciences, 367-368*, 1-13.

IMF. (2014). *External Debt Statistics, Guideline for Compilers and Users.* Washington D. C.: International Monetary Fund.

Javed, A., Lee, B. S., & Rizzo, D. M. (2020). A benchmark study on time series clustering. *Machine Learning with Applications, 1*.

Jorge, M.-B., & Ruben, C. (2024). Time series clustering with random convolutional kernels. *Data Mining and Knowledge Discovery, 38*, 1862-1888.

Kontopoulou, V. I., Panagopoulos, A. D., Kakkos, I., & Matsopoulos, G. K. (2023). A review of ARIMA vs. machine learning approaches for time series forecasting in data-driven networks. *Future Internet, 15*(255), 1-31.

Lewis, C. (1982). *Industrial and business forecasting methods.* London: Butterworth Scientific.

Myttenaere, A., B., G., Grand, B., & Rossi, F. (2016). Mean Absolute Percentage Error for regression models. *Neurocomputing, 192*, 38-48. doi:10.1016/j.neucom.2015.12.114

Rahman, Y. A., & Pujiati, A. (2021). Dynamic forecasting of government foreign debt: Case of Indonesia. *JEJAK: Journal of Economics and Policy, 14*(1), 52-60.

Razinkova, M., Filatova, H., Pylypenko, A., Nebaba, N., Firsov, O., & Zhuravka, F. (2023). Forecasting Ukraine's external public debt under uncertainty conditions. *Financial and Credit Activity: Problems of Theory and Practice, 6*(53), 202-213.

Reinhart, C., & Rogoff, K. (2010). Growth in a Time of Debt. *American Economic Review*.

Sugiyarto, A., & Abadi, A. (2019). Prediction of Indonesian Palm Oil Production Using Long Short-Term Memory Recurrent Neural Network (LSTM-RNN). *2019 1st International Conference on Artificial Intelligence and Data Sciences (AiDAS)* (pp. 53-57). Ipoh, Malaysia: IEEE Xplore. doi:10.1109/AiDAS47888.2019.8970735

Tavenard, R., Faouzi, J., Vandewiele, G., Divo, F., Androz, G., Holtz, C., . . . Woods, E. (2020). Tslearn, A Machine Learning Toolkit for Time Series Data. *Journal of Machine Learning Research, 21*, 1-6.

Xu, W., Peng, H., Zeng, X., Zhou, F., Tian, X., & Peng, X. (2019). A hybrid modeling method for time series forecasting based on a linear regression model and deep learning. *Applied Intelligence, 51*(4), 1-20.

Yang, Y., Fan, C., & Xiong, H. (2021). A novel general-purpose hybrid model for time series forecasting. *Applied Intelligence, 51*(8), 1-17.

Zhao, G., & Zhao, Z. (2024). EMD-DL: A hybrid forecasting approach using empirical mode decomposition and deep learning for time series prediction. *Preprint*.

Zhou, W., Wang, D., Li, H., & Song, W. (2018). Long-Term Forecasting of Time Series Based on Sliding Window Information Granules and Fuzzy Inference System. *2018 5th International Conference on Information, Cybernetics, and Computational Social Systems (ICCSS)* (pp. 375-380). Hangzhou, China: IEEE Xplore. doi:10.1109/ICCSS.2018.8572415

Enhancing External Debt Nowcasting: A Data-Driven Approach Using Advanced Analytics

Bank Indonesia, Statistics Department

February

2025

Table of Contents

Background

Why external debt projection important?

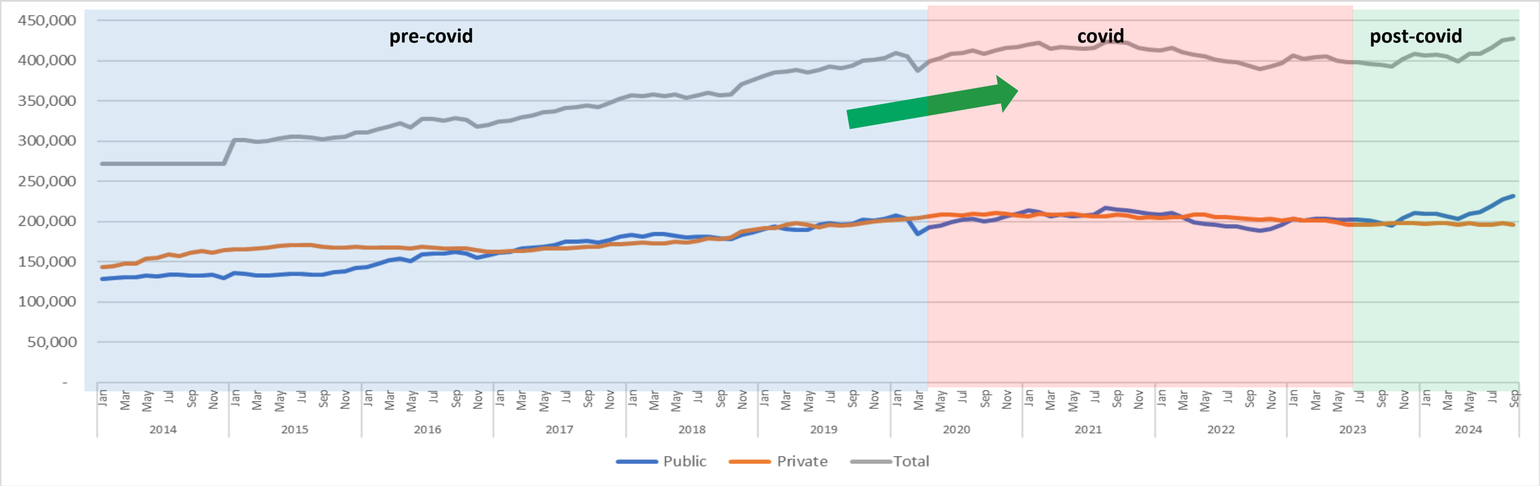
1. External debt is one of the Indonesian primary financing sources.
2. The substantial share of external debt in financing highlights the importance of debt sustainability monitoring (including accurate projection).
3. Projections for external debt disbursements and repayments is a part of foreign exchange supply-demands projection, which are critical for Bank Indonesia's objective of achieving and maintaining Rupiah exchange rate stability.
4. Due to the dependency on corporate report which still lack of good disbursement dan repayment projection, private external debt projection needs improvement in methodology.
5. This research attempt to use data driven approaches (clustering-decomposition, and time series analysis) for private external debt disbursements and repayments projection.

