ATM Cash Withdrawal Forecasting using Calendar Effect and Model Combination

dilan

Muhammad Rasyid Ridha

rasyidstat@gmail.com

Business Intelligence Analyst at GO-JEK Indonesia

ATM CASH OPTIMIZATION

FINHACKS 2018
#DATACHALLENGE

14·11·2018

Overview

Goal

Predict withdrawal amount from 10,626 ATMs for the next 7 days (25 - 31 March 2018)

Data

- Data training: 83 days (1 January - 24 March 2018)
- Data testing: 7 days (25 - 31 March 2018)
- In total: 881,816 rows for training and 74,832 rows for testing
- 16 variables (only withdrawal will be used, univariate time series)
- Data checking: 9 ATMs have missing data
- Data checking: 2001 ATMs (18.83%) have at least one zero data

Evaluation Metrics

- Percent of error below 25% (main focus)
- MAPE
- R-Squared

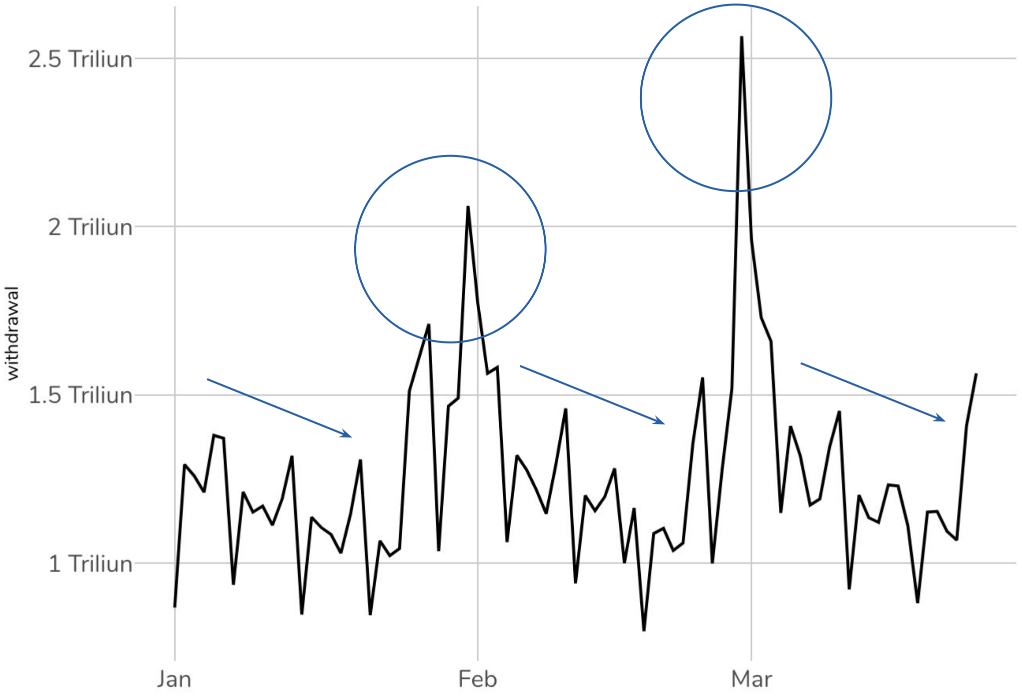
01 Exploratory Data Analysis

General pattern, seasonality, payday date, time series clustering

Exploratory Data Analysis

General Pattern

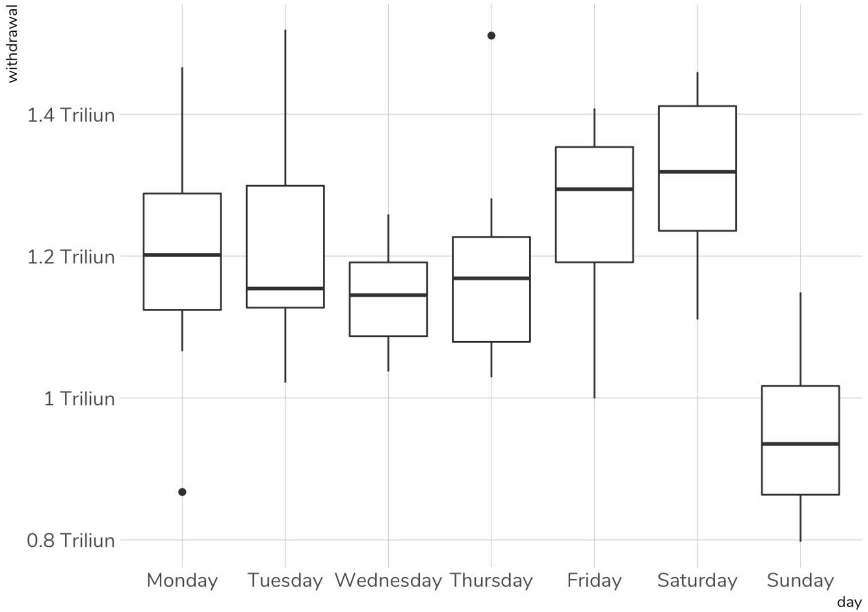
Using aggregated time series from 10,626 ATMs

