

**Steering The Sports Industry**

Technical Assessment: **Next Best Offer**

for the Role of:

**Data Scientist | MLE**

**Solution by**

**Motaseam Yousef**

Contents

[Problem Statement 1](#_Toc178000393)

[Research and Literature Review 2](#_Toc178000394)

[Reference [1] 2](#_Toc178000395)

[Reference [2] 2](#_Toc178000396)

[Methodology 4](#_Toc178000397)

[Conclusion 8](#_Toc178000398)

[Future Work Ideas 8](#_Toc178000399)

[References 9](#_Toc178000400)

[Appendix 10](#_Toc178000401)

[Appendix 1 10](#_Toc178000402)

**Table of figures**

[Figure 1 PKG counts 4](#_Toc178000388)

[Figure 2 evaluation metrics 5](#_Toc178000389)

[Figure 3 feature importance for NBO 6](#_Toc178000390)

[Figure 4 test API 7](#_Toc178000391)

# Problem Statement

Next best offer involves recommending the most suitable product or service to customers based on their behavior and preferences. By predicting the next purchase decision, businesses can enhance customer satisfaction and drive in additional revenue.

In this report, a data analysis will be conducted to better understand the underlying patterns and trends within the data. Following this, machine learning models will be developed to predict customer churn and suggest the next best offer, providing valuable insights for optimizing customer retention and driving sales growth.

# Research and Literature Review

## Reference [1]

- Idea: Tailor Next Best Offer (NBO) for industries with fewer products but rich data, such as banking and telecom.

- Approach: Use supervised classification models to predict the likelihood of each product being purchased by each customer. Calibrate these scores to rank products.

- Data Features: Include customer demographics, interactions with products, payment history, and communication with customer service.

- Models: Use gradient-boosted trees for classification, fine-tuned for each product.

- Pre/Post Processing: Structure and clean the data, calibrate prediction scores for comparability.

- Enhancements: Utilize customer profitability, market share goals, and other strategic factors in decision-making.

- Future: Explore reinforcement learning for dynamic product recommendations based on real-time interactions.

## Reference [2]

- Data Features: Revenue per product, product holdings, customer details, balances, and additional information.

- Pre-Processing: Clean and structure data in seven Flow zones; ensure coherence for accurate predictions and visualizations.

- Model: Classification model trained on 80% of the dataset, uses class rebalancing for imbalanced subscription data.

- Post-Processing: Score predictions, analyze top prospects and campaigns for cross-sell opportunities, visualize impact.

- Enhancements: Integrate the Advisor plugin for real-time customer insights; use dashboards to compare and visualize results.

- Future: Implement Responsible AI to avoid unintended biases and ensure fairness in marketing campaigns.

Summary:

- data used:  
 Use customer demographics, product interactions, payment history, and balances for data features.

- Integrate strategic factors like customer profitability and market share goals in decision-making.

- best models used boosting models.

# Methodology

There are not many references available for this topic.

Here’s the approach I plan to take:

I will subset the data to include only ACTIVE customers.

(Why only ACTIVE customers? Because these are the ones who continue with us and use our services and products effectively.)

After taking this subset, the target will be defined as follows: For the packages, I will select the one each customer uses most frequently.

(For example, if a customer used a 12-month package twice and a 6-month package five times, I will choose the 6-month package as the target.)

For churned users (INACTIVE), the model I build will identify the most similar active users (SUCCESSFUL users -> ACTIVE). This is important because the goal is to convert inactive users into successful ones or active ones!

