

Capstone Project-3

This Case Study has 3 (three) checkpoints defined in it.

Check Point Topics	Remarks	Max Marks
<ul style="list-style-type: none">• Data manipulation and Visualization using Python (30 marks)• Statistical Analysis and Exploratory Data Analysis (50 marks)	Checkpoint 1	80
<ul style="list-style-type: none">• Visualization using Power-BI Dashboard (40 marks)• Model Building using ML algorithms (80 marks)	Checkpoint 2	120
Final Presentation and Viva (50 marks)	Checkpoint 3	50

Domain:

Telecommunication Industry

Title:

Next Best Offer Recommendation System for Revenue Optimization in Telecommunications

Project Description:

The objective of this machine learning capstone project is to develop a recommendation system that suggests the next best offer to customers of a telecommunication company. By leveraging machine learning techniques, the company aims to personalize offers and promotions, ultimately increasing revenue and customer engagement.

Steps:

Data Collection and Exploration: Perform exploratory data analysis to understand the dataset's characteristics and identify relevant variables for offer recommendation.

Data Preprocessing and Feature Engineering: Clean and preprocess the data, handling missing values, outliers, and data inconsistencies. Extract and engineer features that capture customer behavior, preferences, and purchase patterns.

Model Selection and Training: Select an appropriate recommendation algorithm, such as collaborative filtering, content-based filtering, or hybrid approaches.

Split the dataset into training and testing sets.

Train the selected model using the training data and evaluate its performance using appropriate metrics (e.g., accuracy, precision, recall, F1-score).

Model Evaluation and Hyperparameter Tuning: Assess the performance of the recommendation model and fine-tune its hyperparameters using techniques like grid search or random search.

Validate the model's effectiveness through cross-validation and A/B testing.

Next Best Offer Recommendation: Utilize the trained model to generate personalized recommendations for individual customers.

Incorporate customer preferences, purchase history, and contextual information to tailor the recommendations to their specific needs.

Campaign Strategy Design: Design targeted marketing campaigns based on the recommended offers.

Create personalized promotions, discounts, or bundles that align with customer preferences and maximize revenue potential.

Campaign Evaluation and Optimization: Implement the designed campaigns and monitor their performance.

Track key metrics, such as offer acceptance rate, revenue uplift, and customer satisfaction.

Analyze the results and iteratively refine the recommendation strategies for improved outcomes.

Final Model Deployment and Documentation: Deploy the final recommendation model in a production environment to provide real-time next best offer suggestions.

Document the entire project, including data preprocessing steps, feature engineering techniques, model selection, hyperparameter tuning, campaign strategies, and evaluation results.

By completing this capstone project, the telecommunication company will possess a powerful recommendation system that can suggest personalized next best offers to customers. This will lead to increased customer engagement, higher conversion rates, and ultimately, revenue optimization.

Data Dictionary:

Category	Attribute	Description
Customer Profile	MSISDN	Subscriber MSISDN (Primary key)
	Status	Status is committed or non-committed (Pack activated or not)
	Segment	Customer divided into some segments (Gold, silver etc)
	Age on network	Difference between activation date and current date
	Region Type	2G/3G/4G/5G
	Total no. of complaints	MSISDN wise monthly count of complaint (current month, previous month)
	Is VAS subscriber	0/1, whether customer is VAS subscriber or not
Usage Profile	Total no. of outgoing calls	Any outgoing call count monthly (current month, previous month)
	Total no. of incoming calls	Any incoming call count monthly (current month, previous month)
	Total no. of outgoing SMS	Any outgoing SMS count monthly (current month, previous month)
	Total no. of incoming SMS	Any incoming SMS count monthly (current month, previous month)
	Total free data usage	Total free data usage (current month, previous month)
	Total data usage	Total data usage (current month, previous month)
	Total 4G Data usage	Total 4G data usage (current month, previous month)
	4G Upward Tag	0/1, whether customer 4G data usage increasing month on month or not
	Total 3G Data usage	Total 3G data usage (current month, previous month)
	Total 2G Data usage	Total 2G data usage (current month, previous month)
	Data Social Media Usage	Total social media data usage (current month, previous month)
	Data App Usage	Total app data usage (current month, previous month)
	Total incoming call duration	Total incoming call duration in minutes (current month, previous month)
	Total outgoing call duration	Total out going call duration in minutes
	On net outgoing call count	Total on net call count (current month, previous month)

