Prediction Mitigation of Climate Change impacts using Machine Learning

Tanish Chavan

March 15, 2025

1 Objective

The increasing urgency to address climate change necessitates innovative approaches that leverage data analytics and artificial intelligence (AI) to formulate effective strategies for mitigating its impact. This work aims to design an AI model capable of predicting and mitigating climate change impacts by analyzing historical climate data, current environmental conditions, and human activities. By providing valuable insights to policymakers and organizations, this solution seeks to foster the development of informed and actionable climate action plans.

2 Sources

The insights and data for this project are drawn from various reputable sources, including:

1. **World Bank Climate Data**: The dataset sourced from the World Bank provides comprehensive historical climate data, including greenhouse gas emissions, temperature variations, and other relevant environmental factors.

2. **Research Articles**: Relevant studies published in journals focused on climate science and environmental policy.

3. **Sustainable Development Goals (SDGs)**: The framework provided by the United Nations (UN) regarding climate action, particularly focusing on SDG 13, which addresses climate action.

3 Methodology

The proposed solution involves several key steps:

1. **Data Collection**: Collecting and preparing historical climate data, focusing on greenhouse gas emissions, temperature variations, and other relevant environmental factors.

2. **Data Cleaning and Preparation**: Ensuring the dataset is clean and structured for analysis, addressing missing values, and filtering the data for relevant features.

3. **Model Development**: Utilizing machine learning techniques, specifically linear regression, to model the relationship between historical climate data and future emissions.

4. **Model Evaluation**: Assessing the model's performance through metrics such as Mean Squared Error (MSE) and R-squared (R^2) scores.

5. **Visualization**: Utilizing visual tools to convey model outputs and feature importance, facilitating easier interpretation for stakeholders.

6. **Application**: Developing actionable insights and recommendations based on the model's predictions to assist policymakers in formulating effective climate action plans.

4 Focus on Sustainable Development Goals (SDGs)

This project directly aligns with the UN's Sustainable Development Goals, particularly Goal 13, which emphasizes urgent action to combat climate change and its impacts. By creating a predictive model that highlights trends and potential future scenarios, this project aims to support evidence-based decision-making and foster sustainable practices among governments and organizations.

5 Code Explanation

The following code outlines the methodology used to develop the AI model for predicting climate change impacts. It encompasses data loading, cleaning, modeling, evaluation, and visualization.

```
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error, r2_score
import numpy as np

# Load the dataset
file_path = '/mnt/data/climate-change-excel-4-6-mb-.csv'
dataset = pd.read_csv(file_path)
```

5.1 Step 1: Data Loading and Cleaning

```
# Replace placeholders like ".." with NaN and convert year columns to numeric
dataset.replace("..", np.nan, inplace=True)
```

```
yearly_columns = [str(year) for year in range(1990, 2012)]
dataset[yearly_columns] = dataset[yearly_columns].apply(pd.to_numeric, errors='coerce')

# Drop unnecessary columns that might contain non-numeric data
dataset.drop(columns=['Country code', 'Country name', 'Series code', 'SCALE', 'Decimals'],
```

- **Loading Data**: The dataset is loaded using `pandas`, which allows for easy manipulation and analysis of tabular data.

- **Replacing Missing Values**: Any placeholders (like ”..”) are replaced with `NaN` for better handling of missing values.

- **Numeric Conversion**: The yearly columns (from 1990 to 2011) are converted to numeric types to ensure they can be used in calculations.

- **Dropping Non-Numeric Columns**: Columns that do not contribute to the analysis (like country codes) are dropped to streamline the dataset.

5.2 Step 2: Filtering for Target Data

```
# Select rows that contain "CO2 emissions" in the Series name
target_data = dataset[dataset['Series name'].str.contains("CO2 emissions", case=False, na=Fa

# Drop 'Series name' column after filtering
target_data.drop(columns=['Series name'], inplace=True)

# Fill missing values with forward-fill
target_data.fillna(method='ffill', axis=1, inplace=True)
```

- **Target Data Selection**: The dataset is filtered to include only the rows that pertain to CO2 emissions, as this is a primary focus for climate change analysis.

- **Handling Missing Values**: Forward-filling is applied to ensure continuity in the data series, especially for years with missing emission data.

5.3 Final Data Cleaning

```
# Ensure all columns are numeric
target_data = target_data.apply(pd.to_numeric, errors='coerce')

# Drop any rows with remaining NaN values
target_data.dropna(inplace=True)
```

- **Final Data Cleaning**: Any remaining NaN values are dropped to ensure the dataset is ready for model training.

5.4   Step 3: Feature and Label Definition

```
# Define features (historical data from 1990 to 2010) and labels (use 2011 as the target)
features = target_data[yearly_columns[:-1]]  # All years except the target year
labels = target_data['2011']  # Use 2011 as the target variable
```

- **Feature Selection**: The features consist of emissions data from 1990 to 2010, while the label is the emissions data from 2011, which the model will aim to predict.

5.5   Step 4: Train-Test Split

```
# Step 3: Train-Test Split
X_train, X_test, y_train, y_test = train_test_split(features, labels, test_size=0.2, random_
```

- **Data Splitting**: The dataset is divided into training and testing sets, with 80% used for training the model and 20% reserved for evaluation. This ensures the model can generalize well to unseen data.

5.6   Step 5: Model Training and Evaluation

```
# Step 4: Model Training - Using Linear Regression
model = LinearRegression()
model.fit(X_train, y_train)

# Step 5: Model Prediction and Evaluation
y_pred = model.predict(X_test)

# Calculate Mean Squared Error and R2 Score
mse = mean_squared_error(y_test, y_pred)
r2 = r2_score(y_test, y_pred)

print(f"Mean Squared Error: {mse}")
print(f"R2 Score: {r2}")
```

- **Model Training**: A linear regression model is trained using the training dataset.

- **Predictions**: Predictions are made on the test dataset, allowing for performance evaluation.

- **Performance Metrics**: MSE and R^2 scores are calculated to assess model accuracy, where MSE indicates the average squared difference between predicted and actual values, and R^2 reflects the proportion of variance explained by the model.

5.7 Step 6: Visualization of Results

```
# Additional: Display coefficients to show feature importance
import matplotlib.pyplot as plt
import seaborn as sns

# Plot feature importance (coefficients for Linear Regression)
coef_df = pd.DataFrame({
    'Year': yearly_columns[:-1],
    'Coefficient': model.coef_
})
sns.barplot(x='Year', y='Coefficient', data=coef_df)
plt.xlabel('Year')
plt.ylabel('Coefficient')
plt.title('Feature Importance by Year (Linear Regression Coefficients)')
plt.xticks(rotation=45)
plt.show()
```

- **Visualization**: Using `matplotlib` and `seaborn`, the coefficients of the linear regression model are visualized to illustrate the importance of each year in predicting emissions. This helps to identify trends over time and communicate findings effectively.

6 Conclusion

This work presents an AI-driven model for predicting and mitigating climate change impacts by analyzing historical climate data, environmental conditions, and human activities. By employing machine learning techniques, this model provides a robust foundation for policymakers and organizations to develop effective climate action plans. As climate change continues to pose a significant threat to global sustainability, leveraging data and technology in decision-making processes will be crucial for achieving meaningful progress in combatting its effects. Future work may include enhancing the model with additional features, exploring alternative algorithms, and integrating real-time data for dynamic predictions.

By aligning with the Sustainable Development Goals, particularly SDG 13, this initiative underscores the importance of urgent climate action supported by empirical evidence and innovative solutions.