Accenture Applied Intelligence

Predicting Energy Usage in California

UC Irvine Data Science Capstone Project - Engaging and preparing university students by focusing on solving real-world challenge facing local community

Frank Dong (UC Irvine), Yang Weng (UC Irvine)

Agenda

01 **UC Irvine Relationship**

02 **Context**

03 **Framework & Approach**

04 **Output & Recommendation**

05 **Enhancement Opportunities**

Copyright © 2022 Accenture. All rights reserved.

01 UC Irvine Relationship

Copyright © 2022 Accenture. All rights reserved.

UC Irvine Outreach Program

Establish close partnership with a local university for recruiting top talent and working with students on projects, which impact our community and clients and amplify our brand awareness

Recruitment

- Career fairs and information sessions at UCI and specialized schools.
- Goal to increase pipeline of diverse and quality prospects.
- Fall recruitment fair is coming up soon!

Campus Events

- Ways to bring influential leaders to campus and encourage thought-provoking conversations and the exchange of ideas.
- Guest Lecturer opportunities available in Merage's MBA and MSBA programs.
- Build a community of UCI Alumni within Accenture to support all the activities.

Capstone Program

- Accenture sponsors an undergraduate- or graduate-level capstone project in Merage and ICS every Winter Quarter.
- Accenture produces project ideas and students “build” the project over two quarters.
- Student gains hands-on, real-world experience in addition to soft-skills.
- AI for Good initiative.

Value

Increase quality pipeline for recruitment

Create a positive impact in local community

Work with students to coach the future generation

Create opportunities to build AI Powered Solutions

Public speaking events to increase brand awareness

FJORD
Design and Innovation from Accenture Interactive

Copyright © 2022 Accenture. All rights reserved.

Our Journey

Feb 2021

Predicting Wildfire in California

Apr 2021

Merage Analytics Club Speaker Event

Nov 2021

Guest Lecturer for Graduate Students Machine Learning

July 2022

Entry-Level Applications Promoted to UCI

Feb 2022

Recruitment Information Session

Feb 2022

Predicting Risk for Diabetes in California

Oct – Dec 2022

Fall Recruitment Events

Jan 2023

Student Leadership Program Engagements

April 2023

Spring Recruitment (if demand exists)

Copyright © 2022 Accenture. All rights reserved.

Our Team

Martin Hodgett

Manish Dasaur

Vishrut Chokshi

AAI Resources West

Shawna Tuli

AAI Microsoft West

Jeff Lu

T&O West

Ella Liang

CDP West

Moury Bidgoli

Cyber Threat Intelligence Analyst

Copyright © 2022 Accenture. All rights reserved.

02 Context

Copyright © 2022 Accenture. All rights reserved.

California Sets Goal Of 100 Percent Clean Electric Power By 2045

Energy Landscape in California 2022:

- California just ran on 100% renewable energy at peak generation for two days
- In 2022, renewable resources, including hydroelectric power and small-scale, customer-sited solar power, accounted for 49% of California in-state electricity generation. Natural gas fueled another 42%. Nuclear power supplied almost all the rest.

This community challenge led us to asking two key questions:

1. **How can we employ a data and AI based approach to identify and predict energy consumption to plan the future?**
2. **How can we turn this study and predictive model into a research-based credential?**

Copyright © 2022 Accenture. All rights reserved.

UC Irvine Data Science Capstone Course

Undergraduate Senior Capstone Project Design and Development

14 WEEKS

Project-based course for capstone students

FEB 21 – JUNE 5

Starting our engagement with students on Feb 16th

This course allows senior students to solve a substantial real-world problem.

Accenture is a sponsor to a project which helps our local community and people.

FY 2023 Challenge – Predicting Energy Consumption in California

FY 2022 Challenge – Predicting Risk of Type 2 Diabetes with AI in California

42

Undergraduate Computer Science students enrolled in course

50%

Computer Science students are minority, female or first-generation

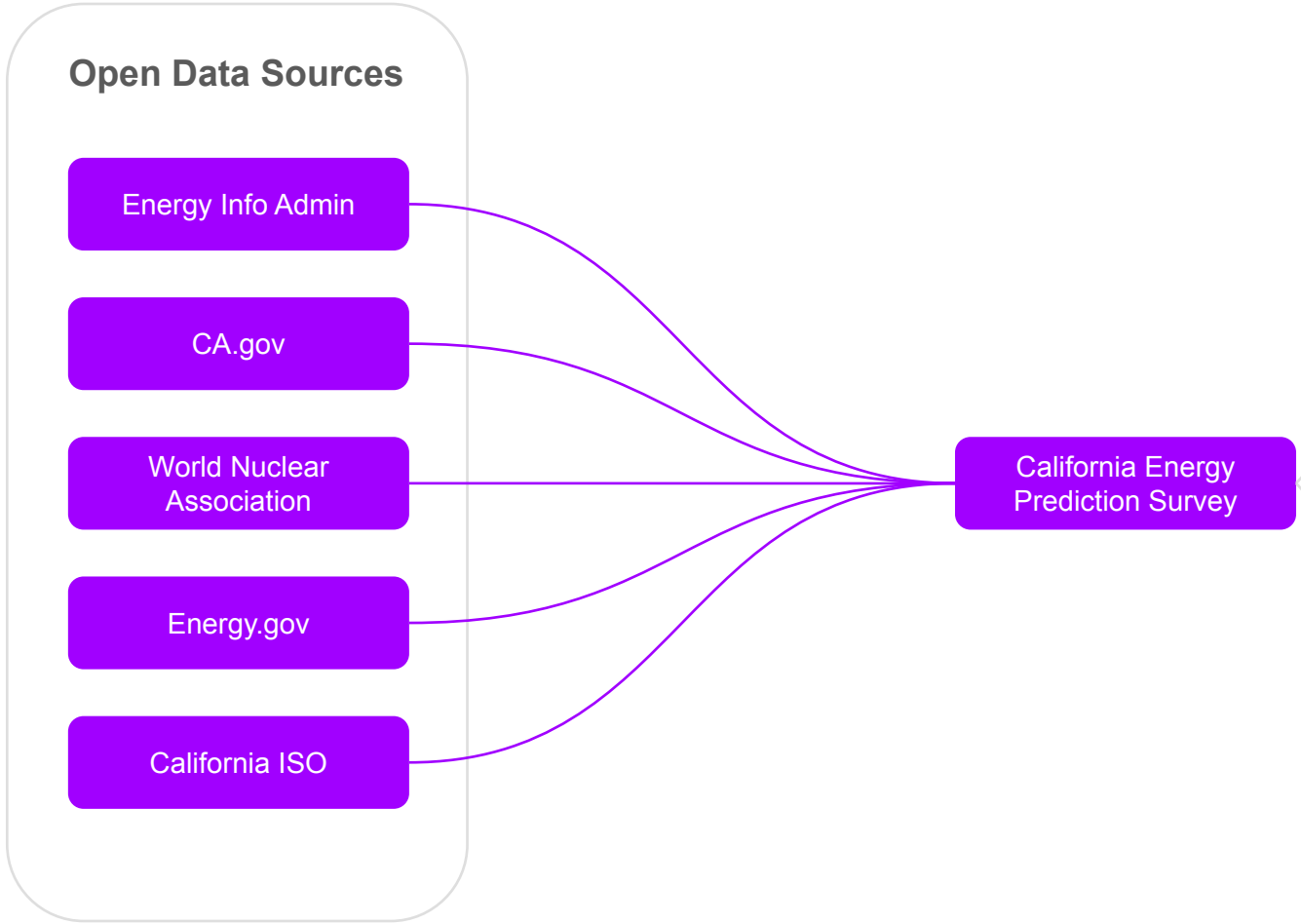
Copyright © 2022 Accenture. All rights reserved.

