Pitney Bowes_Challenge_2021_Team 2

Apr 27, 2021

PROBLEM

1. Pitney Bowes installs devices on client side. Given certain info, can we predict the devices that may fail in next 7 days?
2. Can we classify the features that are most important in predicting such a failure?
3. If we can predict what is the accuracy of our prediction?

EXISTING ALTERNATIVES

We wait for clients to reach out to the installation team when a device fails to run.

SOLUTION

1. Run ML models to validate the training set and measure accuracy in predicting device failures for known instances.
2. Use the feature selection method to identify the most important factors in predicting device failures.
3. Based on different ML models pick the best performing classification technique.
4. Run the chosen model on the unknown test data set.

KEY METRICS

Accuracy is defined as the ratio of correctly defined subjects to the whole pool of subjects.

Accuracy = TP + TN / (TP + FP + FN + TN)

Area Under the Curve (AUC) is the measure of the ability of a classifier to distinguish between classes and is used as a summary of the ROC curve

UNIQUE VALUE PROPOSITION

1. Using device performance features, we are able to predict in advance the possible cases where the devices may fail and need fixing.
2. We are able to make the prediction with over 80% accuracy. That means out of every 5 devices that may fail next week, we can predict 4 of those instances.

HIGH-LEVEL CONCEPT

The best model was XGBoost Classifier:

XGBoost Classifier: It is one of the fastest implementations of gradient boosted trees discussed above. It does this by tackling one of the major inefficiencies of gradient boosted trees: considering the potential loss for all possible splits to create a new branch (especially if you consider the case where there are thousands of features, and therefore thousands of possible splits). XGBoost tackles this inefficiency by looking at the distribution of features across all data points in a leaf and using this information to reduce the search space of possible feature splits.

UNFAIR ADVANTAGE

1. Prior knowledge of device deployment and readings for those devices like average time for charging and discharging, charging rate and discharging rate, etc.

CHANNELS

Email address

Mobile number - call/ text

Billboard advertisement

CUSTOMER SEGMENTS

Housing Societies

Government Institutions

Direct Users like families, shops, etc.

EARLY ADOPTERS

COST STRUCTURE

Based on the accuracy we need to decide how we are going to plan our action. Do we keep a team for a new installation of meters informed? Or do we send a team for fixing. In terms of a device failure what is the cost of delay and how does it affect the users. All these queries don't have clear answer from the data but will help put a cost structure to the entire process. A prior knowledge with prediction algorithm will reduce the planning and execution cost.

REVENUE STREAMS

We can sell services to our clients by asking them to enroll for this study and if they would want Pitney Bowes' team to inform them about future in-time fixes for their meters in advance or a notification for a probable down-time and thus helping them keep alternatives ready or seek services in advance from the engineering team.