ME597: ARTIFICIAL INTELLIGENCE IN THERMAL SYSTEMS

PREDICTING GREENLAND ICE SHEET MELTING: A MACHINE LEARNING APPROACH

AUTHORS:
KARTIK MAHESHWARI, ENGINEERING MANAGEMENT, PURDUE UNIVERSITY, WEST LAFAYETTE, IN 47907

NARMADHA BALRAJ, MECHANICAL ENGINEERING, PURDUE UNIVERSITY, WEST LAFAYETTE, IN 47907

ADVISOR:
DR. VEERARAGHAVA RAJU HASTI, MECHANICAL ENGINEERING, PURDUE UNIVERSITY, WEST LAFAYETTE, IN 47907

6 DECEMBER 2020

I. ABSTRACT

Global warming is affecting our environment at an ever-increasing rate. It is essential to study climate change in Greenland as it has already crossed the point of no return because of its proximity to human civilization. Greenland cannot replenish the amount of ice loss, and sooner or later, the island will be ice free. Greenland ice sheet has the second-largest ice deposits on land just after Antarctica ice sheet. Therefore, understandings developed here could help us tackle the climate change in Antarctica. In this study, we are trying to apply machine learning algorithms to understand the underlying correlations between global warming's cause and effects. Our approach is to develop a reliable machine learning model that can predict the Greenland ice mass loss as a function of global average atmospheric carbon dioxide (CO_2) concentration, global average atmospheric methane (CH_4) concentration, global average atmospheric nitrous oxide (N_2O) concentration, global mean sea level, global average surface temperature, and human population. The idea was to develop a statistical model to predict Greenland ice sheet melting from the year 2021 until the year 2100 based on data from the year 1850 to the year 2020. Linear regression, Polynomial regression, and Polynomial regression using gradient descent were the three different regression-based machine learning algorithms used for data generation. A dataset containing original data points and artificially generated data points were compiled, which then served as the basis for developing final machine learning models. These regression-based machine learning models' performances were then evaluated based on their r-squared scores, training and validation losses and was then compared against each other for cross-checking the predictions. We observed that the linear regression model and the polynomial regression model using gradient descent performed very similarly and followed the expected trends, while the simple polynomial regression model performed poorly. Further observations were made about the repercussions that were postured by the melting of the Greenland ice sheet.

II. KEYWORDS

Global Warming, Climate Change, Machine Learning, Regression, Ice Sheet

III. INTRODUCTION

Life is a sensitive phenomenon. Even the slightest of perturbations in the system can create massive unbalance. One such disturbance capable of wiping out humanity from the surface of the Earth is global warming. Although backed by Science, the majority still holds a rather dismissive attitude towards global warming. Humankind, for now, is shaken by the COVID-19 pandemic, but the danger from global warming is nowhere to go.

The sea levels are rising at an unprecedented rate (currently, the major cause for rising sea level is ocean water expansion; however, in the second half of the century, the dynamics will change), provoking seasonal climate change. This rapid climate change is already affecting us;

such as excessive floods (Indian subcontinent, North-East Africa), forest fires (The Americas, Australia), Hurricanes (USA), heat-waves (Europe), locust attacks (Arabian peninsula, Indian subcontinent), melting ice sheets (Arctic, Antarctica), depleting marine biodiversity, etcetera. Global warming is irreversible; even if we stop emissions, it will take thousands of years for the atmosphere to revert to its preindustrial composition. Therefore, the least we can do is slow it down.

The sudden rise in Earth's average surface temperature throughout the last couple of decades due to the increased greenhouse gas level in the atmosphere warms the Earth beyond the expected standards is called Global Warming. Global warming and cooling are a perennial cycle naturally occurring on the Earth ever since its origin. Earth's climate at any given point of time in history results from the many physical mechanisms (ranging across scales) happening on and beneath the surface of the earth.

The Serbian scientist Milutin Milankovitch proposed that the long-term, collective effects of changes in Earth's position relative to the Sun are a powerful driver of the Earth's climate and are responsible for regulating the ice-ages. He suggested that the shape of the Earth's orbit (eccentricity), the Earth's tilt concerning the orbital plane (obliquity), and the direction of Earth's axis of rotation is pointed (precession). These cyclical orbital movements, referred to as Milankovitch cycles, affect the amount of solar radiation that Earth receives and, in turn, influencing the climate. However, the current global warming state has been sped up by humans using excess fossil fuels and deforestation, causing global average temperatures and global greenhouse emissions to rise at an increased rate. The concentration of CO2 in the air by volume has reached over 400 parts per million (ppm), compared to around 280 ppm in pre-industrial times. Since 1950, global emissions of CO2 from human activity have risen by over 400%. The average global temperature increased by around 3°C to 8°C, over a period of around 10,000 years, after the last ice age that occurred 20,000 years ago.

Small fluctuations in solar output can have a significant effect on the atmosphere of the Planet. A slight reduction in solar output has been shown by satellite observations (since the 1970s). However, over this time, instead of cooling, the Earth has warmed. In general, rather than certain layers of the atmosphere, warmth from the sun should warm the entire atmosphere. However, the results indicate that while the troposphere warms, the stratosphere is, in fact, cooling. This is logical with greenhouse gas heating and not heating due to the sun.

The gas that can trap the heat within the thermal infrared range is referred to as Greenhouse Gas (GHG). The major GHGs present in the Earth's atmosphere are Water Vapor [H_2O], Carbon Dioxide [CO_2], Methane [CH_4], Nitrous Oxide [N_2O], and Chlorofluorocarbon [CFC]. Carbon Dioxide is the prime greenhouse gas emissions contributor with 76%, followed by Methane at 16%. The carbon dioxide enters the atmosphere by burning fossil fuel and is naturally removed from the atmosphere by trees. Since deforestation is ever increasing, the amount of carbon released is beyond the amount absorbed naturally. Significant contributors to the presence of methane in the atmosphere are livestock farming and fossil fuel refineries. In general, transportation, energy production, commercial and residential, and farming are some of the main factors adding to the ever increased greenhouse gas levels. These GHGs stay in the atmosphere for thousands of years and accumulate, in turn creating a thick blanket around the

Earth. This blanket then traps the solar radiation bouncing off the surface of the Earth and warm the atmosphere.

Increased greenhouse gases absorb solar radiation, thus reducing the Earth's albedo's efficacy and warming the planet. Ice melt will increase even with a small rise in solar radiation at northern latitudes. Less sunlight from the bright white surface of the ice is reflected as a result of ice loss, and more is absorbed by the planet, increasing overall warming, creating a feedback loop. The warm Earth warms the oceans, and the warm oceans release carbon dioxide, hence creating a different feedback loop. More carbon dioxide in the atmosphere causes more warming, creating an amplifying effect.