Background

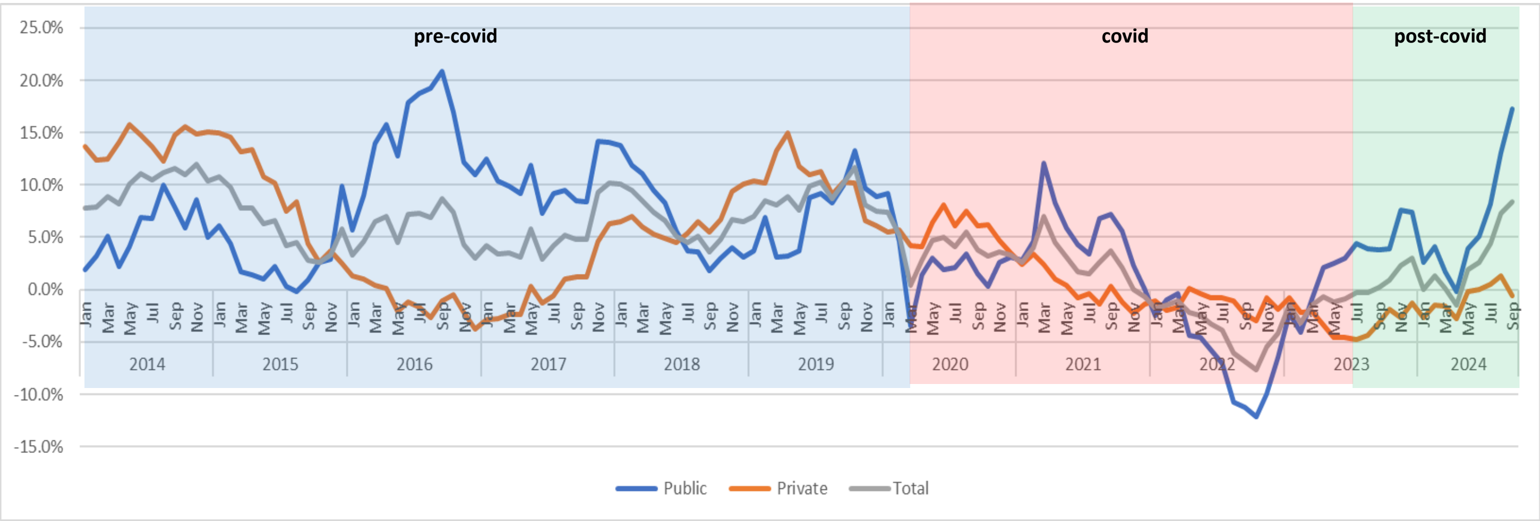
Indonesia's external debt has grown over the past decade

Indonesian External Debt

million of USD



% (yoy)



Indonesia's external debt grew from USD 293.8 billion at the end of 2014 to USD 427.8 billion by the end of third quarter 2024**.** This growth stems from increases in both public and private external debt.

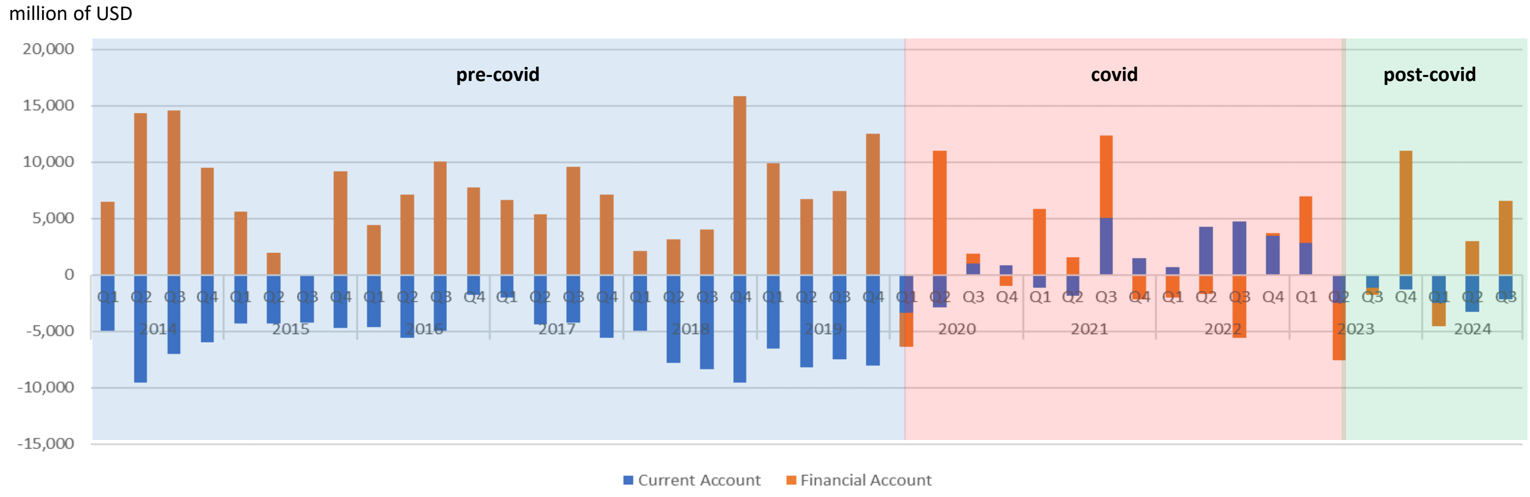
This growth stems from increases in both public and private external debt, with an average annual growth rate of 4.4% (y-o-y). After a contraction during the COVID-19 period (Q1 2020 - Q2 2022), Indonesia’s external debt has resumed growth post-pandemic. However, private external debt growth still under pre-pandemic growth, this affected by post-pandemic economic growth that has not yet fully recovered.

