Capstone Project

&

CleverTap

Introduction

About CleverTap -

Our mission is to help fast growing businesses grow faster. A mobile app is an interface that allows businesses to get a glimpse of who the users are, where they come from, what motivates them to perform a certain activity, what keeps them going over a period of time and what makes them leave your app. CleverTap helps you to build long-term valuable relationships with your customers by not only giving you access to behavioral analytics information in real-time but also providing an Engagement Platform that helps you reach your users via the right channel at the right time and with the most optimal message.

Case Study # 01: Predictive Event Modelling - Propensity to watch a video in the next 2 days

Problem Statement

Segment the audience of a content app based on its user's propensity to watch a video in the next 2 days

Data

The directory contains event details for a Video Content app. As the user engages with the app, some of his actions are recorded in detail.

For example, as soon as the user launches the app, an "AppLaunched" event is recorded which contains details such as timestamp, os, device, country, userId etc.

When the user registers itself for the app, a "Registered" event is raised. Similarly when he views the details of a video or an episode, a "VideoDetails" event is recorded. Also, when he starts watching a video, a "VideoStarted" event is raised.

Many a time, app owners also engage with the customers via Push Notifications, SMS, emails etc. When a user clicks on any such campaign, a UTM Visited campaign is recorded. The details of all these events are provided below in the "Events" section.

Events

The directory contains details about the following 6 events.

1. AppLaunched
2. AppUninstalled
3. VideoStarted
4. VideoDetails
5. UTMVisited
6. Registered

Every event contains the following :

1. UserId: Every app user is assigned a unique identity which is stored in this field.
2. Date: The date on which this event was raised
3. Minute_Of_Day: The minute of that day on which the event was raised
4. Second: The second of the minute on which the event was raised. Date, Minute_of_Day and Second provides the exact timestamp of the event
5. Country: The country Id in which the user was present while doing the event. Note that 255 means "Unknown" country.
6. State: The state Id of that country. For example, India will have state Ids in the range [1,29].
7. OS: The OS of the device from which event was raised. They are coded as :
    a. 0: Others, 1: Android, 2: iOS, 3: Windows, 4: Mac, 5: BlackBerry, 6: Linux
8. Device: The type of device. They are coded as :
    a. 0: Desktop, 1: Mobile, 2: Tablet, 3: TV

The events might also have some custom properties :

1) VideoStarted and VideoDetails :

a) Genre : The genre of the video

b) ProgramType : Records the type - TVShow or Movie etc.

c) Category : Records the category - Video on demand("vod") or not

d) VideoId : The video name

2) Registered :

a) Status: The status of the registration

Further Notes

1. AppUninstalled is done from outside the app. So, this event is not recorded when the user uninstalls the app. It is recorded when we poll all the devices. So, the timestamp of the AppUninstalled event can be 1-2 days after the actual time.
2. Mostly the videos are retained on any Video content app. But some videos(eg live streaming of matches) can be perishable. These perishable videos will not be present in the app the next day.
3. You are given time-wise actions of the users. Make sure you don't use future data to predict data in the past.
4. Having said that, the schedule of most TV Shows, launch dates of Movies and timings of any live-stream videos are usually known a week before the videos are put on the app.

Dataset

- https://drive.google.com/file/d/0Bwjat2USDc2Zb01GSmEtblA3Q1U/edit

Evaluation Metric

The presentation should cover the following

- *Problem definition*
- *Solution overview*
- *High-level approach*
- *Steps involved in solving the problem*
    - *Datasets used*
    - *Data pipeline*
    - *Various models and approaches used*
    - *Model tuning*
    - *Evaluation metric used*
- *Final outcome*
- ***Value delivered to the business***
- *Additional things (optional)*
    - *Next steps (things that you would have done given more time)*
    - *Resources required for deployment*
    - *API and web app for the model*

This is not a comprehensive list, feel free to add more information, however, your presentation should elaborate at least on the above-mentioned points.

If you take up a simple model, which says "Users who watched a video in the last will watch a video in the next 2 days", we get the following stats :

a) Precision ~ 27.8 %

b) Recall ~ 60.7 %

c) F1 score ~ 0.38

Students will be judged on:

1. F1-score: Minimum required F1-score of 0.5
2. Segments: Minimum of 2 segments and a maximum 4 segments can be created.
3. Features: The goal of a marketer is to not only identify users who are likely to convert but also target and convert users who are unlikely to convert. Hence, the features used to create the model should be actionable by a marketer.
4. Overfitting: The ability of the model to minimize overfitting and reproducing the numerical scores of the parameters mentioned above on the new dataset

******