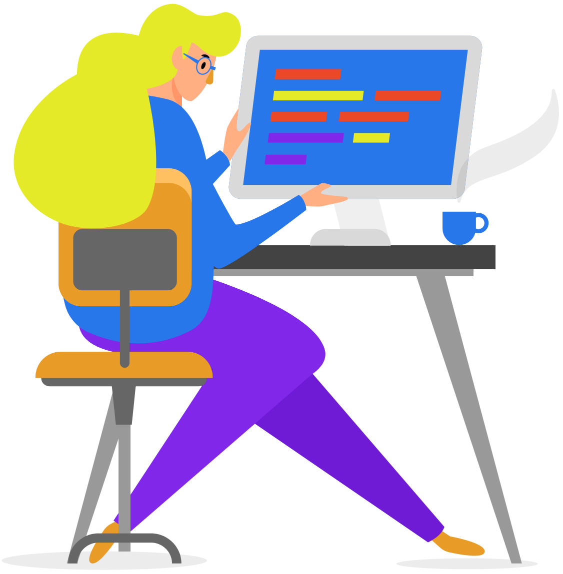
Nowcasting GDP with Machine Learning:

The Case of Indonesia

Ginanjar Utama
Nadira Firinda

Outline



01 Introduction

02 Literature Review

- The Concept of Machine Learning
- Nowcasting GDP Using Machine Learning

03 Data and Methodology

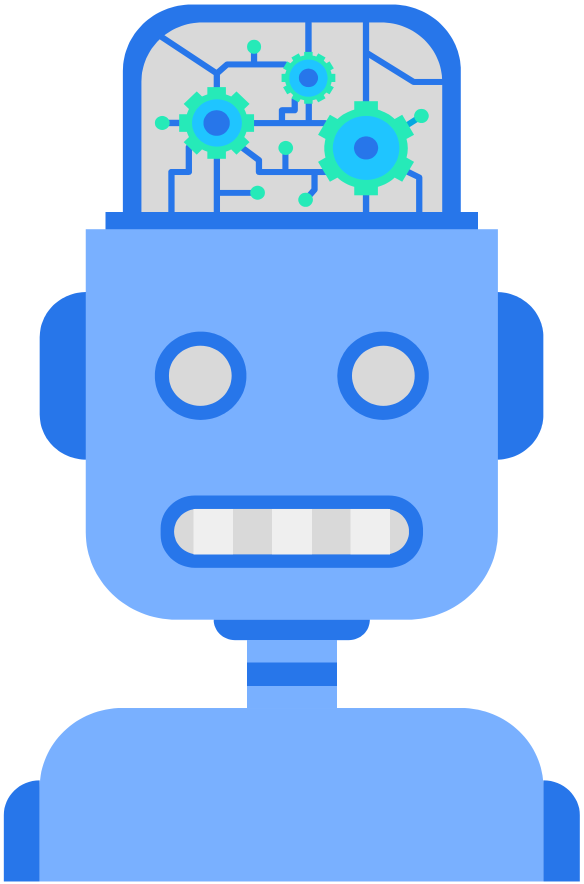
- Data
- Methodology (Variable Selection Criteria & The Models)

04 Results and Discussion

- Evaluation of Model Accuracy
- Nowcasting result
- Model Explanation

05 Conclusion

- Conclusion
- Follow-Up(s)



01 Introduction

Introduction

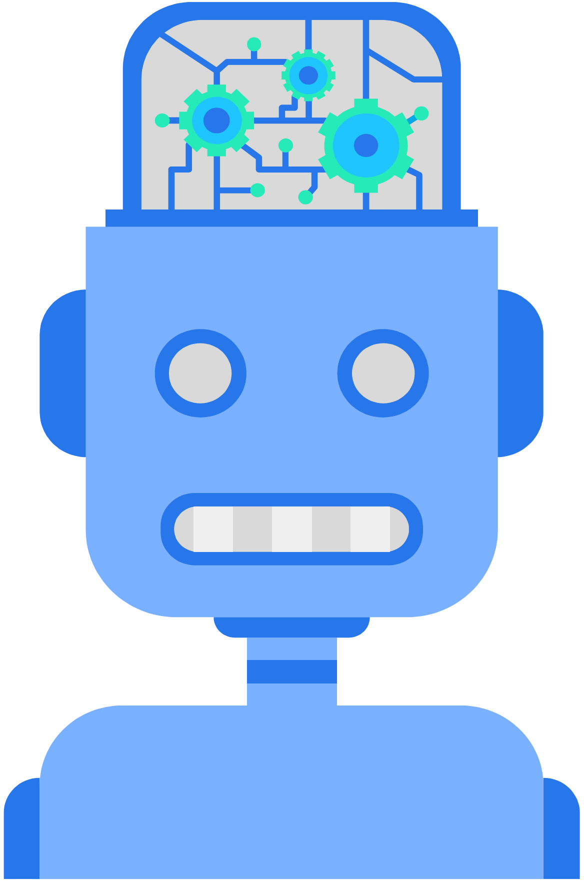
Economic projections, particularly for variables such as GDP, are crucial for policymakers

And yet, factors outside of the government's control, such as pandemics, can have a significant impact on a country's economic growth and make it difficult to accurately predict future trends

This paper describes an effort to develop nowcasting capacity in projecting Indonesia's GDP.