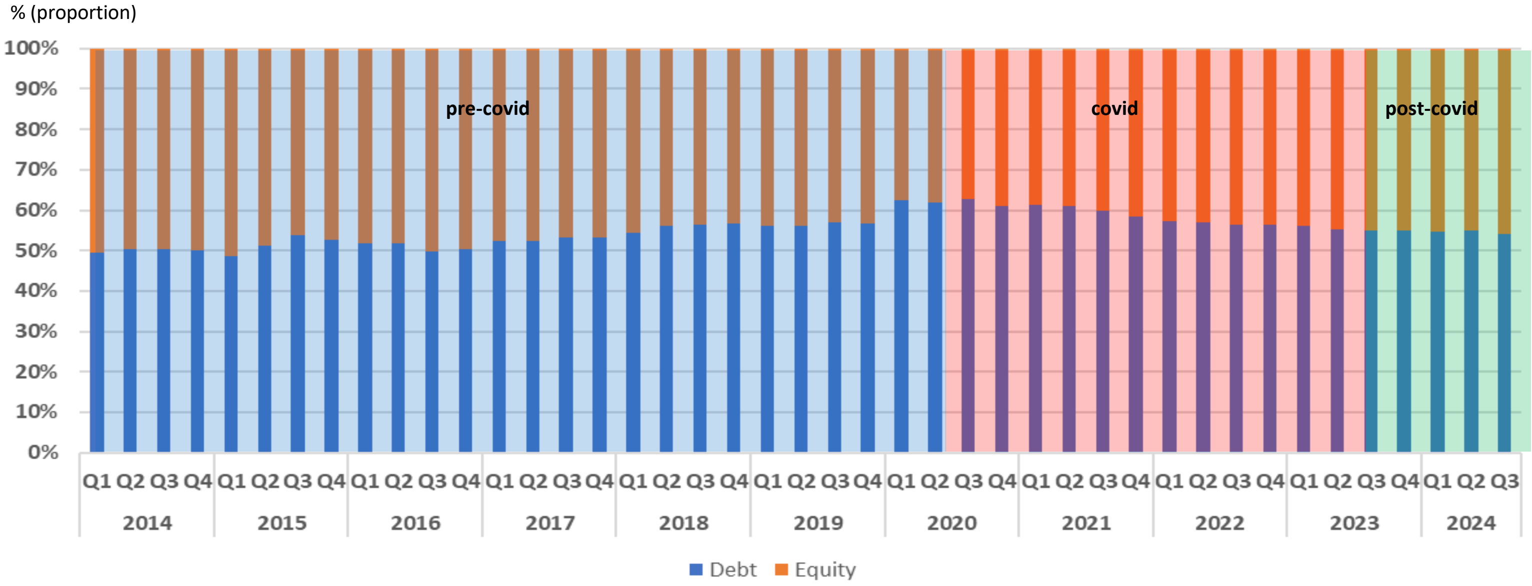
Source: Indonesian External Debt Publication period: September 2024

Background

External debt played important role as part of Indonesia Balance of Payment (BoP)

Indonesian Balance of Payment (BoP) and International Investment Position (IIP)





Indonesia's current account deficit was offset by financial account surplus. This trend reversed slightly during the pandemic, as the current account recorded a surplus driven by reduced imports due to the decline in economic activity. This anomaly ended post-pandemic, with current account returning to deficit, balanced by surplus in financial account.

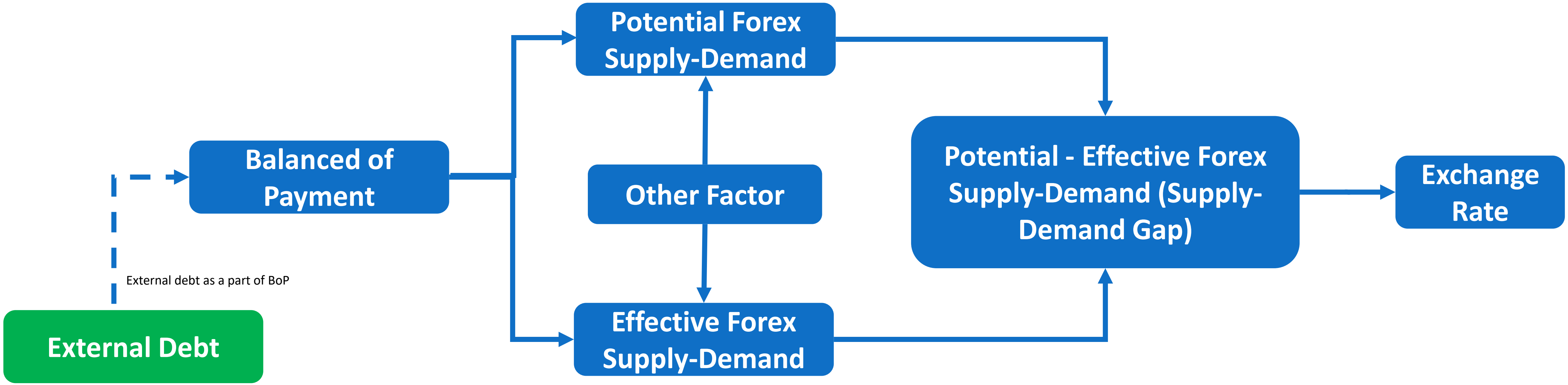
Indonesian external debt also plays an essential role in the Balance of Payments (BoP) and International Investment Position (IIP), especially in financial liability. External debt forms approximately 55% of financial account liabilities position.

Source: Indonesian Balance of Payment and International Investment Position Publication, period: Q3-2024

Background

External Debt can affect Foreign Exchange (Forex) Supply and Demand

- Forecasting foreign exchange demands aligns with the role of Bank Indonesia as the central bank, whose primary objective is to achieve and maintain Rupiah stability. Rupiah stability implies price stability of goods and services as well as rupiah exchange rate stability. To achieve the objective of maintaining rupiah exchange rate stability, Bank Indonesia conduct various strategies that require reliable information, including projections of foreign exchange supply-demands.
- Bank Indonesia projections for external debt disbursements and repayments is an important part of foreign exchange supply-demands projection.



Framework: Working Paper (Piter, 2016)

Research Question

- Is alternative methodology using data-driven approach is suitable for private external debt projection?
- How accurate is data-driven approach (clustering and time series) in projecting private external debt disbursement and repayment?
- What are the advantages in using data-driven approach (clustering-decomposition and time series) for private external debt projection?

Limitation

- This paper is limited to the projection of private external debt, specifically for trade credit and other loan. The other loan type, such as loan agreement and debt securities, needs more deep research on its projection method, including adding external variable such as economic growth, external rate, etc.

Methods: Data Sources

- **Indonesia External Debt** — Joint publication Ministry of Finance (MoF) and Bank Indonesia (Central Bank)
  - **Public**
    - **Government** ← Ministry of Finance
    - **Central Bank** ← Bank Indonesia
  - **Private** ←
    - External Debt Reporting System
    - Bank Report
    - Data Provider
    - Anecdotal Information

Methods:
Research Stages

Data Analysis

This step includes exploring the collected external debt data to identify patterns and relationships that can be used in the development of the nowcasting model. The main objective is to understand the basic behavior of external debt transactions.

