



**Data Glacier**

Your Deep Learning Partner

# Final Project Report

## Retail Forecasting Project

02/01/2023

Gao Mo

# Agenda

Model Deliverables

Model Selection

Model Performance

Feature Importance Summary

# Model Deliverables

- 4 multivariate forecasting model
- One base model, one linear model, one ensemble model, one boosting model
  - Demonstrate explainability in the form of contribution of each variables

# Model Selection

Base Model: Linear Regression

Linear Model: Support Vector Machine with linear kernels

Ensemble Model: Random Forest

Boosting model: Gradient Boosting Regressor

# Model Performance (Base Model)

## Feature Importance

	feat	importance
0	Price Discount (%)	0.019680
1	In-Store Promo	0.083361
2	Catalogue Promo	0.222547
3	Store End Promo	0.265528
4	Google_Mobility	0.002988
5	Covid_Flag	0.343029
6	V_DAY	0.050993
7	EASTER	0.000558
8	CHRISTMAS	0.011315

## Model Accuracy

	Product	Forecast_Accuracy
0	SKU1	0.428766
1	SKU2	0.073225
2	SKU3	0.749226
3	SKU4	0.562450
4	SKU5	0.895168
5	SKU6	0.769993

# Model Performance (Ensemble Model)

## Feature Importance

	feat	importance
0	Price Discount (%)	0.653999
1	In-Store Promo	0.040013
2	Catalogue Promo	0.026686
3	Store End Promo	0.045055
4	Google_Mobility	0.133397
5	Covid_Flag	0.030393
6	V_DAY	0.025220
7	EASTER	0.023251
8	CHRISTMAS	0.021986

## Model Accuracy

	Product	Forecast_Accuracy
0	SKU1	0.510794
1	SKU2	0.226012
2	SKU3	0.969402
3	SKU4	0.873024
4	SKU5	0.956232
5	SKU6	0.677644

# Model Performance (Boosting Model)

## Feature Importance

	feat	importance
0	Price Discount (%)	0.835192
1	In-Store Promo	0.006514
2	Catalogue Promo	0.009072
3	Store End Promo	0.030644
4	Google_Mobility	0.013762
5	Covid_Flag	0.087619
6	V_DAY	0.016612
7	EASTER	0.000434
8	CHRISTMAS	0.000151

## Model Accuracy

	Product	Forecast_Accuracy
0	SKU1	0.908625
1	SKU2	0.294121
2	SKU3	0.772930
3	SKU4	0.379625
4	SKU5	0.721423
5	SKU6	0.808857

# Model Performance (Linear Model)

## Feature Importance

	feat	importance
0	Price Discount (%)	0.653999
1	In-Store Promo	0.040013
2	Catalogue Promo	0.026686
3	Store End Promo	0.045055
4	Google_Mobility	0.133397
5	Covid_Flag	0.030393
6	V_DAY	0.025220
7	EASTER	0.023251
8	CHRISTMAS	0.021986

## Model Accuracy

	Product	Forecast_Accuracy
0	SKU1	0.515940
1	SKU2	0.040987
2	SKU3	0.737935
3	SKU4	0.422220
4	SKU5	0.741711
5	SKU6	0.894752



# Feature Importance

- Among all the features, the percentage of price discount has the highest average feature importance value among all the models, which shows that it is significantly more important than other features
- The second highest feature is whether COVID is in play, which is also worth noting

# Summary

In summary, by comparing the model performance, ensemble model has a higher accuracy among all different model selected for this study. Therefore, the recommendation to replace the in-house forecasting model will be Random Forest ensemble model

# Thank You