Climate change prediction with Artificial Intelligence

1st Atong Abraham Kur
Makerere university
StudentNo: 2100720041, RegNumber: 21/X/20041/EVE
Kampala, Uganda
atongkurabraham@gmail.com

***Abstract*—Climate change prediction system addresses an issue that typically relates to the forecasting of future changes in the Earth's climate.**
This may involve predicting changes in temperature, precipitation, sea levels, and other climate variables over different time scales of global climate trends. Climate changes cause increased heat, drought, flooding, and erosion, a decline in water supplies, reduced agricultural yields and triggered heat-related health impacts in cities.
The climate change prediction model helps improve our understanding of the underlying causes of climate change, providing more accurate and detailed forecasts of future changes and Assisting in the development of policy and decision-making related to climate change.
Here are few examples of possible subtitles in this paper; introduction, Background and Motivation, methodology, Literature Review, dataset description and among others.

***Index Terms*—climate change, Prediction, machine learning, System, Initial conditions and global warming.**

I. Introduction

There seems to be climate changes causing increased heat, drought, flooding, and erosion, a decline in water supplies, reduced agricultural yields and triggered heat-related health impacts in cities, and insect outbreaks which happened unknowingly.
The climate change prediction model is used which helps in forecasting climatic changes on the earth and is used to provide information on historical and upcoming climate changes. The results of this study will enable an understanding of how the climate is changing so that we can prepare for the future.
Climate change prediction system uses AI algorithms which are well-made to analyze large datasets and detect patterns and the accuracy of the predictions which can be increased by incorporating real-time data points such as weather observations.
The type of data to be used is Climatological data; Climatological data: This type of data includes long-term averages and trends of meteorological variables, such as the average temperature for a particular location over a period of time.
These data are used to understand the long-term patterns of the climate and to detect changes over time.

Identify applicable funding agency here. If none, delete this.

II. I. Background and Motivation

A. Maintaining the Integrity of the Specifications

Climate change: Climate change refers to long-term changes in the Earth's climate, including changes in temperature, precipitation, and the frequency and intensity of extreme weather events. Prediction refers to the use of computer models and other tools to forecast how the climate is likely to change over time.
System: A system is a set of interconnected elements that work together to achieve a common goal or purpose.
Initial conditions refer to the state of the climate system at a particular point in time, including atmospheric concentrations of greenhouse gases and other factors that influence the climate.
The background and motivation for climate change prediction system stems from the increasing awareness of the impacts of human activities on the climate system.
Over the past century, the concentration of greenhouse gases in the atmosphere has increased significantly due to the burning of fossil fuels and other human activities. These greenhouse gases trap heat from the sun, causing the Earth's surface temperature to rise. This process is known as global warming.

III. II. Literature Review

A literature review for climate change prediction involves examining the existing research and studies on the subject. These are discussed below;

- The causes of climate change: The primary causes of climate change include increased emissions of greenhouse gases such as carbon dioxide, methane, and nitrous oxide.
- Climate change prediction methods: A review of the various methods used to predict future changes in the Earth's climate, including statistical methods, numerical models, and machine learning algorithms.

- The impact of climate change: A review of the impacts that climate change is likely to have on various regions of the world, including changes in precipitation patterns, sea-level rise, and increased frequency and severity of extreme weather events.

- The challenges of climate change prediction: This includes the complexity of the Earth's climate system, the limited availability of high-quality data, and the difficulty of extrapolating from historical trends.

- The limitations of current climate change prediction methods: A critical review of the limitations of current climate change prediction methods, including their limitations in representing certain aspects of the Earth's climate system, their sensitivity to uncertainties in initial conditions and emissions scenarios.

- Future directions in climate change prediction: A discussion of future directions for research in climate change prediction, including the development of more accurate and comprehensive numerical models, the use of machine learning algorithms to improve predictions, and the use of high-quality data to validate and improve predictions. In climate change prediction, supervised learning methods such as regression and decision trees can be used to predict future temperatures based on historical data. For example, a linear regression model can be trained on temperature data for a specific location over time to predict future temperatures for that location.
  Unsupervised learning methods such as clustering can be used to group similar patterns in temperature data.For example, k-means clustering can be used to group similar temperature patterns for different locations over time, providing insights into regional climate change trends.

