# Comparative Assessment of Cloud-based Machine Learning Models in Production Engineering


Presented by:
Delmar Koshin,
Han Bao,
Ahsen Ejaz

Supervisors:
Dr. Farshid Torabi
Dr. SD Jacob Muthu

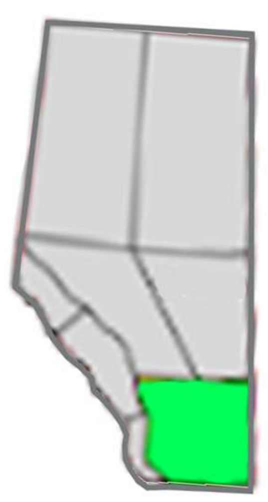
UNIVERSITY OF REGINA


## 1. Background

raw data → Cloud Storage + Processing + Analysis → forecast + analysis

- **Large field data-sets** are fed into commerical cloud services to obtain accurate forcasting or analysis.
- Raw data can be uploaded onto cloud services by users, or live data can be used instead.
- Cloud services perform a multitude of tasks including **Storage, Processing** and **Analysis**.
- Cloud services have **multiple** machine-learning models which can be deployed for **specfic insights.**

### *Experimental Data Source:*

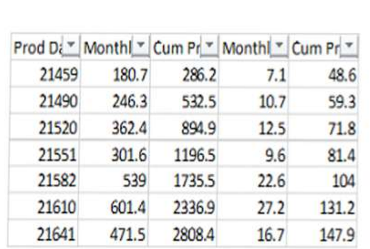


**Southern Plains, AB**

- For this assessment, a **sandstone** reservoir with an adequate homogeneity was selected.
- Production schematic consists of a waterflood project with **135** production wells, and **28** injection wells.
- Approximately ***535*** **unique** data points in **6** distinct categories.