Using

- Various kind of macro-economy and financial data
- Certain variable selection criterias to determine economic indicators that have a strong relationship with GDP
- Four (4) different method of machine learning



02

Literature Review

The Concept of Machine Learning

Nowcasting GDP Using Machine Learning

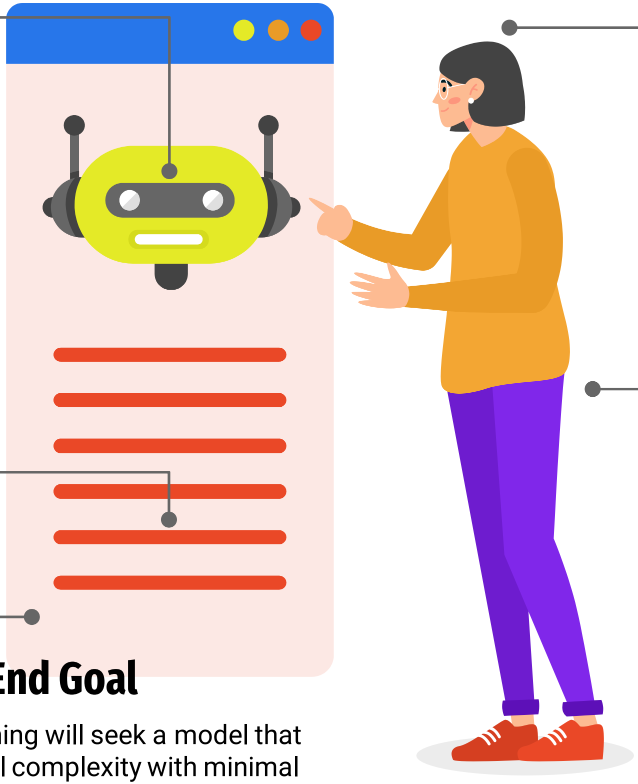
The Concept of Machine Learning

01 Machine Learning

- subfield of artificial intelligence that develops method for teaching machines how to perform tasks by recognizing patterns and relationships
- Its technique divided into 2 main types: Supervised and Unsupervised

02 Supervised Learning

- **Goal:** learning a function that maps the input variables to the output variable
- **Data:** Labeled
- **Example:** Classification, Regression



05 End Goal

Machine Learning will seek a model that has an optimal complexity with minimal total error that comes from variance or bias

Unsupervised Learning 03

- **Goal:** finding patterns or structure in the data without the use of labeled examples
- **Data:** Not Labeled
- **Example:** Clustering

Clustering 04

grouping similar observations together based on their features or characteristics

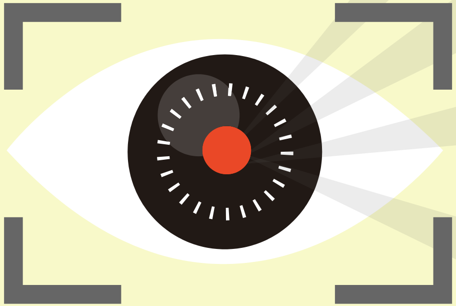
Bias-Variance Trade Off

- **Bias** → error that results from inaccurate assumptions related to data
- **Variance** → error that results from model sensitivity due to changes in the underlying data.

Bias-Variance Trade Off → the more complex the model, the higher the variance and the lower the bias (vice versa)

Nowcasting GDP Using Machine Learning

Some machine learning applications in nowcasting GDP



Tiffin, 2016

Nowcasting of **Lebanese GDP** using the **elastic net and random forest** methods. Elastic net method is better than the random forest method, while them combined provided the best accuracy.

Richardson, 2018

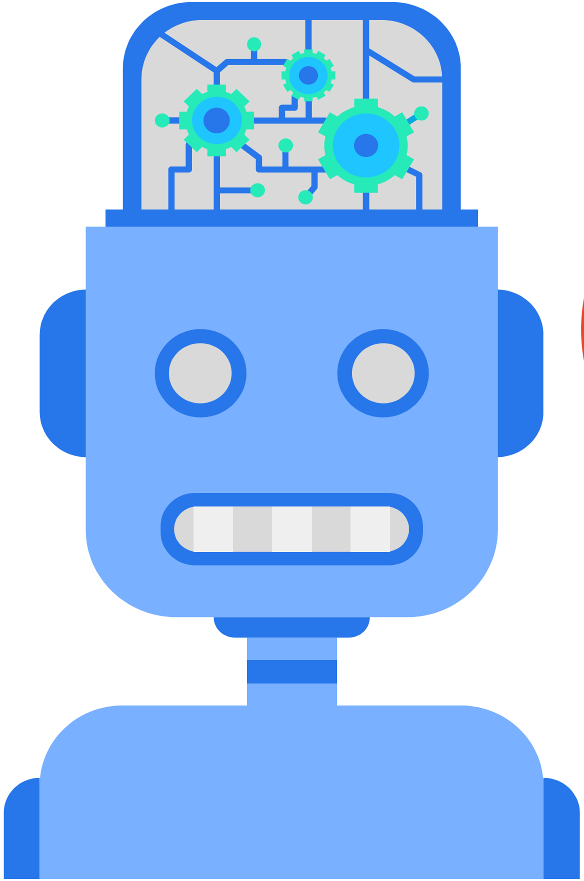
Nowcasting of **New Zealand GDP**. The results show that the **individual machine learning model is better than AR(1)**, and the combination model predicts even better