Although sea-ice is present in the polar region, it plays a major role in influencing the global climate. About 15% of the world's oceans are covered in sea-ice round the year. The temperatures near the poles remain cool relative to the equator because the sea ice's bright surface reflects the sunlight into the atmosphere, therefore avoiding oceans to absorb heat. With the increase in global temperatures, the sea-ice melts gradually, leaving the ocean surface exposed to solar heat. This leads to an increase in ocean temperatures, and a feedback loop is created. The warm water prevents ice formation in winters and fastens the ice melting in summers, hence amplifying the process. The reduction in the amount of sea ice can influence the oceans' natural circulation, resulting in shifts in the global climate. Even a modest temperature rise can lead to increased warming over time, rendering the polar regions the most climate-change-sensitive areas on Earth. Figure-1 below represents the effects of climate change on life on Earth.

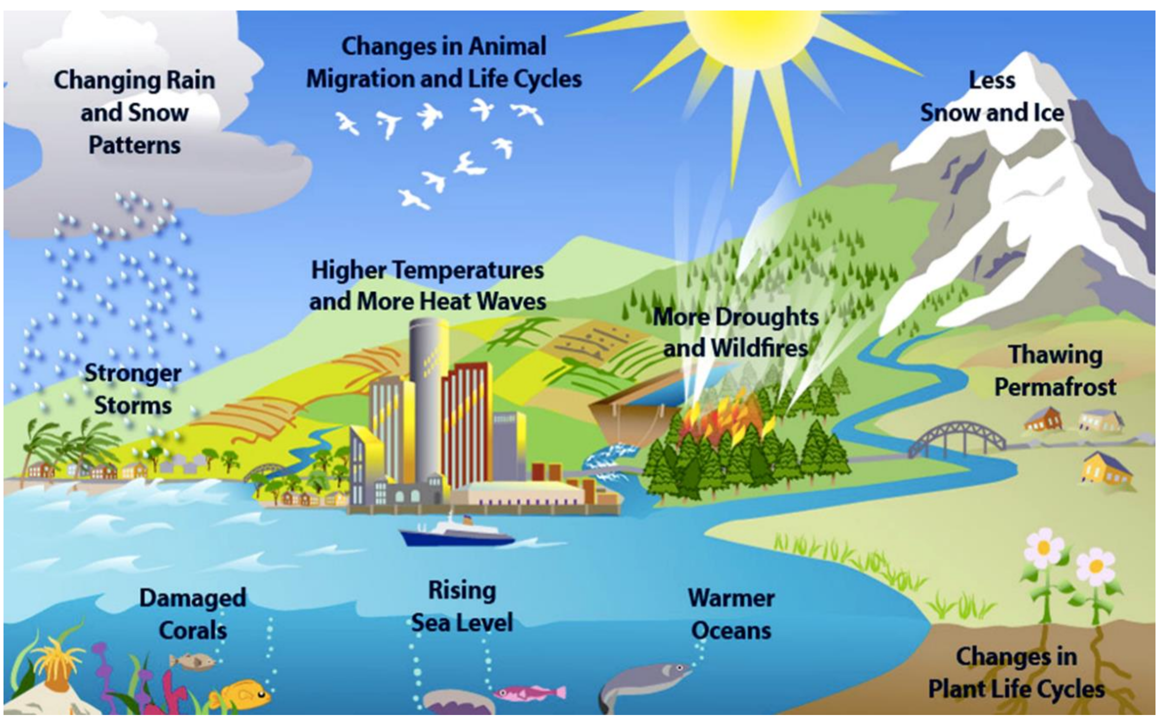


Figure 1: Effects of climate change

Since the industrial revolution, the Earth has already warmed by about 1°C. At the current rate, we are towards 1.5°C by as early as 2030. A primary concern with current climate change is the amplified feedback mechanisms causing a runaway global warming effect that will be incredibly difficult to slow down, let alone halt and reverse. Figure-2 below represents the temperature variations across the globe in the last 50 years. It can be clearly observed that the Arctic region has witnessed a significant temperature rise.

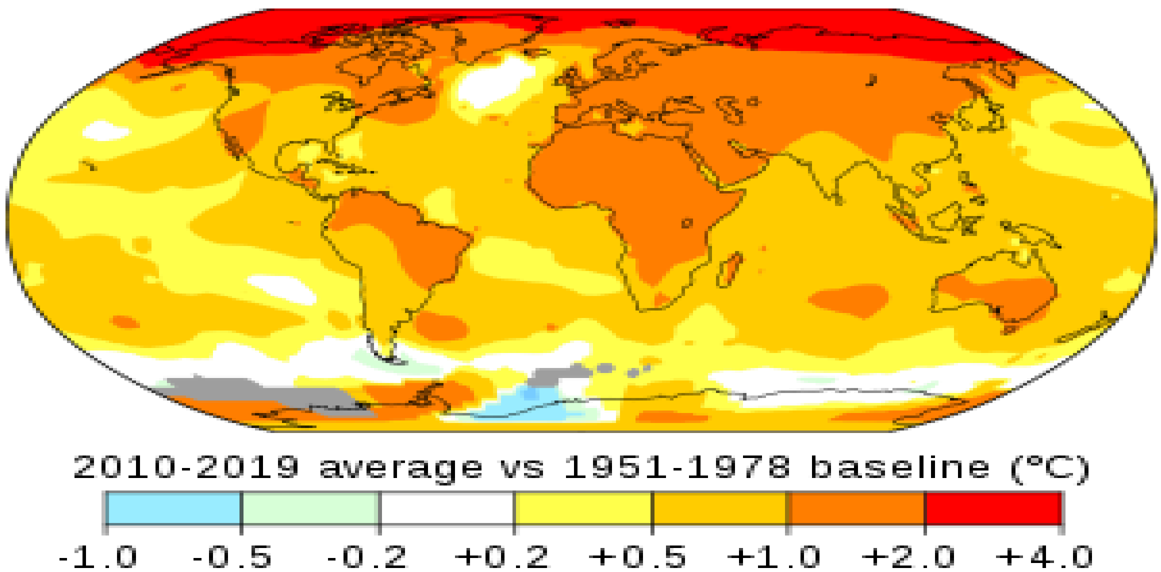


Figure 2: Average temperature change in last 50 years

Machine Learning seems to be the perfect way to take on the climate change challenge. Chi et al. [1] used deep learning approaches to predict monthly Arctic SIC values in a supervised manner. They use the multilayer perceptron (MLP), and long and short-term memory (LSTM) approaches. They use monthly Arctic SIC data that was generated from daily data provided by the National Snow and Ice Data Center.

