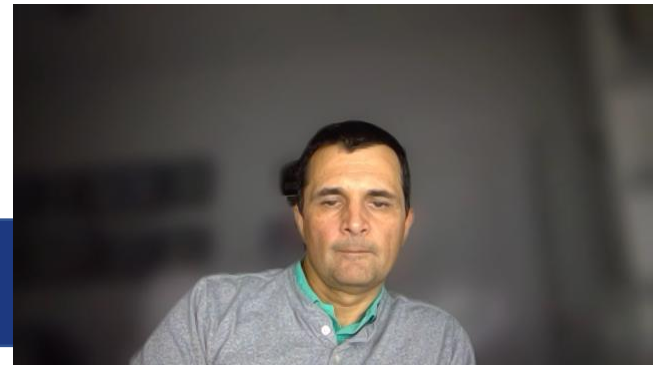
# Using Sentiment Analysis on 10-k Filings to Predict Oil Prices

Peter Kowalchuk

School of Professional Studies, City University of New York

Data 698: Analytics Master's Research Project

Jamiel Sheikh



# Introduction

Can SEC fillings predict the price of a commodity?

Natural Language Processing (NLP) techniques such as Sentiment Analysis are finding their way into Financial Analysis

This study attempts to prove that it is possible to predict the price of a commodity from the content of the Security and Exchange Commission (SEC) 10-K filings.

Commodity: Oil

# Literature Review

Has this been researched already?

Assessment of Sentiment Analysis on company fillings - G. H. Soong and C. C. Tan (2021)

Sentiment Analysis on non-filling data to predict company stock price - Gao, Kampas, and Rinne (2018)

SEC fillings contain broader market information - U.S. Securities and Exchange Commission, "How to Read a 10-K," 2011

Sentiment Analysis on reports used to determine investors' perception of risk - Bao and Datta (2014)

Risk in the financial Sector prediction from report's Sentiment - Rolnicki. (2018)

No studies on predicting the price of a commodity

# Hypothesis

The goal of the study

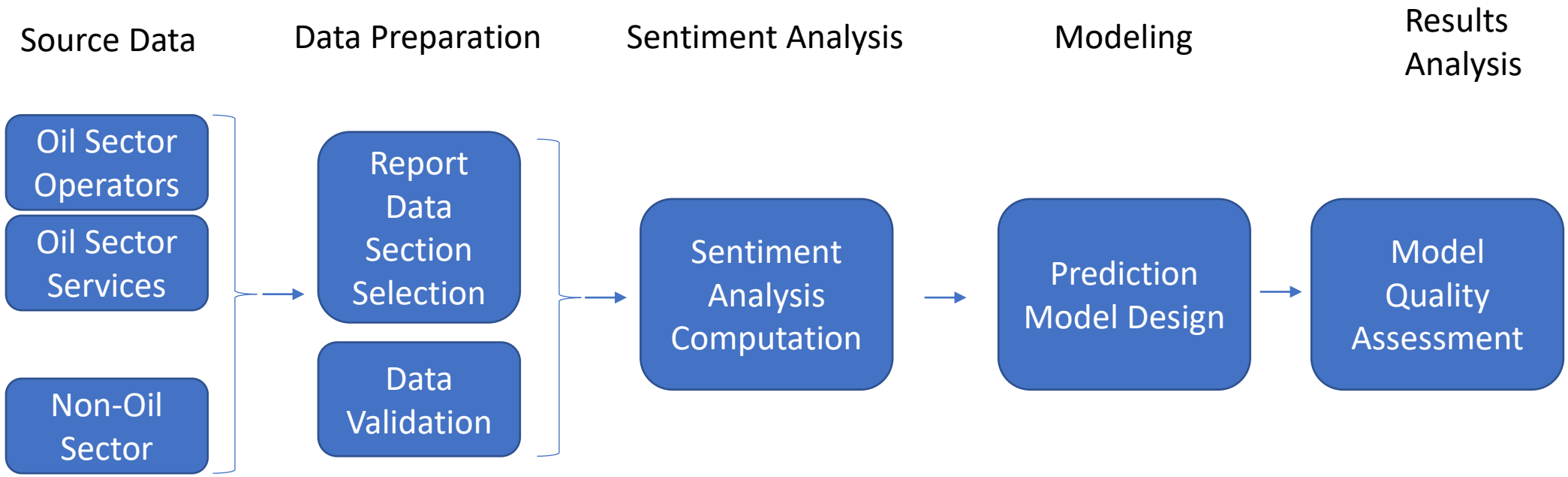
***Hypothesis 1:*** it is possible to design a linear regression prediction model that, with *reasonable accuracy*, can predict the price of oil from SEC 10-K report filings' computed Sentiment Data.