03 Framework & Approach

Copyright © 2022 Accenture. All rights reserved.

Evaluation of Publicly Available Data

Data is at the heart of AI and data quality is key to successful AI



300+ Attributes from 5 different sources and 4 different sectors.

Sources:

- Natural Gas
- Renewable Energy
- Coal
- Nuclear
- Petroleum

Sectors:

- Transportation
- Industrial
- Residential
- Commercial

Copyright © 2022 Accenture. All rights reserved.

Solution Approach

California Energy Consumption Survey

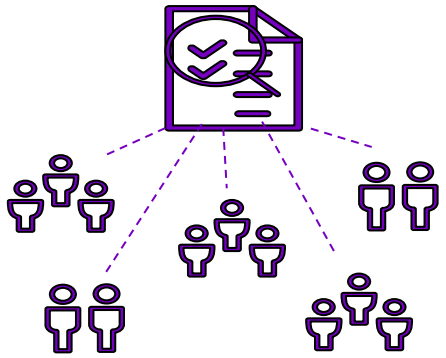
- Energy Price
- Renewable Energy
- 60 Years Long Term Data
- Model Combination and Validation
- Related work

300+ FEATURE GENERATION

MACHINE LEARNING ALGORITHM

Predictive Model with Test Data Validation

Draft for Research Paper



Accenture approves before published to PAPER

***Abstract submitted. Detailed Research Paper in progress.**

Copyright © 2022 Accenture. All rights reserved.

04 Output & Recommendation

Copyright © 2022 Accenture. All rights reserved.

Important Feature Explanation

DFSCB: Distillate fuel oil total consumption adjusted for process fuel

ESSCB: Electricity total consumption adjusted for process fuel

NGRCV: Natural gas expenditures in the residential sector

OPISB: Other petroleum products consumed by the industrial sector excluding refinery fuel and intermediate products

TEPFB: Total energy used as process fuel and other consumption that has no direct fuel costs

TNSCB: Total net energy consumption

WWIXB: Wood and waste consumed by the industrial sector, at no cost

Data Preparation

The first and most important step in an analysis is data preparation. Students followed a 3-step process to obtain the data, format the data, and transform the data so that insightful conclusions could be made.

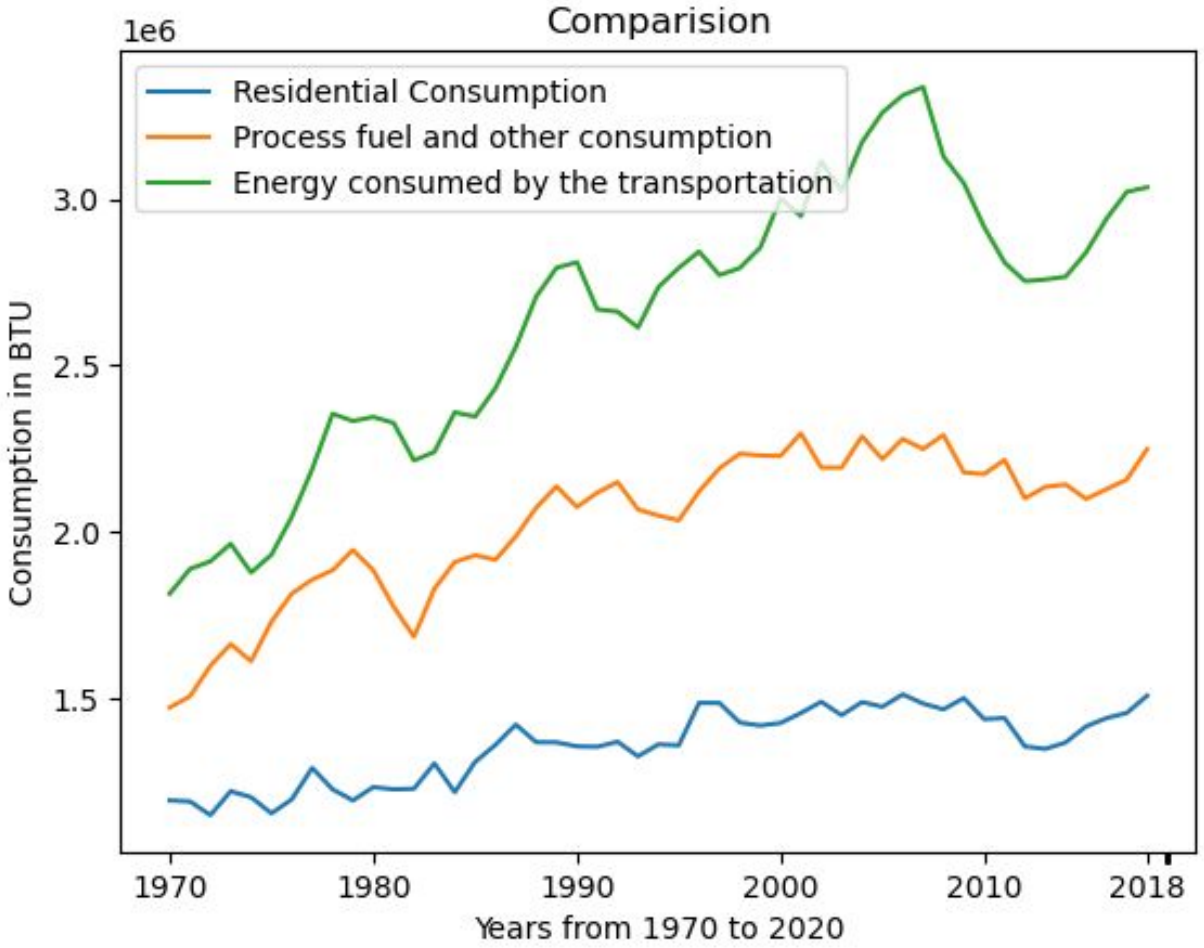
1. **Data Collection**
   - Gathering and Combining Data Together
2. **Data Preprocessing**
   - Formatting
   - Cleansing
3. **Data Normalization**
   - Scaling and normalizing

	1970	1971	1972	1973	1974
ARICD	0.49	0.64	0.64	0.69	1.47
ARICV	39.0	51.9	52.3	60.5	115.9
ARTCD	0.49	0.64	0.64	0.69	1.47
ARTCV	39.0	51.9	52.3	60.5	115.9
ARTXD	0.49	0.64	0.64	0.69	1.47
...	...	...	...	...	...
WXICV	10.5	10.1	11.0	15.5	15.2
ZWCDP	748.0	738.0	748.0	681.0	750.0
ZWHDP	3169.0	3690.0	3278.0	3396.0	3297.0
ResidCA	1,046,492	1,167,142	1,192,848	1,189,801	1,148,722
ResidUS	13,728,645	14,210,554	14,861,651	14,919,481	14,651,895

Copyright © 2022 Accenture. All rights reserved.