	Off net outgoing call count	Total off net call count (current month, previous month)
	On net outgoing call duration	Total on net call duration in minutes (current month, previous month)
Revenue Profile	SMS revenue	Monthly SMS Revenue (current month, previous month)
	Call Revenue	Monthly call Revenue (current month, previous month)
	Data revenue	Monthly data Revenue (current month, previous month)
Recharge Profile	Total recharge amount	Monthly total recharge amount (current month, previous month)
	Total recharge count	Monthly total recharge count (current month, previous month)
	Current balance	Present balance of the subscriber
	Current product id	ID of latest used product
	Current top up value	Current top up value
	Validity days	Current validity days of the customer
	Days since last recharge	No. of days since last recharge
	Last bundle sms Purchased	Last bundle sms Purchased
	Last bundle Purchased	Last bundle Purchased
	Last Recharge Channel	Last Recharge Channel
	Data Top up	0/1 whether the customer has done any data top up
	Bundle pack	0/1 whether the customer is on bundle pack in present month
Handset Profile	Handset category	Handset Network Compatibility (2G/3G/4G/5G)
	Sim support	Multi sim supported Handset
	Smart Phone Tag	0/1, whether the phone is smartphone or not
	Handset change	Count of handset change monthly (current month, previous month)
App Profile	App user	0/1, whether the customer is XYZ company's app user or not
	Days on app	difference between registration date and current date
	Days since last app use	difference between last used date and current date
Activity Profile	Days since last Data Session	Difference between last data session date and current date
	Days since last VAS Session	Difference between VAS session date and current date
	Days since last Voice Session	Difference between last voice session date and current date
Dependent Parameter	Churn Flag	0/1, Whether the customer will churn or not in next 30 days

Check Point 1

Gather the required data: Extract and collect the necessary data from the identified sources using Python. Utilize libraries such as pandas, SQL connectors, or APIs to fetch the data and load it into a suitable data structure (e.g., pandas DataFrame).

Perform data quality checks: Conduct an initial assessment of the collected data to identify and handle any missing values, outliers, or inconsistencies. Use pandas functions and

methods to check for missing values (e.g., `isnull()`, `fillna()`), detect outliers (e.g., box plots, scatter plots), and handle inconsistencies (e.g., data type conversions, data normalization).

Explore the dataset: Utilize Python libraries, such as `pandas`, `matplotlib`, and `seaborn`, to perform exploratory data analysis (EDA) on the collected dataset. Generate summary statistics, visualize distributions, and create meaningful visualizations (e.g., histograms, bar plots, scatter plots, heatmaps) to gain insights into the data.

Identify relevant variables: Based on the EDA findings, determine which variables are most likely to be influential in generating accurate recommendations. These variables can include customer demographics, usage patterns, purchase history, preferences, or any other relevant factors. Select and extract the relevant columns or features from the dataset using `pandas`.

Document the data manipulation and visualization process: Keep a record of the steps taken during data collection, data quality checks, and exploratory data analysis using Python. Document any interesting findings or insights that may be relevant for later stages of the project. Utilize Jupyter Notebooks or other suitable tools to create an organized and well-documented workflow.

By completing Task 1, you will have effectively gathered the necessary data using Python, performed data manipulation and quality checks, and conducted exploratory data analysis to gain insights into the dataset. This will provide a solid foundation for the subsequent tasks, including data preprocessing, feature engineering, and model development.

Data Preparation/Analysis tasks include (but are not limited to) the following.

1. Descriptive statistics for both numerical and categorical and draw a few insights from them. (Univariate Analysis)
2. Bi- Variate Analysis and Multi-Variate Analysis
3. Missing values identification and treatment
4. Outlier analysis and treatment
5. Data scaling using min-max and/or Z-score normalization
6. Data transformation
7. Feature Engineering
8. Perform relevant hypothesis testing (t, chi-Square, Anova tests)

Checkpoint 2

TASK 2.1 (Visualization using Power-BI Dashboard)

Define visualization objectives: Clearly outline the specific objectives and key insights you aim to communicate through the Power BI dashboard. Determine the metrics, trends, or patterns that are crucial for understanding customer behavior and facilitating next best offer recommendations.

Connect data sources: Connect Power BI to the data sources containing the preprocessed and transformed dataset. Power BI supports various data connectors, including Excel files, SQL databases, or online services. Import the necessary data into Power BI.

Design the dashboard layout: Plan the layout of the Power BI dashboard, considering the target audience and the information to be conveyed. Determine the appropriate visualizations and arrange them in a logical and intuitive manner. Common visualizations include charts, graphs, tables, and maps.