A. Identify the Research gaps in the literature.

The literature on climate change prediction systems with AI is a rapidly evolving field, and there are several research gaps that have yet to be fully addressed. Some of the key research gaps include:

- Limited data availability:
  Climate change is a complex and global phenomenon, and there are still significant gaps in our understanding of its causes and impacts. This limited data availability makes it difficult to train machine learning algorithms and develop accurate predictions.

- Lack of interdisciplinary collaboration:
  Climate change is a multifaceted problem that requires interdisciplinary collaboration between experts in the fields of computer science, meteorology, earth science, and more. However, there is currently a lack of interdisciplinary collaboration in this field.

- Inadequate evaluation metrics:
  Evaluating the performance of climate change prediction systems can be challenging, as there is often limited ground truth data available for comparison and no agreement on what metrics should be used to evaluate the performance of the system.

- Lack of transparency and interpretability:
  Climate change prediction systems with AI often use complex algorithms and models, making it difficult to understand how they are making predictions and to assess their reliability.

- Integration with other climate prediction models: Climate change prediction systems with AI are often developed and evaluated in isolation, without being integrated with other existing climate prediction models. This makes it difficult to assess the overall performance of these systems and to compare them to other approaches.

B. Summary of the term paper contributions

Here are some common approaches that can be used to address the technical challenges in this field:

- Data integration: To overcome the challenge of limited data availability, researchers are developing approaches to integrate multiple data sources, such as satellite imagery, weather data, and more, to create a more comprehensive picture of the impacts of climate change.
- Interdisciplinary collaboration: To address the lack of interdisciplinary collaboration, researchers are actively engaging with experts from different fields, including computer science, earth science, and meteorology, to develop a more integrated and comprehensive understanding of climate change.
- Improved evaluation metrics: To address the lack of agreement on evaluation metrics, researchers are developing new metrics and standards to evaluate the performance of climate change prediction systems with AI.
- Increased transparency and interpretability: To address the lack of transparency and interpretability in these systems, researchers are developing new techniques, such as explainable AI, to make it easier to understand how these systems are making predictions and to assess their reliability.
- Integration with other climate prediction models: To address the need for integration with other climate prediction models, researchers are developing approaches to integrate AI-based climate change prediction systems with other existing models, such as general circulation models, to create more comprehensive and accurate predictions.

C. Methodology

Using the supervised learning technique, here is a summary of a typical AI process for climate change prediction systems:

- Data collection and preparation: The first step in the AI process is to collect and prepare the data that will be used to train the model. This may involve cleaning and preprocessing the data, integrating multiple data sources, and more.

- Model selection: Next, the appropriate AI algorithm or model must be selected, based on the type of problem being solved and the characteristics of the data. For example, a supervised learning algorithm may be used for a problem with labeled data, while an unsupervised learning algorithm may be used for a problem with unlabeled data.
- Model training: Once the model has been selected, it must be trained on the data. This involves feeding the model large amounts of data and allowing it to learn the underlying patterns and relationships in the data.
- Model evaluation: After the model has been trained, it must be evaluated to assess its performance. This may involve using metrics such as accuracy, precision, recall, and more, depending on the specific problem and the type of data being used.
- Model deployment: Once the model has been evaluated and found to be effective, it can be deployed in a real-world setting to make predictions about the impacts of climate change. Some common evaluation metrics used in this field include Mean Absolute Error (MAE), Mean Squared Error (MSE), ,R2 Score and more. It is also important to assess the interpretability and transparency of the models, as well as their ability to integrate with other existing climate prediction models.