Data Preprocessing

This stage focuses on preparing raw data before modelling. In the context of this paper, these steps include cleaning up external debt data, handling missing or inconsistent values, and normalizing the dataset. This stage also select perpetrators based on the Pareto 80/20 principle. Only actors who account for 80% of the total contribution (based on historical data) are prioritized for further analysis.

Data Modelling

In this step, the clustering-decomposition method and time series analysis are combined to develop a hybrid nowcasting model. Clustering-decomposition is used to organize debt data into homogeneous groups, thus allowing the application of a specific time series model for each group. This hybrid approach aims to improve prediction accuracy by leveraging the power of both methods.

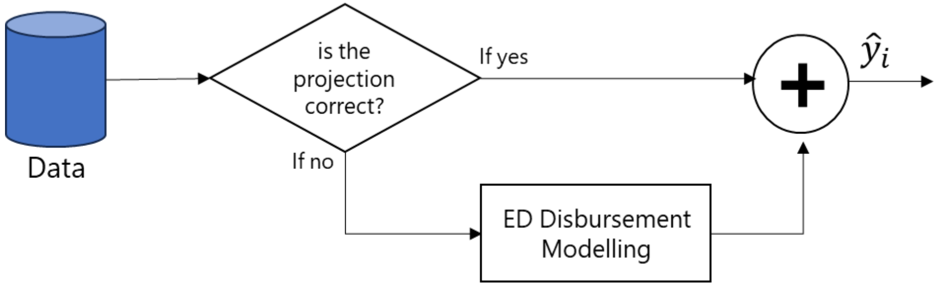
Evaluation

The last stage is evaluation. This stage assesses the performance of the nowcasting model using metrics such as mean absolute percentage error (MAPE). This evaluation ensures that the model is able to provide accurate short-term and medium-term predictions for external debt disbursement and repayment.

Methods

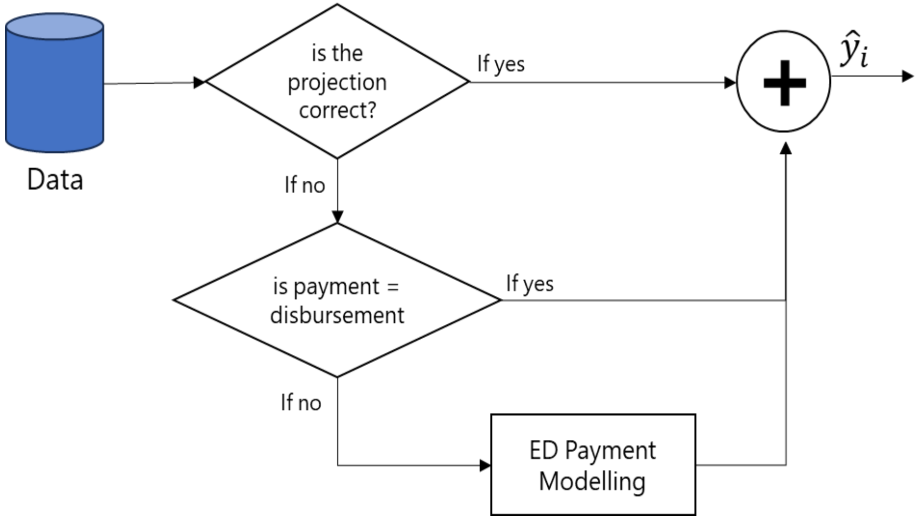
Data analysis:
company selection

Disbursement



- First step is to decide whether the "projection/plan" data reported by company are correct.
- If correct, then use the "projection". If not, then the company is selected to be used in modeling stage.