Yimeng et al. [2] used GAN and a deep convolutional LSTM model to track ice sheet and glacier dynamics. Soroush et al. [3] predicted geothermal heat flux in Greenland (a factor for ice sheet melting); the geothermal heat is predicted using the traditional machine learning model. David et al. [4] talk about various high-impact problems concerned with climate change that can be tackled using Machine Learning.

IV. DATASET

Not all the factors leading to global warming were included primarily due to the lack of publicly available data. Data points from the year 1850 to 2020 were collected from many sources such as environmental agencies, government records, academic institutes, research articles, etcetera, and compiled into one dataset. The dataset used for the model development consisted of factors such as global average carbon dioxide concentration (ppm), global average methane concentration (ppm), global average nitrous oxide concentration (ppm), global mean sea level rise (cm), global average (land-ocean) air temperature (°C), average Greenland ice mass (GT). The dataset also included data for the human population, but this was the only factor having data from the year 1850 to 2100. It is important to note that the dataset had data on an annual basis rather than monthly or weekly. The dataset is described in detail as follows.

A. TEMPERATURE [°C]

Our dataset's average global surface temperature is the average land-ocean surface air temperature collected from Lawrence Berkeley Earth Laboratory. Temperatures are in Celsius (°C) and reported as anomalies relative to January 1951 to December 1980 average. These values were then converted to absolute average temperature values by adding an average surface air temperature of about 14.168 °C. The temperature value for the year 2020 was taken from NOAA's temperature dataset. Figure-3 below represents the increase in the global average surface temperature.

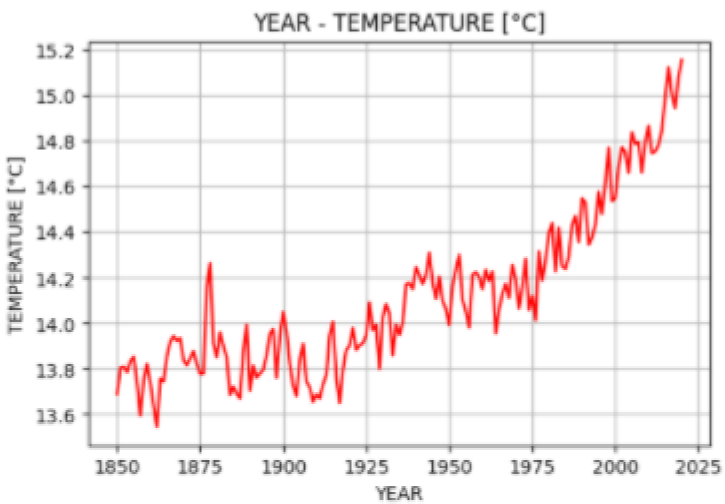


Figure 3: Temperature increase over years

B. GREENHOUSE GAS [PPM]

Most greenhouse gas data were collected from the National Oceanographic and Atmospheric Administration (NOAA's) Paleo Ice Core dataset. This dataset had complete data for carbon dioxide concentration and nitrous oxide concentration. However, it had some missing data points for methane concentration, which were then collected from the United States Environmental Protection Agency (EPA's) dataset. The carbon dioxide concentration level was recorded in parts per million (ppm), whereas the concentration level for nitrous oxide and methane was recorded in parts per billion (ppb) in the original dataset. For consistency, we converted all the concentration levels to parts per million in our dataset. Though these concentration levels are represented as average global concentration levels, they are average Antarctic concentration levels. Figure-4 below represents the increase in the global average concentration level of greenhouse gases.

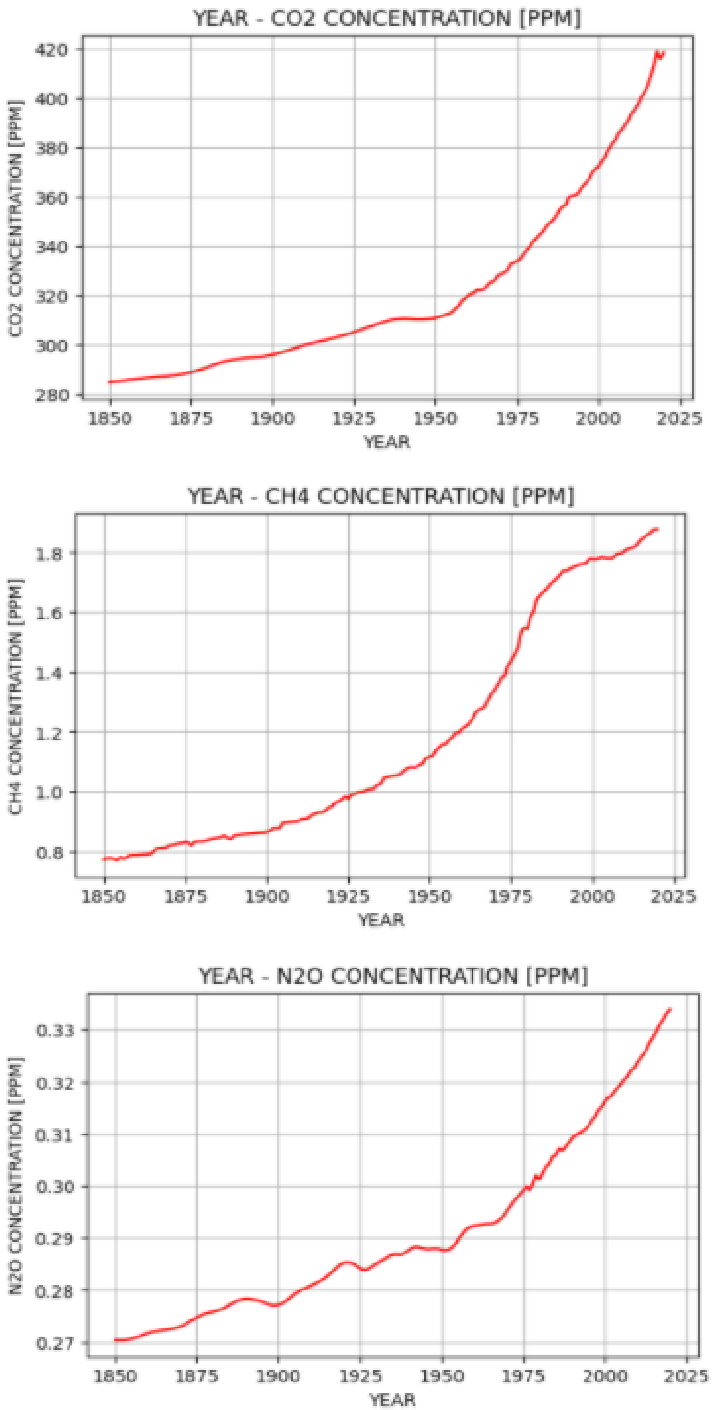


Figure 4: GHG concentration rise over the years

C. MEAN SEA LEVEL [CM]

The global mean sea level data is from the Commonwealth Scientific and Industrial Research Organization (CSIRO's) sea-level dataset. The sea-level values are in centimeter (cm) and reported as anomalies relative to the historic average sea level average. Data for certain missing values in the decade of 1850s, 1860s, and 1870s was curve fitted. The mean sea level data for the year 2019-20 was taken from NOAA's sea-level dataset. Figure-5 below represents the increase in the global mean sea-level rise.

