

Capstone Project-2

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:

Customer Segmentation and Targeted Marketing for Telecommunication Company.

About:

Telecommunication Company XYZ aims to optimize its marketing efforts by implementing customer segmentation techniques and targeted marketing strategies. This capstone project focuses on building a machine learning-based system to segment customers based on their characteristics, behaviors, and preferences. The objective is to enable personalized marketing campaigns that effectively engage different customer segments, leading to improved customer satisfaction, retention, and revenue growth.

Objectives:

- Data preprocessing:* Preprocess the data by handling missing values, encoding categorical variables, and normalizing numerical features.
- Feature engineering:* Extract relevant features from the dataset to capture various customer attributes, such as age, gender, location, usage habits, plan details, customer satisfaction ratings, and response to marketing campaigns.
- Customer segmentation:* Apply machine learning techniques, such as clustering algorithms (e.g., k-means, hierarchical clustering) or dimensionality reduction techniques (e.g., PCA, t-SNE), to segment customers into distinct groups based on their similarities and differences.
- Segment profiling:* Analyze and profile each customer segment by examining their unique characteristics, preferences, and behaviors. Identify key traits and patterns that differentiate each segment from others.
- Targeted marketing campaigns:* Develop personalized marketing strategies for each customer segment, including tailored offers, promotions, and communication channels. Optimize marketing campaign effectiveness by delivering the right message to the right segment.

f. Evaluation: Assess the performance and impact of the customer segmentation and targeted marketing strategies by measuring metrics such as response rates, conversion rates, customer satisfaction scores, and revenue growth.

Data Collection and Preprocessing:

Collect comprehensive customer data from Telecommunication Company XYZ, including demographic information, service plans, call logs, billing records, customer service interactions, and marketing campaign response data. Preprocess the data by handling missing values, encoding categorical variables, and normalizing numerical features.

Feature Engineering:

Extract relevant features from the dataset to capture various customer attributes, such as demographics, usage patterns, plan details, customer satisfaction ratings, and marketing campaign response indicators. Perform necessary transformations or aggregations to create informative features.

Customer Segmentation:

Apply machine learning techniques, such as clustering algorithms (e.g., k-means, hierarchical clustering) or dimensionality reduction techniques (e.g., PCA, t-SNE), to segment customers into distinct groups based on their similarities and differences. Explore different parameter settings and methods to determine the optimal number of segments.

Segment Profiling:

Analyze and profile each customer segment by examining their unique characteristics, preferences, behaviors, and response to marketing campaigns. Identify key traits and patterns that differentiate each segment from others. Visualize and present segment profiles using appropriate charts, tables, or infographics.

Targeted Marketing Campaigns:

Develop personalized marketing strategies for each customer segment based on their profiles and preferences. Design tailored offers, promotions, and communication channels that resonate with each segment. Implement campaign optimization techniques, such as A/B testing or uplift modeling, to refine and optimize the marketing campaigns.

Evaluation:

Evaluate the performance and impact of the customer segmentation and targeted marketing strategies. Measure metrics such as response rates, conversion rates, customer satisfaction scores, and revenue growth. Compare the results to the baseline or non-segmented marketing approach to quantify the effectiveness of the personalized marketing campaigns.

Conclusion:

Summarize the outcomes of the project, emphasizing the successful implementation of customer segmentation techniques and targeted marketing strategies. Discuss the potential benefits for Telecommunication Company XYZ, including improved customer engagement, satisfaction, and revenue growth. Address any limitations or future enhancements that could further enhance the segmentation and marketing strategies.

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

Objective:

Perform data manipulation tasks and create visualizations using Python to explore and understand the customer data for customer segmentation and targeted marketing in the telecommunications industry.

Steps:

a. Load the dataset: Import the customer data into a Python environment (e.g., using pandas library) and create a dataframe.

b. Data exploration: Perform initial exploration of the dataset to gain insights into its structure and content. Use functions such as `.head()`, `.info()`, `.describe()`, and `.shape` to understand the data's dimensions, variable types, and summary statistics.

c. Data cleaning: Identify and handle missing values, outliers, and inconsistent data. Implement appropriate techniques to clean the data, such as dropping or imputing missing values, removing outliers, and addressing inconsistent entries.

d. Data transformation: Apply necessary transformations to the data to make it suitable for analysis. This may include feature scaling, encoding categorical variables, creating derived variables, or aggregating data as required.

e. Data visualization: Utilize Python's data visualization libraries, such as matplotlib or seaborn, to create informative visualizations. Generate various types of plots, such as

histograms, bar charts, scatter plots, or box plots, to understand the distribution, relationships, and patterns within the dataset.

f. Customer profiling: Create visualizations that provide insights into customer characteristics, behaviors, and preferences. Analyze demographic variables, usage patterns, and marketing campaign response indicators to identify meaningful segments within the dataset.

g. Exploratory Data Analysis (EDA): Conduct EDA techniques to uncover patterns, trends, and relationships within the data. Use visualizations to understand the factors influencing customer behavior, such as call duration, plan usage, or campaign response rates.

h. Data summary: Summarize the key findings from the data manipulation and visualization tasks, including notable data trends, patterns, and potential variables of interest for customer segmentation and targeted marketing.

