Forecasting Carbon Emissions for Natural Gas

Adam M. Lang

January 22, 2021

MIT-Applied Data Science Bootcamp

Capstone Project – Final Report

Outline

Part 1

- Problem Overview
- Approach to the Solution
- Key Findings & Insights

Part 2

- Recommendations for Implementation
- Problem and Solution Summary
- Executive Summary

Part 1

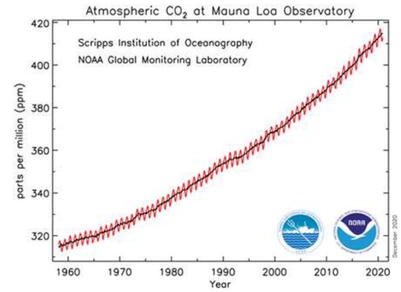
Problem Overview

Approach to the Solution

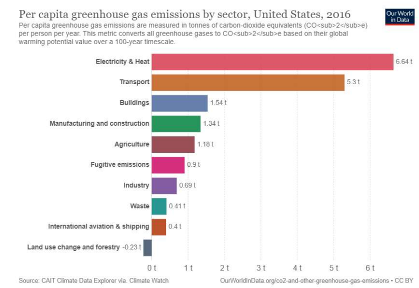
Key Findings & Insights

Problem Overview

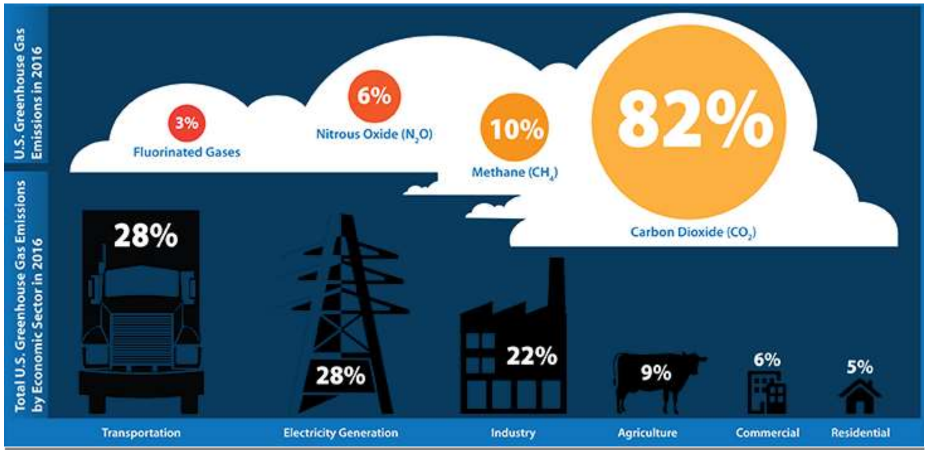
- Climate Change
- Greenhouse Gas Emissions since Industrial Revolution (1760)
- “The CO2 Effect”
  - **82-84% of all greenhouse gases emitted by humans, most damaging to atmosphere**
  - CO2 levels above 400ppm (below 350pm to avoid climate change) (Global Monitoring Lab, 2020)
  - **Fossil Fuels for electricity, transportation, industrial processes generate most CO2 (6.64MMT) - 2016**



CO2 above 400ppm (GML, 2020)



Electricity and Heat dominate CO2 emissions in the USA (Our World in Data, 2016)