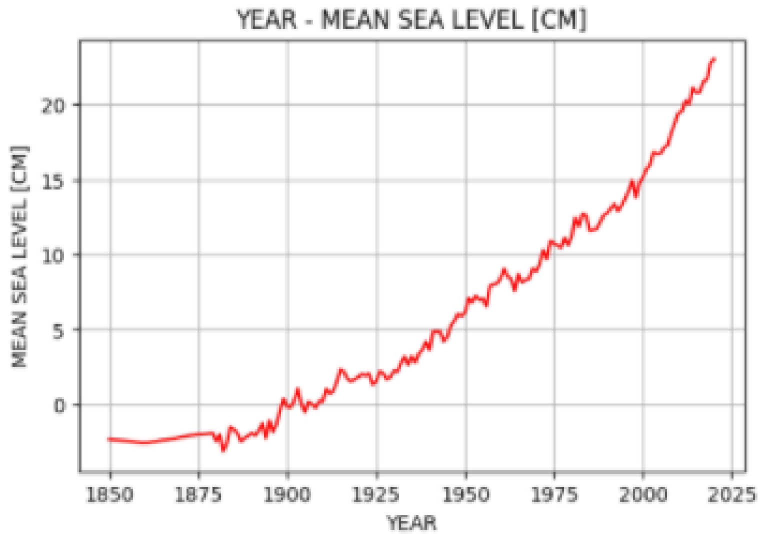


Figure 5: Sea level rise over the years

D. HUMAN POPULATION

The data for the human population was collected from the UN Population Prospects 2019. It is to be noted that the population data from the year 1850 to 2020 is nearly accurate, and the population value for the next 80 years is based on the projection made by the United Nations. The population data for the year 1850 to 1950 was externally compiled from Angus Maddison's population dataset. Figure-6 below represents the increase in the human population.

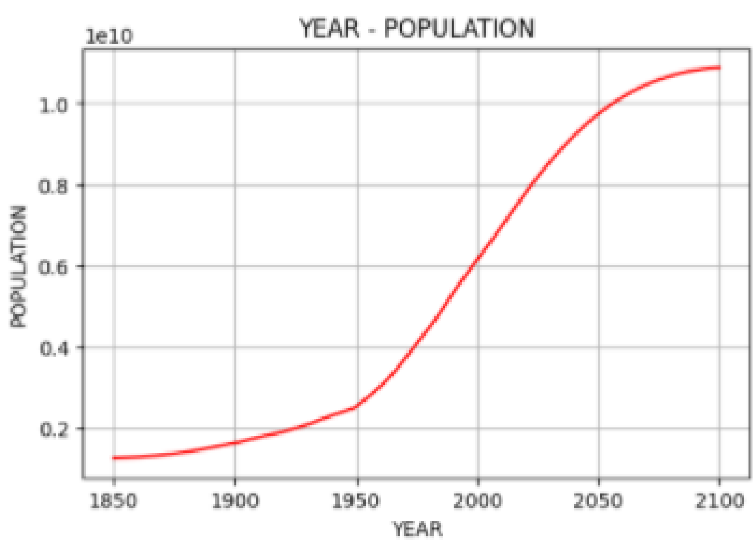


Figure 6: Human population rise over years

E. GREENLAND ICE MASS [GT]

The data for Greenland ice mass from the year 1992 to 2018 is from the IMBIE Greenland Icesheet dataset by European Space Agency (ESA), which is in terms of Gigaton. These data points are recorded as the rate of change in annual total ice mass, which is then converted back to the absolute ice mass value using the total Greenland ice mass, roughly equal to 2.85 million cubic kilometers as the reference value. For the year 2018 to 2020, the data was collected from NASA's GRACE observation dataset. The data for the year 1900 to 1992 is from another Greenland Icesheet dataset developed by Korsgaard et al.. Moreover, the data for the years 1850 to 1900 was curve fitted based on a few known intermittent data points. Figure-7 below represents the decrease in the Greenland ice mass.

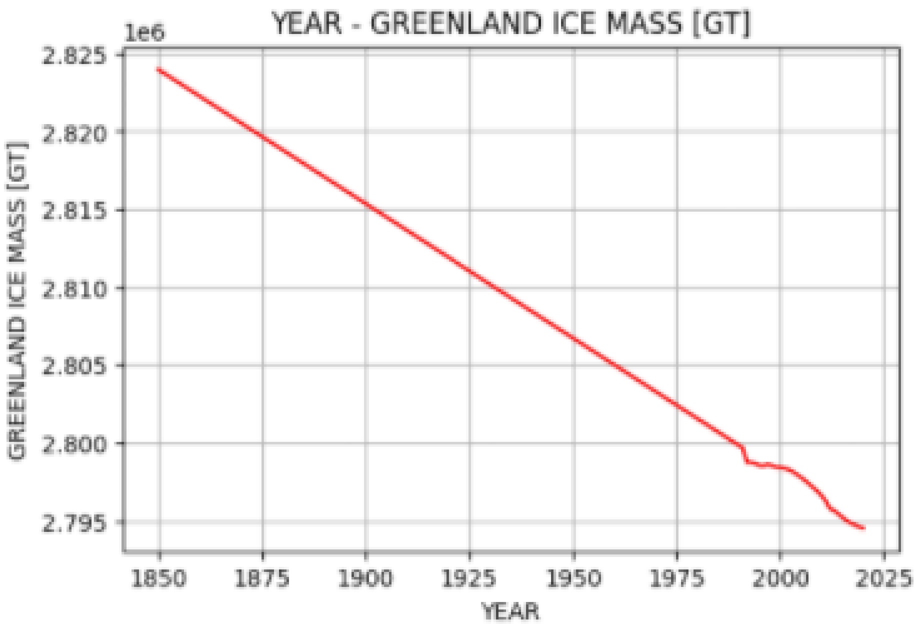


Figure 7: Decrease in Greenland Ice mass over years

F. CORRELATION MATRIX

A correlation matrix is perhaps the best way to represent a linear relationship between variables graphically. It is observed from the heatmap shown in Figure-8 that all the variables have a very high level of correlation with each other. However, the extent of the correlation of temperature with all the other features is relatively low compared to other correlation values in the dataset. It is also recognized that the Greenland ice mass has very strong inverse proportionality to all the other variables.