Exploratory Data Analysis

The below chart show the sample relationship between response variable and target variable (Residential Energy Consumption as shown in example).



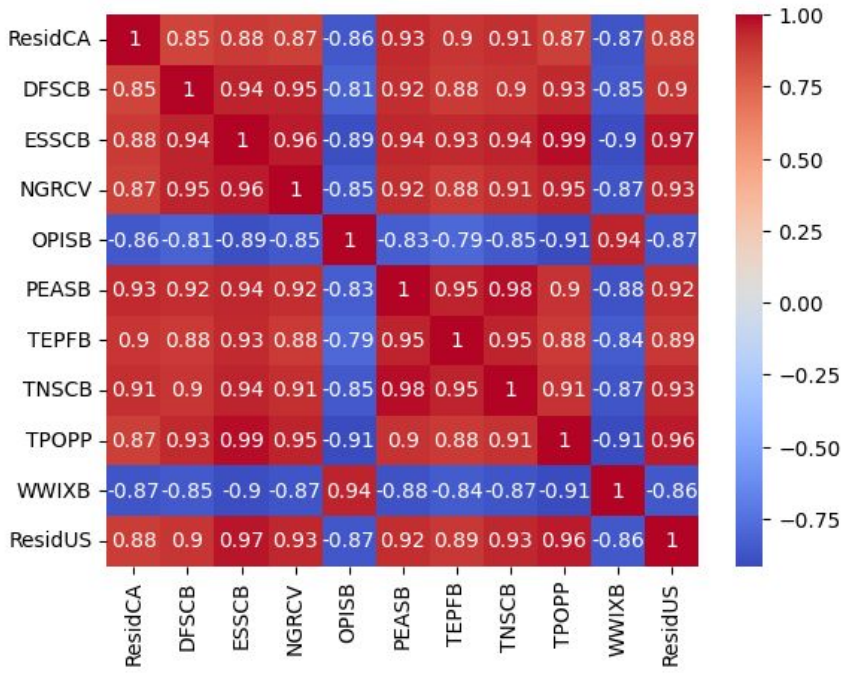
alt text

Copyright © 2022 Accenture. All rights reserved.

Correlation Analysis

Correlation analysis measures the relationship between variables, unveiling connections that help interpret and predict their behavior.

Two year Residential Sector Prediction Sample



Ways to deal with Correlation:

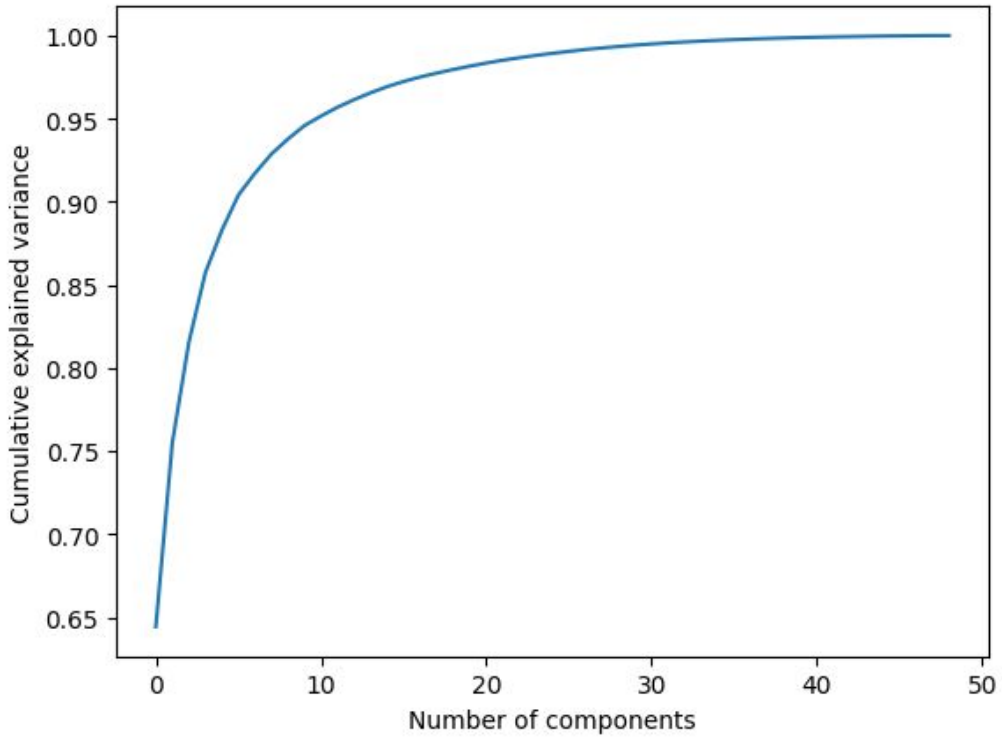
- **High correlation feature selection**
  - Select features with high correlation with target (>0.85)
- **Duplicate feature**
  - - TNASB & PEASB
  - - OPSCB & OPISB
- **Manual clustering**
  - Combine diesel fuel, total consumption …… feature together by PCA.