***Hypothesis 2:*** it is possible to design a regression model that explains oil prices using Sentiment data from SEC 10-K reports as the predictor variable.

***Hypothesis 3:*** SEC 10-K reports from companies that operate directly in the oil sector produce better Sentiment based regression models for oil prices than companies that do not directly operate in the oil sector.

# Methodology

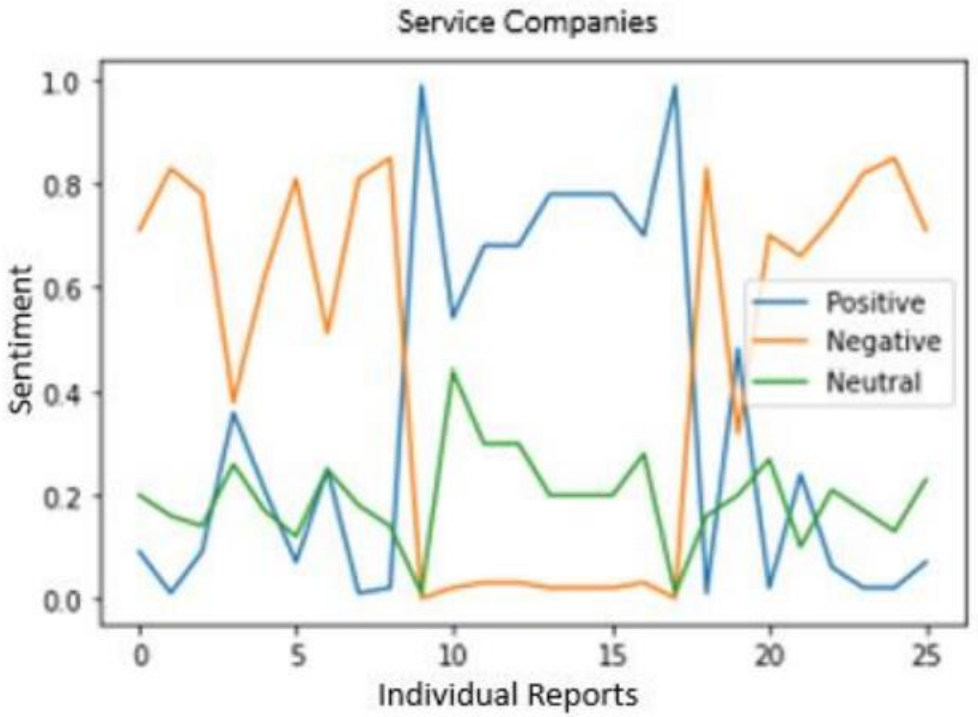
How the study performed the Analysis



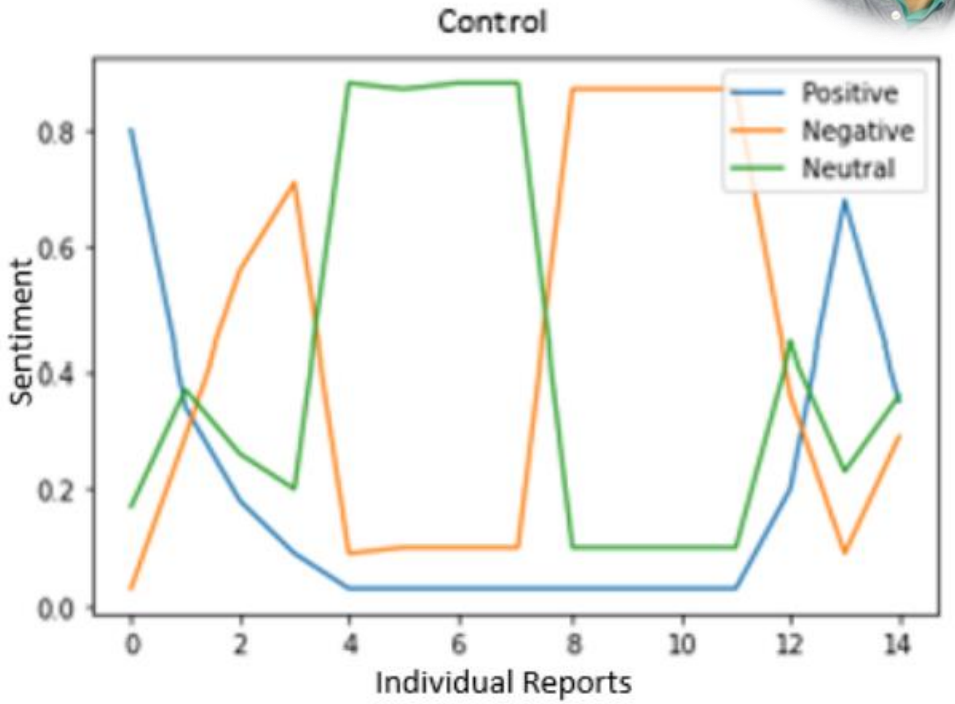
# Study Results

Study findings

A)



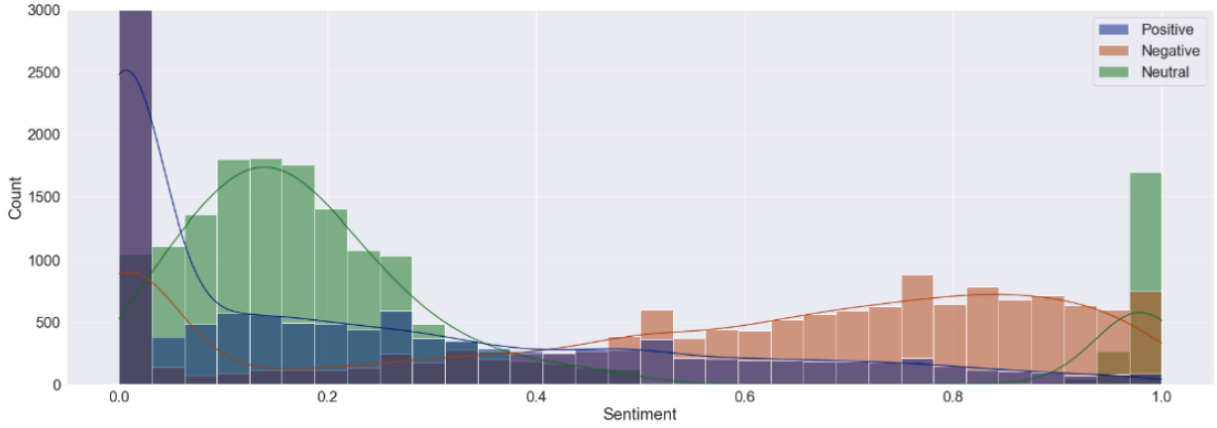

B)