Bulhois, 2020

Used **Random Forest, Gradient Boosted Trees, Support Vector Machine, and the combined version** of those methods to nowcast **Turkey's GDP**. The ensembled forecast method was better than the three methods individually, and the individual model itself was better than DFM.

Hopp, 2022

Examines the performance of **12 different methodologies in nowcasting US GDP growth**. The long short-term memory neural networks (LSTM) and Bayesian vector auto regressian (BVAR) found to be the best.



03

Data & Methodology

Data

Methodology

Variable Selection Criteria

Machine Learning Methods

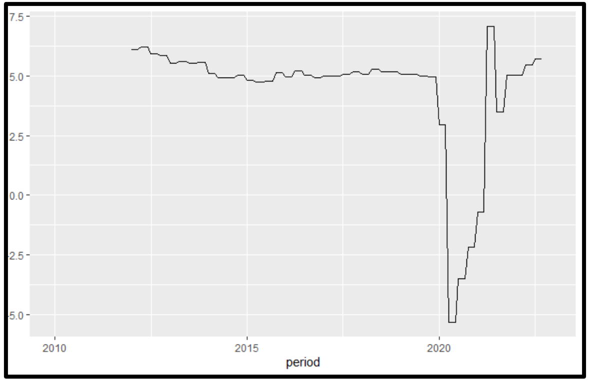
Data

Estimation Data

- Monthly indicator data from 2010 to 2022
- 34 variabel/indicators transformed into Y-o-Y percentage
- Data will have 1 month lag
  (Nowcasting GDP in October will use data indicators up until September and so on)
- **Estimation will be divided into two parts:** in-sample and out-of-sample projections.
  - **In-sample: uses training data**
  (January 2013 – December 2021)
  - **Out-sample: uses testing data**
  (January 2022 – September 2022)

GDP of Indonesia

Quite stable at the range of 5,0 to 7,0 (Y-o-Y % Change) from 2010 before dropping to between -5,0 and -0,6 following Covid-19



Methodology

Variable Selection with clustering based on correlation

01

Determine the relationship between variables by creating a correlation table with six methods, specifically:

1. Maximal Information Coefficient (MIC)
2. Maximal Normalized Mutual Information (Max NMI)
3. Kendall Correlation
4. Spearman Correlation
5. Pearson Correlation
6. Distance Correlation

Selecting variables related to GDP using the clustering method (normal mixture modeling) using correlation mentioned above.

Projection Using Machine Learning Method

02

The projection uses 4 alternative standard machine learning methods, namely:

1. Elastic Net
2. Random Forest
3. XGBoost
4. Support Vector Machine

Exploration of Projected Results

03

- ✓ **Study the projection results through the in-sample and out-sample data.**
- ✓ **The data is divided into two parts to produce in-sample and out-sample projections**

1. In-sample projection will be using training data
2. Out-sample projection will be using testing data

- ✓ **The amount of data on the in-sample side tends to be more than the projection on the out-sample side. This analysis can provide an additional picture of the sensitivity of the training data to the projection results.**
- ✓ **Comparing the RMSE from the projected in-sample and out-sample data for each method in order to find the best GDP Nowcasting model**

Variable Selection Criteria

Association/ Relationship Measures

01 Maximal Information Coefficient (MIC)

capturing functional and non-functional relationships between variables. For functional relationships, can be associated with R2 (coefficient of determination)

02 Maximal Normalized Mutual Information (Max NMI)

- measuring the relationship between two random variables that are sampled simultaneously. Specifically, MI measures how much information is communicated from one variable to another.
- NMI is a measure of MI that is normalized so that it is at a value between 0 and 1