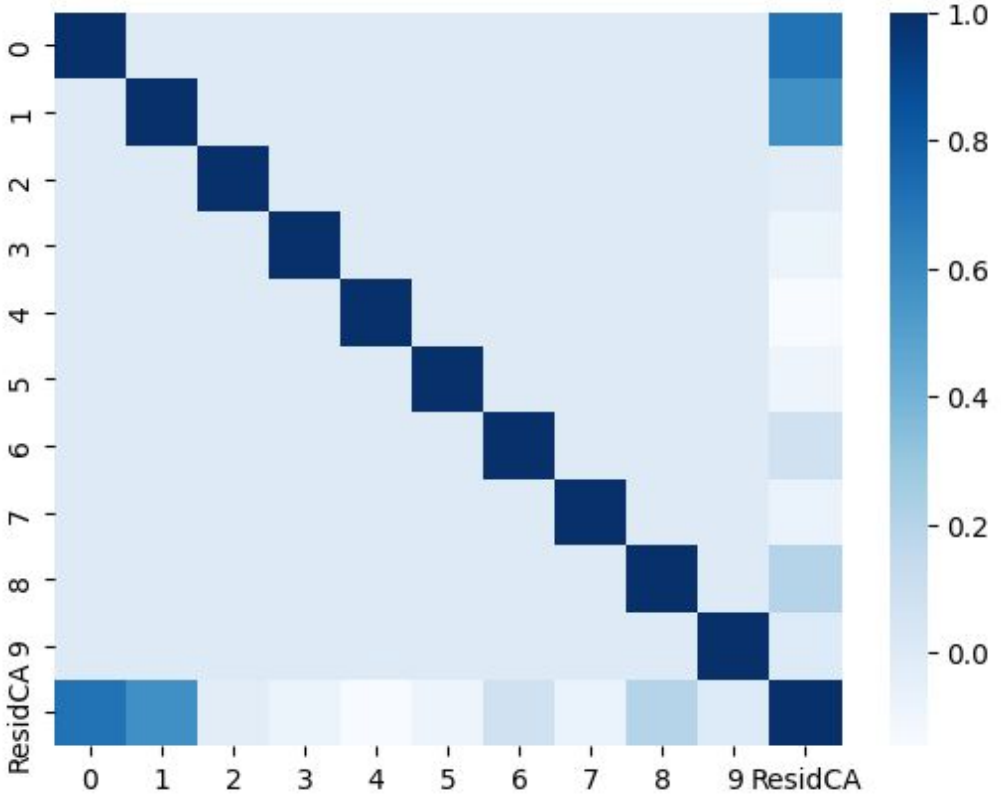
Copyright © 2022 Accenture. All rights reserved.

PCA Analysis

Principal Component Analysis (PCA) is a technique used to simplify complex data sets by reducing their dimensions while retaining key information.

Select 10 as hyperparameter for PCA analysis





Copyright © 2022 Accenture. All rights reserved.

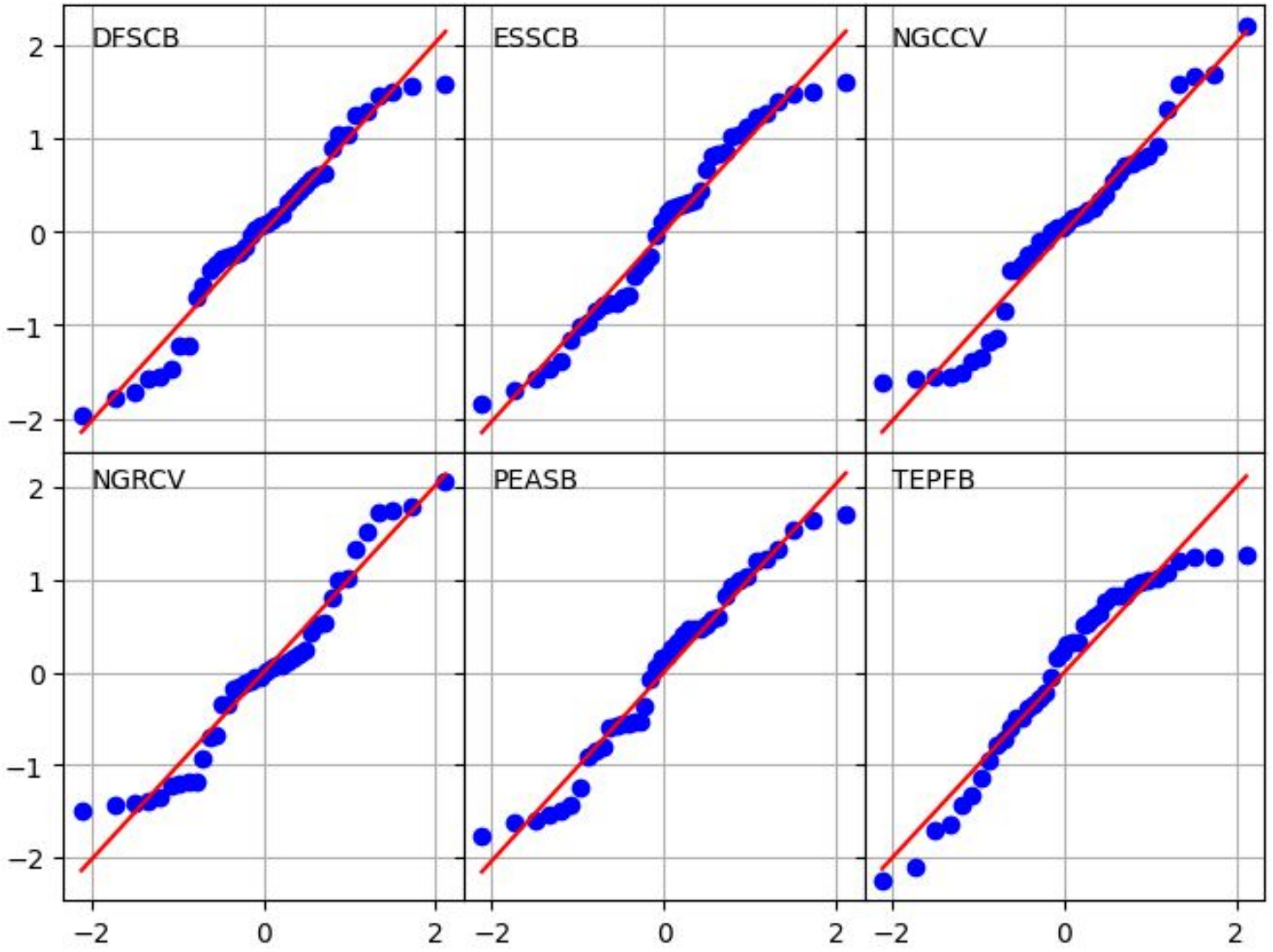
SARIMAX

SARIMAX, short for Seasonal Autoregressive Integrated Moving Average with Exogenous Variables, is a time series forecasting model that extends the traditional ARIMA model to incorporate seasonality and exogenous variables. It is a powerful tool for analyzing and forecasting time series data that exhibits seasonal patterns and is influenced by external factors.

SARIMAX is generally preferred when the data exhibits clear seasonal patterns and there are known external factors that influence the time series.

It relies on the assumption that future values depend on past values and can capture linear relationships.

