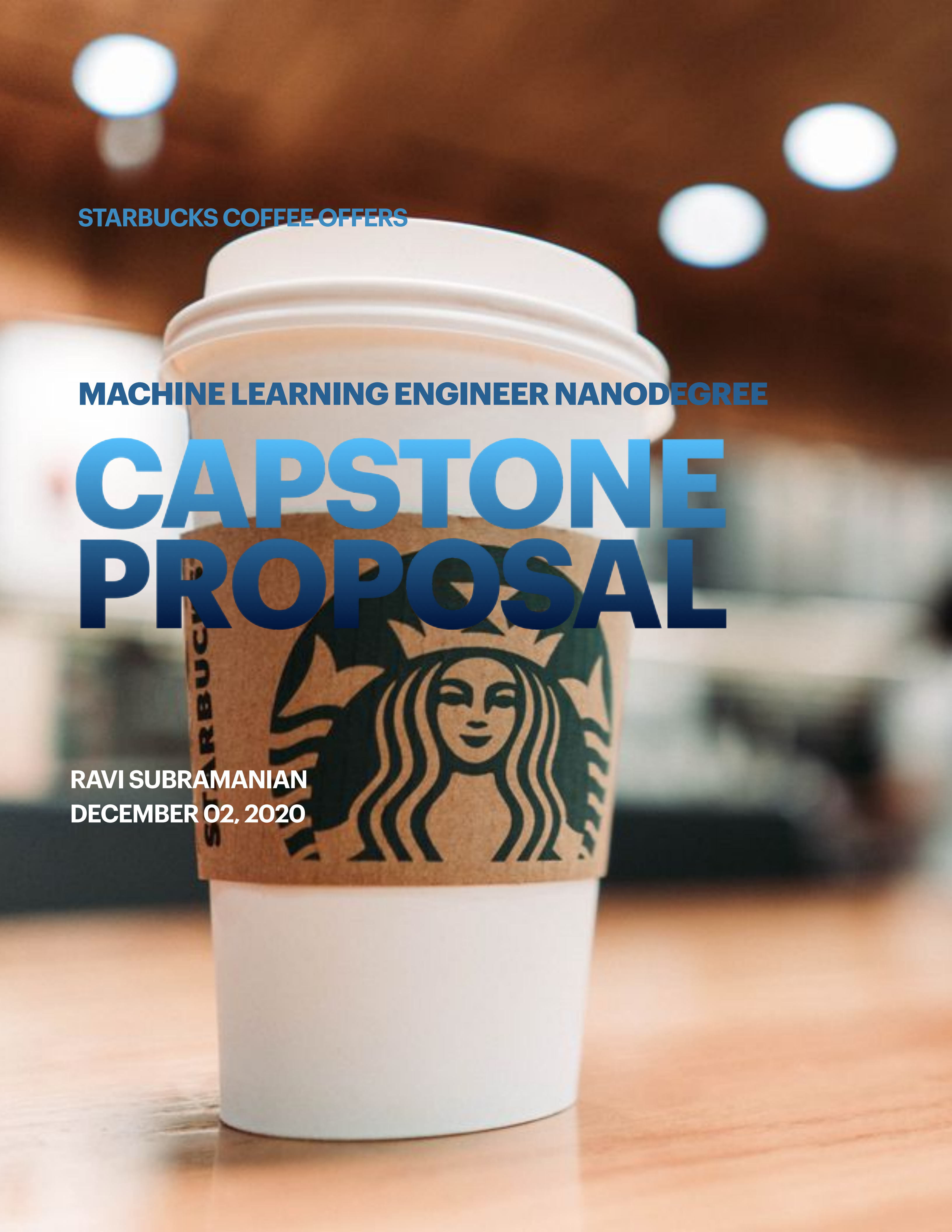

STARBUCKS COFFEE OFFERS

MACHINE LEARNING ENGINEER NANODEGREE

CAPSTONE PROPOSAL

RAVI SUBRAMANIAN

DECEMBER 02, 2020

CAPSTONE PROPOSAL

RAVI SUBRAMANIAN

DOMAIN BACKGROUND

This project is focused presenting personalized offers based on spending habits. I chose this project as it not only applies to Starbucks, but also to many other real world problems on presenting the right offers to the right customers at the right time.