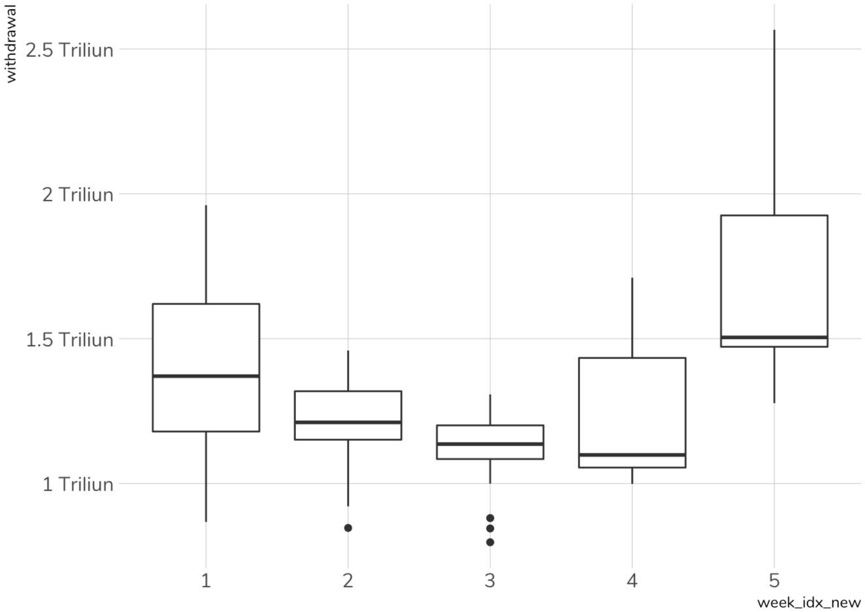


- Monthly seasonality: payday effect
- Monthly seasonality: Decreasing trend over week of the month
- Overall trend is hard to caught since the data is very small
- Weekly seasonality (need to zoom in the data)

Exploratory Data Analysis

Seasonality Exploration





Day-of-the-week

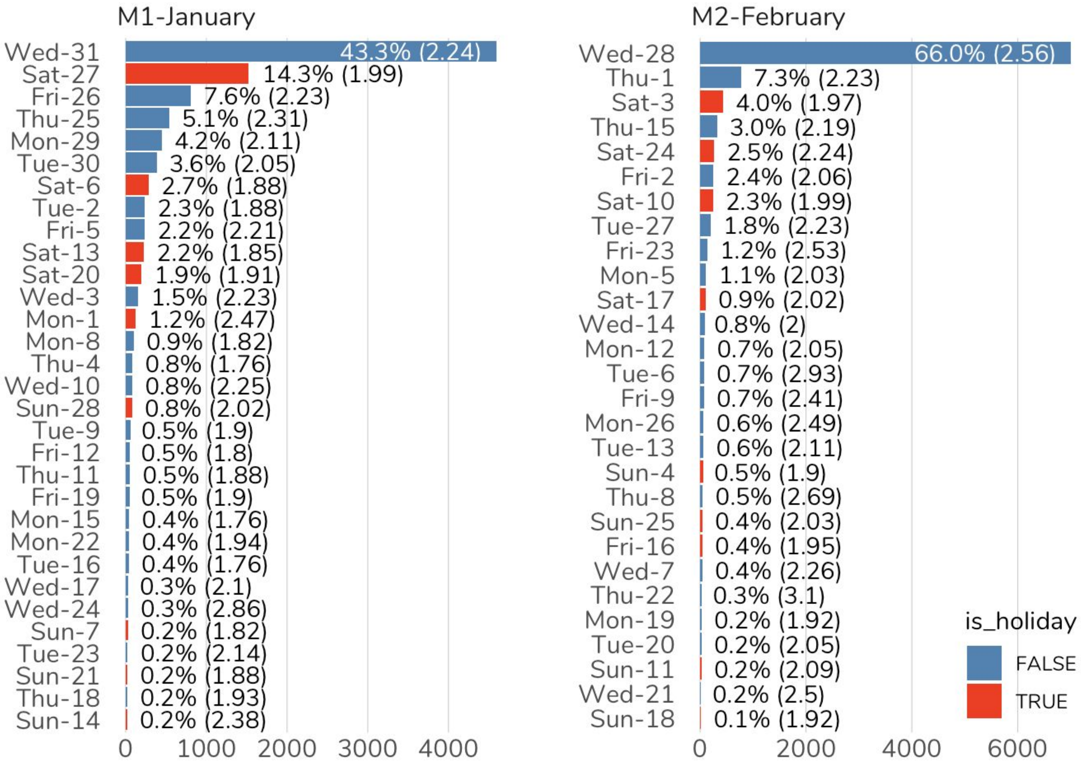
Saturday is the highest,
Sunday is the lowest