D. C. Dataset Description

The data set include the following important features;
Date: The date of the climate data is critical as it provides a timeline for the changes in the climate over time. This information is used to track trends in the climate and to make predictions about future changes.
Average temperature:. Average temperature data can be used to track changes in the climate over time, and to make predictions about future temperature trends.
Country: The country in which the climate data was collected is also an important feature in climate change prediction systems. This information allows us to track regional climate trends and to make predictions about how different regions may be affected by climate change in the future.

Here are a few factors to consider when selecting a dataset for a climate change prediction system with AI:

- Relevance: The dataset should be relevant to the problem being solved, containing information that can be used to make predictions about the impacts of climate change.
- Quality: The data should be of high quality, with minimal errors, missing values, or inconsistencies.
- Volume: The dataset should be large enough to provide the AI model with enough data to learn from, while also being manageable to process.
- Diversity: The dataset should be diverse, representing a variety of regions, climates, and other factors that can impact climate change.

Once the dataset has been selected, it must be preprocessed and prepared for use with the AI model. This may involve normalizing the data, filling in missing values, and more. After the data has been prepared, the AI model can be trained on the data, allowing it to learn the underlying patterns and relationships. The AI model can then be used to make predictions about the impacts of climate change, using the data to inform its decisions.

E. Data Preparation and Exploratory Data Analysis

1. Data cleaning In performing data cleaning, we check if there are null, missing and inconsistent values in the dataset and then remove them.
(i) Dealing with missing or null values.
There are many ways of dealing with them such as;

- Deleting the columns with missing data
- Deleting the rows with missing data
- Filling the missing data with a value
- Imputation with an additional column
- Filling with a Regression Model

However, im going to illustrate only deleting the columns with missing data

(a) Checking and removing null values
We first check for null values and then remove them since available.

Checking null values

```
demo.isnull().sum()
```

```
dt                                   0
AverageTemperature               32651
AverageTemperatureUncertainty    31912
Country                              0
dtype: int64
```

Removing null values

```
demo=demo.dropna(how='all',axis=0)
demo=demo.dropna().reset_index(drop=True)
```

```
demo.isnull().sum()
```

```
dt                               0
AverageTemperature               0
AverageTemperatureUncertainty    0
Country                          0
dtype: int64
```

(b) Renaming the columns and converting data type of dt into datetime and then into index.

Before.

```
demo.head()
```

	dt	AverageTemperature	AverageTemperatureUncertainty	Country
0	1743-11-01	4.384	2.294	Åland
1	1744-04-01	1.530	4.680	Åland
2	1744-05-01	6.702	1.789	Åland
3	1744-06-01	11.609	1.577	Åland
4	1744-07-01	15.342	1.410	Åland

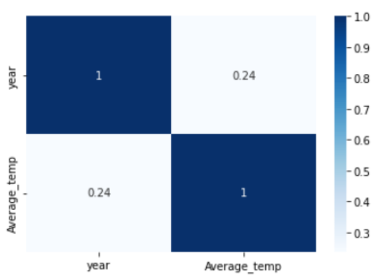
After.

Date	Average_temp	AverageTemp_uncertainty	Country	year
1743-11-01	4.384	2.294	Åland	1743
1743-12-01	NaN	NaN	Åland	1743
1744-01-01	NaN	NaN	Åland	1744
1744-02-01	NaN	NaN	Åland	1744
1744-03-01	NaN	NaN	Åland	1744

2. Exploratory Data Analysis.

(i) Heatmap

```
demo= demo[['year','Average_temp']]
sns.heatmap(demo.corr(),cmap='Blues', annot=True)
sns.set(rc={'figure.figsize':(10,10)})
```



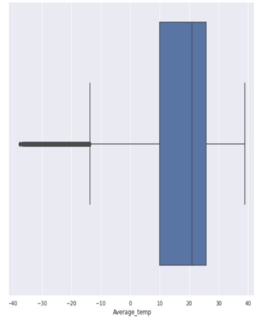
The heat-map above visualizes the relationship between temperature and year where darker color represents the higher values.