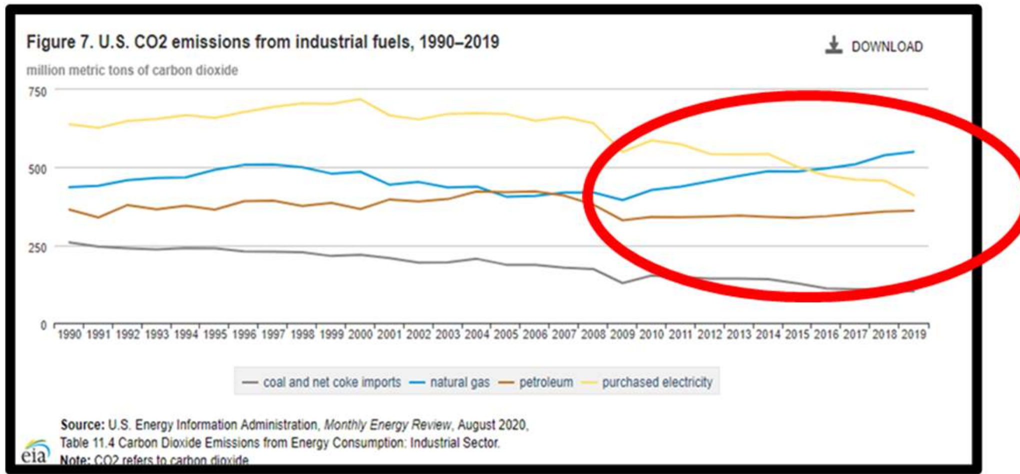


SEIA, 2020

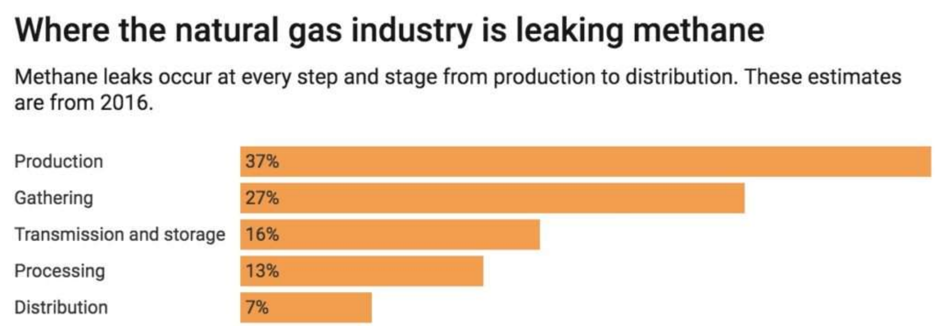
Problem Overview

Natural Gas "Enigma"

- **Industrial natural gas CO2 emissions in the United States have risen most in years since 2009.**
- Increased use has reduced overall US CO2 emissions – least carbon intensive fossil fuel used for electricity generation
- Drop in prices, highly efficient natural gas technology = increased use
- **Methane leakage in gas production and transportation is high**
- → Methane warming effect 80-90 times more powerful than CO2 over 20 year timescale!
- **Pipelines, liquefied natural gas (LNG), hydraulic fracking**
- → CO2 emissions in water, air, land



The Red circle highlights the steady rise in use of natural gas and its CO2 emissions since 2009. Other fossil fuels have plateaued or dropping.



Methane Leakage sources from Natural Gas. (PBS, 2018)

Approach to Solution

Problem Statement:

1. Forecast the carbon emissions value for **natural gas (NNEIEUS)** fuel type for next 12 months (2016-07 to 2017-07)
2. Propose certain measures that can be adopted as policies to reduce these emissions.

Approach

- EIA.gov Data – CO2 Emissions from Energy Consumption in Electric Power Sector
- Forecasting Models in Python:

1. Exponential Smoothing (Holt-Winters)
2. ARMA, AR, MA
3. SARIMA
4. Prophet (Facebook)

- Train, Test, and Cross Validate models on 10-15 years of data
- Compare most accurate model(s) to historical/raw data
- Implementation of Forecast Model Dashboard
- Evaluate Trends in CO2 Emissions and evaluate measures to reduce future CO2 emissions from natural gas

Key Findings & Insights

- Best Models:

1. Prophet
2. SARIMA

*see appendix for accuracy/cross validation metrics

- Prophet predicted 571.440 MMT in CO2 emissions for July 2016-July 2017

➢ **Raw Historical data from EIA.gov: total CO2 emissions for July 2016-July 2017 was 571.537**

