6105 Proposal

Author:

Qiuchi Chen : 001448400
Jing Ren : 001447030

TOPIC :

Store Management

KEYWORDS :

TOP 10 Popular Products Prediction
Rush Hour
Busy Day

GROUP MEMBER :

Hansen Guo : 001828028
Xiangyu Liu : 001498478
Qiuchi Chen : 001448400
Jing Ren : 001447030
Yunyi Wu : 001490936

INTRODUCTION

With the online shopping rising, retailing is severely impacted. Layoffs and reasonable purchase have become the top priority options for river closure. In this program, we aim to find the best schema of layoff and purchase by using machine learning to create the optimal model for predicting future trendy with related history dataset.

GOALS

As the trendy of benefit maximization, store management should be optimized with machine learning. We choose history data information about Orders, Products, Hour Of a Day, The Day of a Week and so forth as the dataset, to analyze and predict two things to help reducing the cost of managing a store.

First, we will predict the top 10 most popular products in order to have the store purchasing and selling more targeted.

Second, according to dataset we can predict in one week which day is the most busy and which day order number is the least. Accordingly, store supervisor would like to arrange schedule more reasonable according to this prediction.

METHODOLOGY

Data Processing:

I Multiple table merging
II Data cleaning
III Dimension reduction

Algorithms:

I Linear regression
II Decision tree

ALGORITHMS

Linear Regression:

Linear Regression is usually the first machine learning algorithm that every data scientist comes across. It is a simple model but everyone needs to master it as it lays the foundation for other machine learning algorithms.
Linear regression performs the task to predict a dependent variable value (y) based on a given independent variable (x). So, this regression technique finds out a linear relationship between x (input) and y(output).

Decision tree:

In decision analysis, a decision tree can be used to visually and explicitly represent decisions and decision making. As the name goes, it uses a tree-like model of decisions. it is also widely used in machine learning, which will be the main focus of this article.

Decision Tree models are created using 2 steps: Induction and Pruning. Induction is where we actually build the tree i.e set all of the hierarchical decision boundaries based on our data. Because of the nature of training decision trees they can be prone to major overfitting. Pruning is the process of removing the unnecessary structure from a decision tree, effectively reducing the complexity to combat overfitting with the added bonus of making it even easier to interpret.