03

04 Kendall & Spearman Correlation

Non-parametric test to measure the strength of dependency between two variables

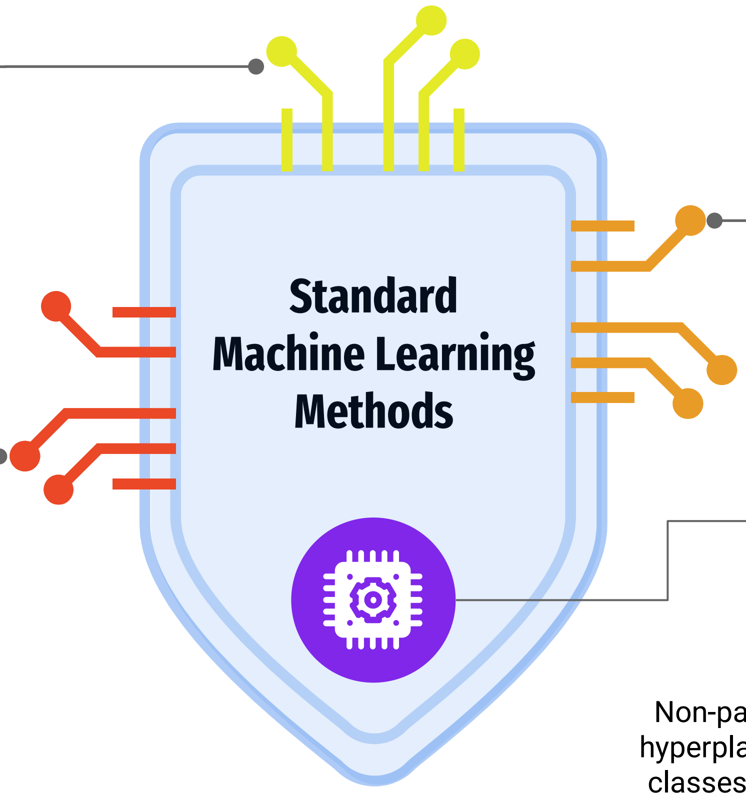
05 Pearson Correlation

Parametric test to measure the strength of a linear relationship between variables (may not be appropriate for data that are not normally distributed). variables must be normally distributed, linear, and homoscedastic

06 Distance Correlation

measuring the relationship between non-linear random variables. It can spot more than linear associations and work multi-dimensionally

Machine Learning Methods



01

Elastic Net

A form of penalized regression by combining dimension reduction and variable selection in one step

02

Random Forest

Combines multiple decision trees with each decision tree is trained on a random subset of the training data and a random subset of the predictor variables

03

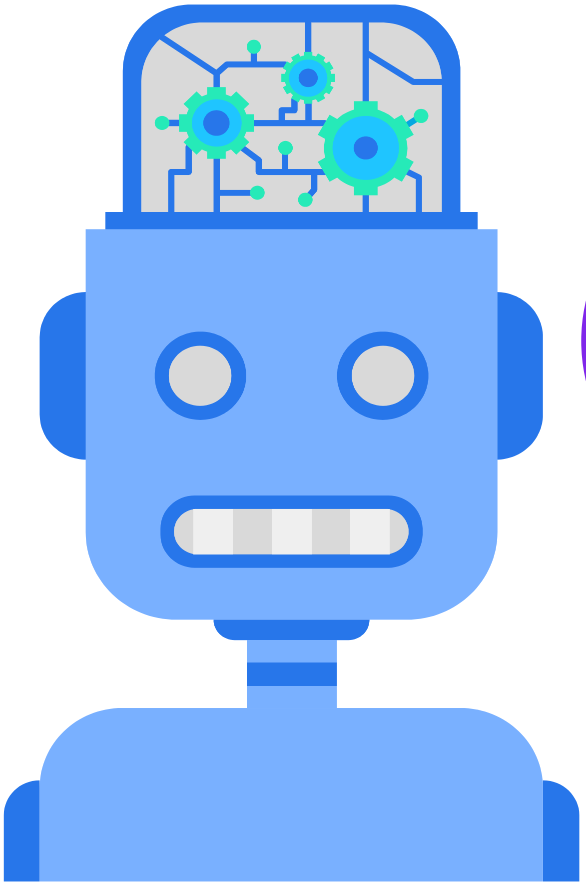
XGBoost

Similar to RF, with each decision tree is trained to correct the errors of the previous trees, in order to improve the overall accuracy of the ensemble

04

Support Vector Machine

Non-parametric technique that aims to identify hyperplanes that maximize the distance between classes in the feature space while ensuring that the perpendicular distance between two adjacent points of different classes is maximized



