**Problem Statement**

The Travel company wants to enable and establish a viable business model to expand the customer base. One of the ways to expand the customer base is to introduce a new offering of packages. Currently, there are 5 types of packages the company is offering - Basic, Standard, Deluxe, Super Deluxe, King.

Looking at the data of the last year, we observed that 18% of the customers purchased the packages. However, the marketing cost was quite high because customers were contacted at random without looking at the available information.

The company is now planning to launch a new product i.e. Wellness Tourism Package. Wellness Tourism is defined as Travel that allows the traveler to maintain, enhance or kick-start a healthy lifestyle, and support or increase one's sense of well-being.However, this time company wants to harness the available data of existing and potential customers to make the marketing expenditure more efficient.

We need to analyze the customers' data and information to provide recommendations to the Policy Maker and Marketing Team and also build a model to predict the potential customer who is going to purchase the newly introduced travel package.

**We will build best model with different Classification Models using Bagging, Boosting, stacking and tunning Model using Hyperparmeters**

**Architecture**

Start🡪 Data Ingestion 🡪 Data Validation 🡪

Data Tranformation 🡪Model Trainer 🡪

Hyperparameter Tuning 🡪Model Evaluation

🡪Prediction🡪 Model Pusher 🡪

Save Model🡪Pushing App to Cloud 🡪

Deployement 🡪END

**Data Description**

* CustomerID: Unique customer ID
* ProdTaken: Whether the customer has purchased a package or not (0: No, 1: Yes)
* Age: Age of customer
* TypeofContact: How customer was contacted (Company Invited or Self Inquiry)
* CityTier: City tier depends on the development of a city, population, facilities, and living standards. The categories are ordered i.e. Tier 1 > Tier 2 > Tier 3
* Occupation: Occupation of customer
* Gender: Gender of customer
* NumberOfPersonVisiting: Total number of persons planning to take the trip with the customer
* PreferredPropertyStar: Preferred hotel property rating by customer
* MaritalStatus: Marital status of customer
* NumberOfTrips: Average number of trips in a year by customer
* Passport: The customers own a car or not (0: No, 1: Yes)
* NumberOfChildrenVisiting: Total number of children with age less than 5 planning to take the trip with the customer
* Designation: Designation of the customer in the current organization
* MonthlyIncome: Gross monthly income of the customer

Customer interaction data:

* PitchSatisfactionScore: Sales pitch satisfaction score
* ProductPitched: Product pitched by the salesperson
* NumberOfFollowups: Total number of follow-ups has been done by the salesperson after the sales pitch
* DurationOfPitch: Duration of the pitch by a salesperson to the customer

**Data Ingestion:**

The whole Project is attached in File and folder

In first step we create data ingestion folder to insert url data and to split data in train and test.

**Data Validation**

In this step, we perform different sets of validation on the given set of training files.

1. We craete folder Validation. We validate the name of the files based on the given name in the schema file
2. Number of Columns - We validate the number of columns present in the files,
3. Name of Columns - The name of the columns is validated and should be the same as given in the schema file.
4. The datatype of columns - The datatype of columns is given in the schema file.
5. Null values in columns - If any of the columns in a file ,we use datadrift and report html file to remove it.
6. Report file give us alle Graphic about data and deep analysis about correlation of Data.

**Data Transformation**

We transform data in train and test data and saving prossessing File

Data Export from Transformation Data - The data in a stored in Data

Transformation Folder and is exported as a Train and Test file to be

used for model training.

**Model Trainer**

- After clusters are created, we find the best model for each cluster. We are using two Classification algorithms, "Random Forest" and "AdaBoost". For each cluster, both the algorithms are passed with the best parameters derived from GridSearch. We calculate the AUC scores for both models and select the model with the best score. Similarly, the model is selected for each cluster. All the models for every cluster are saved for use in prediction.

Model Evaluation

Model can make two kinds of wrong predictions:

Predicting that the customer will purchase a Travel Package when they don"t. - False Positive.

Predicting that the customer will not purchase a Travel Package when they do. - False Negative. The Travel company"s objectives are:

Make Marketing Expenditure more efficient and focused on the customers that would actually purchase the product. Predict and Identify all potential customers who will purchase the newly introduced travel package. Metric for Optimization:

For the above objectives, its important that both False positive and False negative values are low. Hence we would want the F1-Score to be maximized. The greater the F1-Score, greater the chances of predicting both classes correctly.

We will build following models, tune them and compare the outcome of all the models:

Decision Tree model.

Bagging Classifier.

Random Forest Classifier.

ADABoost.

GradientBoost.

XG Boost.

Stacking Classifier.

Model Pusher

We save best model in pkl file and we use it by prediction in new dataset

**Deployment**

We will deploying the best model into the cloud Platform Hiroku.

This is a workflow diagram for the prediction of using the trained model.

#### Conclusion

A Key missing variable is if the Product pitched was the same product that was bought.

Basic and Deluxe are the most poppular packages.

There was imbalance in data, as only 18% of customers bought any product. This must be fixed for future analysis.

NumberofChildrenVisiting does not seem to impact the performance of model a lot.

The company can run the model to achieve desired performance levels for new data, also to offer better packages to customers.

Young and single people are more likely to buy the offered packages.

Age and Income have a correlation and we see that higher age groups and higher Monthly Income groups lean towards the expensive packages.

Recommendations

The marketing team can curate the individual packages to the specific business designation

The marketing team can create product and customer segment specific sale pitch to reduce the DurationOfPitch.

The WELLNESS TOURISM PACKAGE should be curated considering the features of existing packages that customers have purchased.

The company can run various campaigns and offers for customers with family to increase sales.

The data shows customers with passport has higher buying ratio and business can curate international packages for such customers.

Specific packages can be created for different income groups.