Case Study Worksheet: Predictive Analytics

Name:
Date: 12 December 2023
Course: Machine Learning

Understanding the Business Problem

- **Contextualizing the Challenge**: In the context of a UK-based online retailer specializing in unique gift items, many of which are wholesalers, discuss potential challenges in customer retention. What factors might influence customer loyalty in this sector?
- **Problem Presentation**: How might you frame the problem of customer profiling in a data-driven way? What types of data would be crucial for understanding trends?
  - **Hypothetical Example**: Imagine a scenario where an online retailer observes a decline in repeat purchases over several months. How might they frame this problem in a data-driven way? Consider what types of customer data could be crucial for understanding this trend, like frequency of purchases, average transaction value, or the diversity of products purchased.
  - **Abstract Approach**: Without diving into the specific dataset yet, brainstorm how you might use data to uncover patterns or indicators of customer loyalty or attrition in an online retail setting.

Data Source Introduction

- **Dataset Overview**: The Online Retail II dataset includes transactions from 01/12/2009 through 09/12/2011, including product details, quantities, prices, customer information, and more.
- **Dataset Characteristics**: This dataset contains unique identifiers for transactions (InvoiceNo), products (StockCode, Description), along with Quantity, InvoiceDate, UnitPrice, CustomerID, and Country. Note: Cancellations are marked with a 'c' in InvoiceNo.
- **Data Set Access**: Download the dataset from the UCI Machine Learning Repository - Online Retail II.

Data Exploration and Problem Identification

- **Customer Behavior Analysis**: Analyze patterns to identify signs of potential customer loss (or any other task you depicted in the previous step).
- **Key Variables**: Identify key variables in the data set that may influence customer retention (e.g., purchase frequency, order value).

Data Representation and Model Selection

- **Data Preparation**: Discuss how you would prepare this data set for a machine learning model (data cleaning, feature engineering).
- **Model Selection**: Propose an appropriate machine learning model for predicting customer behaviour based on this data set. *What kind of problem is this? What kind of model would you use? What are the pros and cons of this approach?*

Implementing the Solution

- **Training and Testing the Model**: Outlines the steps for training and testing the model.
- **Interpreting Results**: How will you interpret the model's predictions? *What are the potential risks of this approach? How might you mitigate these risks?*

Business Insights and Decisions

- **Gain insights from the data**: Discuss potential business insights that can be derived from the model's predictions.
- **Strategic Actions**: Suggest strategic actions the store can take based on these insights to improve customer loyalty.

Reflection

- **Learning Takeaways**: What did you learn about the role of machine learning in solving real-world business problems?
- **Real-World Application**: Consider how this approach can be applied in your business or industry. What are the potential benefits and risks of this approach?