Deliverables:

a. Python code: Provide well-documented Python code showcasing the data manipulation and visualization steps performed on the customer dataset.

b. Visualizations: Include visualizations generated during the data exploration and EDA processes, such as plots, charts, or graphs, that provide insights into customer characteristics, behaviors, and preferences.

c. Data summary: Prepare a concise summary highlighting the important findings and observations derived from the data manipulation and visualization tasks. Summarize any data cleaning or transformation steps undertaken to ensure data quality.

Optional Enhancements:

Depending on the dataset and specific project requirements, you can consider additional data manipulation and visualization techniques, such as:

a. Interactive visualizations: Utilize libraries like Plotly or Bokeh to create interactive visualizations that allow for deeper exploration and interactivity.

b. Geospatial visualization: If the dataset contains location information, create geospatial visualizations using libraries like GeoPandas or Folium to understand the geographical patterns of customer behavior.

c. Temporal analysis: Analyze temporal patterns and trends by creating time series plots or heatmaps to identify seasonality or changes in customer behavior over time.

d. Network analysis: If the dataset contains network-related information (e.g., call logs or social connections), perform network analysis and visualize customer relationships or influence networks.

Note: The specific data manipulation and visualization techniques may vary depending on the dataset and project requirements. Adapt the steps and enhancements accordingly.

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)

Objective:

Create an interactive Power BI dashboard for customer segmentation and targeted marketing in the telecommunications industry. This task aims to leverage Power BI's capabilities to visualize customer data, segment profiles, and marketing campaign performance, enabling effective decision-making and personalized marketing strategies.

Steps:

- a. Data import: Import the preprocessed and cleaned customer dataset into Power BI. Connect to the appropriate data source and load the data into the Power BI environment.
- b. Data modeling: Perform any necessary data modeling tasks within Power BI to define relationships between tables, create calculated columns, or apply other transformations required for analysis.
- c. Dashboard design: Design the layout and structure of the Power BI dashboard. Select appropriate visualizations, arrange them logically, and customize their appearance to ensure a cohesive and visually appealing dashboard.
- d. Customer segmentation visualizations: Create visualizations to showcase the different customer segments identified during the project. Utilize charts, graphs, or maps to represent the characteristics, behaviors, and preferences of each segment. Enable interactive features to explore and compare segment profiles.
- e. Marketing campaign performance: Visualize the performance of marketing campaigns targeting different customer segments. Create visualizations, such as line charts, bar charts, or funnel charts, to display metrics such as response rates, conversion rates, or campaign ROI for each segment. Enable interactive filtering and slicing to analyze campaign performance across different dimensions.

f. Customer profiling: Develop visualizations to present customer profiles based on demographics, usage patterns, or preferences. Use charts, tables, or infographics to highlight key customer attributes for each segment and identify opportunities for personalized marketing strategies.

g. Insights and storytelling: Create narrative-driven visualizations and storytelling elements within the Power BI dashboard. Use text boxes, images, or tooltips to provide context, highlight key findings, and guide users through the insights derived from the customer data. Present actionable recommendations for targeted marketing campaigns.

h. Dashboard interactivity: Set up interactions between different visualizations within the Power BI dashboard. Define how one visualization affects or filters another to create a seamless and interactive user experience. Enable users to drill down, filter, or highlight specific data points for deeper analysis.

i. Testing and refinement: Test the Power BI dashboard functionality, responsiveness, and user experience. Refine and optimize the visualizations, interactions, and overall performance as needed.

Deliverables:

a. Power BI dashboard: Provide the Power BI dashboard file (.pbix) containing the interactive visualizations, customer segment profiles, and marketing campaign performance analysis created for customer segmentation and targeted marketing.

b. Documentation: Document the design decisions, visualizations used, and any notable insights or observations derived from the Power BI dashboard. Include a brief guide explaining how to navigate and interact with the dashboard for other users.

