**Project Title**

**Global Temperature Change Prediction**

**School of Computer Science Engineering and Technology**

**A logo of a university

Description automatically generated**

|  |
| --- |
| Bennett University  Greater Noida, Uttar Pradesh |

Submitted by- Submitted to-

Ishita Diwedi (E22CSEU0575) Dr. Mala Saraswat

Aannya Gupta (E22CSEU0632)

Dev Vishnu Srivastav (E22CSEU0574)

**Introduction and Objective**

Global climate change has already had observable effects on the environment. Glaciers have shrunk, ice on rivers and lakes is breaking up earlier, plant and animal ranges have shifted and trees are flowering sooner.

Effects that scientists had predicted in the past that would result from global climate change are now occurring: loss of sea ice, accelerated sea level rise and longer, more intense heat waves and so on.

So far, global warming has been limited to a rise of around 0.75C since the end of the 19th century. This sounds like a small change, but scientific evidence suggests it is already leading to a range of impacts around the world.

More importantly, however, is that the temperature rise observed so far is not the chief cause of concern. More worrying is the significantly larger temperature rise (and associated changes in rainfall, snowfall, sea level and other phenomena) that scientists expect the world to experience in the coming decades and centuries.

Temperature plays an important role in shaping weather patterns, guiding the life cycle of various organisms and maintaining ocean levels. Shifting the temperature a couple of degrees can throw an entire ecosystem into chaos.

We, as Computer Science students, can contribute to climate change and global warming by predicting future temperature changes and making others aware of the consequences they will need to face. We have taken up this topic to dwell upon the rising concern of global warming and to gain insights from data that is available to us in the real world.

**Data Collection, Processing and Preprocessing**

To accomplish this, we used data from the [*Climate Change: Earth Surface Temperature Data*](https://www.google.com/url?q=https%3A%2F%2Fwww.kaggle.com%2Fberkeleyearth%2Fclimate-change-earth-surface-temperature-data) dataset on Kaggle.

We have used Pollution data from Kaggle :- <https://www.kaggle.com/sogun3/uspollution/data>

We have used Greenhouse Gases data from Kaggle :- <https://www.kaggle.com/unitednations/international-greenhouse-gas-emissions>

**Pre-Processing:**

1. Converting the 'dt' (date) column to DateTime format.
2. The city-wise monthly temperature data had several missing values for certain cities. Instead of replacing these NaN values with the mean temperature, we chose to remove rows having missing values.

**Processing:**  
ARIMA models need the data to be stationary i.e. the data must not exhibit trend and/or seasonality. To identify and remove trend and seasonality, we used the following methods:

1. Plotting the time series to visually check for trend and seasonality
2. Checking if the histogram of the data fits a Gaussian Curve, and then splitting data into two parts, calculating means and variances and seeing if they vary
3. Calculating the Augmented Dickey-Fuller Test statistic and using the p-value to determine stationarity

If the data was not stationary, we performed **differencing** to make it stationary.  
  
**Fitting the ARIMA model:**  
We performed a grid-search to estimate the best p, q values for the model, for the given data.  
We then fit the ARIMA model using the calculated p, q values.

**Time Series Modelling and Diagnostics**

Model:-

Since we have time series data, we have chosen to use an ARIMA (AutoRegressive Integrated Moving Averages) model for time series forecasting. We have automated the process of estimating the best p, q parameters for the ARIMA model using a grid-search technique, for any given time series data. The p,q values were chosen based on the pair that returned the lowest AIC (Akaike Information Criterion) values.

The model was used to successfully forecast the temperatures for a given city across a specified time period.

Diagnostic:-

The crux of our problem lies in Time Series Forecasting. We have made use of ACF and PACF plots, in combination with AIC values, to estimate the best p, q values for the ARIMA model and have forecasted the temperatures for cities over a given period of time. We have used libraries including pandas, numpy, matplotlib, sklearn, seaborn, plotly and statsmodels. Additionally, we have also used Pearson’s Correlation Coefficient to analyze the pairwise correlations between Pollution, Greenhouse Gas emissions and temperature change.

