Mira Wadehra

Data Science Project: Customer Churn Analysis in the Telecommunications Industry

1. Define the question or hypothesis for the project:
    - Question: What main factors contribute to customer churn in the telecommunications industry?
    - Hypothesis: The quality of customer service, pricing competitiveness, and service reliability are significant factors affecting customer churn.
2. Method of communication of the results:
    - The results will be communicated through a comprehensive report, including visualizations and key findings, presented to stakeholders in the telecommunications company.
3. Identify the industry, data science question, hypothesis, and method of communication:
    - Industry: Telecommunications
    - Data Science Question: What main factors contribute to customer churn in the telecommunications industry?
    - Hypothesis: The quality of customer service, pricing competitiveness, and service reliability are significant factors affecting customer churn.
    - Method of Communication: Comprehensive report including visualizations and key findings
4. Identify the skills required for each step:
    - Domain Knowledge: Understanding the telecommunications industry, customer relationship management, and customer churn analysis.
    - Applied Statistics: Statistical analysis techniques, hypothesis testing, correlation analysis, and regression modeling.
    - Computer Science: Data cleaning and preprocessing, database management, programming languages (e.g., Python or R), and data visualization.
    - Machine Learning: Classification algorithms (e.g., logistic regression, decision trees, random forests) for prediction and feature importance analysis.
5. Steps and data required:
    - a. Problem Definition: Define the objective of the analysis and the specific churn prediction problem.
    - b. Data Collection: Gather customer data, including demographics, service usage, customer interactions, and churn status.
    - c. Data Preprocessing: Clean the data, handle missing values, and transform variables if needed.
    - d. Exploratory Data Analysis: Analyze the data to identify patterns, correlations, and potential factors influencing churn.
    - e. Feature Selection/Engineering: Select relevant features and create new features based on domain knowledge and exploratory analysis.

- f. Modeling: Build predictive models using machine learning algorithms to identify the most important factors contributing to churn.
- g. Model Evaluation: Assess the performance of the models using appropriate evaluation metrics (e.g., accuracy, precision, recall).
- h. Communication: Prepare a comprehensive report presenting the findings, including visualizations, insights, and recommendations.

The success of this project relies on collaboration between domain experts, statisticians, computer scientists, and machine learning practitioners to effectively analyze customer churn in the telecommunications industry.