

Basics of Machine Learning

1.2 Ethics and Direction of Machine Learning Programs

Kendra Jackson

1. **Task:** Read or watch one of the examples of machine learning ethics below. Each example looks at how machine learning was used to address different technical or societal issues. What potential is there for bias or ethical issues when dealing with climate change data? Where would ClimateWins need to be cautious about using machine learning to develop answers? Write about 200 words showing what pitfalls ClimateWins should avoid
2. **Questions to Address:**
 - a. Is there personal information that may be exposed?
 - b. Are there regional or cultural biases in climate change that might be made worse by machine learning?
 - c. Is there human bias in climate change that might be propagated while training machine learning?
 - d. Could machine learning potentially make incorrect decisions about where weather conditions might worsen and cause harm?

Article: [What are Ethics and Why Do They Matter? Machine Learning Edition](#)

[Machine Learning and Human Bias](#)

Exposure of Personal Information:

Using climate data typically does not risk the exposure of personal information. That being said, using geo-specific data has the potential to expose details about certain areas and potentially even the people living there. Data anonymity and excluding geo-specific details can assist in preventing this risk. At a glance, there does not appear to be any PII within there data set.

Regional and Cultural Biases:

Data gathered from different weather stations across Europe may not represent the geographical areas evenly; distribution of stations could be skewed causing bias. Data collection from earlier periods of time may have bias towards regions with larger population and infrastructure. To mitigate this risk, data could be weighted from regions with less historical data (worth more) to equalize the data set and adjust for these potential biases. Additionally, machines trained on a specific type of region may not adjust well to prediction for other regions.

Human Bias Impact:

As the data set is recorded over an extended period, earlier data collection methods may not be on par with more recent methods. Older collection methods may not have the precision that new methods do which can cause bias in earlier data records. The data set could be restricted to analyze periods of time in which the measurement techniques and devices that were used were the same to account for this potential bias. Additionally, if using supervised machine learning, since an individual is setting parameters, rules, and restrictions, their bias can be transferred to the model.

Potential for Incorrect and Harmful Decisions:

Machine Learning models trained on older data may not be able to accurately predict due to the change in recording technology as well as climate change. Old data does not include recent events of climate change that have and will continue to lead to extreme weather events. Should the model be trained on this old data, and not be able to account for this, it could potentially not predict these dangerous events, thus leading to inaccurate predictions with potential risk of harm. In order to prevent this, ClimateWins needs to incorporate models that can adjust to changing conditions and should test model predictions for recent and future global environmental extreme events.

Summary:

Machine learning can assist in the prediction of weather patterns, but it is crucial to understand its limitations and potential for bias. ClimateWins must ensure thorough checks of its data to ensure ethical concerns and potential biases are addressed. ClimateWins must handle the data they have carefully and consider weighting newer data more heavily than older data. To ensure their model remains accurate and does not become outdated, ClimateWins should perform ongoing monitoring and updating of the models it is using for weather prediction. In doing so, ClimateWins should be able to mitigate areas of bias that could potentially affect the integrity of their predictions.