Week-of-the-month

Last week-of-the-month is the highest followed by the first week-of-the-month where payday usually occurs, Mid week-of-the-month is the lowest

Exploratory Data Analysis

When Payday Usually Occurs?



For every 10,626 ATMs, withdrawal peak in January and February are identified

- End of month (31 Jan, 28 Feb)
- First of month (2 Jan, 5 Jan, 1 Feb, 5 Feb)
- Proportion of withdrawal peak in 28 Feb is far higher than 31 Jan because there is no 29,30,31 in February (it might accumulate to 28 Feb)
- 1 Jan is not the peak like 1 Feb due to holiday (New Year)
- Saturday in the last week or first week of month is also high

Exploratory Data Analysis

Time Series Clustering

- **Cluster 1** (4745 ATMs)
  Low withdrawal amount
- **Cluster 2** (372 ATMs)
  Inactive during weekend, containing many zero value (most likely ATM in CBD area)
- **Cluster 3** (2137 ATMs)
  High withdrawal amount
- **Cluster 4** (3363 ATMs)
  High withdrawal during weekend

02 Pre-processing

Features engineering, payday effect, time series cross-validation

ATM CASH OPTIMIZATION

FINHACKS 2018 #DATACHALLENGE

Pre-processing

Features Engineering

Lag 7 days

Dummy variables / regressor

$$\underbrace{Y_t}_{} = \overbrace{Y_{t-7}}^{\text{Lag 7 days}} + \overbrace{\textit{Main Features} + \textit{Payday Effect}}^{\text{Dummy variables / regressor}}$$

- Week-of-the-month
- Day-of-the-week
- Public holiday
  - Mon, 1 January 2018 (New Year)
  - Fri, 16 February 2018 (Chinese New Year)
  - Sat, 17 March 2018 (Nyepi Day)
  - Fri, 30 March 2018 (Good Friday)

6 different

features engineering scenario

f5: payday at 1,25,28,31

f51: payday at 1,25,28,31 + pre and after effect during saturday

f52: payday at 1,2,5,25,26,27,28,29,30,31

f53: payday at 1,2,5,25,26,27,28,29,30,31 + pre and after effect during saturday

f54: payday at 1,25,28,31 + separated pre and after effect during saturday

f55: payday at 1,2,5,25,26,27,28,29,30,31 + separated pre and after effect during saturday

Total: 18-26 features

Pre-processing

Payday Effect in Indonesia

From previous EDA, peak of withdrawal (payday) of each month do not occurs in the same day-of-the-month.