Repayment

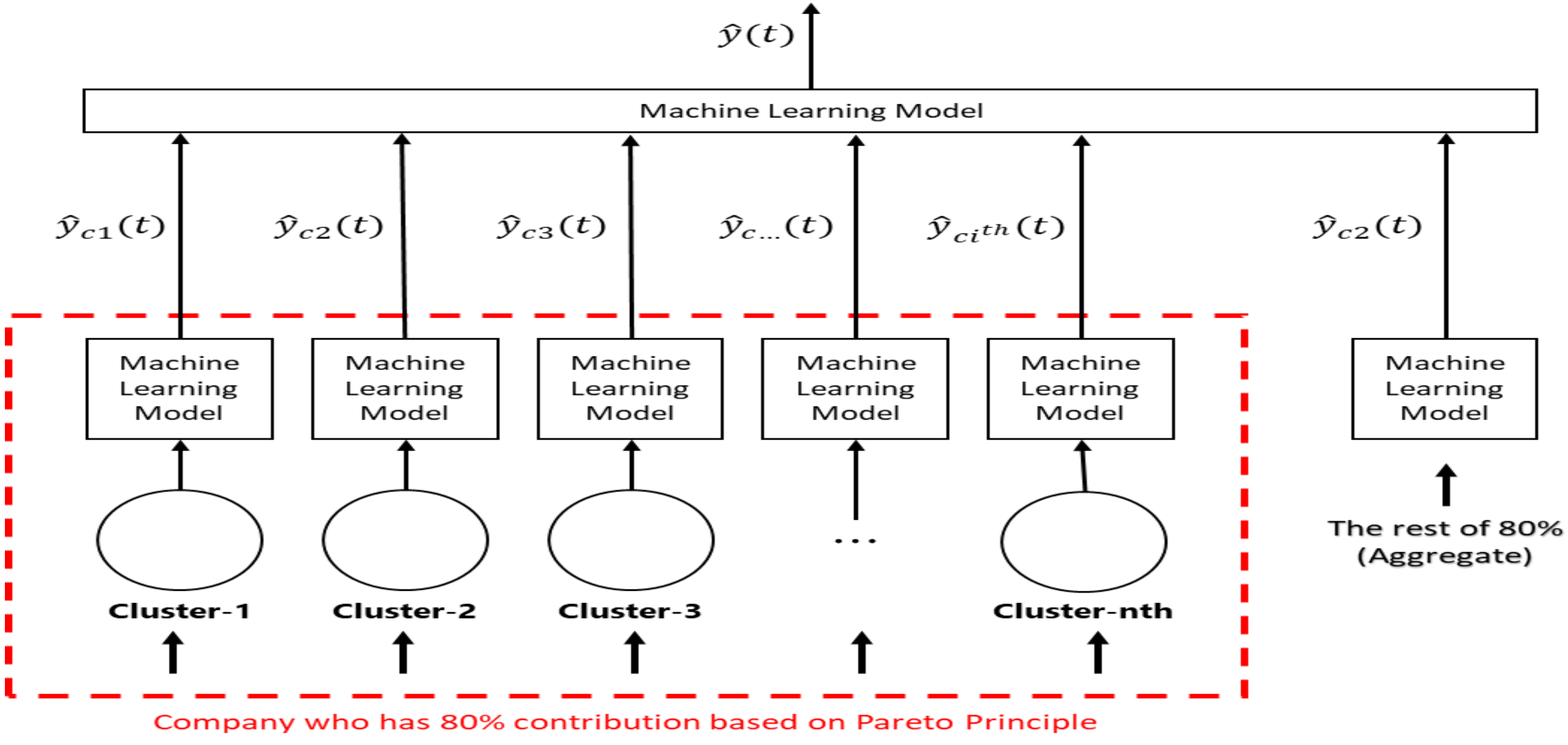


- First step is to decide whether the "projection/plan" data reported by company are correct.
- If correct, then use the "projection". If not, then decide whether the payment is mirroring the disbursement in the same period.
- If correct, then use the disbursement data for repayment. If not, then the company is selected to be used in modeling stage.

Methods

Data preprocessing and modelling:
clustering and time series analysis

Clustering and time-series analysis

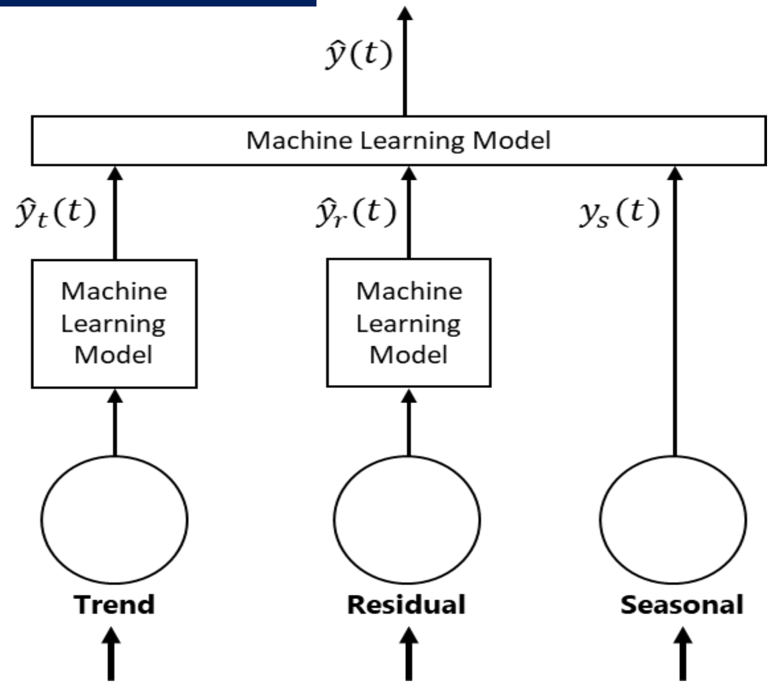


1. Select list of company that have 80% contribution based on nominal value.
2. Using machine learning to divide group of company into several cluster, based on their behaviour on external debt disbursement and repayment.
3. Construct machine learning model (time-series model) for each cluster.
4. Combining result from each model to produce a result.

Methods

Data preprocessing and modelling: decomposition and time series analysis

Decomposition and time-series analysis



1. Divide the company population into 3 categories based on their behaviour in debt disbursement and repayment. The category are:
   - Trend, if there is a pattern on company disbursement and repayment
   - Seasonal, if company make disbursement and repayment in certain period.
   - Residual, other company that doesn't have pattern on disbursement and repayment.
2. Construct machine learning model (time-series model) for each category.
3. Combining result from each model to produce a result.

Methods:

Evaluation

Mean Absolute Percentage Error (MAPE)

For evaluation we use mean absolute percentage error (MAPE). This evaluation ensures that the model is able to provide accurate short-term and medium-term predictions for external debt disbursement and repayment. The formula for the metric is as follows:

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{\hat{y}_i - y_i}{y_i}\right| \times 100$$

$\hat{y}_i$: predicted value for i-th data point

y_i : actual value for the i-th data point

n : number of observations

In order to conclude whether the model is reliable or not, Lewis (1982) divide MAPE criteria as follows:

MAPE	Forecasting Power
<10%	Highly accurate forecasting
10% - 20%	Good forecasting
20% - 50%	Reasonable forecasting
>50%	Weak and inaccurate forecasting

Result

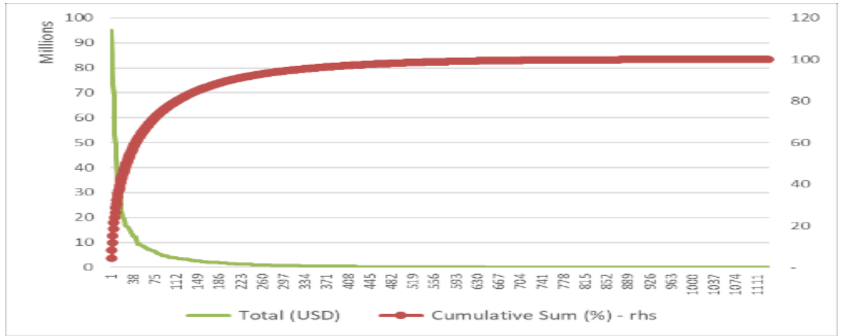
This research show that data-driven approach using clustering-decomposition and time series can project external debt disbursement and repayment specifically for trade credit and other loan

Result : Projection on Trade Credit

1. Data Analysis

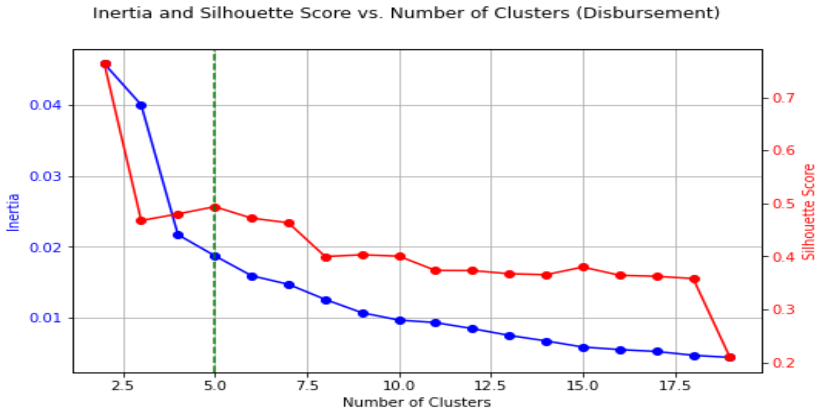
For the trade credit, we use clustering and time-series analysis. The first step is to identify if the projection from company reporting is correct. We found that for all company that have trade credit external debt, the projection reporting is not reliable. Next step, for repayment, we identified 13 company have identical repayment and disbursement patters. There are 4 (four) debtors from the Direct Investment-Trade Credit (DI-TC) group and 9 (nine) debtors from the Non-DI-TC group with such patterns.