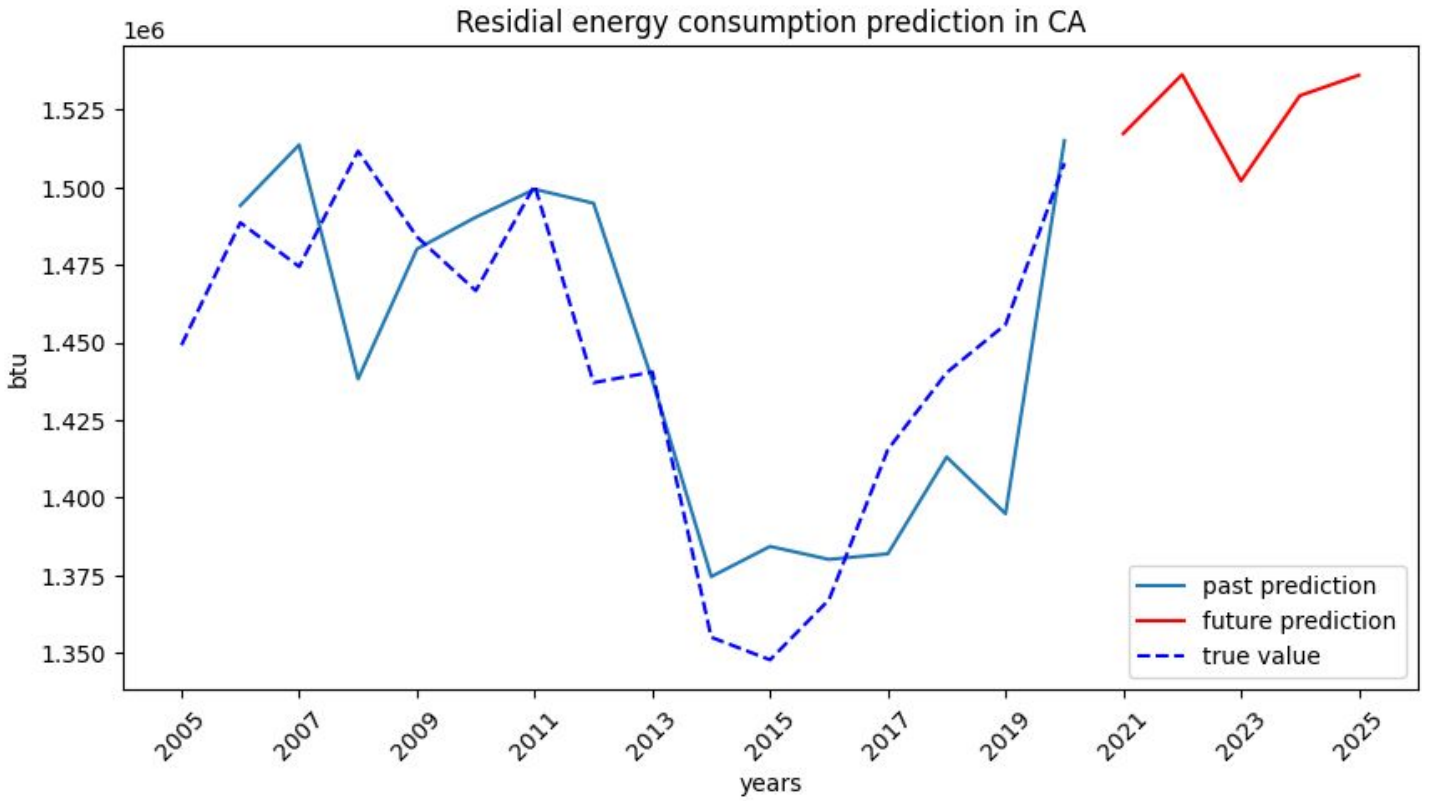
SARIMAX Assumption



SARIMAX Prediction

5 year prediction with validation

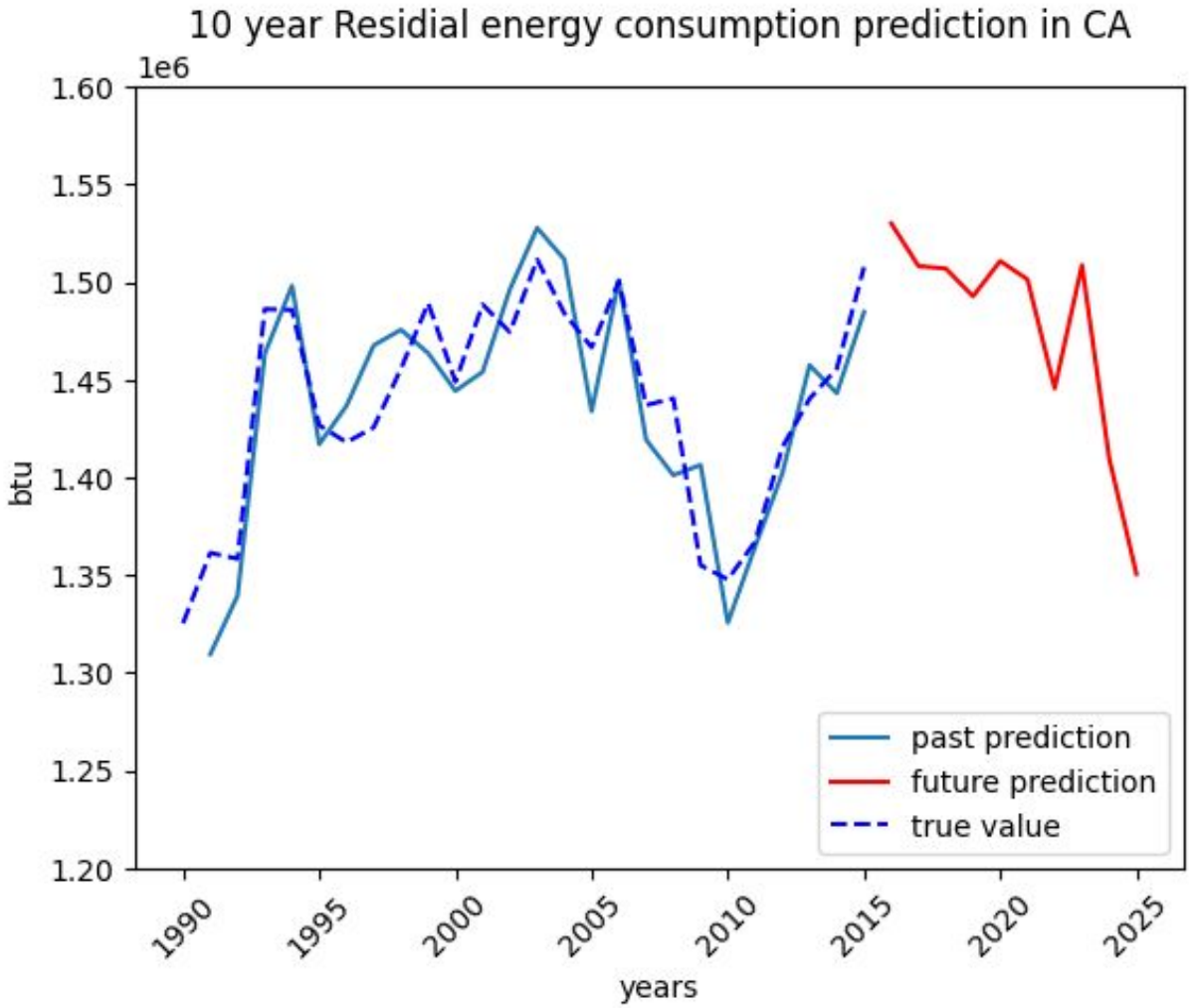
 # Validation error rate 1%



SARIMAX Prediction

10 year prediction and validation

 # Validation error rate 3%



Key Outcomes

This analysis identified an accurate model for yearly data prediction with a seasonal pattern and identified top factors.