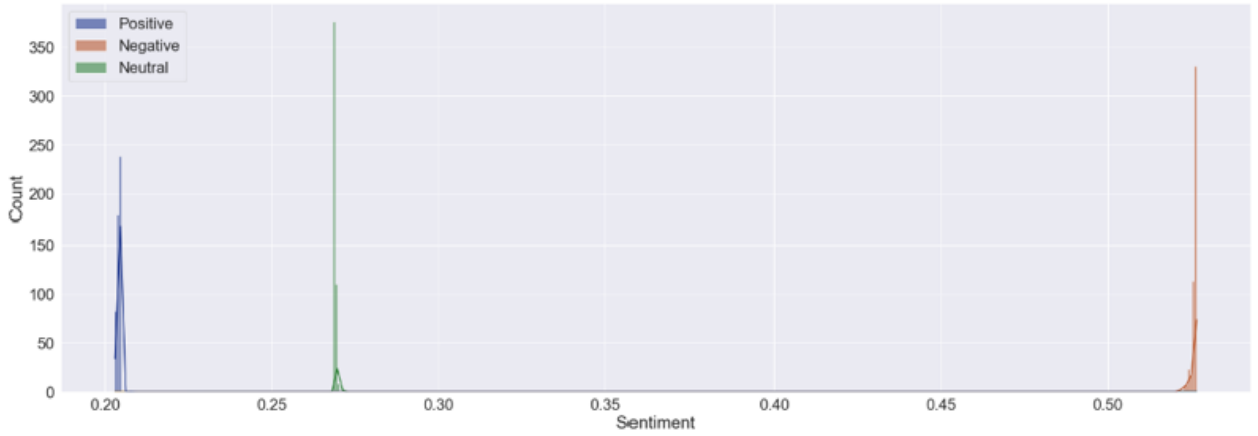

Issues found when analyzing entire reports

# Study Results

Study findings



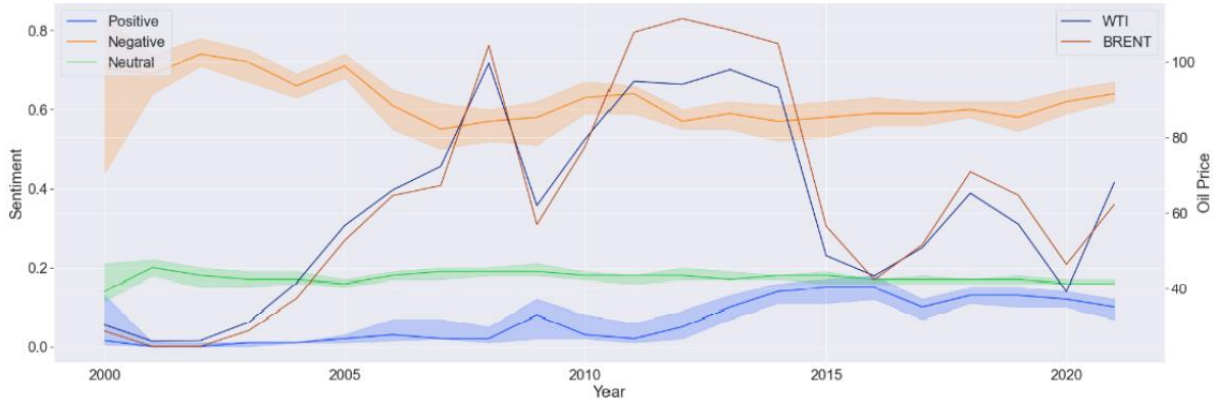

Histogram plot of Sentiment Data in study



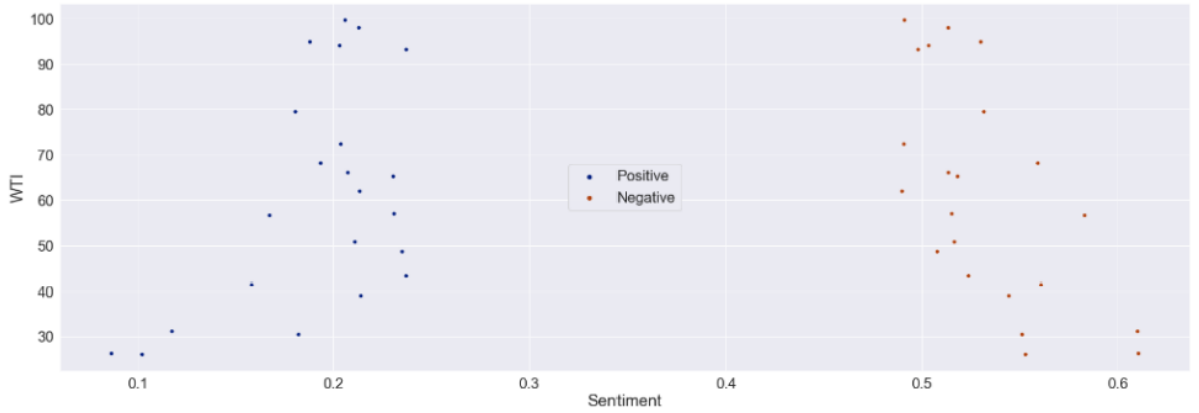

Bootstrap Sentiment data distributions for 2015

# Study Results

Study findings



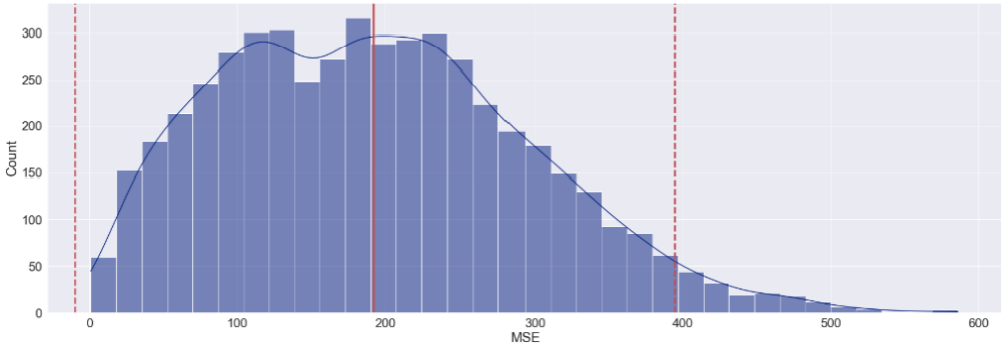

Histogram plot of Sentiment Data in study



