MARCO PEROTTI

DEPT Assignment

Churn prediction & Tailored Email Messaging using GenAI

Objectives of This Project:

✅ Build a machine learning model that identifies customers at risk of churn with actionable insights.

✅ Use Large Language Models (e.g. GPT-4o) to generate email content that:

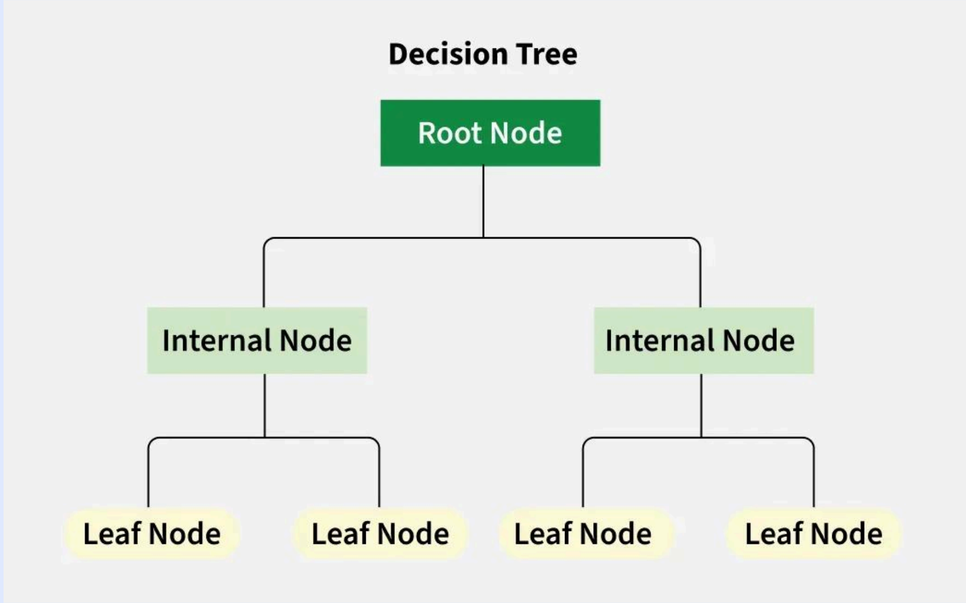
- Aligns with the brand tone and guidelines
- Resonates with each customer segment
- Encourages customers to stay

Dataset Overview:

- 7043 customers, with 5174 that have not churned and 1869 who did.
- Aside from customerID and Churn, 19 different features to describe each case.

Proposed Model:

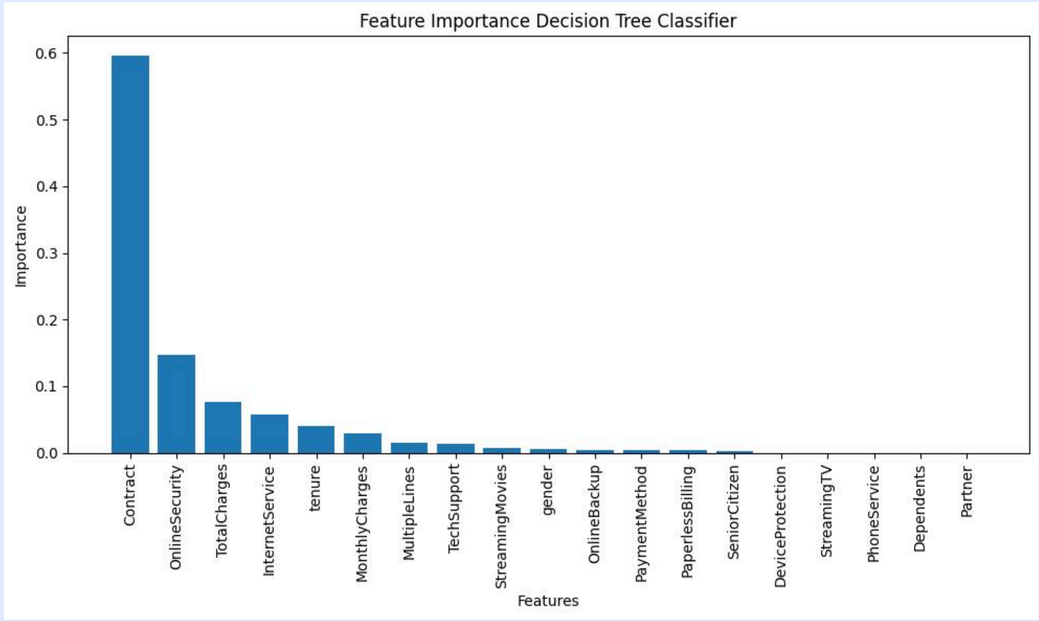
- The chosen model for the task is a DecisionTreeClassifier. With minimal preprocessing and GridSearch is able to obtain an accuracy of 80% on the dataset.
- Additional methods (PCA) and models (KNN, LogisticRegression Classifier) have been tried, but they achieve similar performance.



