Final Presentation

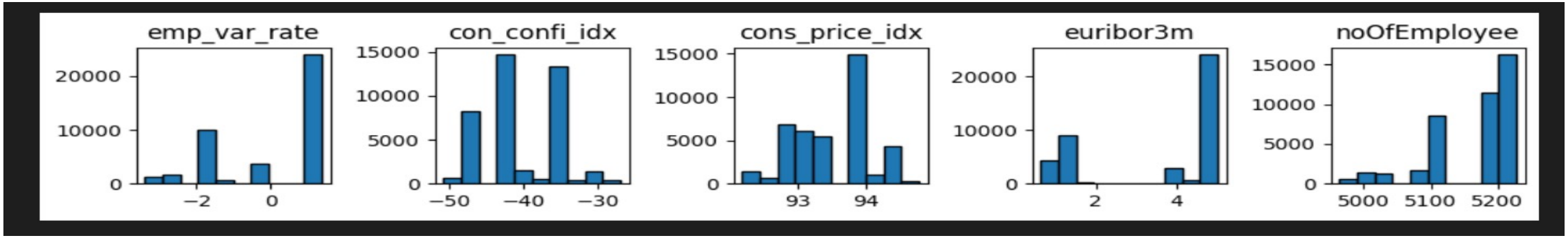
Project Name: Bank Campaign Marketing

Preprocessing

1. Label encode : Encode the target values
2. One-hot encode: encoded the categorical values.
3. Total Number of Features : The features which were used while implementing a model were 16 features.
4. The target value is highly imbalanced with ratio of 9:1.
5. To avoid bias in the model due to imbalance data following techniques were applied:
    - > Assigning weights to minority class
    - > Applying cross- validation and stratified kfold technique

Feature Engineering

- There were 4 data features which were irrelevant and creating more bias due to a few skewed data or few unscaled data points.
    - > Employment variation rate (emp_var_rate)
    - > Constant price index (cons_price_idx)
    - > euribor3m
    - > Constant confidence index (con_confi_idx)



Model Evaluation Matrix

Business Goal: To increase number of people subscribing the deposit scheme product.

Technical : The goal is to improve false positive values to decrease the chances of loosing the vital subscriber.

It can be fulfilled by improving recall score and also accuracy must be stable which can help us maintain the prediction quality.

Mode 1: Linear Model

Logistic Regress is one of the simple and effective linear model.

i) Type 1: Assign weights to the target class:

```
Final Model - Classification Report:
              precision    recall  f1-score   support

           0       0.96      0.92      0.94      7303
           1       0.51      0.70      0.59       935

    accuracy                           0.89      8238
   macro avg       0.74      0.81      0.76      8238
weighted avg       0.91      0.89      0.90      8238
```

Accuracy : 89% and recall: 70%

> Type 2 : No weights assigned

```
Final Model - Classification Report:
              precision    recall  f1-score   support

           0       0.97      0.86      0.91      7303
           1       0.43      0.82      0.56       935

    accuracy                           0.85      8238
   macro avg       0.70      0.84      0.74      8238
weighted avg       0.91      0.85      0.87      8238
```

Accuracy : 85% and recall: 82%

Low Accuracy and High recall (comparing accuracy and recall with the above model)

Helps in predicting more false positive values which reduces the chances of losing even a vital subscriber.

Model 2: Ensemble Method

Random Forest:

> Type 1 : Applied Cross validation and balanced class weights:

```
Cross-Validation Accuracy Scores: [0.90591806 0.90728376 0.90880121 0.90698027 0.90955994]
Mean CV Accuracy: 0.9077086494688922
Cross-Validation Recall Scores: [0.42240216 0.4439946  0.44264507 0.41565452 0.42240216]
Mean CV Recall: 0.42941970310391364
```

Accuracy = 90% and recall = 42.9%

> Type 2 : Applied Cross validation and assigned varied class weights:

```
Cross-Validation Accuracy Scores: [0.91016692 0.91198786 0.90864947 0.90804249 0.91168437]
Mean CV Accuracy: 0.9101062215477997
Cross-Validation Recall Scores: [0.4682861  0.47638327 0.46018893 0.43994602 0.46018893]
Mean CV Recall: 0.4609986504723347
```

Accuracy = 91% and recall = 46%

Observation:

Performs poorly compared to Linear model though the accuracy is higher compared to linear model

As, Recall values are slightly increasing for Type 2 with comparison of Type 1 but it has been worst while comparing with linear model

Model 3: Boosting Method

Light Gradient Boosting:

> Applied Stratified Fold and under sampling the class to balance class:

```
Mean Accuracy: 0.8932458138145425
Mean recall score: 0.9045258620689655
```

The Linear Model and Light Gradient model are more reliant and predictive for analyzing people behavior on whether they will subscribe for the product or not?

However, LGB is more efficient and effective in recommending the behavior of people as the accuracy and Recall has been high and significant.

Conclusion

- The Ensemble Technique has varied accuracy with different tehniques applied on the predicting the target variable, but recall remain the constant.

- The Linear Model improved with different techniques and helped the model to predict falser positive to increase the chance of campaigning for more reliable subscriber, but accuracy was being very low.

- However, Boosting technique had overall high accuracy and high recall which had made more varied to the business goal and can predict the

Total number of must be contact = True Positive + False Positive