(ii) Correlation matrix

```
cor=demo.corr()
cor
```

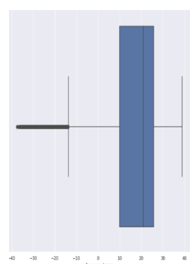
	year	Average_temp
year	1.000000	0.236765
Average_temp	0.236765	1.000000

(iii) Boxplots

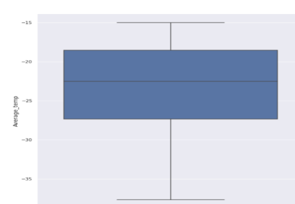


The boxplot above represents the interquartile range (IQR), which contains the middle 50(iv) Outlier plots
Using a box plot, outliers are defined as data points that fall outside the upper and lower hinges (whiskers) of the box plot as shown below.

(iv) Outlier plots
Using a box plot, outliers are defined as data points that fall outside the upper and lower hinges (whiskers) of the box plot as shown below.



After removing them, there are no values now outside the range of upper and lower hinges (whiskers) of the box plot as shown below.



3. Data splitting
Let's say we want to split the data in 80:10:10 for train: valid: test dataset.
In the first step we split the data into training and remaining dataset.
Now since we want the valid and test size to be equal $(10Wenowhavetodefinevalid_size = 0.5$ (that is 50The outputs are shown below:

```
9]:  print('Train data'), print(X_train.shape), print(y_train.shape)
     print('-----------------------------------------------------')
     print('Test data'), print(X_valid.shape), print(y_valid.shape)
     print('-----------------------------------------------------')
     print('Valid data'),print(X_test.shape), print(y_test.shape)

     Train data
     (4584, 1)
     (4584,)
     -----------------------------------------------------
     Test data
     (573, 1)
     (573,)
     -----------------------------------------------------
     Valid data
     (574, 1)
     (574,)
9]:  (None, None, None)
```

F. AI model selection and optimization.

- Linear Regression. This is used to model the relationship between a dependent variable and one or more independent variables. It can be used to model the relationship between average temperature and other features such as date, country, etc. Linear regression uses a linear equation to predict the target variable based on the independent variables. The equation takes the form of Y = b0 + b1X1 + b2X2 + ... + bn*Xn, where Y is the target variable, X1, X2, ..., Xn are the independent variables and b0, b1, b2, ..., bn are the coefficients that are estimated from the data.

- Random Forest Regression: Random Forest Regression is an ensemble learning method that uses multiple decision trees to predict the target variable. This reduces the variance in the model, making it more robust and less prone to overfitting.

- MLPRegressor: MLPRegressor is a type of artificial neural network that can be used for regression problems. It is a multi-layer perceptron that takes input features and maps them to the target variable through multiple hidden layers. In the context of climate change prediction system, MLPRegressor can be used to model the relationship between average temperature and other features such as date, country.

In the figure above, linear regression model gives the smallest coefficient of determination which tends to be the best model out all.

G. AI model selection Accountability.

AI accountability refers to the process of ensuring that the decisions and actions taken by AI systems are fair, transparent, and can be explained and justified.

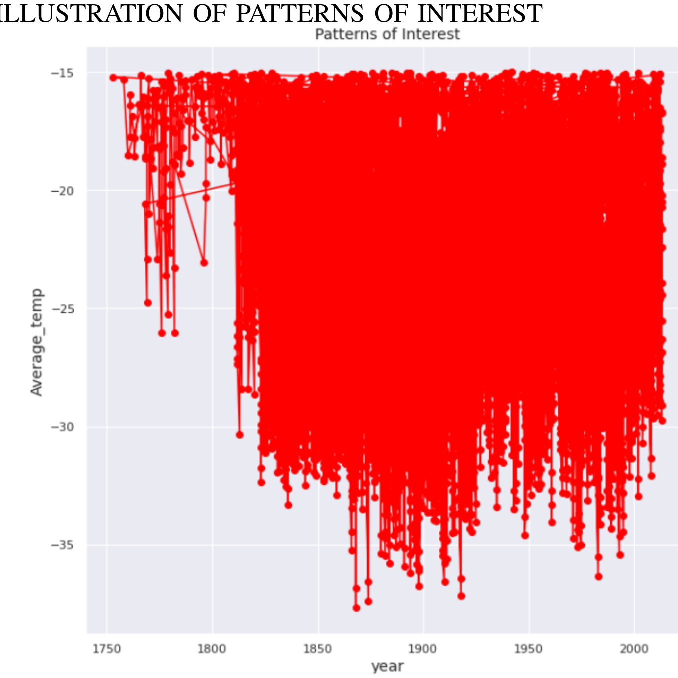
In the context of climate change prediction systems, AI accountability can be applied in several ways:

- Algorithmic transparency: The models used in climate change prediction systems must be transparent and understandable, so that the predictions they make and the factors that influence them can be easily explained.
- Fairness and bias mitigation: Climate change prediction systems must avoid any form of bias, including demographic, cultural, or socioeconomic bias. This can be done by carefully selecting and preprocessing the data used for training and validating the models, and by implementing fairness constraints in the model development process.
- Model evaluation and validation: Climate change prediction systems must undergo rigorous evaluation and validation to ensure their accuracy and robustness in predicting future climate change patterns. This includes using performance metrics such as mean squared error, accuracy, and F1-score.

Explainable AI (XAI) is a subfield of AI that focuses on developing models and algorithms that are transparent and interpretable, allowing their decision-making processes to be understood by human stakeholders.
Some of the XAI techniques that could be selected for climate change prediction systems include:

- Model interpretation techniques such as partial dependence plots, feature importance analysis, and local interpretable model-agnostic explanations (LIME), which allow stakeholders to understand the contribution of individual features to the model's predictions.

ILLUSTRATION OF PATTERNS OF INTEREST



- Model transparency through the use of simple and interpretable models, such as linear regression or decision trees, that can be easily understood by stakeholders.
- Model visualization techniques, such heat maps, which can provide an intuitive representation of the model's decision-making process.

```
9]: print('Train data'), print(X_train.shape), print(y_train.shape)
    print('-----------------------------------------------------')
    print('Test data'), print(X_valid.shape), print(y_valid.shape)
    print('-----------------------------------------------------')
    print('Valid data'),print(X_test.shape), print(y_test.shape)

    Train data
    (4584, 1)
    (4584,)
    -----------------------------------------------------
    Test data
    (573, 1)
    (573,)
    -----------------------------------------------------
    Valid data
    (574, 1)
    (574,)
9]: (None, None, None)
```

- Model explanation methods, such as instance-based explanation or case-based reasoning, which can provide a detailed explanation of why a particular prediction was made, including the factors that influenced the decision and the weight given to each feature.

RESULTS AND DISCUSSION

Since climate change prediction system uses regression model instead of classification metrics (Accuracy, Recall, Precision, F1-Score).
The most commonly used evaluation metrics for this regression problems are:

- Mean Absolute Error (MAE): It is the average absolute difference between the actual and predicted values.
- Explained Variance Score (EVS): It measures the proportion of variance in the target variable that is explained by the model. The score ranges from -1 to 1, where a score of 1 indicates that the model perfectly captures the variance in the target variable, and a score of -1 indicates the model is doing worse than just predicting the mean value.
- Mean Squared Error (MSE): It is the average of the squared difference between the actual and predicted values. MSE gives a high penalty to large errors, which makes it more sensitive to outliers compared to MAE.
- Root Mean Squared Error (RMSE): It is the square root of the mean squared error.
- R2 Score: It is a measure of goodness of fit and is defined as the proportion of variance in the target variable that is explained by the model. The score ranges from 0 to 1, where a score of 1 indicates a perfect fit and a score of 0 indicates that the model does not explain any variance in the target variable.
  Their details are given below

TABLE OF RESULTS

Since climate change prediction system uses regression model instead of classification metrics no table is reguired for (Accuracy, Recall, Precision, F1-Score) because they are not used.
The most commonly used evaluation metrics for this regression model are : Mean Absolute Error, Mean Squared Error, Root Mean, Squared Error, Explained Variance Score and R2 Score.

