International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 13 Issue 9, September 2024

B. Methodology

We first Visualize the data as follows:

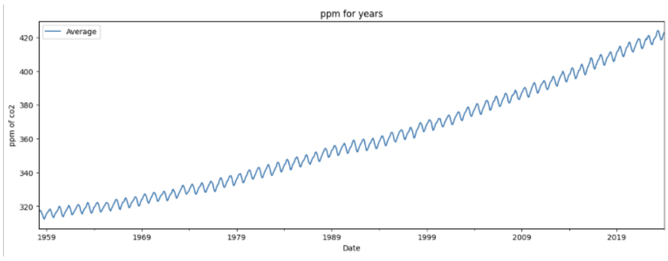

Figure 1: plot of the average levels of ppm levels for co2 from 1959 to 2004

As we have observed from the dataset, we need a methodology to predict the future levels of co2 in the atmosphere. In order to do this, we make use of the time seriesconcept in machine learning. The main objective of time seriesanalysis is predicting the future while training it on the presentdata on hand, which the predicted data can be used by various industries to help predict their future outcomes or to prevent future collapses based on the future predictions.

We are predicting thefuture CO2 levels of the atmosphere.

Two models related to time series forecasting in this paper namely, “ARIMA” and the “SARIMAX” model. The Arima model needs nonstationary data, while the SARIMAX model uses the Seasonal data.

We can check the Seasonality of the data using the stationarity test, here the ADF (Augmented Dicky - Fuller) test.When we apply for this test, we get the values of p as 1.0. So, this isnon-Stationary as for it to be stationary the p value should be less than 0.05. So, we apply the stationarity tests for it but shifting the order by a value of 12 (this 12 indicates the data has a 12-month duration in a year) and when we apply the test on the new data again, it indeed shows the data is stationary now.

But the ARIMA model doesn't give a proper prediction to the data on hand and gives an inaccurate prediction.

So, we turned to the SARIMAX model. The main advantageof the SARIMAX model is that it doesn't depend on the Seasonality of the data as mentioned before. That means we don't need to convert to stationary data. We can just give the data to the model, and it will be able to give us the required output.

An important tool we used here is auto Arima from the pmdarima library. With the Advancements in the Machine Learning process, instead of manually testing the various p q and d values, we can use the auto_arima code to run and generate all the various combinations in the SARIMAX model. It then gives us the most accurate model parameters to run in the model. We get the parameters for this, using the summary to get the information .

```
                                     SARIMAX Results
==========================================================================================
Dep. Variable:                            Average   No. Observations:                  632
Model:             SARIMAX(2, 1, 1)x(1, 0, 1, 12)   Log Likelihood                -166.909
Date:                            Sun, 21 Jul 2024   AIC                            345.818
Time:                                    16:22:39   BIC                            372.502
Sample:                                03-01-1958   HQIC                           356.182
                                     - 10-01-2010
Covariance Type:                              opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1          0.4003      0.040     10.050      0.000       0.322       0.478
ar.L2          0.0991      0.038      2.587      0.010       0.024       0.174
ma.L1         -0.7332      0.038    -19.497      0.000      -0.807      -0.660
ar.S.L12       0.9997      0.000   3112.585      0.000       0.999       1.000
ma.S.L12      -0.8753      0.022    -38.918      0.000      -0.919      -0.831
sigma2         0.0911      0.005     18.306      0.000       0.081       0.101
===================================================================================
Ljung-Box (L1) (Q):                   0.07   Jarque-Bera (JB):                 0.58
Prob(Q):                              0.79   Prob(JB):                         0.75
Heteroskedasticity (H):               1.05   Skew:                            -0.04
Prob(H) (two-sided):                  0.73   Kurtosis:                         3.13
===================================================================================

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).
```

Figure 2: Result obtained after executing the sarimax model on the chosen dataset

We get the parameter values from this dataset namely

- p = 2 (Auto Regressive Component)
- q = 1 (Moving Average Component)
- d = 1 (Integrated Component)
- P = 1 (Seasonal Auto Regressive Component)
- D = 0 (Seasonal Integrated Component)
- Q = 1 (Seasonal Moving Average Component)
- s = 12 (Seasonal Period)

Now we split the data into train and test where 80% is for train and 20% is for test. We then train the train part using he model we made and then we apply this model to the testdataset. We then plot the test dataset and the actual dataset which when plotted we get a pretty accurate plot

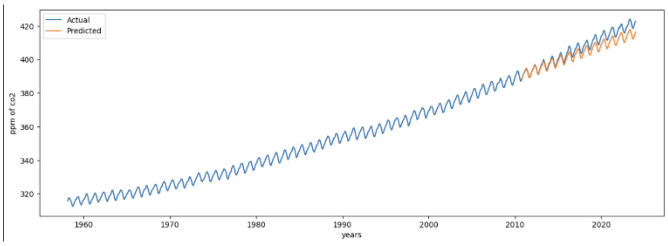

Figure 3: plot of the taken dataset with the predicted data using the sarimax model

As now we get a proper prediction, we then make the future predictions using the model on hand. We make the a few future dates and then assign then the values according to the time series forecast of the SARIMAX model.