Optional Enhancements:

Depending on the specific project requirements and available data, consider additional enhancements for the Power BI dashboard, such as:

a. Predictive analytics: Incorporate predictive visualizations or forecasting models within the Power BI dashboard to estimate customer behavior or campaign outcomes.

b. Dynamic reporting: Implement dynamic reporting features, such as slicers, filters, or bookmarks, to enable users to customize and personalize the dashboard based on their specific needs.

c. Real-time data integration: Connect Power BI to real-time data sources to enable real-time monitoring and reporting of customer behavior, campaign performance, or other relevant metrics.

d. Exporting and sharing: Configure the Power BI dashboard for exporting or sharing capabilities, allowing stakeholders to download or share specific visualizations or reports with others.

Note: Adapt the steps and optional enhancements according to the specific requirements of the project and the available features and capabilities of Power BI.

TASK 2.2 (Model building using ML algorithms)

Objective:

Build machine learning models to predict customer behavior and optimize marketing strategies in the telecommunications industry. This task focuses on applying various ML algorithms to develop predictive models that can segment customers, predict customer churn, or identify factors influencing customer behavior. The objective is to select the most accurate and reliable ML model(s) for decision-making and targeted marketing campaigns.

Steps:

- a. Data preparation: Split the preprocessed dataset into training and testing sets. Identify the target variable (e.g., customer churn, response to marketing campaigns) and select the relevant features (independent variables) for model building.
- b. Select ML algorithms: Choose a set of ML algorithms suitable for the project's objectives. Common algorithms include logistic regression, decision trees, random forests, gradient boosting, or support vector machines (SVM). Consider ensemble techniques like stacking or boosting for improved model performance.
- c. Model training: Train each selected ML algorithm using the training dataset. Fit the models to the training data and adjust the hyperparameters to optimize performance. Use techniques like cross-validation to assess model performance and avoid overfitting.
- d. Model evaluation: Evaluate the trained models using the testing dataset. Calculate evaluation metrics such as accuracy, precision, recall, F1-score, and area under the ROC curve (AUC-ROC) to assess the models' predictive performance. Consider additional evaluation measures relevant to the project, such as customer segmentation metrics or campaign performance metrics.
- e. Model comparison: Compare the performance of different ML algorithms based on the evaluation metrics. Identify the most accurate and reliable models for the project's objectives. Consider factors like interpretability, scalability, and computational requirements in the model selection process.
- f. Hyperparameter tuning: Fine-tune the hyperparameters of the selected ML algorithm(s) to further improve their performance. Utilize techniques such as grid search, random search, or Bayesian optimization to find the optimal hyperparameter configurations. Repeat the model training and evaluation process with the tuned hyperparameters.
- g. Model interpretation: Interpret the trained ML models to gain insights into the factors influencing customer behavior or marketing campaign effectiveness. Analyze feature

importance, coefficients, or decision rules to understand the variables' impact on the target variable. Identify actionable insights that can inform marketing strategies.

h. Final model selection: Select the best-performing ML algorithm based on the evaluation metrics, interpretability, scalability, and business requirements. Document the chosen model, including the hyperparameter configuration and any important considerations.

Deliverables:

a. Model building code: Provide well-documented code showcasing the implementation of ML algorithms, including data preparation, model training, evaluation, hyperparameter tuning, and interpretation.

b. Evaluation results: Present the evaluation metrics, such as accuracy, precision, recall, F1-score, AUC-ROC, or other relevant metrics, for each trained model. Compare and summarize the results to identify the best-performing algorithm(s) for the project's objectives.

c. Model interpretation summary: Summarize the key insights derived from the interpretation of the ML models, including feature importance, coefficients, or decision rules related to customer behavior or marketing strategies.

d. Final model documentation: Document the selected ML algorithm, along with the optimal hyperparameter configuration, as the final model for customer behavior prediction or marketing optimization. Explain the rationale behind the model selection and its potential implications for the telecommunications industry.

Optional Enhancements:

Depending on the project requirements and available resources, consider the following enhancements:

a. Ensemble modeling: Explore ensemble techniques, such as stacking, bagging, or boosting, to combine multiple ML algorithms for improved prediction accuracy.

b. Feature selection: Implement feature selection techniques, such as recursive feature elimination or feature importance ranking, to identify the most influential variables for customer behavior prediction or marketing optimization.

c. Model deployment: Develop a pipeline or framework to deploy the selected ML model into a production environment, allowing real-time or batch predictions. Consider using cloud-based services or containerization technologies for scalability and ease of deployment.

d. Model monitoring and maintenance: Establish a system for monitoring and maintaining the deployed ML model, ensuring its performance remains optimal over time. Implement techniques for detecting concept drift or model degradation and plan for regular model retraining and updates.

Note: Adapt the steps and optional enhancements based on the specific requirements of the project, available data, and resources. Consider the computational resources required for training and deploying ML models and choose algorithms accordingly.

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