2. Data Preprocessing



Example of DI-TC Disbursements

The first step of data preprocessing is to determine company who contribute 80% of the total external debt. The company contributing 80% for the DI-TC loan type (disbursement) are 113 debtors, Non-DI-TC (disbursement) are 115 debtors, DI-TC (payment) are 114 debtors, and Non-DI-TC (payment) are 115 debtors. After that, data normalization is performed to align variable scales to equalize, improve model performance, reduce bias, and facilitate analysis.



Example of Selecting the Optimal Number of Clusters for DI-TC Disbursement

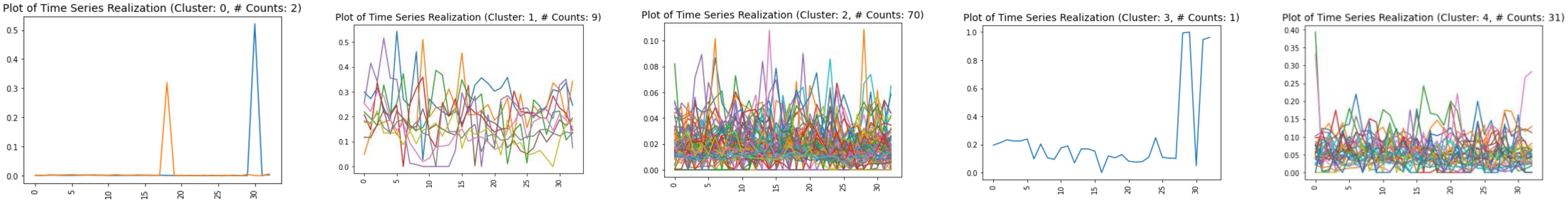
The next step is the clustering process using DTW, selected based on the Within-Cluster Sum of Squares (WCSS) and Silhouette Score values. The results for the WCSS and Silhouette Score TC are as follows:

		# Best Clusters	Silhouette Score	WCSS Score
Disbursement	DI-TC	5	0.4936	0.0187
	Non-DI-TC	4	0.7000	0.0007
Repayment	DI-TC	5	0.4206	0.0202
	Non-DI-TC	6	0.4668	0.0137

Result : Projection on Trade Credit

3. Data Modelling

In this stages, we use a clustering-based hybrid model to forecast Trade Credit (TC) Loans. This method involves grouping time series data using clustering algorithms to form clusters with similar characteristics. The clustering process aims to identify localized patterns within each cluster, enabling a more focused and customized analysis. In this modelling approach, we separate the data based on investment status (Direct Investment (DI) and Non-Direct Investment (Non-DI)) categories for further analysis.



Example of Clustering Results from DI-TC Disbursement

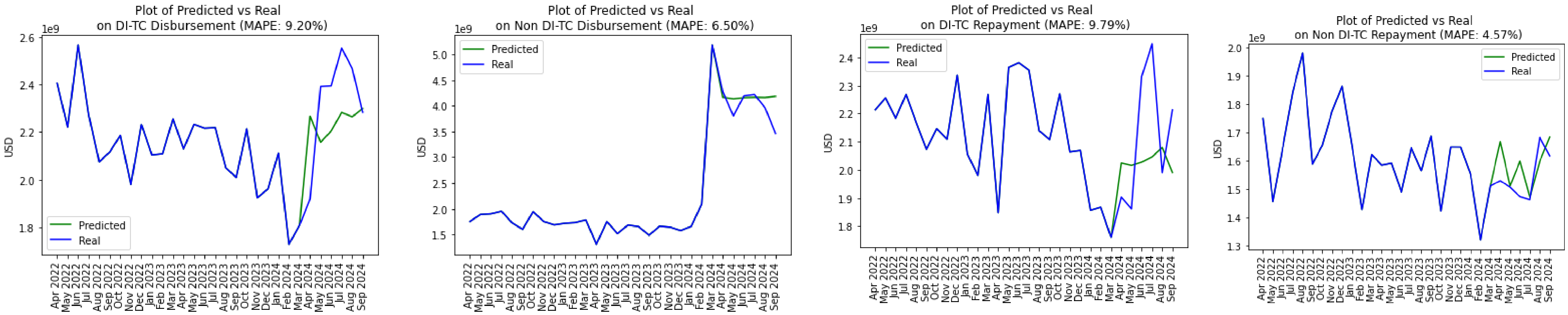
Five clusters were generated for DI-TC Disbursement, each with distinct pattern characteristics:

1. Cluster 0, exhibits a simple pattern where most values remain flat with a significant spike toward the end.
2. Cluster 1 displays a more complex and fluctuating pattern, with values rising and falling variably over time.
3. Cluster 2, shows highly volatile and seemingly random patterns, indicating significant variation within this cluster.
4. Cluster 3 stands out with its unique pattern, where values remain stable for most of the period before experiencing a sharp increase at the end.
5. Cluster 4 demonstrates relatively more organized fluctuations compared to Cluster 2, although variations are still present.

Result : Projection on Trade Credit

4. Data Modelling and Evaluation

Next, data modelling is performed for all cluster based on the results from the previous steps. Next, we combined result form each cluster for: 1) DI TC Disbursement; 2) Non-DI TC Disbursement; 3) DI TC Repayment; and 4) Non-DI TC Repayment.