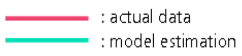
04

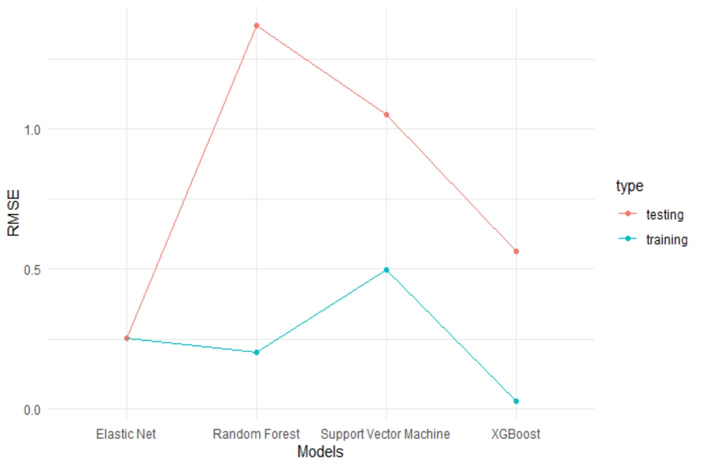
Results

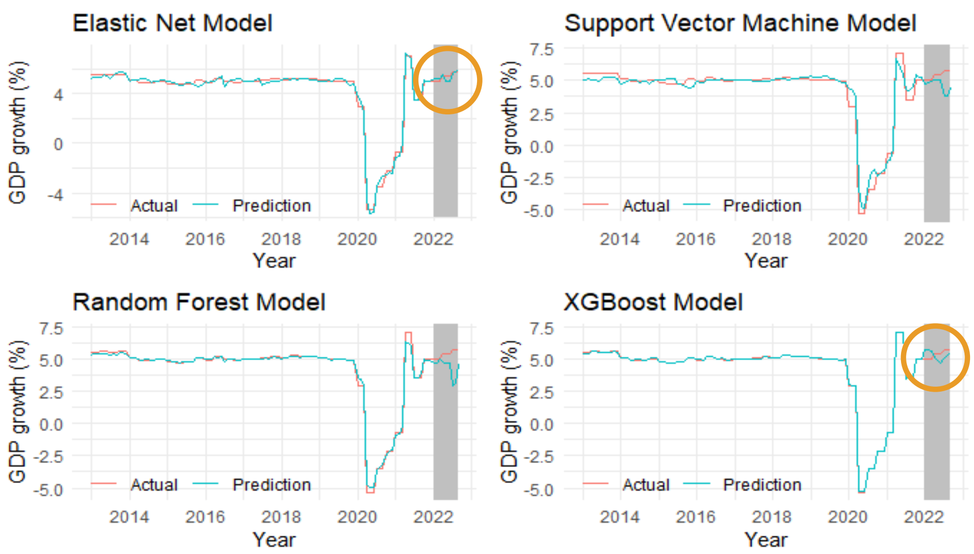
Results

Discussions

Evaluation of Model Accuracy







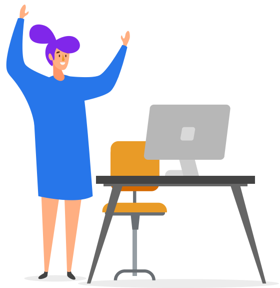
RMSE Value for Training and Testing Data

Model	Training	Testing	Difference
Elastic Net	0.2541254	0.2531369	0.001
Random Forest	0.2041514	1.3689247	1.165
XGBoost	0.0296590	0.5628105	0.053
Support Vector Machine (SVM)	0.4971251	1.0499439	0.522

Elastic Net and XGBoost

Are proven as the best models to nowcast GDP. This is because they:

- Have the smallest difference in RMSE values for training and testing data
- Have their actual data and model estimation graphs drawn close to each others



Nowcasting Result (2022Q4)

Indicator Data Periode	Elastic Net	Random Forest	XGBoost	SVM	Average Ensembled
October	5.384883	4.766341	4.868336	4.974105	4.998416
November	5.292687	4.854379	4.689212	4.965604	4.950471
December	5.188386	4.898498	4.886504	4.948184	4.980393
Average	5.288652	4.839739	4.814684	4.962631	4.976427

Key Findings:

01

Selected variables from Variable Selection Criteria include

- Consumer Confidence Index
- Index of Current Income
- Purchase of Durable Goods Index
- Current Economy Index
- PMI
- Information & Communication equipment
- Food & Beverages
- Vehicle Parts
- Clothing Goods
- Other Goods

02

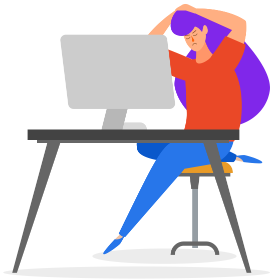
Elastic Net and XGBoost provide the best performance in projecting Indonesia's GDP

Compared to SVM and Random Forest

03

The average projection of GDP in Indonesia over the last quarter of 2022 (2022 Q4)

The averafe projection of GDP from Elastic Net and XGBoost are 5,29 and 4,81. While the average projection that ensembles multiple results from the models is 4,98.