The most accurate **model** for yearly prediction

- SARIMAX performed the best over other stats models and machine learning models
- Able to apply this model on other energy source and sectors and even other states. (Data trained based on the US.)
- Residential energy consumption remain on a high level, for 5 years and will have a sharp decrease in 10 years.

Top **features** for Residential Energy Consumption

- Top 6 features:
  - Resident population
  - Electric total consumption
  - Natural gas expenditures
  - Total net energy consumption
  - Natural gas expenditures in residential sector
  - Natural gas expenditures in commercial sector

Natural Gas is still the most important energy source for CA residents' daily lives.

Copyright © 2022 Accenture. All rights reserved.

BMJ Publication

We are working on submitting a paper on this analysis to the Journal. If accepted, Accenture will gain major credentials for our work in Energy Prediction analytics.

- 3,000-word manuscript
- 300-word abstract
- Reviewed by a scientific panel where we submitted 3 reviewer preferences
- Shawna Tuli is first author with Vishrut Chokshi, Marty Hodgett, Manish Dasaur, Dalal Sachin, Michelle Smith, Praneet Raj, Bryan S. Price, Erik Kronstadt, Anthony Agbasi, Mei-Chi Chen, Nile Hanov, Jeffrey Vu, Jeffrey Lu, Ella Liang, Moury Bidgoli, Frank Dong (UCI) , Yang Weng(UCI) as co-authors

Copyright © 2022 Accenture. All rights reserved.

05 Client Engagement Opportunities

Copyright © 2022 Accenture. All rights reserved.

How this matters for clients

This analysis shows how energy data can be compiled and analyzed to provide actionable insights for policy maker and Energy company.

Develop and deliver a comprehensive Energy saving target solution

- Combines several different datasets both publicly and privately available that encompasses the many factors in Energy
  - EX: government data, Energy company data, emissions data
- Predicts several different sources and sectors outcomes

Value for clients

- Enable clients to reach their strategic goals by approaching Energy data holistically and innovatively
- Empower departments to better utilize their human resources to focus on specific areas
- Public Services use case – use for city planning
- Company use case – conduct different sectors and sources segmentation analyses to create new frameworks, strategies, and processes for Energy generate

Copyright © 2022 Accenture. All rights reserved.

Question & Answer

Copyright © 2022 Accenture. All rights reserved.

APPENDIX

Copyright © 2022 Accenture. All rights reserved.

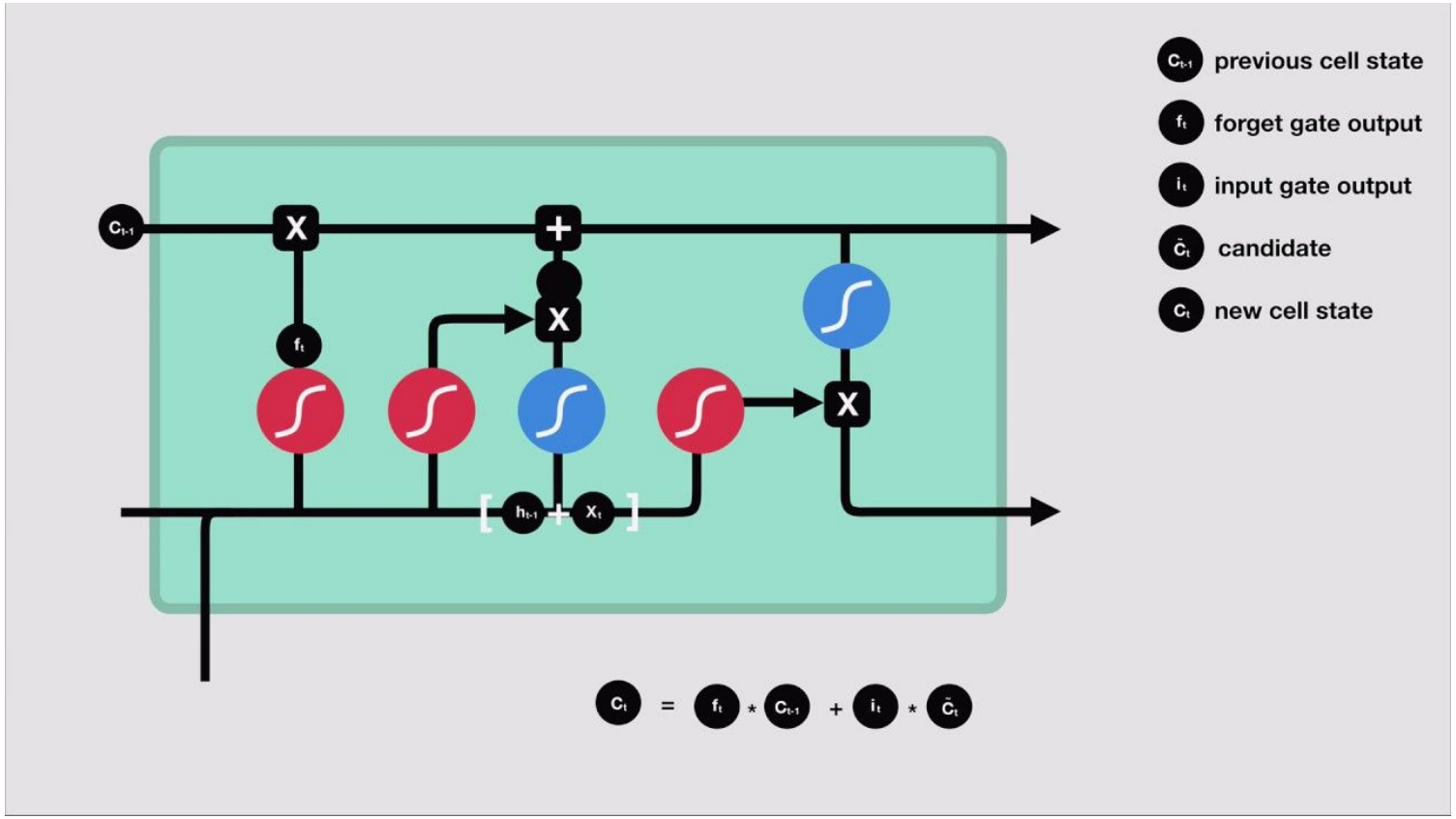
LSTM

LSTM, short for Long Short-Term Memory, is a type of recurrent neural network (RNN) architecture that has gained significant popularity in the field of deep learning, particularly for modeling sequential data. It is designed to address the vanishing gradient problem of traditional RNNs, allowing it to capture long-term dependencies in time series or sequential data.