It happens because when a payday falls in weekend/holiday, it moves to the previous working day

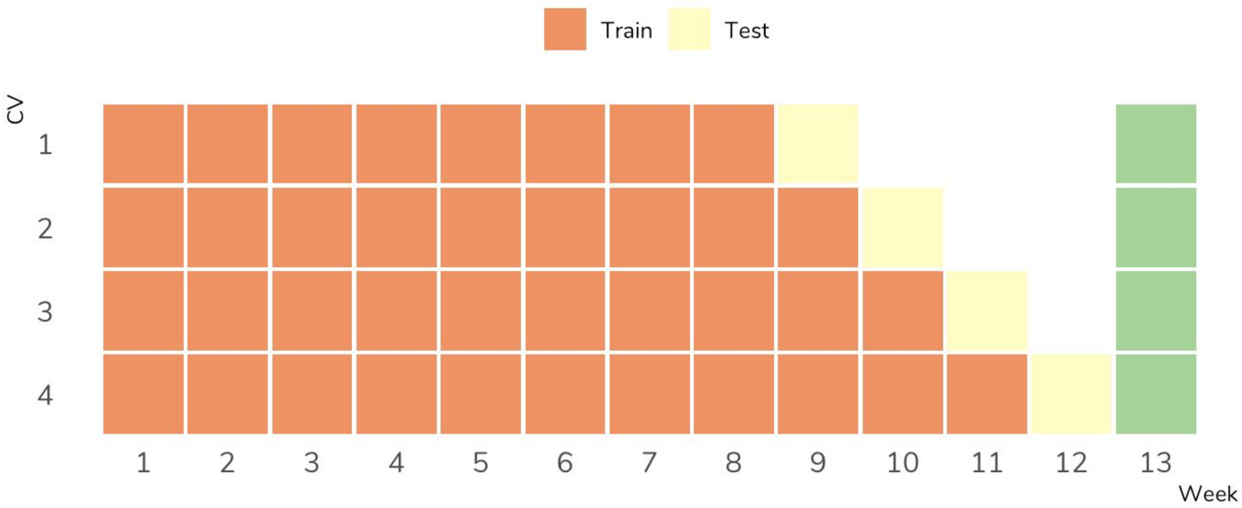
JAN 2018	23 Tue	24 Wed	25 Thu	26 Fri	27 Sat	28 Sun	29 Mon	30 Tue	31 Wed
FEB 2018	23 Fri	24 Sat	25 Sun	26 Mon	27 Tue	28 Wed	29 NA	30 NA	31 NA
MAR 2018	23 Fri	24 Sat	25 Sun	26 Mon	27 Tue	28 Wed	29 Thu	30 Fri	31 Sat

Test Set

Pre-processing

Time Series Cross-Validation

The method for time series cross validation is using ***expanding window*** approach where the 1st CV (data test: 25 Feb-3 Mar) used as the main focus since it will reflect the similar pattern with the purpose of the forecast (data test: 25 Mar-31 Mar). The rest CV is used for model stability checking



https://robjhyndman.com/hyndsight/tscv/
https://eng.uber.com/forecasting-introduction/

03 Modeling

Literature review, model framework, prediction result

Literature Review

Learning from the past winners of similar competitions

1. **Kaggle Web Traffic Time Series Forecasting** (2017)
   a. ~145K time series
   b. 1st winner use RNN
   c. 2nd winner use *model combination* between NN and XGBoost
   d. Simple median forecasting is a hard-to-beat benchmark

2. **M4 Forecasting Competition** (2018)
   a. ~100K time series
   b. 1st winner use hybrid model between RNN and ETS
   c. 2nd winner use *model combination* with XGBoost as the meta-learning classification
   d. Most of top models use *model combination* of statistical method

Model Framework

Single Model

- Naïve
  `naive()`
- Exponential Smoothing
  `ets()`
- ARIMA
  `auto.arima()`
- Linear Regression
  `lm()`
- XGBoost (learning rate = 0.01, 6 different features set)
  `xgboost()`

