

# Project 3: Check In

Team: my (ana)conda don't

Members: Aditya Srivastava , Harsh Gupta , Rachna Konigari

## Method

We have outlined in detail the approach that we are using on our project;

## Data Analysis

We are performing the following to analyze our data;

1. Data visualization using TSNE
2. Checking for feature correlation using Pearson's R

## Feature Engineering

We are performing the following to experiment with feature engineering;

1. Feature normalization and scaling
2. Comparing model performance on stratified vs non-stratified sampling of data
3. Checking feature importance using ablation studies
4. Using correlation scores to remove redundant features
5. Dimensionality reduction

## Modeling

We are performing the following to model the data;

1. Unsupervised modeling using KNNs
2. Supervised modeling using gradient boosted decision trees with XGBoost
3. Supervised modeling using Neural Networks

The models will be evaluated using F1 score as the primary metric.

## Revisions

1. Using F1 score to evaluate models.
2. Narrowing the number of models we are trying and focusing more on qualitative analysis of the models and the data itself.

## Response to feedback

### 1. *"Need def of success" - review 2*

One key piece of criticism was that we hadn't mentioned our evaluation strategy or a success metric for the modeling. We now plan on using the F1 score to evaluate our models

### 2. *"so considering feature engineering might be a good idea" - review 1*

We plan on performing ablation studies, normalizing and scaling of features, and experimenting with other forms of feature engineering. We will also study feature correlation to rule out redundant features.

### 3. *"focus on the explanatory value of your objectives - mostly objectives 3 and 4 - that can provide new insight into both methods and features of the dataset" - review 6*

*"I think this group should focus on optimization using fewer techniques, as they may be overwhelmed with all the methods mentioned" - review 5*

*"it might make more sense to narrow the scope given the amount of time remaining in the semester" - review 4*

*"Lots of methods are stated and it seems that the group has to decide how to narrow down their project" - review 3*

We are now planning to do more qualitative analyses of the data and will focus on fewer modeling methods - one each from supervised traditional ML algorithms, supervised neural networks and unsupervised ML algorithms. We will be picking our modeling approach after studying past approaches on [Kaggle](#) as well.

## Confidential Feedback on the Group Members

No issues to report.