LSTM, on the other hand, is a type of recurrent neural network (RNN) that can model complex nonlinear relationships in time series data. It can capture long-term dependencies and has the ability to learn and extract features from raw data without manual feature engineering.

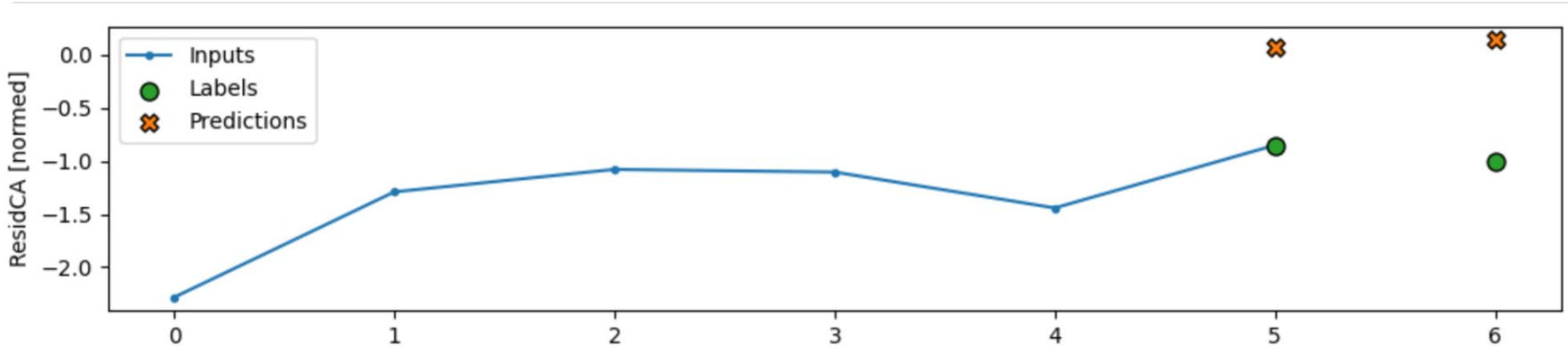
Does not fit for small seasonal data, causes underfitting due to yearly data.

LSTM


C_{t-1} previous cell state
f_t forget gate output
i_t input gate output
$\tilde{C}_t$ candidate
C_t new cell state
$[h_{t-1}, x_t]$
$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$