Train models for every ~10K ATMs

Iteration using `purrr`, functional programming in R

Model Combination

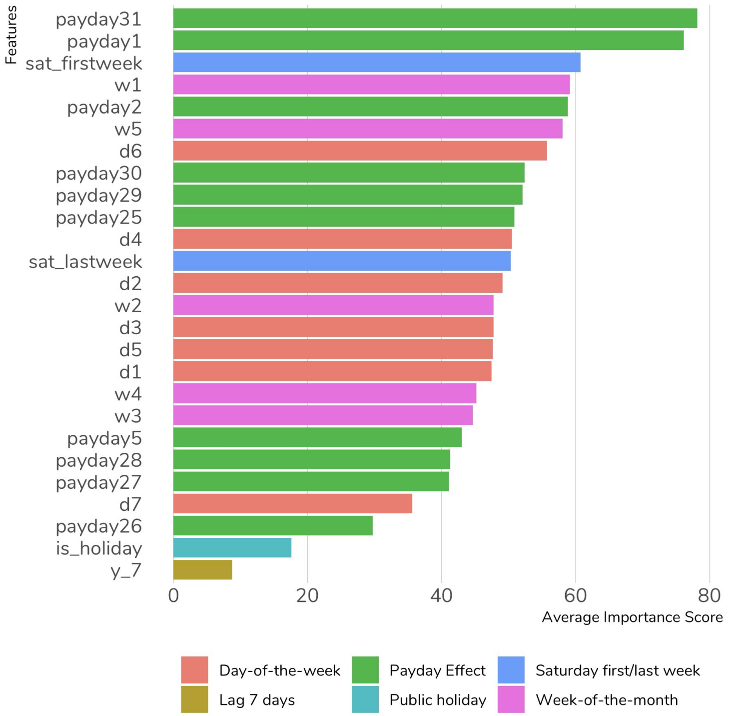
- Ensemble averaging
  Average prediction result from chosen models
- Best model selection
  (Bucket of models technique)
  Select best model for each ATM based on majority of best evaluation metrics in 4 cross-validation sets. If the result is tie, default model will be selected

? *The blue one is the models used in submission using best model selection*

Modeling

Prediction Result

Below is one of the example of aggregated prediction result and features importance using XGBoost model (f55)
The peak of the prediction falls on Thursday, 29 March 2018, not in the last date of the month, Saturday, 31 March 2018



04 Evaluation

Single model, correction, best model selection, bad prediction fixing

Evaluation

Single Model

With features engineering mentioned before, linear regression can give decent result. The best model is XGBoost which tends to not overfit compared to linear regression.

Model	PE25 Focus	PE25 Train	PE25 Test	PE25 st.dev.	sMAPE Train	sMAPE Test
naive	41.83%	47.62%	40.06%	2.23%	30.02%	43.01%
ets	39.52%	56.45%	46.27%	7.53%	35.53%	35.67%
regressionf5	**57.50%**	73.54%	61.20%	2.58%	20.83%	27.88%
xgboostv4f5	**59.06%**	70.33%	62.40%	2.44%	22.70%	27.00%
xgboostv4f52	**60.66%**	72.37%	63.00%	1.69%	21.78%	26.80%
xgboostv4f53	**60.68%**	73.33%	63.58%	2.71%	21.40%	26.52%
xgboostv4f55	**60.25%**	73.78%	63.45%	2.79%	21.19%	26.61%

ARIMA and regularized regression are also tried but none of them can outperform linear regression

?

PE25 Focus: percent of error below 25% using 1st CV (test data: 25 Feb-3 Mar)

PE25 Train/Test: average of percent of error below 25% using all CV

Prediction with Correction

Prediction with correction is multiplying prediction with a factor greater or lower than 1.

For focused CV, correction using 1.05 factor can help improve accuracy for XGBoost model with fewer features (f5 and f52). However, it does not improve much for XGBoost model with more features (f53 and f55).

Model	PE25 Focus	PE25 Train	PE25 Test	PE25 st.dev.	sMAPE Train	sMAPE Test
xgboostv4f5c5	**61.05%**	68.72%	61.87%	1.02%	22.97%	26.92%
xgboostv4f52c5	**61.45%**	70.76%	62.54%	0.84%	22.10%	26.70%
xgboostv4f53c5	**60.87%**	71.73%	63.08%	1.93%	21.69%	26.43%
xgboostv4f55c5	**60.60%**	72.15%	62.97%	1.92%	21.50%	26.51%

In overall CV, correction does not improve the accuracy. It only happens for 1st CV since there are many peak values in the test set where XGBoost model tends to underestimate.