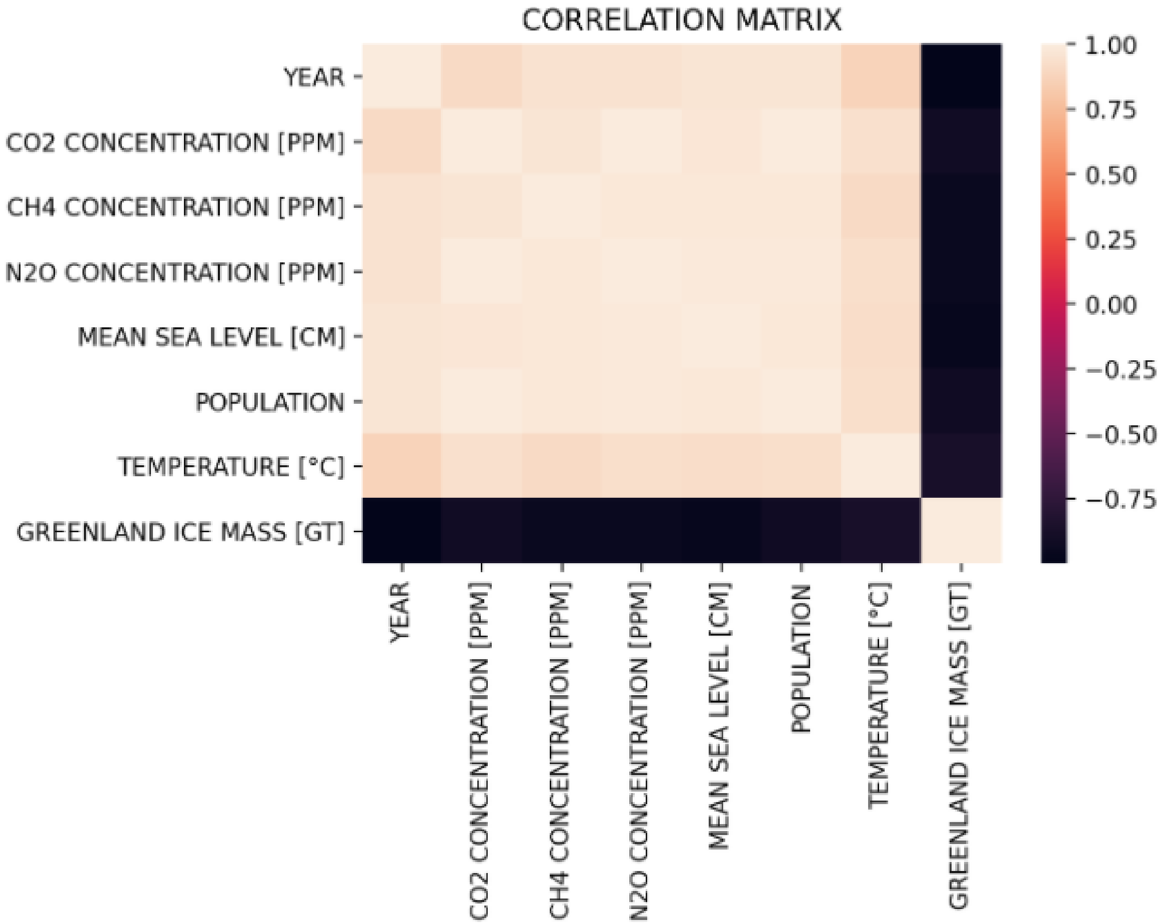


Figure 8: Correlation Matrix represented using a heatmap

V. METHODOLOGY

To predict the Greenland Ice Mass Concentration, we are using three supervised learning models. As this problem is a regression problem, we use 1. Linear regression, 2.Polynomial Regression and 3.Polynomial regression with SGD.

A. LINEAR REGRESSION MODEL

As the dataset has values for population from 2020 till 2100. We use that to predict the CO2 data for the years ahead until 2100. This is reasonable as increased Population is what causes more CO2 emissions. More population leads more usage of energy and hence more release of harmful gases such as CO2. It's also justified from the heat map that they are highly positively correlated. Next, we predict the methane data for future years using the population and predicted values of carbon dioxide. Nitrous oxide is predicted in the same fashion.

As these gases are greenhouse gases, they directly affect the global average temperature. Hence to predict the temperature for the years 2020 to 2100, we use predicted values of all the greenhouse gases and the population data. Mean sea level is then predicted using regressed values of temperature and greenhouse gases. Finally the Ice Mass Concentration in Greenland is predicted using all the regressed data of the factors affecting the melting of Ice. In the same way we repeat using polynomial regression and with SGD.

A loop is created to predict the values for the years 2020 till 2100. The data is split into train and test data in the ratio 80% and 20%. A random state is set to a constant value across all models to have a standardized base for comparison among the three models.

Now, the model is defined as linear regression and data is trained using the linear regression model. The trained model then predicts the values of CO_2 for the upcoming 80 years. The predicted values get appended to the dataset as loop continues to run till 2100 years. Then, the next feature or factor is predicted utilizing the previously predicted feature/s (like the CO_2 data for Methane prediction, CO_2 and methane for N_2O prediction) and population data. This way, we end up with values for Greenland Ice Mass Concentration for the next 80 years. The same way other regression models are trained and compared with.

B. POLYNOMIAL REGRESSION MODEL

The Polynomial regression model is generated in a similar fashion as the linear regression. In polynomial regression, we use a polynomial model of degree 2. The R-Squared score, training loss and validation loss are calculated and compared with other models.

C. POLYNOMIAL REGRESSION USING GRADIENT DESCENT

The Polynomial regression model using gradient descent is generated in a similar fashion as the other models. In polynomial regression using Gradient Descent method, the following parameters have been used after altering them and comparing with their training and validation losses. We use the squared loss function as we don't have outliers in the data. We used ridge regression penalty function. Alpha is set to 0.1. A constant learning rate with eta equal to 0.01 is set. The R-Squared score, training loss and validation loss are then calculated and compared with other models.

