**Capstone Project 1: First year well production prediction for onshore oil field**

Problem Statement:

Many industries are faced with optimization problems where processes, methodologies, and actions are sought to be as effective as possible with balance between minimal cost and resources required, and best quality and performance. Oil and Gas industry is no exception. The ultimate goal of any oil/gas company is to produce hydrocarbons as effectively as possible. That usually requires production control of different reservoir fluid types (oil, gas, and water). While maximizing oil production, one might want to minimize income of reservoir brines.

Oil production optimization is a very complex problem, since there are many parameters that effect the overall production rates. Oil production from a well depends on many aspects that include but not limited to the following: geology of the subsurface (interval thickness, rock properties or producing and surrounding rocks, etc.), well bore specifics, completion and perforation parameters, etc. For a human it is a problem with too many degrees of freedom (tens and even hundreds of parameters). Finding the optimal combination for all these controls is not straightforward and usually requires a lot of experience in the field. Moreover, production controls may be quite different between different h/c fields. The same parameter that is associated with production increase in one field may hinder it in another.

This problem is ideal for Machine Learning. When trained on historical production data it is possible for an algorithm to find a function that will predict oil production from controls. This can be an extremely useful tool for new well planning as well as quality control of existing wells.

Dataset:

I’ll be working with a publicly available dataset which consists of three .csv files containing production data and some of the control variable.

'well-index.csv’ file lists all the wells in the dataset and some of the meta data associated with each well such as location, operation company, well type, producing interval, etc. Each well in the dataset has a unique identifier (API).

In order to produce from tight formation, wells are fractured hydraulically by injecting pressurized fluids and other additives. Information about well treatment procedures is given in ‘completion.csv’ file. It describes the amount and type of fluid injected, number of injection zones (stages), depth interval, etc. Each well in the database has API for reference.

‘monthly-production.csv’ contains well production data for different fluids over some period. The rows in the dataset are associated with a reported month of production. Number of days that a well was producing in that month given in a separate column along with the volumes of fluids produced. I will be predicting a well production for the first year. This number can be derived from ‘monthly-production.csv’ for each well.