Unhealthy Jackets

Team Members:

Rhea Mathew, Tianqi Sun, Ronith Yalamanchili, Parth Shinde, Dristi Shah

Problem

Determine correlation between certain climate events, such as decreasing forest area and temperature, malaria spread, and commodity prices.

Datasets Used

- Forest Cover:

  https://raw.githubusercontent.com/pshinde612/Team-Unhealthy-Jackets/main/Forest_Area_ShareOf_Total_Area.csv

- Malaria Atlas:

  https://raw.github.com/pshinde612/Team-Unhealthy-Jackets/main/MalariaAtlas_Data_Vectors_1990-2009.csv

- Annual Corn Prices:

  https://raw.githubusercontent.com/pshinde612/Team-Unhealthy-Jackets/main/cornpricesglobalimf.csv

- Malaria Sets:

  https://raw.githubusercontent.com/pshinde612/Team-Unhealthy-Jackets/main/MALARIA_2000_2019_PV_ADMIN2_ALL_AGES.CSV

  https://raw.githubusercontent.com/pshinde612/Team-Unhealthy-Jackets/main/MALARIA_2000_2019_PF_ADMIN2_ALL_AGES.CSV

- Annual Rate of Anomalies in Temperature:

  https://raw.githubusercontent.com/pshinde612/Team-Unhealthy-Jackets/main/temperature_anomaly.csv

Preprocessing

1. Deleted irrelevant columns in large datasets
2. Cleaned up NaN values using function .na()
3. Narrowed down a range of years covered in all datasets
4. Had quarterly values for corn price, so used average to annualize and make compatible with other datasets
5. Grouped column data by year into one new dataset
6. Dropped year column

Mishaps

- Set year as x-value, malaria as the y-value
- Original linear regression error: ~500,000

Fixes

- Dropped year column since it was irrelevant
- New Linear regression error: ~170,000

- Set corn prices as the y-value, malaria as an x-value
- New SVM error: **0.0272**

Results

- X-values: temperature anomaly, forest cover, malaria species
- Y-value: corn price

Model	Mean Squared Error
SVM	0.0272
Linear Regression	0.0381

Current State

- Final product are good SVM and Linear Regression models with respectable values for error and accuracy
- Clearly shows a correlation between environmental/climate factors and corn prices

Future Plans

- More data!
- X-values: Add more climate/environment factors such as pollution/emissions of various greenhouse gases
- Y-Values: Add price data of other crops, and of different financial products of agricultural commodities, including derivatives, futures contracts, and bonds