Modeling Challenge

In this challenge we are looking to assess your programming and data science skills. You can find the dataset and other files attached to the email.

As a media company, helping local small and medium sized businesses through digital advertising is one of our main revenue streams. We do so by running advertising campaigns on behalf of our clients on platforms such as Google and Facebook. In this exercise the task is to predict if a client will stop running advertising campaigns (churn).

The provided dataset has ten columns as follows:

1- CPL_wrt_BC: CPL stands for cost per lead. It is defined as the average amount of advertising dollars spent in order to get a lead (phone call or website click). CPL_wrt_BC shows the change in CPL of a client with respect to the business category they are in.
2- client_state: Client's location
3- duration: The length of time a business has been our client (in months)
4- num_products: The number of distinct advertising products the client has bought
5- calls: Number of calls the client received as a result of their advertising efforts
6- CPL_wrt_self: Change in client's cost per lead in past three months
7- churn: Target column showing if a client churned (1 for churn, 0 for retention)
8- avg_budget: Average monthly budget the client spends on advertising
9- BC: Client's business category
10- clicks: Number of clicks the client received as a result of their advertising efforts

You can find the definition of the columns in the dataset in *dict.txt* as well.

Treat this challenge as if it were a real business project. Perform data cleaning and exploration if necessary. Train a model that predicts churn and describe its performance. Justify the metrics you chose for evaluation and present the results in whichever way you think best communicates the insights. We are mindful of your time and so this challenge has been designed to take no more than 4 hours to complete at most. Please do not spend over 4 hours on this challenge and return the deliverables no later than 48 hours from the time you receive this email.

Deliverables:

1- A Jupyter notebook (or any other format) containing the entire modeling workflow. It is up to you to determine how to organize the file and what to include. Exercise programming best practices. If you prefer, you can upload it to Github and provide the link.

2- A technical report no more than two pages long that includes your thought process, steps taken and final findings.

3- A non-technical report no more than two pages long that summarizes your approach, key findings and recommended next steps for non data science oriented audiences.

(please send all deliverables via email and refrain from posting any part of this exercise online)