Feature Importance:

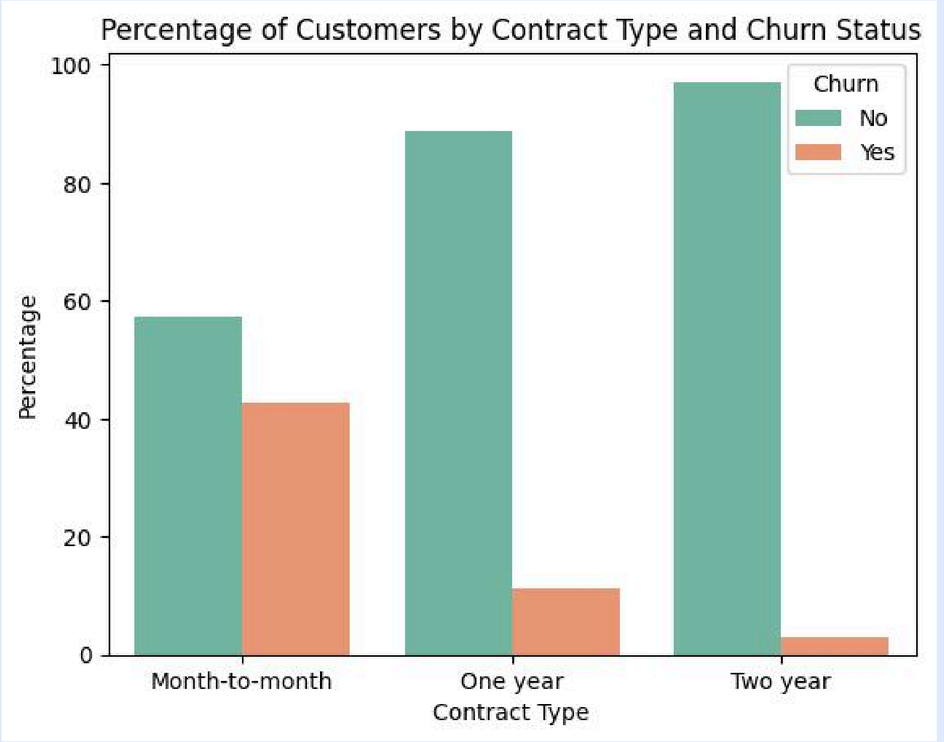
- Using the decision tree, it is possible to see which features are more determinant to establish whether a customer will be more likely to churn or not.

- It seems that the contract type plays a major role, with OnlineSecurity a distant second. Some other features play a part, but a number of them are not relevant.



Insights from the dataset:

- senior citizens churn more (natural causes?)
- customers without dependents churn more
- customers with fiber optics churn way more
- customers without online security, without online backup and without device protection churn more
- customers without tech support churn more
- month-to-month customers churn more
- electronic check customers churn more
- customers tend to churn less the longer their tenure is
- customers tend to churn more the higher their monthly payment is



Proposed LLM: OpenAI

- GPT-4o is able to handle instructions very well, making its use through API a valuable and fast choice to implement.
- It is possible to feed the guidelines to the model before making it generate the email responses.

Next steps:

- Obtain more customer data in order to train more accurate models
- Use dataset insights to tailor to specific customers and their possible churn reasons, e.g. offering tech support or not month-to month contracts
- Explore fine-tuning of custom LLMs.