As a customer of many products myself, I often get frustrated with irrelevant ads, offers and rewards that doesn't apply to me. I often feel these companies are bombarding with spam mails, offers and rewards that lead me to ignore reading the offer even sometimes if it totally applies to me. (missed offers)

Personalized timely offers is one of the best way to engage customers using the products that not only benefits the customer needs but also adds business value as the customer is engaged, motivated and use more product offerings creating a win-win situation.

PROBLEM STATEMENT

In this project, we would like to eventually **engage customers,** meaning we would like to **determine the right offer** that is sent and used by customer based on their past purchases and interaction with previous sent offers. We also want to avoid sending inappropriate offers. One possible key metric is to increase the efficiency of the offers *(Increase the percentage of USED OFFERS/SENT OFFERS per customer)*

DATASETS AND INPUTS

The data provided contains 3 files. It data was captured over a 30-day period.

I. Portfolio.json

This json file has information about the different kinds of offers (Total 10 offers)

- **Reward:** The USD given to the customer for completing an offer

- **Channels:** The medium in which the offer is sent (Email, Mobile, Social, Web)

- **Difficulty:** The minimum USD that the customer must spend in order to complete that offer

- **Duration:** The number of days that the offer will last

- **Offer Type:** Buy One Get One (BOGO), Discount, Informational

- **ID:** A unique id for a particular offer

II. Profile.json

This json file has information about the customer demographics

- **Gender:** M (Male), F (Female), O (Other) & None if not available

- **Age:** Birth year defaults to 1900 if no age is available

- **ID:** of a customer (unique)

- **Became a member on:** Date in YYYYMMDD

- **Income:** USD annual, NaN if not available

III. Transcript./ json

This json file has information about the customer purchases and offer interactions.

- **Customer ID:** Unique customer id for this transaction/interaction

- **Event:** Transaction, Offer Received, Offer Viewed, Offer Completed

- **Value:** For the type of event it contains the amount spent if transaction, offer id if offer received or views, offer id or reward id if compare completed.

- **Time:** The time in hours since the start of this test (approximately 30 days)

I am planning to explore data, clean the data using OHE as necessary and remove all NaN values wherever required to make it a usable ML ready data set. Then I'll sklearn.model_selection import train_test_split to split the customer information into appropriate train, test and validation. I will also balance the dataset in each of the datasets.

Interaction with offers: Receiving of offers, viewing of offers, completing of offers. Also it is crucial to note that a customer can compete an offer without ever having to view it.

SOLUTION STATEMENT

There are numerous ways to approach this problem and zone in a solution.

The end goal is to use appropriate ML model that predicts what is the best offer type for each customer based on their profile and transactions data so that they engage or use the offer (Increase the percentage of used offers/sent offers)

To start with I am planing to use between 2-4 models (Eg. XGBoost, Linear Learner, SVM, Logistic Regression). I will analyze the confusion matrix between these 2-3 models and choose the one that performs best.

BENCHMARK MODEL

A possible benchmark model can be Logistic Regression, I'll compare the conversion rate between the all my chosen models (% of Offers Used/Offers Received)

EVALUATION METRICS

Confusion matrix is the best evaluation metrics

TP: Offer sent and Offer Used

FP: Offer sent however users does’t need or doest use it.

TN: Offer not sent and user doesn’t want or doesn’t use it

FN: Offer not sent however user will use it.

The goal is to increase TP and TN and decrease FP and FN.

PROJECT DESIGN

- High level approach to the project:
- Explore the data:
- Clean/Pre-Process the data:
- Build the bench mark model:Logistic Regression
- Build other 2-3 models: XGBoost, SVM and/or LL
- Perform HP tuning (on Validation Set)
- Run the models agains the test set
- Compare each model agains the evaluation metrics (CF Matrix)