ECE 417 - Final Project Proposal Team A

Argiris Adam, Koffas George, Chatzistefanidis Ilias

March 2019

The problem Most of our energy comes from fossil fuels, and burning these fuels causes environmental problems; in particular, global warming. Specifically, global warming:

- raises the sea level[1][2][3]
- brings drought in tropical regions near the equator
- increases the frequency of hurricanes, tornadoes, and floods[4]
- causes the spread of diseases[5][6]

The consequences are serious and have the potential to bring tremendous unrest in the world. Over the past 50 years, the average global temperature has increased at the fastest rate in recorded history. Additionally, experts see the trend is accelerating: All but one of the 16 hottest years in NASA's 134-year record have occurred since 2000.

The data There are a range of organizations that collate climate trends data. The three most cited land and ocean temperature data sets are 1. NOAA's MLOST, 2. NASA's GISTEMP and 3. UK's HadCrut. Kaggle has repackaged the data from a newer compilation put together by Berkeley Earth, which is affiliated with the Lawrence Berkeley National Laboratory. The Berkeley Earth Surface Temperature Study combines 1.6 billion temperature reports from 16 pre-existing archives. It is nicely packaged and allows for slicing into interesting subsets (for example by country). In this way, we will use several files from Kaggle's dataset "Climate Change: Earth Surface Temperature Data" for our survey. In specific:

- Global Average Land Temperature by Country (GlobalLandTemperaturesByCountry.csv)
- Global Average Land Temperature by State (GlobalLandTemperaturesByState.csv)
- Global Land Temperatures By Major City (GlobalLandTemperaturesByMajorCity.csv)
- Global Land Temperatures By City (GlobalLandTemperaturesByCity.csv)
- Global Oceanic and Land Temperatures (GlobalTemperatures.csv)

The model Data in Time Series Forecasting (TSF) surveys are highly correlated because of the time factor. This is why it is of great importance for special models to be applied. Thus, we will use Seasonal Auto Regressive Integrated Moving Average (SARIMA), Long Short-Term Memory (LSTM), XGBoost and AutoML models. Moreover, we consider it necessary to observe how accurate some traditional ML methods would be, so in addition we will use Support Vector Machine (SVM), Kernel SVM, Ridge and Linear Regression models.

How the model will be evaluated We will split every dataframe resulting from our initial dataset in training and testing set. In this way we will fit our models with the training set and make predictions using the testing set. Next, we can compare the predictions with the real values using Root Mean Square Error (RMSE) and each model's score/accuracy to conclude how valid every model's results are.

Anticipated challenges It is understandable that the big challenge in this dataset is the correlation between its features. There is yearly, seasonal and daily correlation among the temperatures and we should handle it very carefully in order to achieve appreciable results.

The promise Through our analysis, we hope to

1. Raise awareness over the real threat of global warming.
2. Actively contribute to the ever-growing research field of global warming by determining a list of high-precision models for wider use in the global warming time-series forecasting problem.

References

[1] S. Jin, L. Zhang, and G. Feng, "Earth's surface fluid variations and deformations from gps and grace in global warming," 2011.

[2] L. Hinnov, K. N. Ramamurthy, H. Song, M. Banavar, and L. Spanias, "Interactive tools for global sustainability and earth systems: Sea level change and temperature," in *2013 IEEE Frontiers in Education Conference (FIE)*, pp. 1965–1970, 2013.

[3] W. G. S. L. B. Group, "Global sea-level budget 1993–present," *Earth System Science Data*, vol. 10, no. 3, pp. 1551–1590, 2018.

[4] J. Chen and Y. Gao and L. Wen, "A study of the relationship between global warming and hurricane activity based on time series," in *2018 International Conference on Robots Intelligent System (ICRIS)*, pp. 460–463, 2018.

[5] B. Andrick, B. Clark, K. Nygaard, A. Logar, M. Penaloza, and R. Welch, "Infectious disease and climate change: detecting contributing factors and predicting future outbreaks," in *IGARSS'97. 1997 IEEE International Geoscience and Remote Sensing Symposium Proceedings. Remote Sensing - A Scientific Vision for Sustainable Development*, vol. 4, pp. 1947–1949 vol.4, 1997.

[6] V. Petraşcu, G. Czibula, D. Sîrbu, M. Popa, and D. Curşeu, "Identifying the impact of global warming on population using a clustering based approach," in *2010 IEEE International Conference on Automation, Quality and Testing, Robotics (AQTR)*, vol. 2, pp. 1–6, 2010.