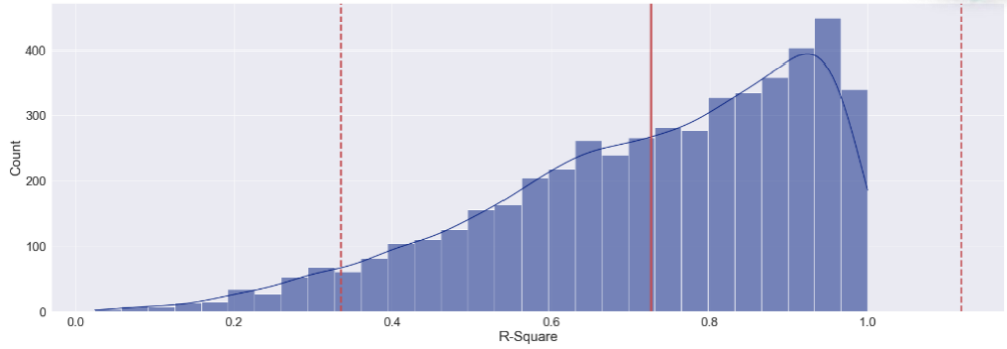

Positive and Negative Sentiment vs. WTI oil price

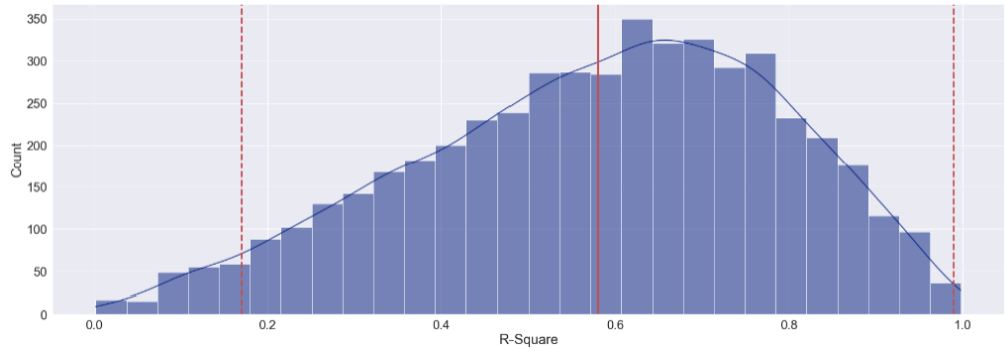
# Study Results

Study findings




No model with an MSE of $2 was found

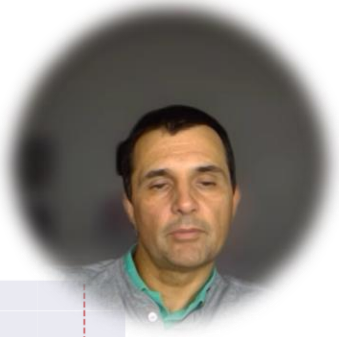


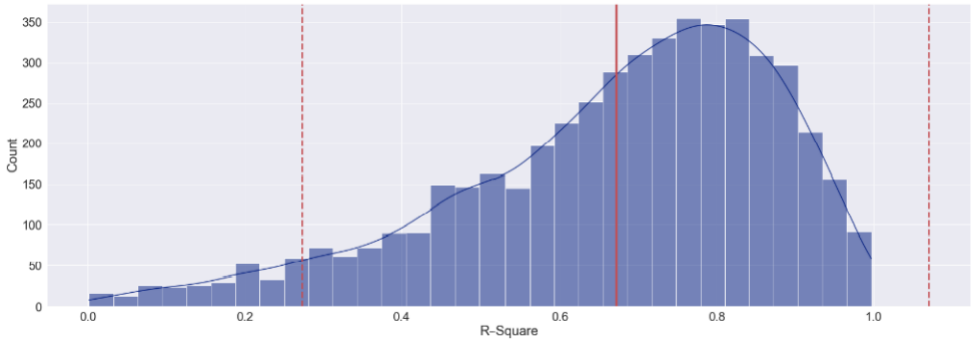




Certain combinations of Sentiment produced better correlations

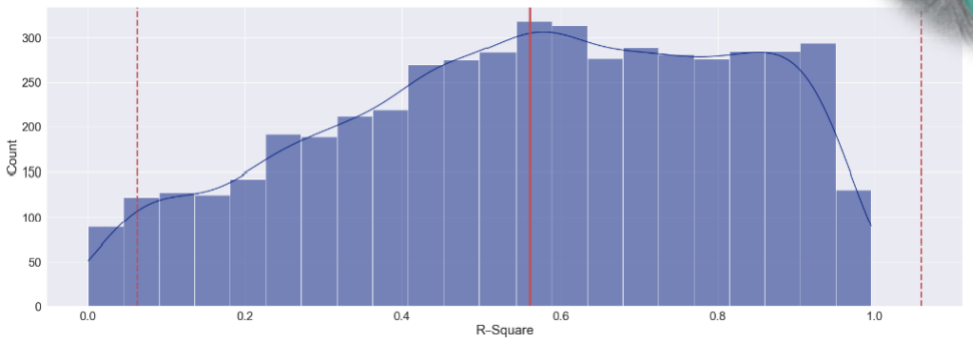
# Study Results

Study findings

