**Plan:**

* 3 minutes
  + 1st minute:
    - intro,
    - problem,
    - aims
  + 2nd minute: what we did/our approach
    - Data
    - Model selection
    - Changing Y?
  + 3rd minute:
    - Graph,
    - improvements,
    - review aims,
    - conclusion/thanks

**Text:**

|  |  |
| --- | --- |
| 1 | Hello, we’re Paul, Fernando, and Rodrigo and this is our presentation for our machine learning project for the Master for Data Science programme at the Hertie in Fall 2022. |
| 2 | For our project, we’ve chosen to apply machine learning models to a crucial and complex question for policy makers. How do we predict the consequences of climate change on mortality worldwide? |
| 3 | As the effects of man-made climate change are felt, excess deaths due to extreme temperatures and weather events are expected to rise. Climate change can directly and indirectly impact human health in multiple ways.  Increasing food insecurity, the spread of disease vectors, armed conflicts, and extreme weather events can all be attributed to our changing climate. |
| 4 | Our aims:  1. Predict the consequences of rising temperatures in terms of mortality  2. Test different machine learning models and their efficacy for this problem  3. Demonstrate the different outcomes between predicted emission scenarios |
| 5 | Two emissions scenarios:  RCP 45 - Realistic-emissions scenario (medium concentration of CO2)  RCP 85 - High-emissions scenario (high concentration of CO2) |
| 6 | **Data we used:**   * ***Our World in Data: causes of death*** * ***World Bank: global population*** * ***EU Copernicus Climate Data Store: Temperature & emissions data***   The end result of transforming the data lead to:   * Country * Temperature * Population * Deaths due to temperature variation |
| 7 | We  tested and compared predictive regression models to anticipate the number of excess deaths due to climate change. |
| 8 | Models we used:  We used the following metrics to evaluate our models' performance:   * R2 Score * Adjusted R2 * Mean Squared Error * Root-Mean-Square Error   Decision Tree  XGBoost  Gradient Boost  Linear Regression  Gaussian Process Regression  After tuning, the best performing model was Gradient Boost. |
| 9 |  |
| 10 | In the context of our chosen policy domain, we believe that we illustrated the varying effectiveness of widely-used machine learning models in answering our guiding question.  We were able to make predictions about temperature-related mortality, but our predictions could be more accurate.  We need more variables to see more variation in our predictions, particularly for the different emissions scenarios  We hope to make more accurate predictions and our results show that complex models perform better than traditional models, like logistic regression |

Predicting the consequences of our changing climate for public health is a crucial area where machine learning can assist policy makers. As the effects of man-made climate change are felt, excess deaths due to extreme temperatures and weather events are expected to rise.

Climate change is a massive problem and will impact the world in a number of ways

One key prediction is rising temperatures

* + Rising temperatures has immediate implications for mortality and deaths

One key variable is emissions

* + The outcomes for public health under the two different scenarios could look very different
* Our project was about three things:
  + Predicting the consequences of climate change
  + Demonstrating the different outcomes between predicted emission scenarios
  + Demonstrating the difficulty of applying ML to this question

We applied our selected models to two separate, yet credible climate scenarios, further demonstrating the complexity of climate-change mitigation and adaptation for public health authorities and bodies.

**Complication:**

* **Data:**
  + Causes of death from Our World in Data
  + Global population from the World Bank
  + Temperature data from the EU Copernicus Climate Data Store
* **Models:**
* In sum, we used the following five models on our data sets:
  + Linear Regression
  + Decision Tree
  + Gaussian Process Regression
  + Gradient Boosting Regressor
  + Extreme Gradient Boosting Regressor
* **Tuning:**

???

* **Results:**
* Best performing model was:
  + Gradient Boosting Regressor
* Our model predicted a causal relationship between temperature and mortality.
* However, it didn’t recognize a difference between the emissions scenarios.
  + We need more data to make this work
  + Further tuning of models
  + And time lag built into models

**Conclusion:**

* We were able to make predictions about the outcomes of the different scenarios.
* We need more variables to see more variation in our predictions
* But at least we have a start and we’ve demonstrated the difficulty of applying machine learning to this complex public policy space.

We hope to make more accurate predictions and it turns out, that complex ML models perform better than traditional models, like logistic regression

**450 words**