VI. RESULTS

A. Correlation Matrix

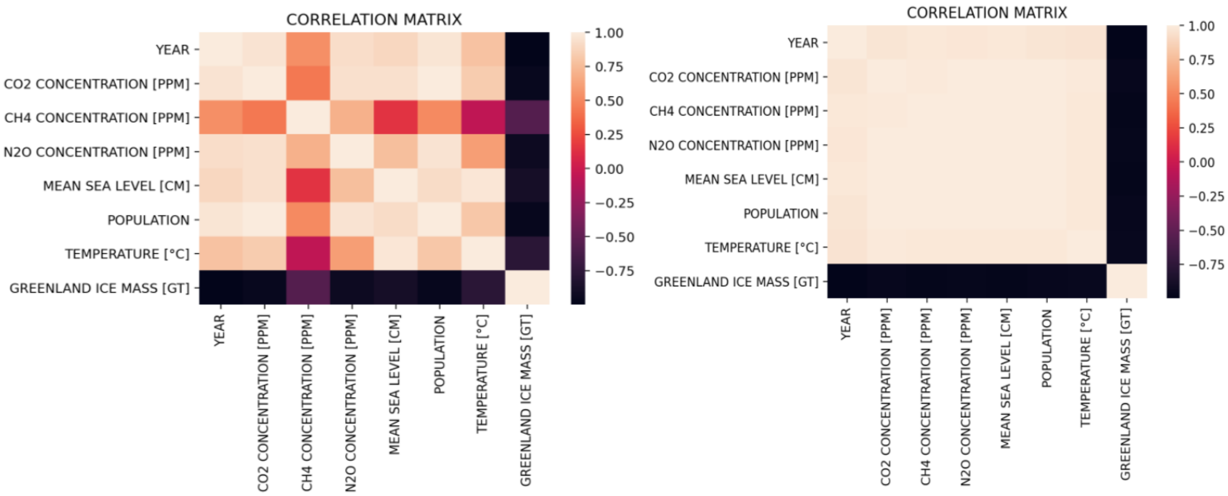
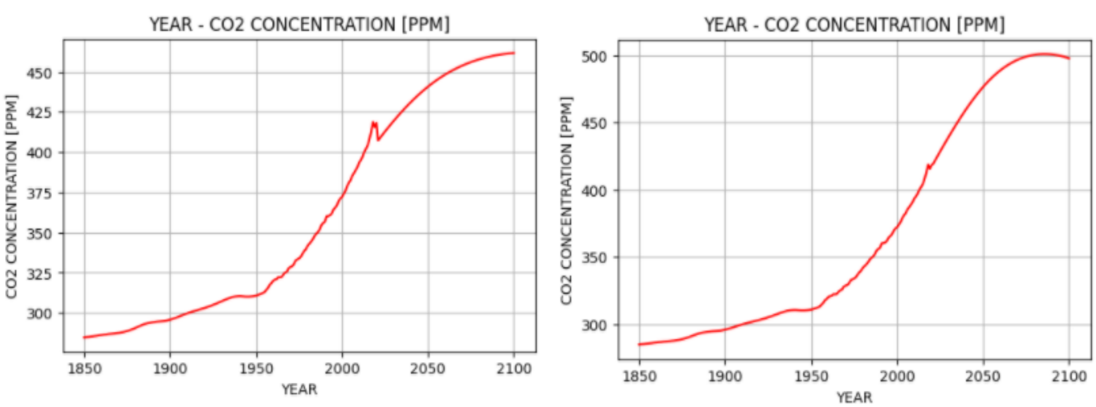


Figure 9: Correlation Matrix of predicted data using a heatmap for (a)Polynomial Regression (b) Polynomial Regression[SGD] and Linear Regression

A heatmap of predicted data using the polynomial regression shows methane data to be not correlated to other factors. While the heatmap of predicted data using linear regression and gradient descent shows perfect positive correlation among the factors and negative correlation with the ice mass concentration.

B. YEAR – CO_2 CONCENTRATION [PPM]



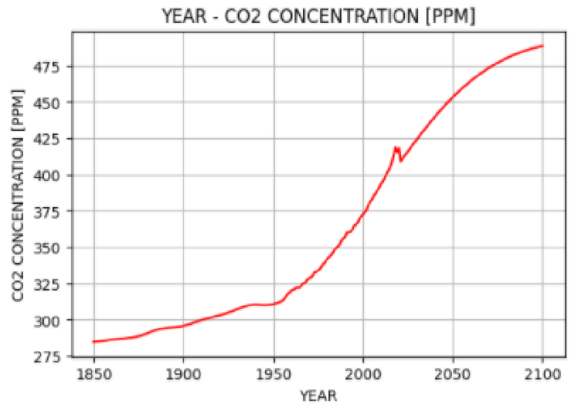


Figure 10: Plot of CO2 Concentration over the years for different models (a) Linear Regression (b)Polynomial Regression (c) Polynomial Regression[SGD]

It can be seen from the plots of CO_2 concentration that predicted data by all the models follow a similar trend. But, the polynomial regression model has a slight deviation.