Create visualizations: Utilize Power BI's wide range of visualization options to create compelling visuals that effectively convey insights. Select the appropriate visualization types (e.g., bar charts, line charts, scatter plots) based on the data and the message you want to convey. Customize the visuals by adjusting colors, labels, and formatting options.

Implement interactivity: Enhance the user experience by incorporating interactive features in the Power BI dashboard. Use Power BI's built-in capabilities to create slicers, filters, and drill-through actions, allowing users to dynamically explore the data and focus on specific segments or time periods of interest.

Add calculated measures: Utilize Power BI's DAX (Data Analysis Expressions) language to create calculated measures that provide additional insights. For example, you can calculate metrics such as revenue per customer, churn rate, or average usage per customer. These calculated measures can enrich the understanding of customer behavior and aid in the next best offer recommendations.

Incorporate dynamic visuals: Utilize Power BI's capabilities to create dynamic visuals that update based on user interactions or selected filters. For example, you can create visuals that display different segments of customers based on demographics or visualize how offers affect customer behavior dynamically.

Test and refine the dashboard: Validate the visualizations and interactivity by testing the Power BI dashboard with sample data and user scenarios. Gather feedback from stakeholders and iterate on the design and functionality to ensure a seamless and effective user experience.

Publish and share the dashboard: Once the dashboard is finalized, publish it to Power BI service or share it with relevant stakeholders. Determine the appropriate sharing options, such as embedding the dashboard in a website or sharing it with specific individuals or groups.

By completing Task 2, you will have designed and created a visually appealing and interactive Power BI dashboard that effectively communicates insights about customer behavior and facilitates next best offer recommendations. The dashboard will provide a powerful tool for analyzing and presenting data to stakeholders, driving informed decision-making in the telecommunication company.

TASK 2.2 (Model building using ML algorithms)

Select the ML algorithm: Choose a suitable machine learning algorithm for next best offer recommendation based on the project requirements. Common algorithms include collaborative filtering, content-based filtering, or hybrid approaches that combine multiple algorithms. Consider factors such as the available data, scalability, interpretability, and performance metrics.

Split the data: Divide the preprocessed dataset into training and testing sets. Typically, a 70-30 or 80-20 split is used, where the larger portion is used for training the model and the smaller portion is reserved for evaluating its performance.

Feature selection/engineering: Identify the relevant features that will serve as inputs to the ML algorithm. These can include customer demographics, usage patterns, purchase history, preferences, or any other relevant variables. Perform any necessary feature engineering, such as one-hot encoding categorical variables or scaling numerical features.

Train the ML model: Apply the selected ML algorithm to the training data. Use a suitable library in Python (e.g., scikit-learn) to train the model on the training set. Adjust the hyperparameters of the algorithm, if applicable, to optimize its performance.

Evaluate the model: Assess the performance of the trained model using appropriate evaluation metrics for next best offer recommendation, such as precision, recall, F1-score, or area under the receiver operating characteristic (ROC) curve. Use the testing dataset to evaluate the model's performance on unseen data.

Fine-tune the model: Conduct hyperparameter tuning to optimize the model's performance. Utilize techniques such as grid search, random search, or Bayesian optimization to find the optimal combination of hyperparameters that yields the best results.

Cross-validation: Perform cross-validation to validate the model's generalization capability. This involves splitting the data into multiple subsets and training/evaluating the model on different combinations of training and validation sets. Cross-validation helps assess the model's stability and robustness.

Model interpretation: Depending on the selected ML algorithm, explore ways to interpret the model's decision-making process. This can involve feature importance analysis, generating explanations for individual predictions, or visualizing the model's internal workings (e.g., decision trees, feature importance plots).

Model performance documentation: Document the chosen ML algorithm, its hyperparameters, and the evaluation results. Record the model's performance metrics and any insights gained from its interpretation.

Iterative improvement: Iterate on the model-building process by experimenting with different algorithms, feature combinations, or data preprocessing techniques to improve the model's performance and recommendation accuracy.

By completing Task 3, you will have built and evaluated a machine learning model for next best offer recommendation. The model will utilize customer data and generate personalized recommendations, contributing to revenue optimization in the telecommunication company.

Checkpoint 3

Prepare a crisp Final presentation including all the Checkpoint achievements and appear for the Q&A session.

The above three Checkpoints completes the Capstone Project