The most package the users pay it’s a MONTH\_3\_PKG

MOST\_ACTIVE\_PKG

MONTH\_3\_PKG 223833

MONTH\_12\_PKG 103920

MONTH\_6\_PKG 63208

MONTH\_1\_PKG 5997

MONTH\_9\_PKG 3056

DAYS\_1\_PKG 148

As shown below in the count plot

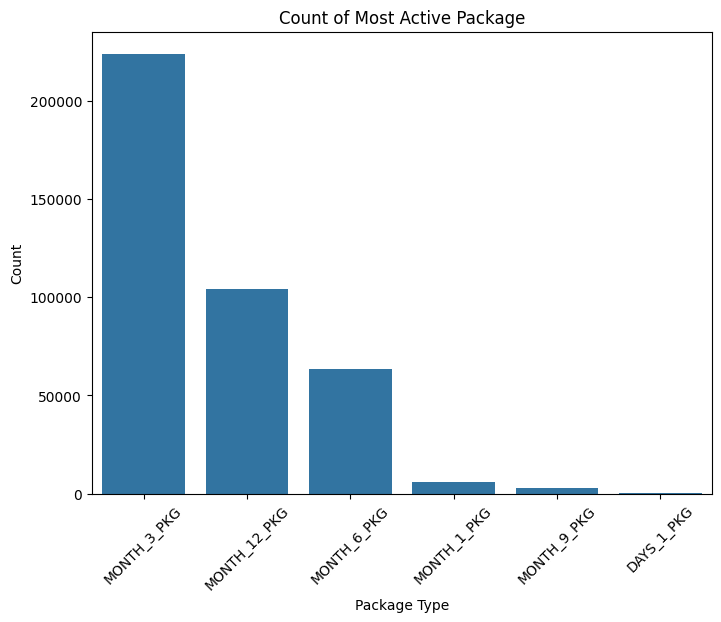


Figure PKG counts

This will be the target!

Now:

* Split data to train, test 80 20
* Scale data
* Build model -> xgboost
* Evaluate

Model result

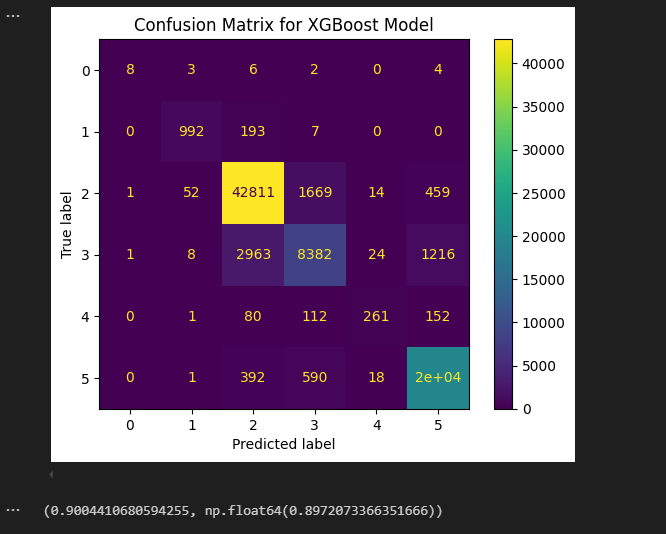


Figure evaluation metrics

As shown the model give a great results -> accuracy of 90%

And the f1\_score = ~90% too!

The feature importance, shown below

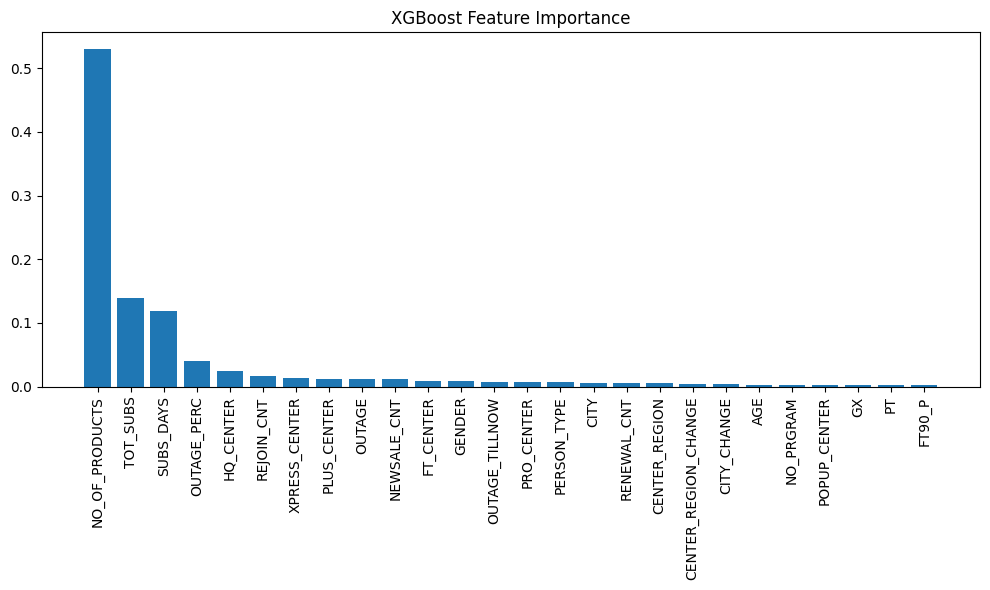


Figure feature importance for NBO

As shown the number of product is the highest feature, then the total subscription, also, the outage have importance!