C. YEAR – MEAN SEA LEVEL [CM]

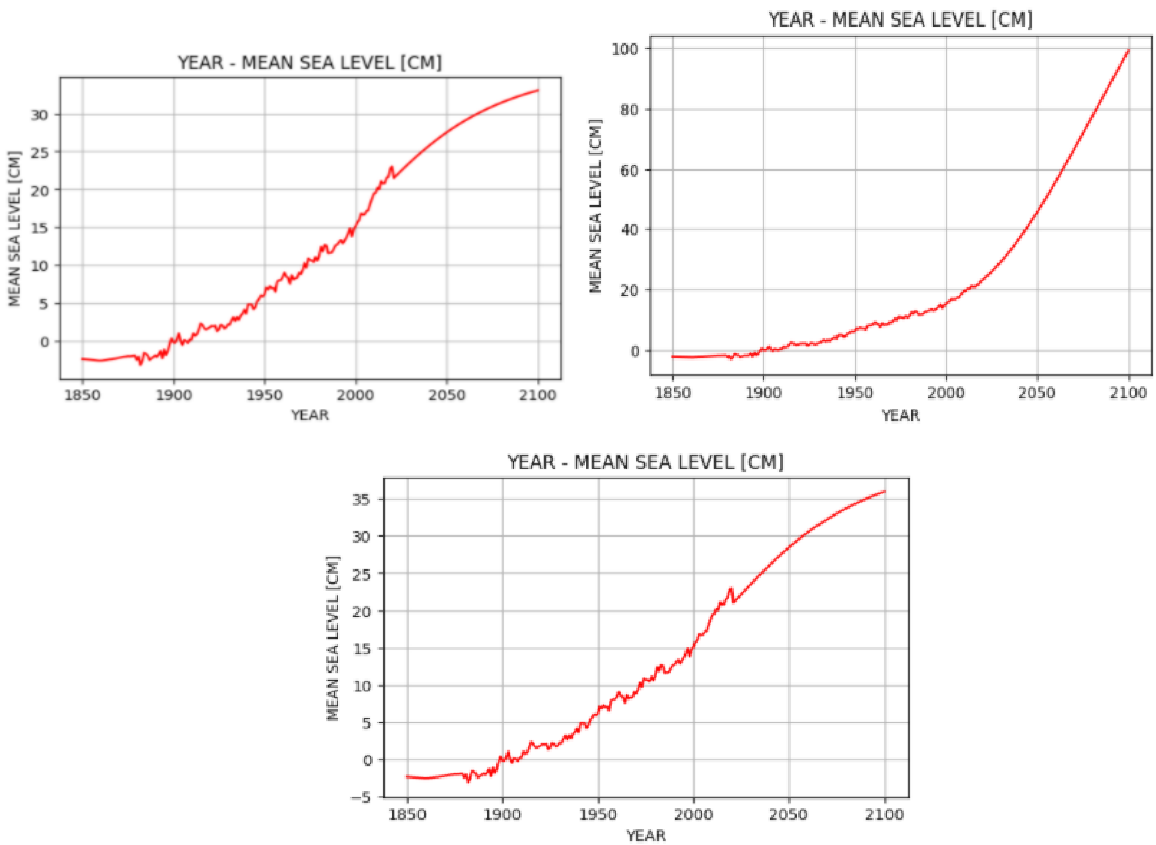


Figure 11: Plot of Mean Sea Level over the years for different models (a) Linear Regression (b)Polynomial Regression (c) Polynomial Regression[SGD]

From the mean sea level plots, it is evident that the simple polynomial regression model overshoots and the predicted sea level reach a whopping 100cm in the year 2100. Whereas the Polynomial SGD and Linear regression models have a similar trend and reasonable sea level prediction of around 32cm for the year 2100

D. YEAR – TEMPERATURE [°C]

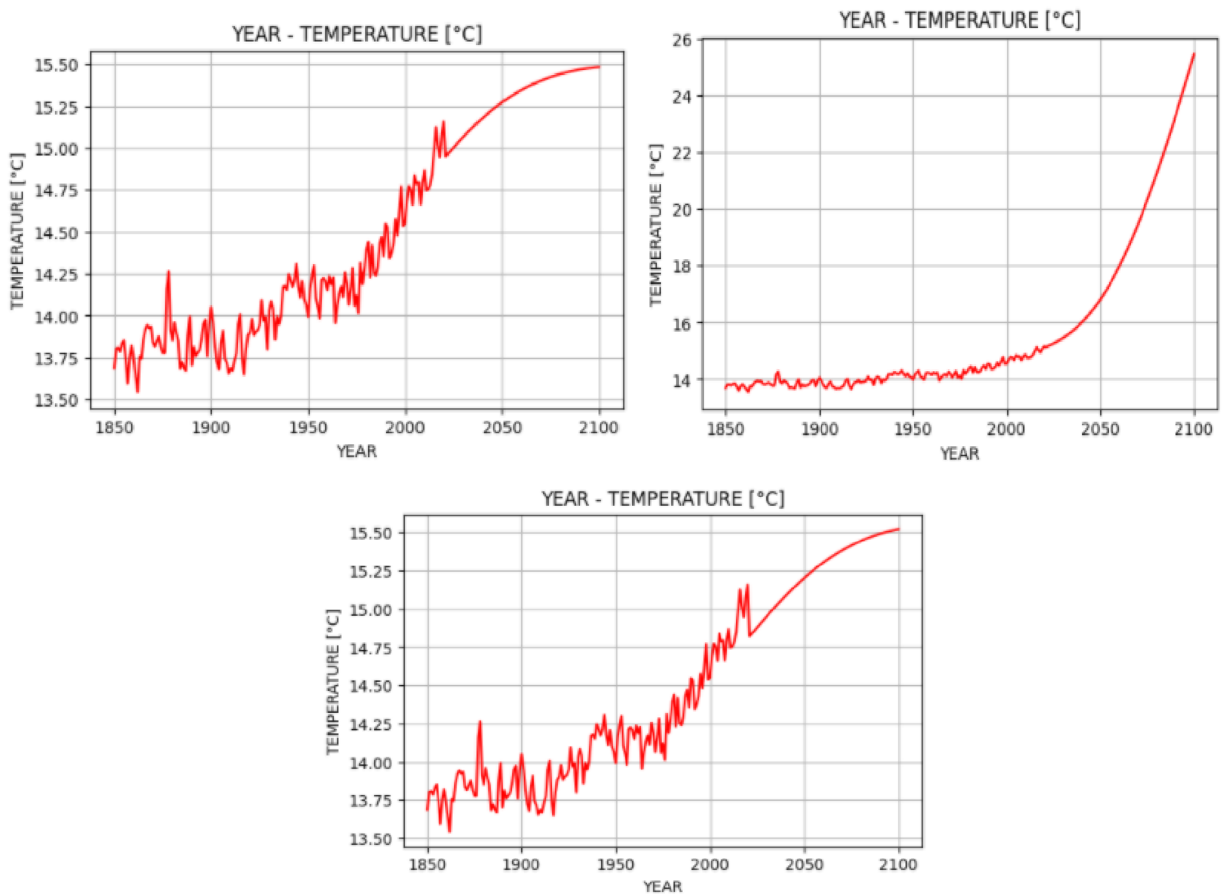


Figure 12: Plot of Global Average Temperature over the years for different models (a) Linear Regression (b)Polynomial Regression (c) Polynomial Regression[SGD]

The Global average temperature values as predicted by the linear regression and the polynomial regression with gradient descent models are fairly similar and predict a temperature of 15.5 degrees Celcius in year 2100. But, the simple Polynomial regression model has predicted a very steep increase in the Global average Temperature.