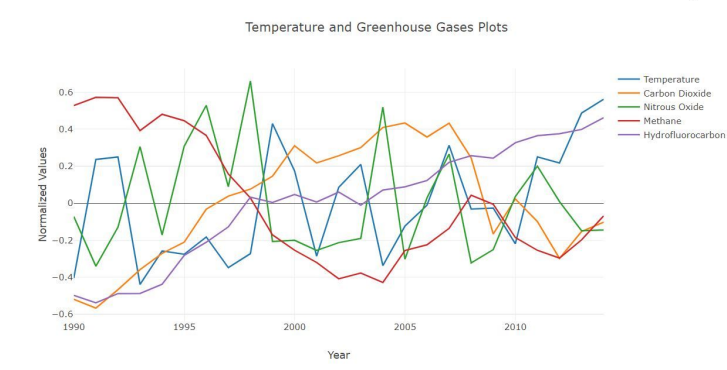
A screenshot of a graph

AI-generated content may be incorrect.

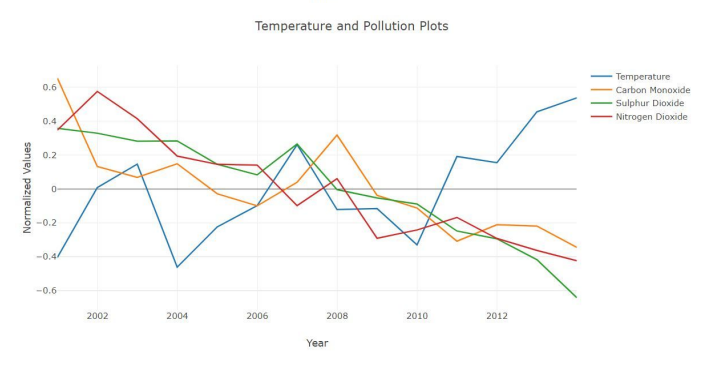
From the above correlation matrix, we can infer that HCFCs (HydroChloroFluoroCarbons) will still have the highest impact on temperature.  
Steps must be taken to reduce the emission of HCFCs to reduce their impact on temperature.

**Forecasting and Evaluation**

Temperature and Greenhouse Gases Plots:



Temperature and Pollution Plots:



Predicted temperature and Greenhouse Gases Plots:

A graph showing the difference between greenhouse gases and greenhouse gases

AI-generated content may be incorrect.

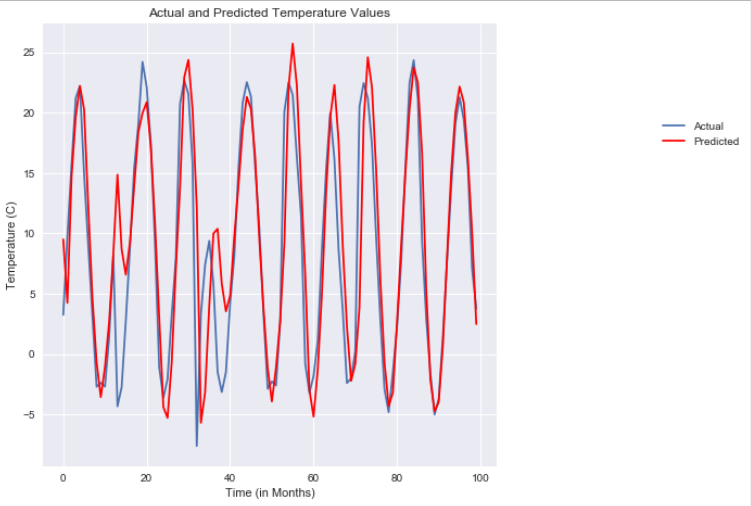
Top 10 cities with most Temperature Change in 10 years:A map of the united states

AI-generated content may be incorrect.

Evaluation:-

We have used Mean Squared Error (MSE) and Mean Absolute Error (MAE) to evaluate the performance of the ARIMA model. For New York, the ARIMA model achieved: -

* Mean Squared Error: 4.67802989468
* Mean Absolute Error: 1.57799576152



**Discussion and Conclusion**

In this project, we:

 Developed a predictive model using historical climate data to analyze and forecast global temperature change.

Applied ARIMA time series modeling, focusing on ensuring data stationarity for accurate forecasting.

 Forecasted future temperature trends for various cities, identifying patterns of consistent warming in major urban areas.

 Analyzed and found a strong correlation between temperature rise and greenhouse gas emissions as well as pollution data.

 Highlighted the impact of human activities on climate change, emphasizing the urgency for sustainable policy-making.

 Demonstrated the value of data science in tackling environmental challenges and informing climate action.

**Future Improvements**

* Include seasonal effects using SARIMA (Seasonal ARIMA).
* Add more recent or real-time temperature data.
* Use advanced models like LSTM (deep learning) or Facebook Prophet for better accuracy.