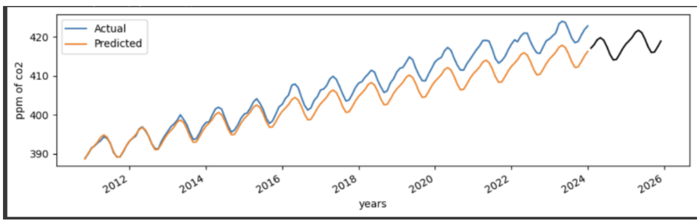

Figure 4: plot of the chosen dataset and the predicted data from sarimax model with the future predictions of the same predicted data

IV. RESULTS AND DISCUSSION

We tried applying both SARIMAX and ARIMA for the CO2 forecasting. From our experiments, we found that ARIMA doesn't give accurate predictions, with SARIMAX proving to be notably better.The application of the SARIMAX(Seasonal Autoregressive Integrated Moving Average with Exogenous Factors) model for forecasting has demonstrated its effectiveness in capturing the underlying patterns and seasonality in the historical data. Through careful preprocessing and differencing, the data was made stationary, which is a prerequisite for effective time series modeling.

The SARIMAX model with parameters (2,1,1)x(1,0,1,12) was fitted to the training dataset and subsequently validated using a test dataset. The model's predictions showed a good fitwith the actual observed values, indicating that it successfully captured the seasonal trends and cyclical behavior of CO2 levels over time. Additionally, the model successfully forecasted future CO2 levels up to December 2025, providing valuable insights into potential future trends.

V. CONCLUSION AND FUTURE PROSPECTS

In our study on CO2 forecasting using the SARIMAX and ARIMA models, we showed that SARIMAX and ARIMA models can capture historical CO2 data trends and seasonality in our CO2 forecasting investigation. Both models successfully predicted future CO2 levels after thorough adjustment and verification. Through this we have achieved the main objective of our research.

In the future, we plan to improve the accuracy of SARIMAX and ARIMA models by experimenting with different seasonal and non-seasonal parameters. We will consider incorporating additional exogenous variables, such as economic indicators, industrial activity data, and policy changes, to improve predictive power. Extending the forecast horizon beyond 2025 will provide long-term insights crucial for policy planning and climate action strategi

To capture subtle data patterns and provide more reliable predictions, we plan to integrate these models with Prophet, LSTM, or other deep learning methods. Our objective is to distribute the findings of our CO2 prediction to the general public in order to increase understanding of the current patterns in greenhouse gas discharges and the critical need to tackle climate change. Our model's projections can be utilized by educational campaigns to demonstrate hypothetical future scenarios and underscore the significance of sustainable behaviors.

REFERENCES

[1] M. Satish, Prakash, S. M. Babu, P. P. Kumar, S. Devi and K. P. Reddy, "Artificial Intelligence (AI) and the Prediction of Climate Change Impacts," 2023 IEEE 5th International Conference on Cybernetics, Cognition andMachine Learning Applications (ICCCMLA), 2023, pp. 660- 664.

[2] J. Cowls, A. Tsamados, M. Taddeo and L. Floridi, "The AI gambit: leveraging artificial intelligence to combat climate change—opportunities, challenges, and recommendations.", Ai& Society, 2021, pp. 1-25.

[3] S. Fathi, R. S. Srinivasan, C. J. Kibert, R. L. Steiner and E. Demirezen, "AI-based campus energy use prediction for assessing the effects of climate change", Sustainability, vol.12,no.8. 2020, pp. 3223.

[4] A. Dikshit, B. Pradhan and M. Santosh, "Artificialneural networks in drought prediction in the 21st century–A scientometric analysis." Applied Soft Computing, vol.114, 2022, pp. 108080.

[5] M. Algarni, "Deploying Artificial Intelligence for Optimized Flood Forecasting and Mitigation", 2023 20th ACS/IEEE International Conference on Computer Systems andApplications (AICCSA), IEEE, 2023, pp.1-6.

[6] A. Ilapakurti, S. Kedari, R. Vuppalapati, S. Kedari, J. S.Vuppalapati and C. Vuppalapati, "Human-AI Symbiosis: Decode Climate Change to Prevent Heat-Related Mortalities and to Protect Our Most Vulnerable Population", 2019 IEEE International Conference on Computational Science and Engineering (CSE) and IEEE International Conference on Embedded and Ubiquitous Computing (EUC), IEEE, 2019, pp.331-338.