Result of Projection on Trade Credit

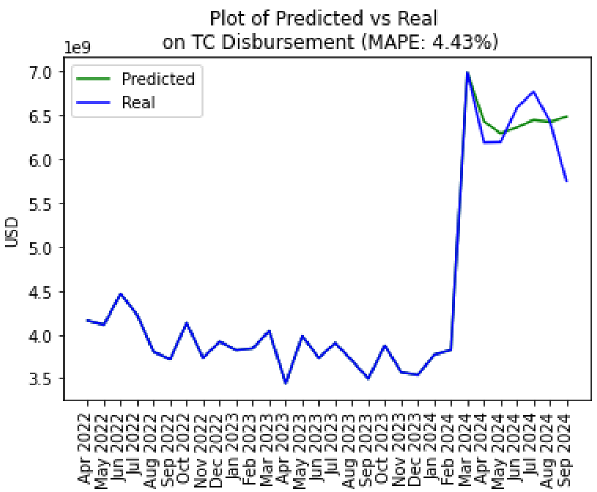
Final step is evaluation, this process use mean absolute percentage error (MAPE). The result is as follows:

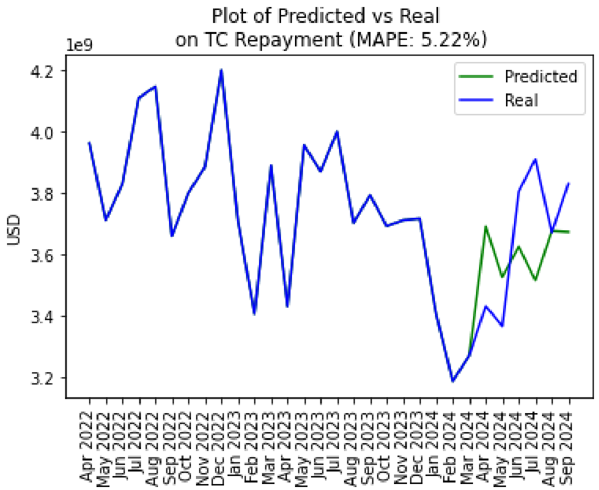
No	Loan Type	MAPE score	Description
1	DI TC Disbursement	9,20%	<10%, highly accurate forecasting
2	Non-DI TC Disbursement	6,50%	<10%, highly accurate forecasting
3	DI TC Repayment	9,79%	<10%, highly accurate forecasting
4	Non-DI TC Repayment	4,57%	<10%, highly accurate forecasting

Result : Projection on Trade Credit

5. Data Modelling and Evaluation

Furthermore, we tried to combined result form DI and Non-DI trade credit, both disbursement and repayment.





Result of Projection on Trade Credit

The combined result give better evaluation score. The evaluation use mean absolute percentage error (MAPE), as follows:

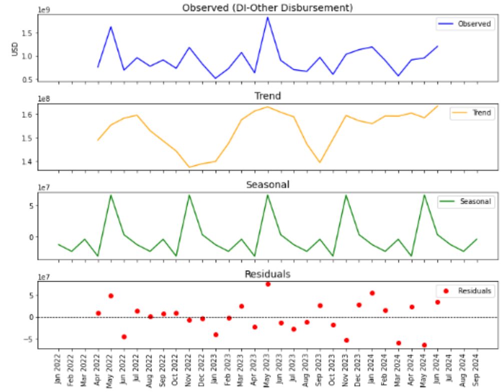
No	Loan Type	MAPE score	Description
1	TC Disbursement	4,43%	<10%, highly accurate forecasting
2	TC Repayment	5,22%	<10%, highly accurate forecasting

Result : Projection on Other Loan

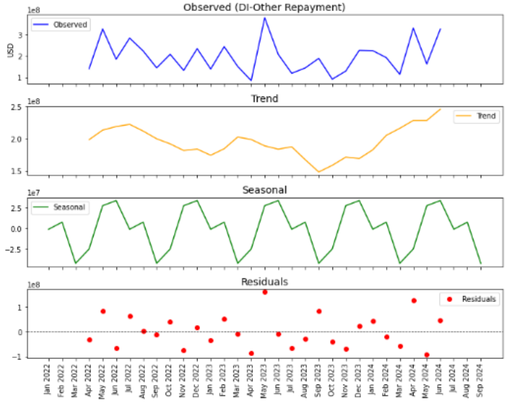
1. Data Analysis

For the other loan, we use decomposition and time-series analysis. Clustering is not used in this type of loan, because the behaviour of company in disbursement and repayment external debt often change over time. The first step is to identify if the projection from company reporting is correct. We found that for all company that have other loan external debt, the projection reporting is not reliable.

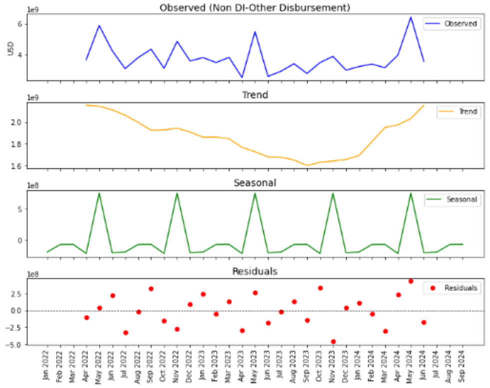
2. Data Preprocessing



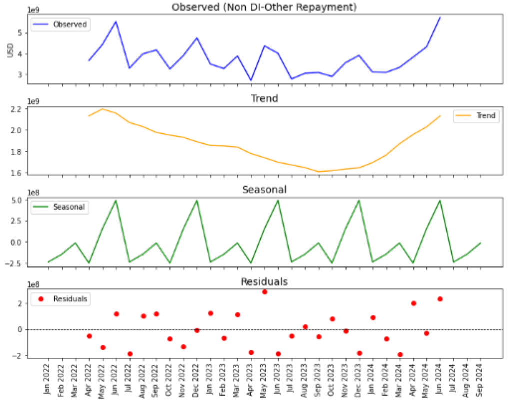
Decomposition results of DI Other Loan Disbursement



Decomposition results of DI Other Loan Repayment



Decomposition results of Non-DI Other Loan Disbursement



Decomposition results of Non-DI Other Loan Repayment

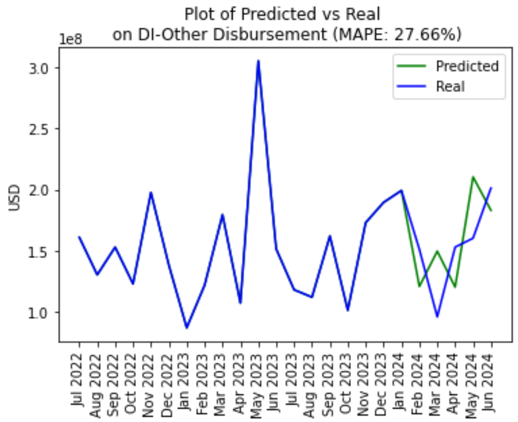
For each other loan type component (DI disbursement, DI repayment, non-DI disbursement and non-DI repayment), we formed 3 models which is trend, seasonal, and residual. The trend and residual components will be modelled separately, while the seasonal component will be studied for its seasonal pattern and directly added to the predictions of the trend and residual components. This approach is taken because the seasonal patterns formed through additive decomposition are already clear.