Then the center type,

The Age, city and region and the program less important!

Convert to API, and test using Postman

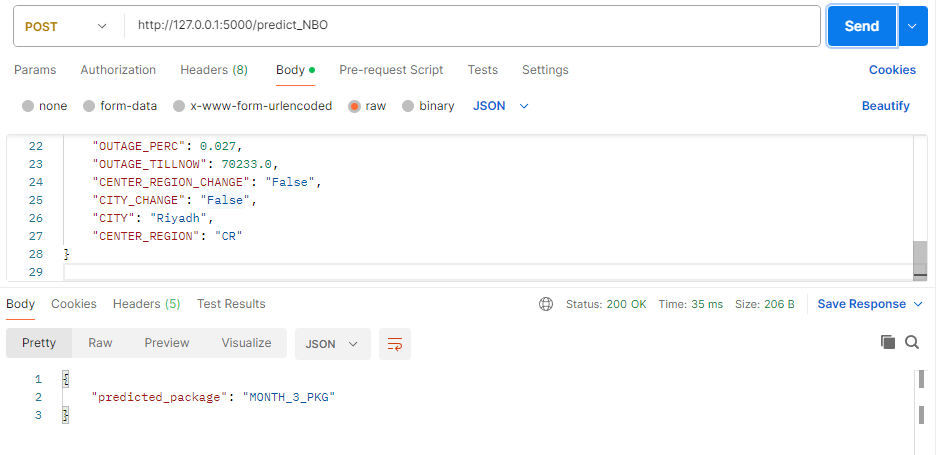


Figure test API

The used data in appendix.

# Conclusion

- Made research about best practice for such models as NBO.

- Made an approach depends on the given features.

- Created a Flask API to predict the most active package using a XGBoost model.

- Applied necessary mappings, encodings, and feature scaling before prediction.

- Integrated model and scaler loading with Flask to handle POST requests.

- Tested the API using Postman.

# Future Work Ideas

-**Customer Segmentation**: Group users into different segments to target with specific marketing campaigns.

-Predict the number of products.

- Predict best center and training programs not only the best offer.

# References

[1] Prabhu, A., 2019. Next Best Offer: When You Have Few Products but Lots of Data. Medium. Available at: <https://medium.com/swlh/next-best-offer-when-you-have-few-products-but-lots-of-data-521349035a9d> [Accessed 22 September 2024].

[2] Dataiku, 2024. *Next Best Offer Solution*. Dataiku Knowledge Base. Available at: <https://knowledge.dataiku.com/latest/solutions/financial-services/solution-next-best-offer.html> [Accessed 22 September 2024].

# Appendix

## Appendix 1

Used input to test JSON:

{ "AGE": 41.0,

"PERSON\_TYPE": "CORPORATE",

"GENDER": "M",

"GX": "No Showup",

"PT": "GX Showup",

"FT90\_P": "No Showup",

"NO\_PRGRAM": "At least one program",

"TOT\_SUBS": 3.0,

"REJOIN\_CNT": 0.0,

"RENEWAL\_CNT": 2.0,

"NEWSALE\_CNT": 1.0,

"FT\_CENTER": 3.0,

"PRO\_CENTER": 0.0,

"PLUS\_CENTER": 0.0,

"XPRESS\_CENTER": 0.0,

"POPUP\_CENTER": 0.0,

"HQ\_CENTER": 0.0,

"NO\_OF\_PRODUCTS": 3.0,

"OUTAGE": 4.0,

"SUBS\_DAYS": 452.0,

"OUTAGE\_PERC": 0.027,

"OUTAGE\_TILLNOW": 70233.0,

"CENTER\_REGION\_CHANGE": "False",

"CITY\_CHANGE": "False",

"CITY": "Riyadh",

"CENTER\_REGION": "CR"}