➢ Prophet successfully accounted for significant changepoints in time series

➢ 15 years of training data was better than 5 or 10

- Interesting Trends

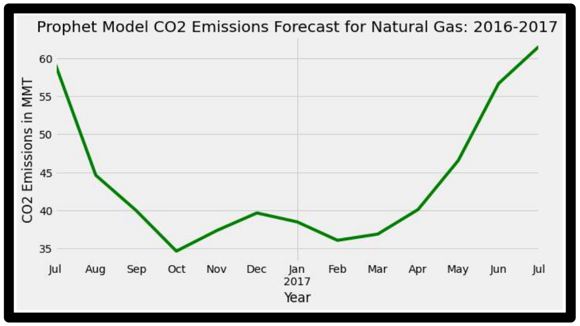
➢ High values in Spring, Summer

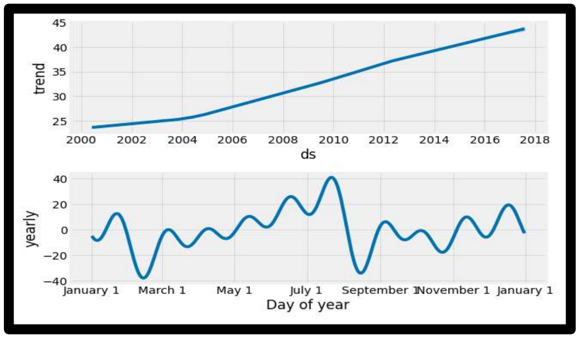
➢ Low values in Fall, Winter

➢ **CO2 emissions will continue to rise!!!**

➢ Peaks and Valleys with Exponential Growth

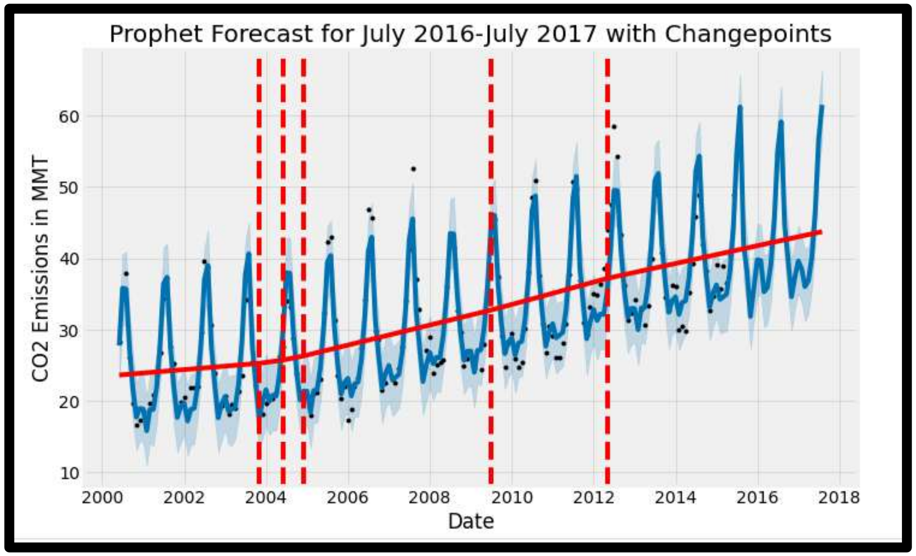
➢ July to July trend different than January to January





