Jefferson Fernandez

1/28/-2022

Capstone Three: Project Proposal

Problem Statement Formation

Predicting the Future Production of an Oil Well is one of the key tasks of a reservoir engineer which is tied to the profitability of the well. The ability to predict the future production rates of an oil well gives the stakeholder the ability to decide early in the production life of the well whether the well is profitable to keep or is more profitable to sell.

Context

The Volve field was an on oil producing field that is in the Norwegian North Sea that was active between 2008 and 2016. There are total of 7 offshore wells drilled in the area. These wells were operated by Statoil (Equinor today). Along with the daily oil production rates for each well there are also 12 exogenous variables available in the data set which are average physical properties like choke size, wellhead temperature, etc.

Criteria for Success

In this project we will build multiple time series models for each well encompassing classical time series, forecasting methods, deep learning algorithms and proprietary packages such as FbProphet and NeuralProphet. We will then select the best model for each well and use that to predict future production for the next three months for each well. These models will be compared with respect to appropriate performance metrics, such as R-Squared, Mean Absolute Percent Error (MAPE), and upper/lower bound of worst-case residuals.

Scope of solution space

Machine Learning models capable of predicting future oil production for each well.

Constraints within solution space

There are considerable number of days with no oil production (gaps) affecting the granularity of the data therefore the models will be based on a down sampled data set.

Stakeholders to provide key insight

Well operator who own the producing assets.

Key data Sources

CSV file from the volve data set which contains the production history for each well and 12 other averaged physical properties. Each row represents the values for each variable. There are missing dates and as a result a considerable amount of data preparation must be done.