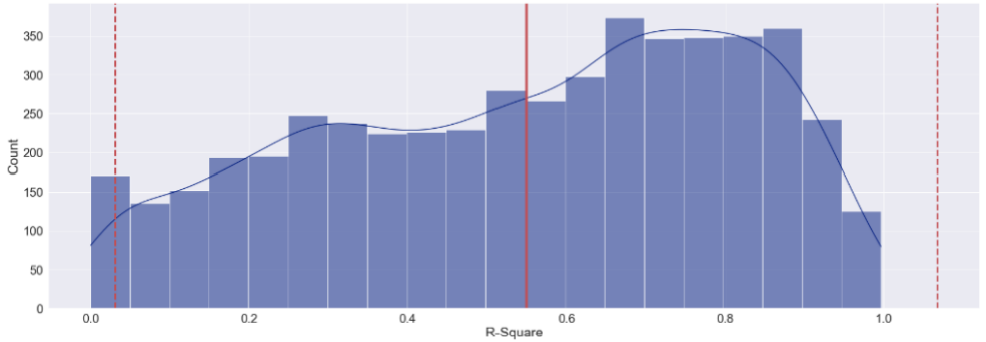





Linear Model generated with Operator and Service Companies




Linear Model generated with Operator companies




Linear Model generated with Service Companies

# Conclusions

- Models to predict the price of a commodity were not generated in this study
- Model that explain the price of a commodity were produced, but of low quality
- Model using commodity sector specific companies produced lower quality models