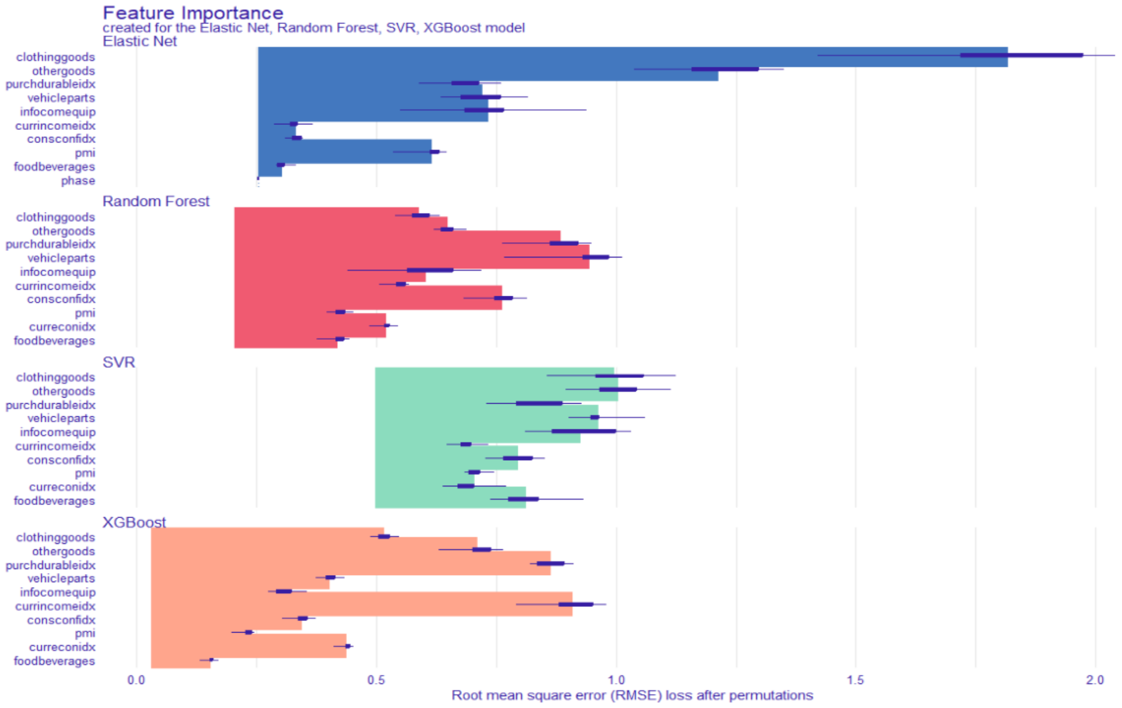
Model Explanation – Feature Importance

Analyzing the importance of predictor variables in model

If a predictor variable is important for projections, then it should have a noticeable effect in model performance..

The GDP nowcasting results are most influenced by the consumption aspect; Clothing and Other Goods.
Some of the variables that also have a significant role in GDP nowcasting including:

- Purchase of Durable Goods Index
- Information and Communication equipment
- Index of Current Income



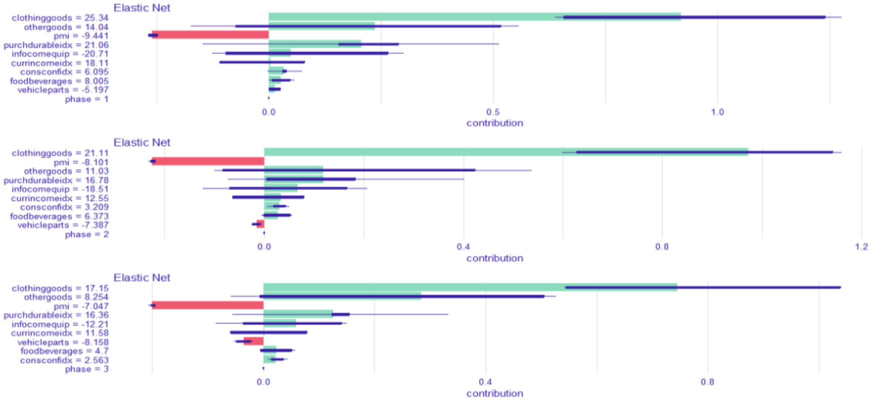
Model Explanation – Shapley Value

Analyzing the value of a variable's attribution over possible orderings

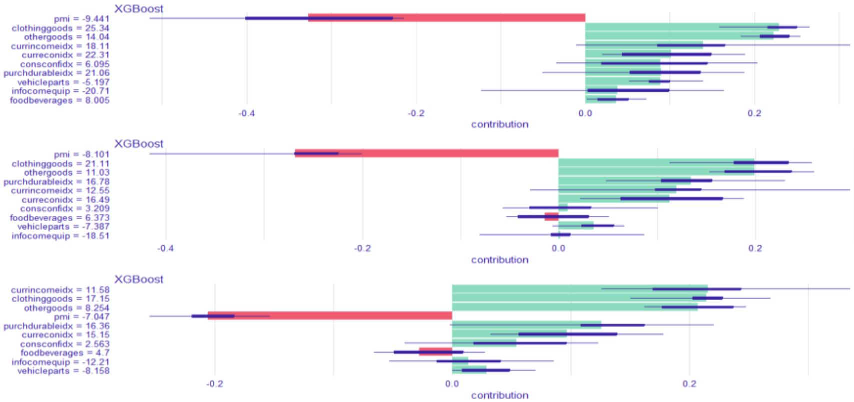
using the data between October to December 2022, some particular variables have constant importance in order, either derived from Elastic Net or XGBoost

- Clothing Goods
- Other Goods
- Purchase of Durable Goods Index
- Information and Communication equipment
- Index of Current Income

Meanwhile, PMI is inspected to have a negative contribution in nowcasting GDP compared to other variables



