Windfall Data Science Challenge

WINDFALL

PROBLEM STATEMENT

The Windfall Children's Center is having a fundraising drive. They are interested in building a Propensity to Give model that targets major donors. A Propensity model is used to predict how likely someone is to make a contribution to your organization. For this fundraising drive, Windfall has defined a major donor as any donor who is likely to give at a level of at least $20,000 over a 5 year period. Windfall has approximately 130,000 potential donors in their database, some who have donated significant amounts previously and others who have donated nothing at all. Windfall's fundraising team has a limited capacity for targeting donors, so it is important to properly identify only those who are most likely to meet the targeted donation amount. Using the attached historical donation data in donations.csv and Windfall's unique wealth features in windfall_features.csv, train a model that is able to identify which constituents are likely to give at a level of > $20,000 over the next 5 years.

Donation data in **donations.csv** is a straightforward dataset of donations, including the date of the donation and the candidate's current age. The Windfall data has 22 features. Those containing "ClassA", "ClassB", etc, ("isClassADonor", "sumClassADonation") represent different sources of donation data. Those that contain "CauseA", "CauseB", etc ("sumCauseADonations") represent charity groupings - e.g. environmental non-profits. This is simply an FYI and not necessarily something you need to account for while developing your model.

Here are some hints for data processing and labeling:

- The target donation variable of interest that you may use for labeling is the "**amount"** column in **donations.csv**. We recommend using this column in conjunction with "**trans_date"** to create aggregate variables that can be used for both labeling and feature engineering (if needed).
- The other donation-related variables in **windfall_features.csv** are independent from "amount" and can be safely used as features in your model.
- The target population for this model is "**all the candidates"** in **donations.csv** for whom the model should be able to predict a propensity score. Please note that some of these individuals do not have any records in **windfall_features.csv.**
- The "age" column in donations.csv shows the current age of the donor in 2020.

Once you have developed your model, please provide your source code and summarize in a document that includes the following:

- Instructions for running your training script
- Rationale for construction of labels
- Summary of data preprocessing and feature engineering
- Explanation of your model algorithm choice and chosen model parameters
- Metrics/figures of your model's performance
- Advice for how to use the model for decision making

Keep in mind that we are more interested in the overall approach you take defining labels/scores and the model you build than we are in the actual accuracy of the model.

Thanks for your interest in Windfall Data.

Best of luck.