? *Additional `c5` name in the model means that it uses 1.05 factor correction*

Evaluation

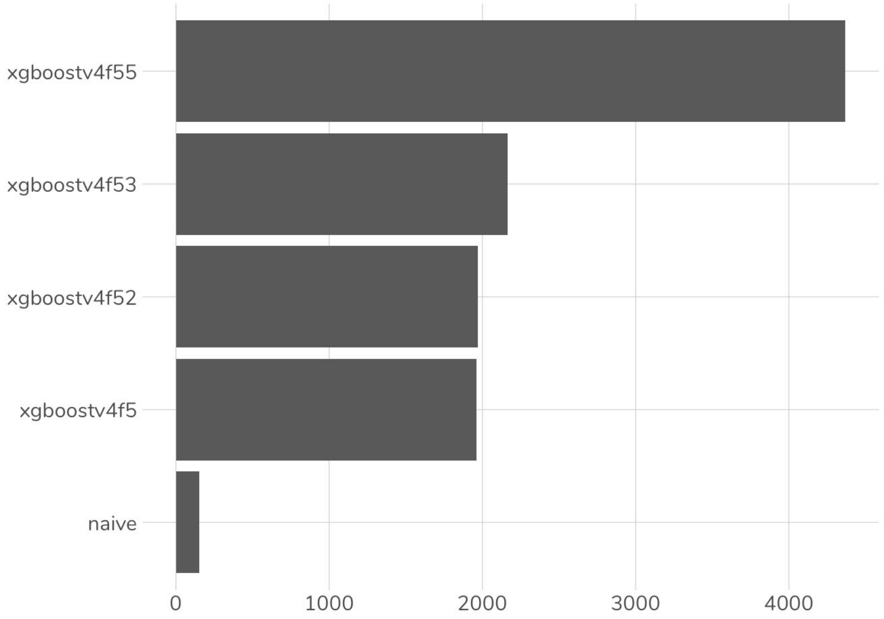
Best Model Selection

Pool models used for best model selection are Naive, XGBoost with features engineering: f5, f52, f53 and f55. The default model is XGBoost with features engineering f55.

For each ATM, one best single model will be selected from those 5 models.

CV	PE25 Train	PE25 Test	sMAPE Train	sMAPE Test
1	74.47%	**63.31%**	20.40%	27.17%
2	73.49%	64.57%	21.16%	25.76%
3	73.05%	65.17%	21.42%	26.32%
4	72.36%	67.26%	21.82%	24.43%
AVG	*73.34%*	*65.08%*	*21.20%*	*25.92%*

Best models used for ~10K ATMs



Best model selection can improve accuracy quite significantly. However, it has tendency to overfit since it select the best model from local CV, not in the unseen test set.

Evaluation

Fix ATMs with Bad Prediction

After doing additional evaluation and post-exploration, ATMs in Cluster 2 have poor accuracy. It happens because *XGBoost fails to predict zero value* where most of withdrawals during weekend/holiday in Cluster 2 are zero value.

Cluster	# of ATM	PE25 Train	PE25 Test
1	4745	73.30%	64.79%
2	372	51.35%	**47.30%**
3	2137	76.40%	68.25%
4	3363	73.90%	65.44%

Solution

Replace predictions in weekend/holiday with median of past value of weekend/holiday days

Result Impact

It can improve upto 1,116 predictions or about **1.49%** difference of percentage error below 25%

? *372 x 3 = 1,116 is calculated since March test data have three holidays: Sunday, Friday and Saturday*

Final Submission

*The final model is the model combination based on best model selection of **1 Naive + 4 XGBoost** models with different features engineering (f5, f52, f53, f55) and 1.02 correction (0.95 correction for Good Friday, 30 March 2018 and no correction for Thursday, 29 March 2018) trained using full dataset.*

Models used for 9 ATMs with missing value are XGBoost (6 ATMs), median forecasting (2 ATMs) and Naïve (1 ATM).

- **1st submission** (28-09-2018) → achieve **60.43%** in evaluation metrics (get 1st rank in leaderboard with 1.33% or ~996 data test difference compared to 2nd/3rd rank)
- 2nd submission (07-10-2018) → failed, using assumption that 1 April 2018 effect moves to 29 March 2018
- 3rd submission (13-10-2018) → failed, same like 1st submission but use single XGBoost model

Conclusion

Learning from model buildings

- EDA is very important and helps in deciding proper features engineering
- Domain knowledge about the real world is very important
- Start with simple model with proper features engineering (even linear regression works quite well!)
- Model combination can improve accuracy

Real world application

- Single model can be preferable due to simplicity and faster implementation/calculation
- Use sMAPE as evaluation metrics since MAPE does not have upper limit and the value can be infinite
- Use correction factor greater than 1 since underestimated forecast (out of stock) is avoided
- Confidence interval can be used to include uncertainty of the forecast
- Forecasting result can be used for optimization problem (minimize the operational cost)

THANK YOU

dilan

Muhammad Rasyid Ridha

rasyidstat@gmail.com

Business Intelligence Analyst at GO-JEK Indonesia