Elastic Net

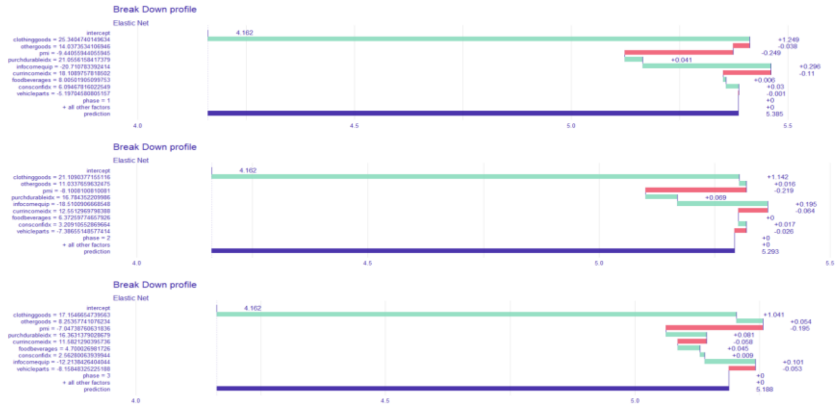


XGBoost

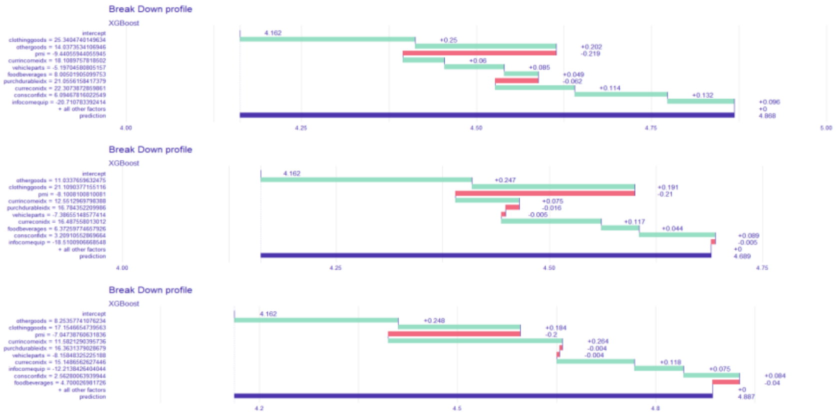
Model Explanation – Breakdown Profile

Analyzing which variables contribute most to the result

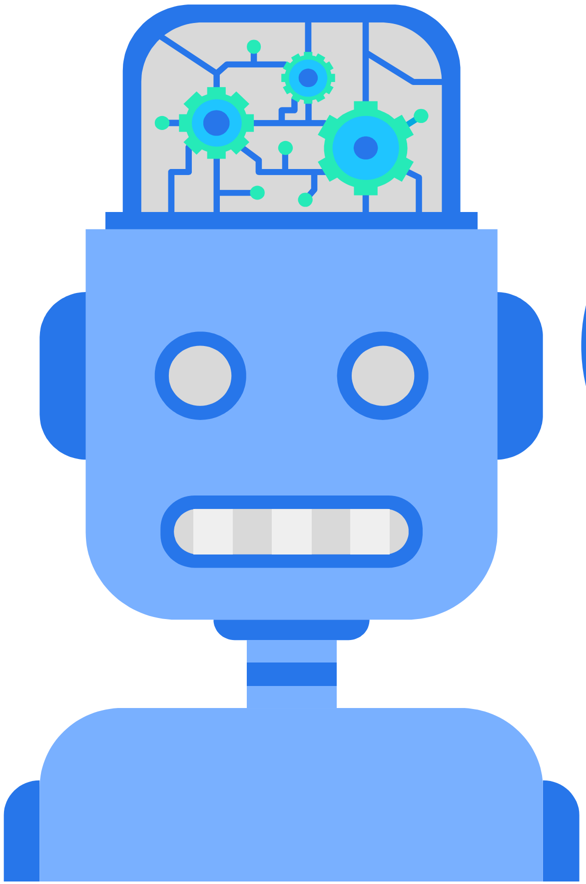
- **intercept → the mean value that estimate of the expected value of the model's predictions over the distribution of all variables**
- **The green and red bars → positive and negative changes in the mean predictions**
  - Elastic Net
    Clothing Goods and Purchase of Durable Goods Index as variables with highest contribution
  - XGBoost
    Clothing Goods and Other Goods as variables with highest contribution



Elastic Net



XGBoost



05 Conclusion

Conclusion and Follow-Up

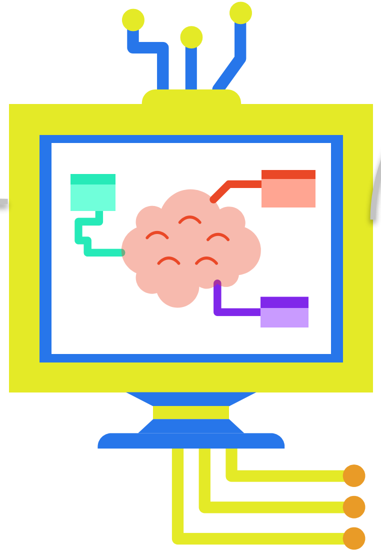
Conclusion

- **GDP growth projection for 2022 Q4 in Indonesia**

  Ranges from 4.81% to 5.29%, with an average of 4.98% using an ensemble method

- **RMSE**
  - Training Data (2013M01 – 2021M012) 0.03 to 0.49
  - Testing Data (2022M01 – 2022M09) : 0.25 to 1.36