CONCLUSION AND FUTUREWORKS

Conclusion:
Climate change prediction is a complex and challenging task that requires advanced modeling techniques and large amounts of data. While significant progress has been made in recent years, there is still much to be done to improve the accuracy and reliability of climate change predictions.

Future works:

- Develop more accurate and efficient climate models that can better simulate the complex interactions between different components of the Earth system, such as the atmosphere, oceans, and biosphere.
- Integrate climate change predictions with socio-economic models to better understand the impacts of climate change on human populations and to develop effective adaptation and mitigation strategies.
- Increase public awareness and education about climate change and its potential impacts, and promote policies and actions that reduce greenhouse gas emissions and support a sustainable future.
- Improve the quality and quantity of data used for climate change prediction by increasing the number of observational networks and developing new technologies for measuring key variables such as temperature, precipitation, and greenhouse gas concentrations.
- Enhance our understanding of the underlying physical and chemical processes that drive climate change, including the feedback loops and tipping points that can amplify or dampen the effects of global warming.

DATASET AND PYTHON SOURCE CODE

LINK to the Final Python Source Code (Notebook) https://www.kaggle.com/code/atongabraham/assignment5
LINK to the used dataset (Notebook)

REFERENCES

[1] G. Eason, B. Noble, and I. N. Sneddon, “On certain integrals of Lipschitz-Hankel type involving products of Bessel functions,” Phil. Trans. Roy. Soc. London, vol. A247, pp. 529–551, April 1955.

[2] C. Yin, F. Yang and J. Wang, ”Analogs of Future Climate in Chinese Cities Identified in Present Observations,” in IEEE Access, vol. 8, pp. 219151-219159, 2020, doi: 10.1109/ACCESS.2020.3042548.

[3] C. Forgotson et al., ”How Satellite Soil Moisture Data Can Help to Monitor the Impacts of Climate Change: SMAP Case Studies,” in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 13, pp. 1590-1596, 2020, doi: 10.1109/JSTARS.2020.2982608.

[4] T. Jia, K. Yang, Z. Peng, L. Tang, H. Duan and Y. Luo, ”Review on the Change Trend, Attribution Analysis, Retrieval, Simulation, and Prediction of Lake Surface Water Temperature,” in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 15, pp. 6324-6355, 2022, doi: 10.1109/JSTARS.2022.3188788.

[5] R. Bamdale, S. Shelar and V. Khandekar, ”How to tackle Climate Change using Artificial Intelligence,” 2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT), 2021, pp. 1-7, doi: 10.1109/ICCCNT51525.2021.9579674.

[6] J. Serra et al., ”Validation of AI models for ITCZ Detection from Climate Data,” 2022 5th International Conference on Data Science and Information Technology (DSIT), 2022, pp. 1-7, doi: 10.1109/DSIT55514.2022.9943879.
[7] Z. Yuan, L. Mou, Z. Xiong and X. X. Zhu, ”Change Detection Meets Visual Question Answering,” in IEEE Transactions on Geoscience and Remote Sensing, vol. 60, pp. 1-13, 2022, Art no. 5630613, doi: 10.1109/TGRS.2022.3203314.
[8] S. Brink, ”AI for Datacenter Optimization (ADOPT'22),” 2022 IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW), 2022, pp. 1-1, doi: 10.1109/IPDPSW55747.2022.00119.
[9] W. Jeong and S. Oh, ”Senior healthcare testbed design and construction: focusing on the temperature control system,” 2021 International Conference on Information and Communication Technology Convergence (ICTC), 2021, pp. 669-672, doi: 10.1109/ICTC52510.2021.9620907.
[10] M. Kim, J. Jun, N. Kim, Y. Song and C. S. Pyo, ”Sequence-to-Sequence model for Building Energy Consumption Prediction,” 2018 International Conference on Information and Communication Technology Convergence (ICTC), 2018, pp. 1243-1245, doi: 10.1109/ICTC.2018.8539597.