LSTM Prediction & Validation

2 year prediction

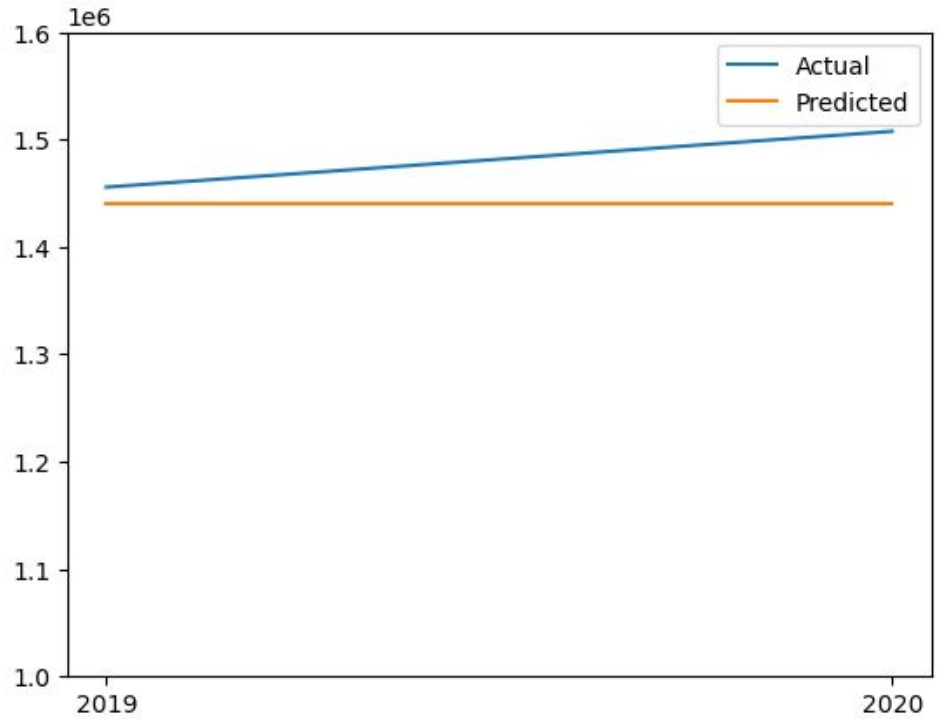


Decision Tree Prediction & Validation

2 year prediction

MAPE: 0.027555422500944136

	ResidCA	predictions
2019	1455751	1440463.0
2020	1507721	1440463.0

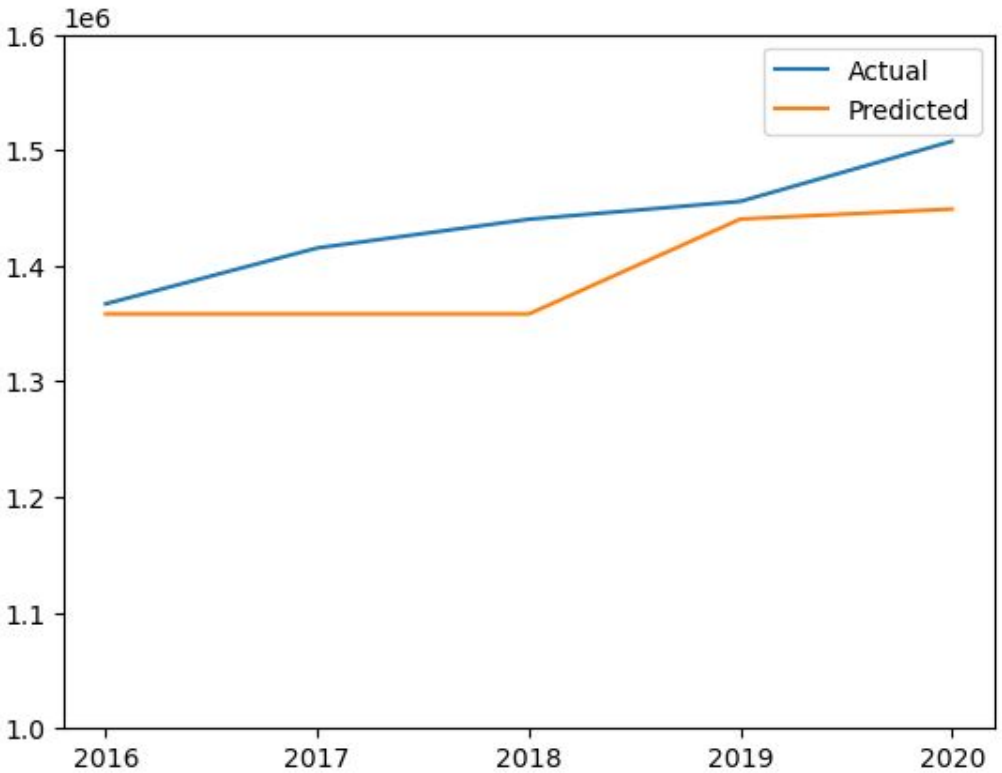


Decision Tree Prediction & Validation

5 year prediction

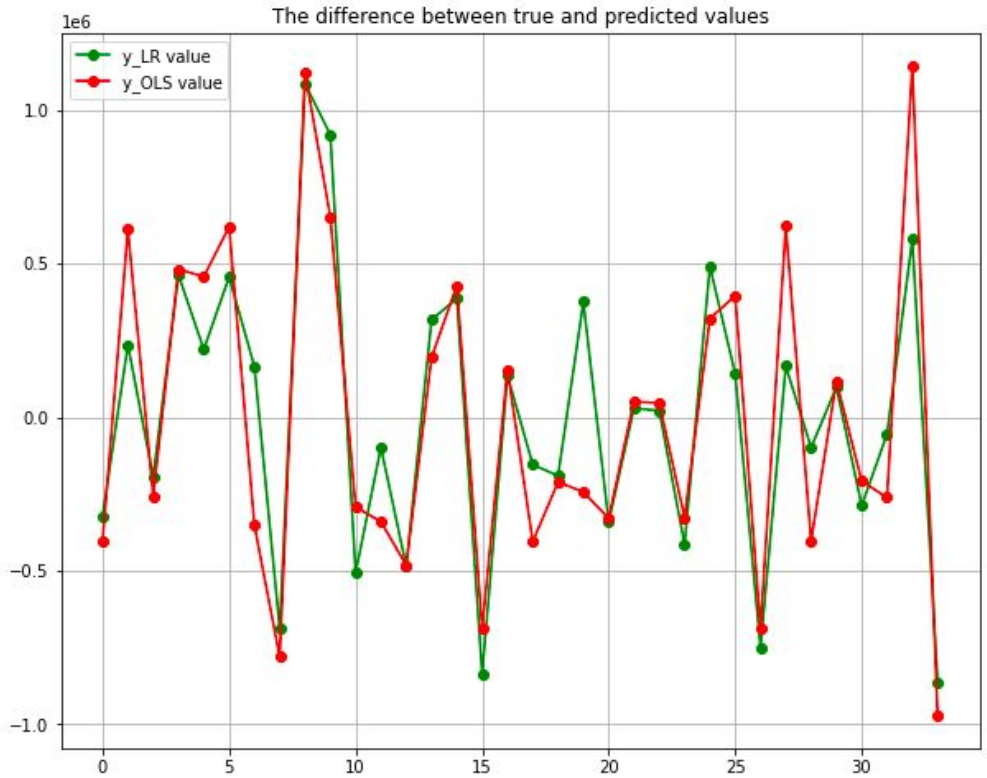
MAPE: 0.030588452843416 23

	ResidCA	predictions
2016	1367103	1358381.0
2017	1415455	1358381.0
2018	1440273	1358381.0
2019	1455751	1440463.0
2020	1507721	1449101.0

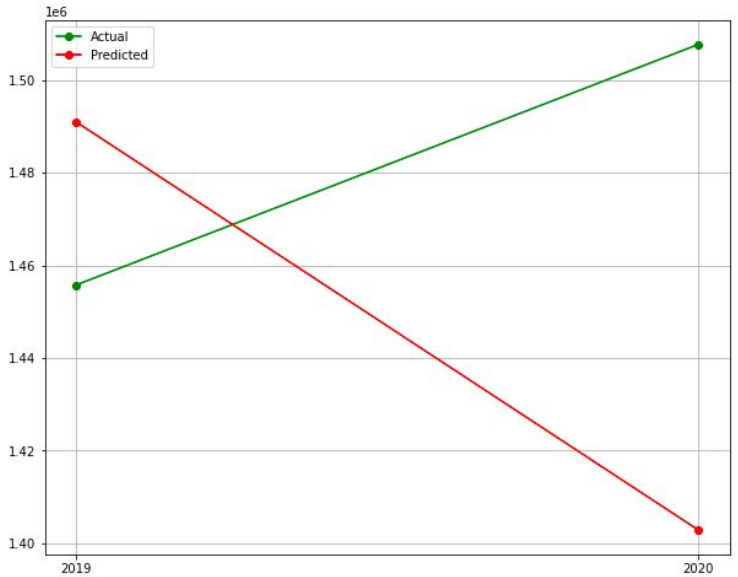


LINEAR REGRESSION PREDICTION & VALIDATION

- The regression was built using the OLS model that is the statsmodel and compared with a LinearRegression model.
- We compare the difference between the predicted and true values of the two models and draw line graphs of the values.
- The difference between the predicted and true values of the LinearRegression model is smaller and the model is more accurate.



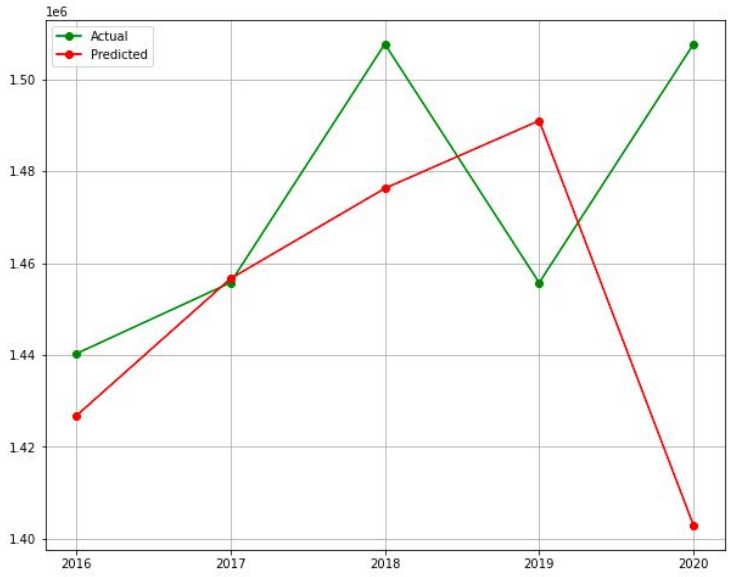
LINEAR REGRESSION PREDICTION & VALIDATION



	ResidCA	predictions
2019	1455751	1491031.12
2020	1507721	1402875.88

2 year prediction

MAPE: 0.04688689937953451



	ResidCA	predictions
2016	1440273	1426727.09
2017	1455751	1456631.79
2018	1507721	1476283.08
2019	1455751	1491031.12
2020	1507721	1402875.88

5 year prediction