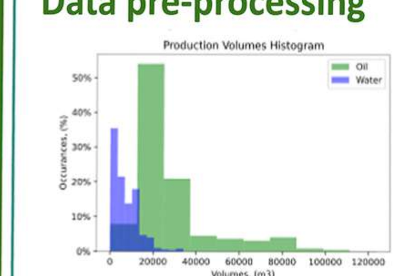
## 2. Methodology

**20%** Data Collection



- Defining inputs Data Ingestion
- QA/QC checks

**60%** Data pre-processing



- Formatting Data
- Cleansing Data
- Visualization & Descriptive Statistics

**20%** ML Model & Assessment

- Model training
- Fine tuning
- Cost Analysis & Assessment

## Ordinary Least Square Regression Model

Split Data
Completed

Linear Regression
Completed

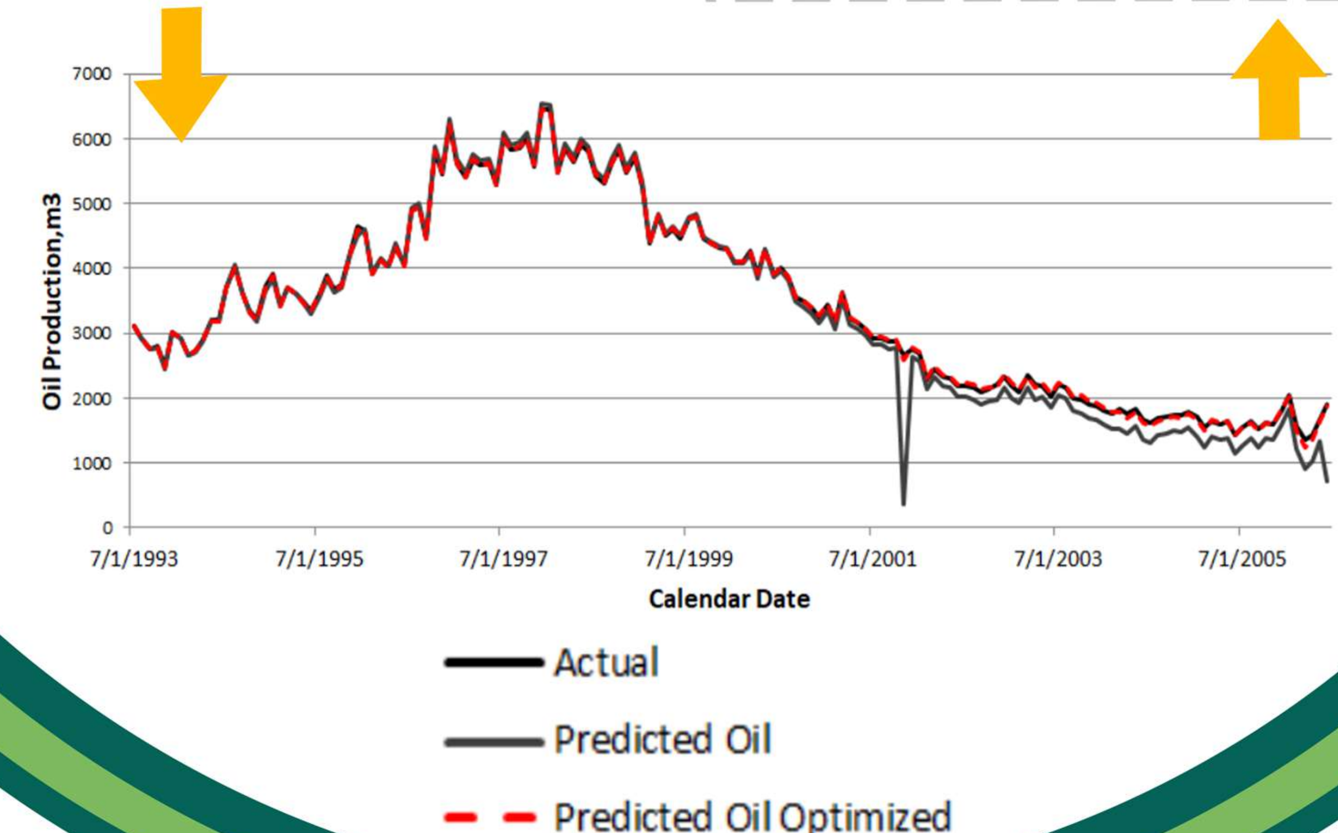
Tune Model Hyperparameters
Completed

**Mean Absolute Error**

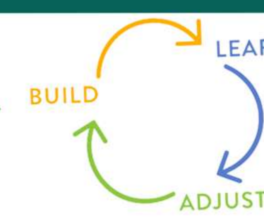
| | |
|---|---|
| Before optimization | 135.27 |
| After optimization | 18.89 |
| Reduction in mean absolute error | ↓ 86.03% |

**Mean $R^2$ Value**

| | |
|---|---|
| Before optimization | 0.967 |
| After optimization | 0.999 |
| Increase in $R^2$ value | ↑ 3.18% |

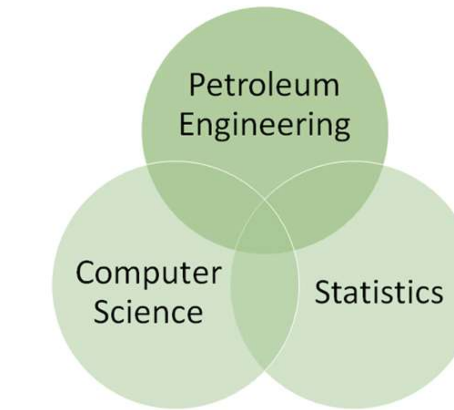



## 3. Engineering Design

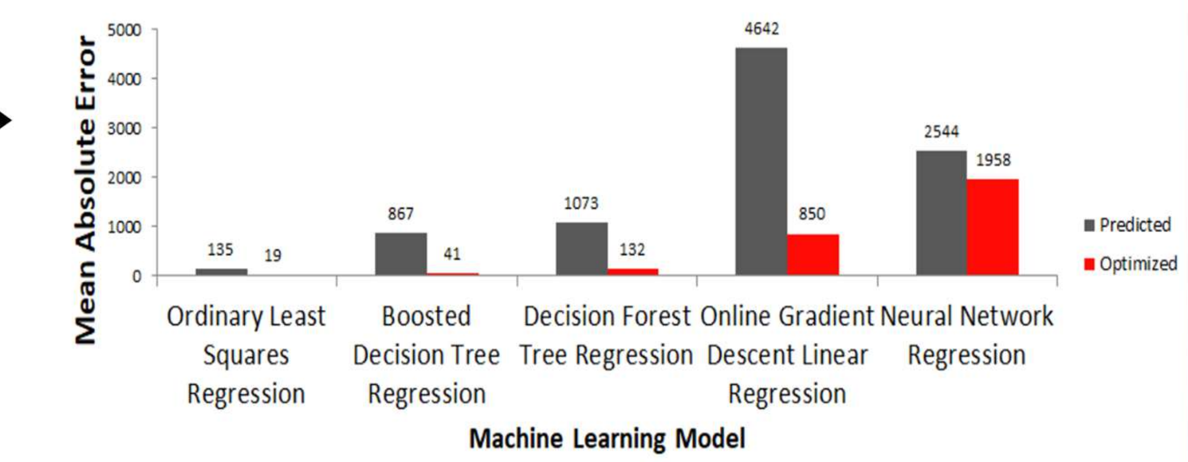



**A Dual Design Approach to Maximize the Accuracy of the Models Used in this Project**

- **Data-centric Approach:** This approach is centered around **retreiving**, and/or **improving** the quality of an existing dataset — either through **filtering erroneous data**, or **enhancing training data** through use of synthetic data. Data quality is often regarded as more important than the selection and optimization of a model. For this project, specific queries were used to acquire a higher quality dataset from **AccuMap**.




- **Model-centric Approach**: An **iterative,** and a **'trial and error'** approach for **training, testing,** or **tuning** of models. Once complete, the **optimum hyperparameters** are selected. For this assessment, hyperparameters of **5** ML models were iteratively optimized, and subsequently finalized based on the optimum error criterion.




## 4. Conclusions

- **Pressure depletion** decline analysis is not typically used for **waterflood applications**.
- Traditional analytical methods, such as the **Buckley-Leverett theory,** require specific reservoir properties to predict, e.g. lab analysis for water saturation values.
- **Azure Cloud Services** are capable of constructing models and computing predictions with rudimentary field data.

- Model considerations:
  - > Memory Usage
  - > CPU /GPU Usage
  - > Interpretability
  - > Predicted costs

**Buckley-Leverett Equation 1-D**

$$\frac{\partial S_w}{\partial t} + \frac{\partial}{\partial x}\left(\frac{Q}{\phi A} f_w(S_w)\right) = \phi\frac{\partial S_w}{\partial x},$$

## Did you know?

In order to increase **efficiency** and **operational sustainability**, energy companies are collborating with **Azure**, as strategic partners.

SUNCOR ENERGY Imperial bp eni

Azure solutions on **Internet of things (IoT)** delivers an intelligent edge to operators, to enhance **production** and optimize surface facilities **operation**.

## Team