**Arcadium coding challenge**

Data:   
- This data set contains transactions occurring in an online store (E-commerce).   
- Out of the 12,330 customer samples in the dataset, 84.5% (10,422) were negative class samples (i.e. customers who did not end up buying the product), and the rest (1908) were positive class samples (i.e. customers who ended up buying).   
- The dataset consists of 10 numerical and 8 categorical attributes.   
- The 'Revenue' attribute can be used as the class label.   
- "Administrative", "Administrative Duration", "Informational", "Informational Duration", "Product Related" and "Product Related Duration": These represent the number of different types of pages visited by the visitor in that session and total time spent in each of these page categories. The values of these features are derived from the URL information of the pages visited by the user and updated in real time when a user takes an action, e.g. moving from one page to another.   
- The "Bounce Rate", "Exit Rate" and "Page Value" features represent the metrics measured by "Google Analytics" for each page in the e-commerce site.   
    Bounce Rate: The value of "Bounce Rate" feature for a web page refers to the percentage of visitors who enter the site from that page and then leave ("bounce") without triggering any other requests to the analytics server during that session.   
    Exit Rate: The value of "Exit Rate" feature for a specific web page is calculated as for all pageviews to the page, the percentage that were the last in the session.   
- Page Value: The "Page Value" feature represents the average value for a web page that a user visited before completing an e-commerce transaction.   
- Special Day: The "Special Day" feature indicates the closeness of the site visiting time to a specific special day (e.g. Mother’s Day, Valentine's Day) in which the sessions are more likely to be finalized with transaction.   
    The value of this attribute is determined by considering the dynamics of e-commerce such as the duration between the order date and delivery date.   
    For example, for Valentina’s day, this value takes a nonzero value between February 2 and February 12, zero before and after this date unless it is close to another special day, and its maximum value of 1 on February 8.   
- The dataset also includes operating system, browser, region, traffic type, visitor type as returning or new visitor, a Boolean value indicating whether the date of the visit is weekend, and month of the year.

Task:   
- Build a machine learning model to predict whether a customer will buy a product or not.

- Please reply to this email with a link to your GitHub repo with the solution-code. Use git/GitHub from the very beginning of your solution and not just to upload the final solution.   
- You will be assessed not solely on the accuracy of your final model but also on steps you have taken to arrive at the final solution. We would like to look through the commit history of this solution-repo to evaluate your approach.   
- Please reply within 7 days from today.

Good luck.

Regards,

Srini