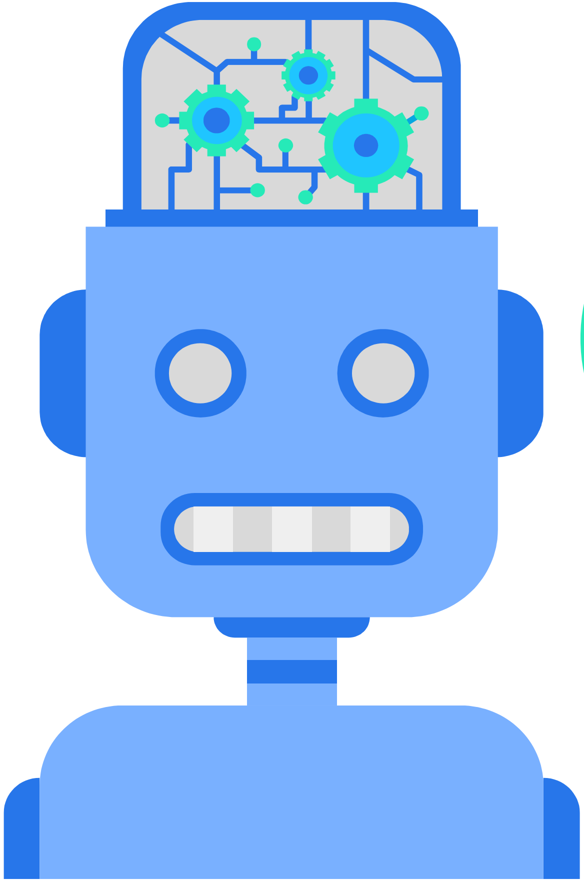
- **Variables that have significant contributions**
  - Current Incomes Index, Consumer Confidence Index, Purchase of Durable Goods Index
  - Real retail sales indicators; Information and Communication equipment, Food and Beverages, Vehicle Parts, Clothing, and Other Goods



Follow-Up

What do we need to develop this research to its maximum capacity?

- **Continual evaluation and development of models**
- **Four areas of Responsible AI principles**
  - internal governance structures and measures
  - determining the level of human involvement in AI-augmented decision-making
  - operations management
  - stakeholder interaction and communication



06 References

References

- Biecek, P. & Burzykowski, T. (2021). Explanatory Model Analysis: Explore, Explain, and Examine Predictive Models. Chapman and Hall/CRC, New York
- Biecek, P (2018), DALEX: Explainers for Complex Predictive Models in R. Journal of Machine Learning Research 19(84):1−5, 2018.
- Bolhuis, M., & Rayner, B. (2020). Deus Ex Machina? A Framework for Macro Forecasting with Machine Learning. IMF Working Papers, WP/20/45. International Monetary Fund.
- Dauphin, Jean-Francois and Dybczak, Kamil and Maneely, Morgan and Taheri Sanjani, Marzie and Suphaphiphat, Nujin and Wang, Yifei and Zhang, Hanqi (2022). Nowcasting GDP - A Scalable Approach Using DFM, Machine Learning and Novel Data, Applied to European Economies. IMF Working Paper No. 2022/052
- Hall, A. S. (2018). Machine Learning Approaches to Macroeconomic Forecasting. The Federal Reserve Bank of Kansas City Economic Review, 103(63), 2.
- Hopp, D (2022). Benchmarking econometric and machine learning methodologies in nowcasting. UNCTAD Research Paper No. 83
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An Introduction to Statistical Learning. New York: Springer.
- Jung, J. K., Patnam, M., & Ter-Martirosyan, A. (2018). An Algorithmic Crystal Ball: Forecasts Based on Machine Learning. IMF Working Paper 18/230.
- Kuhn, M, and K Johnson. 2013. Applied Predictive Modeling. Springer.
- Kuhn, M, and K Johnson. 2020. Feature Engineering and Selection: A Practical Approach for Predictive Models. CRC Press.
- Kuhn, M, and Silge, J. 2022.Tidy Modeling with R: A Framework for Modeling in the Tidyverse, O'Reilly Media, Inc.
- Loermann, J. & Maas, B. (2019). Nowcasting US GDP with Artificial Neural Networks. MPRA Paper 95459, University Library of Munich, Germany.
- Molnar, C. (2021). Interpretable Machine Learning: A Guide for Making Black Box Models Explainable. https://christophm.github.io/interpretable-ml-book/
- Richardson, A., & Mulder, T. (2018). Nowcasting New Zealand GDP Using Machine Learning Algorithms,” Mimeo.
- Richardson, A., Mulder, T., Vehbi, T. (2019). Nowcasting GDP using Machine Learning Algorithms: A Real-Time Assessment. Discussion Paper, 2019-03. Reserve Bank of New Zeland.
- Page, S.E. (2010). Diversity and Complexity (Primers in Complex Systems). Princeton University Press
- Tiffin, A. (2016). Seeing in the Dark: A Machine-Learning Approach to Nowcasting in Lebanon. IMF Working Paper 16/56.