E. YEAR – GREENLAND ICE MASS [GT]

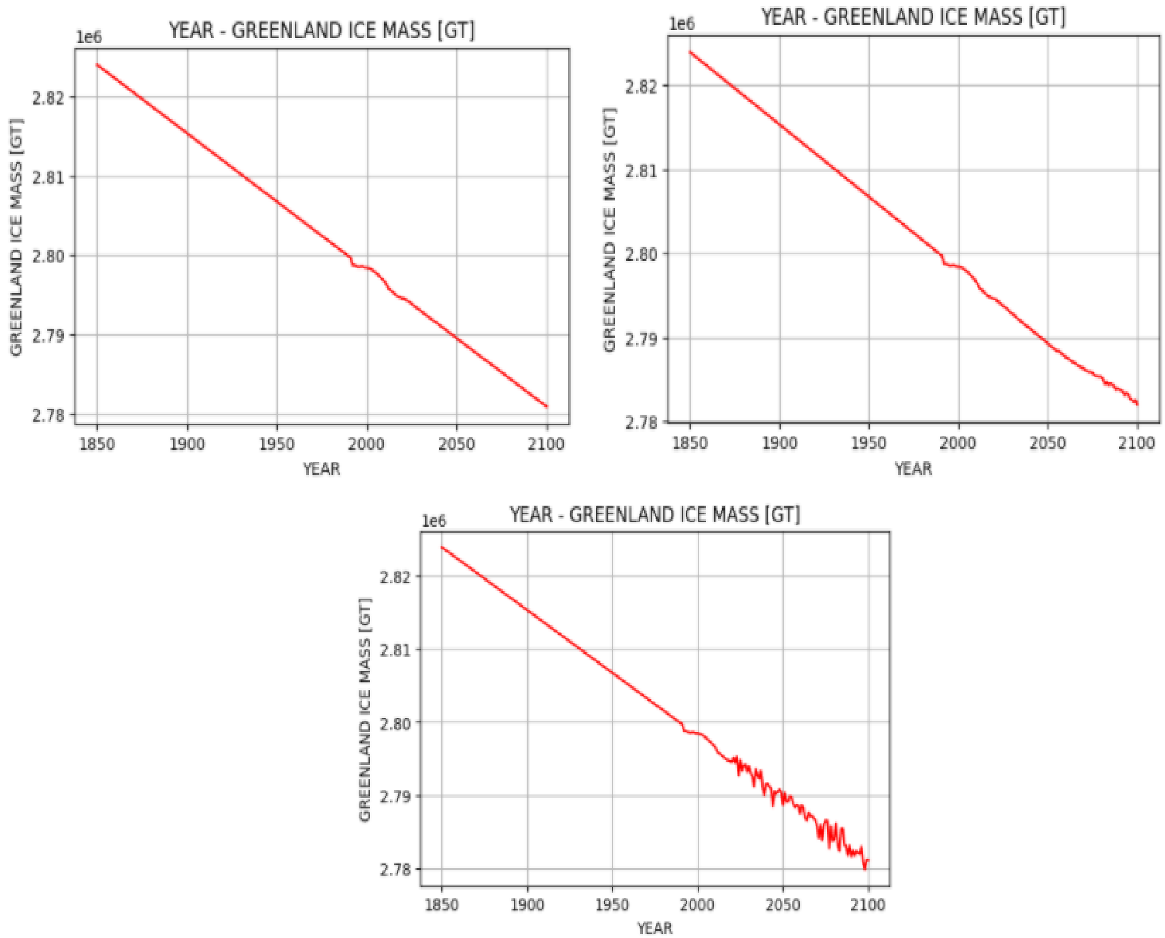


Figure 13: Plot of Greenland Ice Mass over the years for different models (a) Linear Regression (b)Polynomial Regression (c) Polynomial Regression[SGD]

Despite the deviation in predicted values of the factors in simple polynomial regression model, all three models have similar trend in the prediction of Greenland ice mass concentration. They predict the ice concentration to be around 2.78 Giga Tones by the year 2100.

F. MODEL EVALUATION

	LINEAR REGRESSION	POLYNOMIAL REGRESSION	POLYNOMIAL REGRESSION [SGD]
R-Squared Score	0.99996	0.99984	0.99674
Training Loss Greenland Ice Mass [GT]	82.70976	118.65132	2550.54848
Validation Loss Greenland Ice Mass [GT]	118.33943	237.99721	6747.88796
Model Fit	Near Perfect Fit	Slight Overfit	Overfit

The R-Squared Score is a good measure to test the accuracy of any model. All three models have shown good accuracy of 0.99. Comparing the training loss with validation loss of each model we find the linear regression model to be a near perfect fit as the difference between the losses is quite small. Following which the simple polynomial regression with a slight over fit provides a good R-squared score. The polynomial regression model using gradient descent has huge losses and the difference between the training and validation loss is also huge with validation loss being the highest. Hence, it's an over fit model.

VII. CONCLUSION

Humans have long underestimated climate change. This study proves that we do not have much time to react. A few decades from now, we will be at a point of no return.

It is evident from the predicted data that the factors contributing to global warming will also increase with an increase in the human population. According to our models, the amount of ice melted over the last 100 years is equal to the projected amount of ice that will melt in the next

80 years, from the year 2021 to 2100. This ice melt would increase global mean sea levels by as much as 21 cm (this value is derived from the fact that the total amount of ice on Greenland (if melted) can raise global mean sea level by as much as 7.2 meters).

The artificially generated data using the linear regression model and polynomial regression using the gradient descent model follow expected trends. However, Methane and Nitrous Oxide data points derived using the simple polynomial regression model do not follow the expected trend.

Nuanced hyperparameter tuning and standardizing the data before model development could yield better results. The models employed for the current study are relatively simple, and often simple models do not process the complex data effectively. Therefore, going forward, implementing Neural Networks, Physics-based machine learning algorithms, etcetera could help predict more realistic data.

To develop better performing models, we can increase the number of features in the dataset. We can include factors such as atmospheric water vapor concentration, atmospheric currents, global ice sheet covers, Earth's albedo levels, the natural rate of global warming, deforestation, etcetera. Our dataset is small; it contains data points for yearly anomalies and using bigger datasets with data points for monthly or even daily anomalies could yield better results.

We cannot stop climate change, but we can surely slow it down, and the greatest tool at our disposal to help us achieve this is Machine Learning and Artificial Intelligence. Addressing global warming will allow us to be ready for the future (in terms of power generation), reduce the runaway costs of climate change (saving billions of dollars in post-disaster management). At the same time creating new employment opportunities, improving global public health, helping sustain critical habitats and organisms, and avoiding the breakdown of the food chain, etcetera.

VIII. ACKNOWLEDGEMENTS

We would like to thank our advisor Dr. Veeraraghava Raju Hasti for organizing a wonderful course with lots of helpful webinars and for his help in the course project. We would like to thank the TA and our fellow classmates for prompt help in home works and resource sharing on piazza.

IX. REFERENCES

[1] Chi, J.; Kim, H.-C. **Prediction of Arctic Sea Ice Concentration Using a Fully Data Driven Deep Neural Network.** *Remote Sens.* 2017, *9*, 1305.

[2] Yimeng Min, S. Karthik Mukkavilli, Yoshua Bengio. **Predicting ice flow using machine learning.** arXiv:1910.08922

[3] Soroush Rezvanbehbahani, Leigh A. Stearns, Amir Kadivar, J. Doug Walker, C. J. van der Veen. **Predicting the Geothermal Heat Flux in Greenland: A Machine Learning Approach.** Geophysical Research Letters, 2017; DOI: 10.1002/2017GL075661

[4] David Rolnick et al. **Tackling Climate Change with Machine Learning**.arXiv:1906.05433

[5] Young Jun Kim, Hyun-Cheol Kim, Daehyeon Han, Sanggyun Lee, and Jungho Im **Prediction of monthly Arctic sea ice concentrations using satellite and reanalysis data based on convolutional neural networks**. The Cryosphere, 14, 1083–1104, 2020