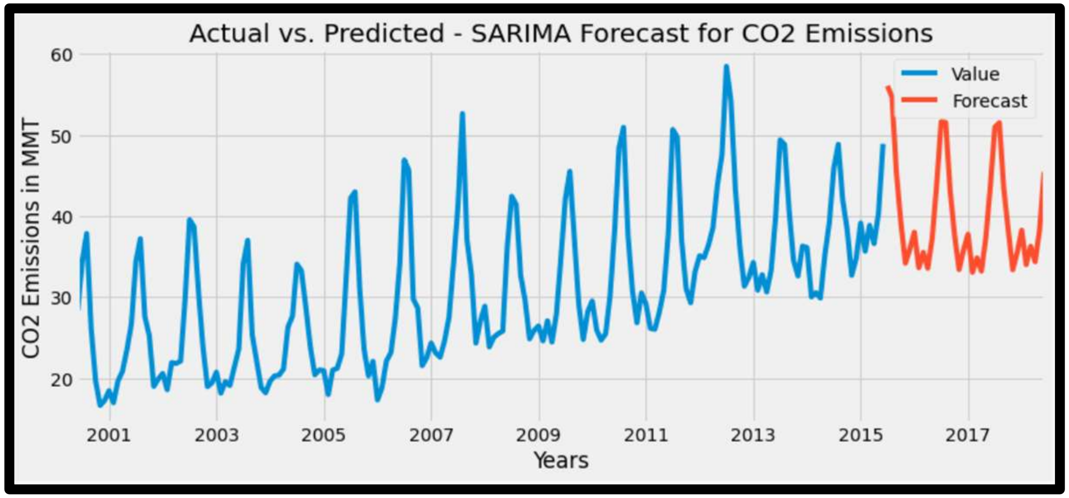
July 2016-July 2017 Forecasts

- Prophet Model



Prophet Model shows continual upward trend with identified changepoints. Notice there are multiple changepoints in the training set.

- SARIMA Model



SARIMA Model forecasted a slight decrease for 2016-2017 compared to the other models.

Part 2

Recommendations for Implementation

Problem and Solution Summary

Executive Summary

Recommendations for Implementation

Key Actions

- Build Dashboards in Tableau
- Build Website/Portal for deployment
- Model Deployment

1. Implementation phase: 1-3 months collaborate with IT
2. Data Warehousing, Code translation for production
3. Production diagnostics on new incoming data
4. Pilot Phase

- Deploy
- A/B Testing
- Troubleshooting

5. Track Model Drift – update data: daily, weekly, monthly?

Cost/Benefit Analysis

- Dashboards, Website construction
- Personnel – Data Science Team, IT Team, Environmental Scientists
- Electrical Energy Sector Private Companies vs. Government (local, state, Federal)
- Educational Resources

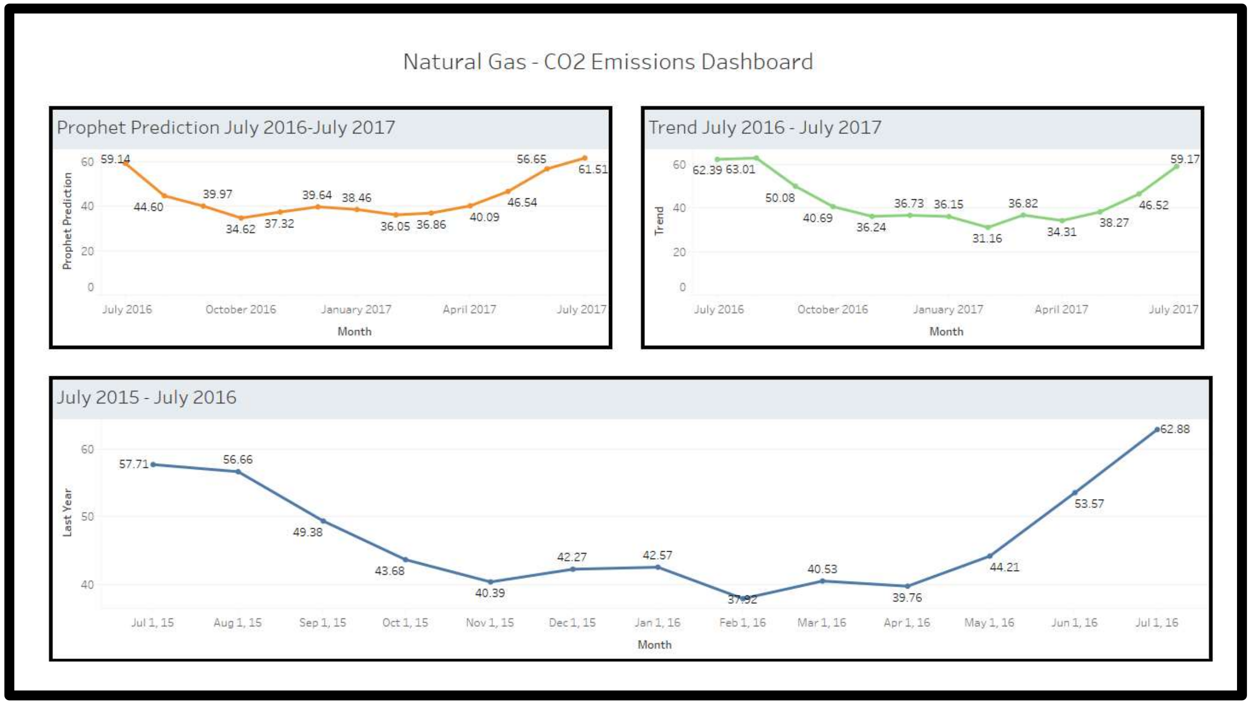
Risks/Challenges

- **Model Drift/Decay**
- Chosen Models lose accuracy/validation on new data
- Code doesn’t transfer to production

