Your name
Date

McKesson Final Project Deep Azure

Topic: Microsoft Azure's Machine Learning

Problem Statement: Consumers and Financial Institutions have been grappling with fraud for decades. With recent advancement in technology one might expect the trends to reverse. Rather, the conmen seemed to have become smarter and have started using cutting edge technologies to conduct fraud. So, could the data trail usually left behind, be a means of supplementing a system built to be intelligent enough to look for the patterns and flag transactions as Fraud? In this project, we will attempt to build one such system using Microsoft Azure Machine Learning.

Overview of Technology: Azure Machine Learning is Microsoft's cloud based ML platform that allows building, deploying and publishing predictive analytics solutions all from the ease of a web browser. It provides an easy-to-navigate GUI that enable users to design a simple ML experiment in a matter of minutes.

Overview of Steps:

1. Define problem statement
2. Obtain data
3. Exploratory analysis
4. Clean and massage data
5. Implement and execute experiment
6. Score and evaluate model

Big Dataset: https://www.kaggle.com/ntnu-testimon/paysim1

Size: 470MB (decompressed file) | Sample size: 9.45MB| Format: csv

Hardware: Windows 10 on i7-6700, NVIDIA 1060, 16GB

Software: Python2.7, EmEditor, Github Atom, Azure ML

Lessons Learnt: Although Azure ML provides an amazing capability to implement ML models, Python integration is very limited

Pros: Ease of design and implementation, powerful cluster of machines on the cloud, no local installations required, suite of training, scoring and evaluation methodologies, free tier for learning, simplicity of Microsoft GUI

Cons: Insufficient Python integration

YouTube URL: Short Video | Long Video

Over $16 Billion lost to fraud in year 2016 and more being lost year over year.

It's time we got smarter than the smartest con out there.

Azure ML – A platform to build efficient and scalable ML solutions. Drag and drop modules of standard ML via GUI

Machine Learning lifecycle. Iterative and performance driven

Over 6 Million transactions to train and test the models

Azure ML runs on cloud. Any machine with a modern browser and internet can run the experiments

Powerful suite of ML tools with free tier for exploration

Limited Python integration