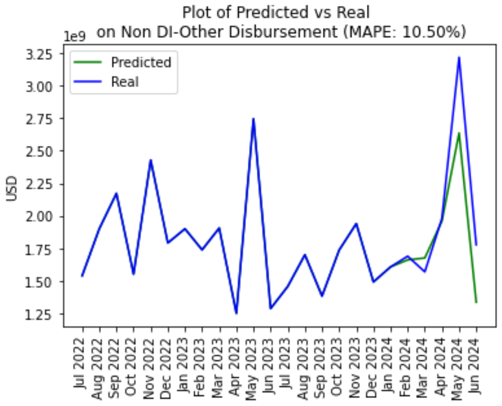
Example of Selecting the Optimal Number of Clusters for DI-TC Disbursement

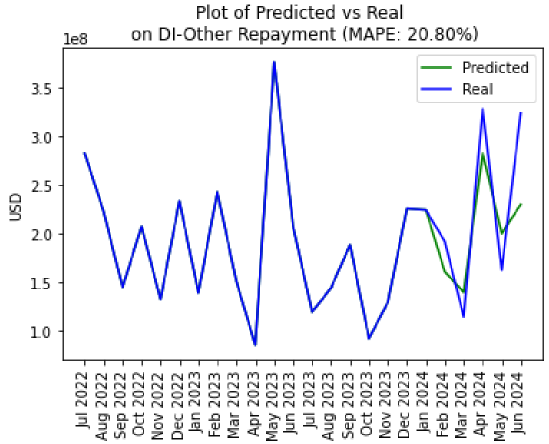
Result : Projection on Other Loan

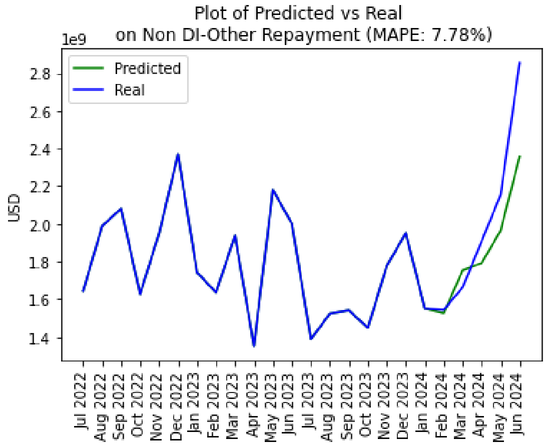
3. Data Modelling and Evaluation

Next, data modelling is performed for all category based on the results from the previous steps. Next, we combined result form each cluster for: 1) DI OL Disbursement; 2) Non-DI OL Disbursement; 3) DI OL Repayment; and 4) Non-DI OL Repayment.









Result of Projection on Trade Credit

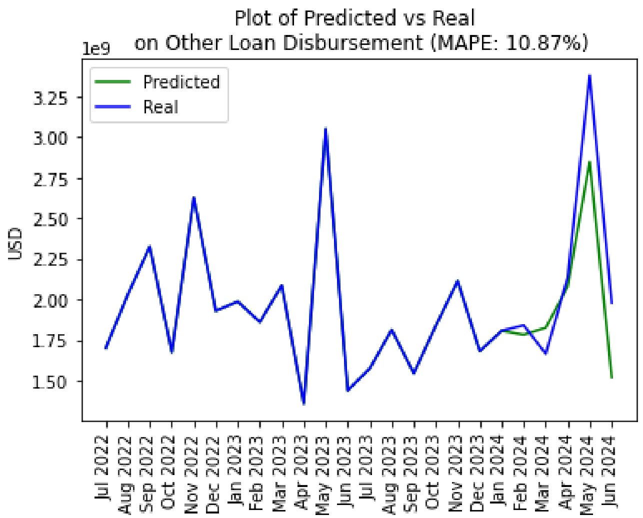
Final step is evaluation, this process use mean absolute percentage error (MAPE). The result is as follows:

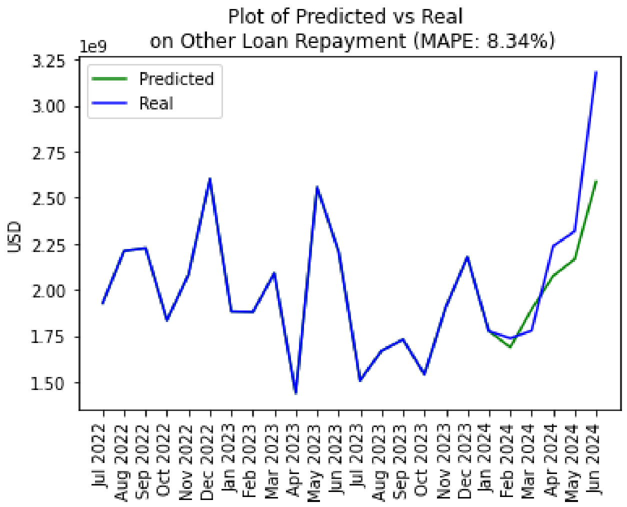
No	Loan Type	MAPE score	Description
1	DI OL Disbursement	27,66%	20-50%, Reseonable forecasting
2	Non-DI OL Disbursement	10,50%	10-20%, Good forecasting
3	DI OL Repayment	20,8%	20-50%, Reasonable forecasting
4	Non-DI OL Repayment	7,78%	<10%, highly accurate forecasting

Result : Projection on Other Loan

4. Data Modelling and Evaluation

Furthermore, we tried to combined result form DI and Non-DI other loan, both disbursement and repayment.





Result of Projection on Other Loan

The combined result give better evaluation score. The evaluation use mean absolute percentage error (MAPE), as follows:

No	Loan Type	MAPE score	Description
1	OL Disbursement	10,87%	10-20%, Good forecasting
2	OL Repayment	8,34%	<10%, highly accurate forecasting

Conclusion

- Advanced analytics, specifically clustering-decomposition and time series methodology **can be used for projecting** private external debt (trade credit and other loan) disbursement and repayment.

- Implementing clustering-decomposition and time series methodology increase **the reliability of projection**. This proves by projection evaluation using Mean Absolute Percentage Error (MAPE).

- Implementing advanced analytics for projection private external debt, **can increase work process efficiency**. This achieved by reducing dependencies to company reporting data, thus eliminating coordination process such as the data confirmation

THANK YOU