Troubleshooting/Issues

- Data – Changepoints, Non-Stationary
- New Data – source? Integrity? Processing?
- Deployment and Production

Sample Dashboard Implementation - Tableau



- Model forecast, Active Trend, Last Year's trend
- Consider interactive plots/filters?

Problem and Solution Summary

Problem Analysis

- **Natural Gas – Electrical Sector CO_2 emissions continue to increase yearly since 2009**
- **CO_2 trend: high in spring/summer, low in fall/winter – not changing!**
- Natural Gas prices are dropping → increased use in USA
- CO_2 leads to increased Methane release - increase climate change!

Proposed Solutions

- **Solar, Wind, Hydroelectric Power (EIA.gov)**
- Carbon Capture and Storage (CCS)
- Advances in drilling and production technology (i.e. fracking)
- **Reduce Methane Leakage in extraction and transportation of natural gas**
- Blending hydrogen with natural gas in pipelines
- **IoT Enabled Gas Sensors (Gomes et al. 2019)**

Why is this a valid solution?

- Evidence Based Solutions to reduce greenhouse case emissions!
- Data Scientists can monitor these solutions and their effects over time

Executive Summary

Key Findings

- CO2 seasonal Trends
- Prophet and SARIMA models ideal for forecasting seasonal and changepoint effects
- **Natural Gas continues to be the only energy source for electrical power that is INCREASING CO2 emissions**
- Dashboard Implementation
- Renewable energy sources (Solar, Wind, Hydro) should be viable alternatives for climate change prevention

Key Next Steps

- Compare Forecast Models to Machine Learning Models?
- Testing and Validation on new data as available
- Multivariate Energy Source Analysis
- **Methane Gas Leakage Data Analysis**

1. **Primary component of natural gas**
2. **Global warming potential 21 times higher than CO2**

References

- Altexsoft (2018) Machine Learning Project Structure: Stages, Roles, and Tools. Retrieved from: https://www.altexsoft.com/blog/datascience/machine-learning-project-structure-stages-roles-and-tools/
- Bhattacharjee, J. (2019) Common metrics for Time Series Analysis. Retrieved from: https://joydeep31415.medium.com/common-metrics-for-time-series-analysis-f3ca4b29fe42
- Center for Climate and Energy Solutions (2020) Natural Gas. Retrieved from: https://www.c2es.org/content/natural-gas/
- EIA.gov (2020) Natural Gas Consumption for Electricity Generation by Sector. Retrieved from: https://www.eia.gov/electricity/data.php
- EIA.gov (2020) Solar, natural gas, wind make up most 2016 generation additions. Retrieved from: https://www.eia.gov/todayinenergy/detail.php?id=25172
- Global Monitoring Laboratory (2020) Monthly Average Mauna Loa CO2. Retrieved from: https://www.esrl.noaa.gov/gmd/ccgg/trends/
- Gomes et al. (2019) IoT-Enabled Gas Sensors: Technologies, Applications and Opportunities. Journal of Sensor and Actuator Networks. Vol. 8(57):1-29.
- Luigi (2019) The Ultimate Guide to Model Retraining. Retrieved from: https://mlinproduction.com/model-retraining/#:~:text=Rather%20retraining%20simply%20refers%20to,t%20involve%20any%20code%20changes.
- Our World in Data (2016) Sector by sector: where do global greenhouse gas emissions come from? Retrieved from: https://ourworldindata.org/emissions-by-sector
- SEIA.org (2020) Climate Change. Retrieved from: https://www.seia.org/initiatives/climate-change
- U.S. Energy Information Administration (2020) Electricity CO2 Emissions Data. Retrieved from: https://www.eia.gov/electricity/data.php

Thank You

Questions?

