

Ideas: Capstone Project 2

Below are couple of datasets that I am considering for capstone project 2.

1. Acquire Valued Shoppers Challenge

This is kaggle dataset, to predict which shoppers, when presented an offer, will return back buy a new item.

This is a huge dataset with 350 million rows of transactional data for over 300,000 shoppers.

My capstone project 1 was a regression problem, and for project 2 I would like to do something around classification.

This dataset seemed interesting because of its size and also because most of the companies run offers to their customers to boost sales etc and see how effective their offers were and this problem will help me gain better understanding of such businesses.

<https://www.kaggle.com/c/acquire-valued-shoppers-challenge>

2. Human Activity Recognition with Smartphones

This is also a Kaggle dataset, this is a classification problem to classify the events into one of the six activities.

Interesting part of this dataset to me is the number of features, there are around 500 features.

This problem apart from giving me the exposure to classification, will give me an opportunity to deal with the curse of dimensionality and PCA.

Also many of the tech companies today would like to record user behaviors using the gadgets like smart phone, this will be a very relevant problem to solve.

<https://www.kaggle.com/uciml/human-activity-recognition-with-smartphones/home>