Appendix

Appendix A: Accuracy and Cross Validation Metrics

Accuracy Measures for all Forecast Models

	RMSE	MAPE
Triple Exp Smoothing	1.296	17.511
SARIMA	2.641	20.176
ARMA	2.565	14.676
Prophet	2.328	19.762

Fig 1. Accuracy Measures for training data

- Exp Smoothing had the best RMSE
- ARMA had the best MAPE

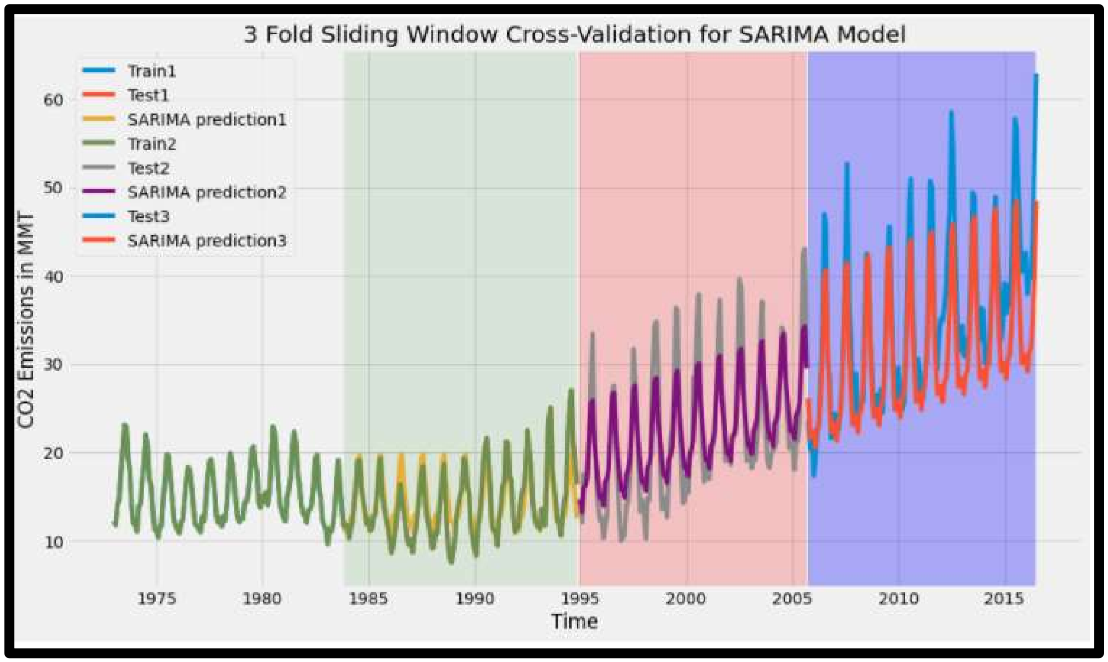
Cross Validation Results

	RMSE	MAPE
Triple Exp Smoothing	4.87	14.60
SARIMA	3.69	12.24
ARMA	9.56	29.88
Prophet	2.16	11.37

Fig. 2. Cross Validation Results

- Prophet had the best RMSE with SARIMA close behind
- Prophet had the best MAPE with SARIMA a close second again

Appendix B: Cross Validation SARIMA



- 3-Fold Sliding Window Cross Validation is seen for SARIMA Model
- Prediction for all 3 folds/windows was nearly spot on to the train and test sets
- **Of note this plot was not able to be produced for the Prophet model as it has its own plotting schema

Appendix C: Total CO2 Emissions Forecast for July 2016-July 2017

	Total CO2 Emissions Forecast July 2016-July 2017
Triple Exp Smoothing	613.229
Actual 2016-2017	571.500
Prophet	571.440
ARMA	562.680
SARIMA	524.420

- Prophet was the most accurate when it came to forecasting the total CO2 emissions for the next year, you can see it was nearly the same as the Actual raw data from EIA.gov.
- SARIMA under predicted while Triple Exp Smoothing overpredicted