CAREERFOUNDRY

Machine Learning with Python: Weather Conditions and Climate Change with ClimateWins

Objective

In this project, you'll use machine learning to help predict the consequences of climate change while working as a data analyst at a European nonprofit organization.

Introduction and Company Profile

Congratulations! You've been hired as a data analyst by a European nonprofit organization, ClimateWins. Because ClimateWins is a nonprofit, it doesn't have a lot of funds to hire a dedicated data scientist and engineering team. This means you'll have a chance to train on some more advanced topics and do some work normally done by data scientists.

Cimate Wins is interested in using machine learning to help predict the consequences of climate change around Europe and, potentially, the world. It's been sorting through hurricane predictions from The National Oceanic and Atmospheric Administration (NOAA) in the U.S., typhoon data from The Japan Meteorological Agency (JMA) in Japan, world temperatures, and a great deal of other data. However, it's hard to grasp how everything is changing in the world at once. This is where machine learning comes in! Because ClimateWins is still a small organization without a lot of resources, you'll be working as a data analyst, data scientist, researcher, and trainee all at the same time to help them with this project.

The first task you'll tackle is getting up to speed with machine learning methods, then identifying the best resources for predicting climate changes where people live. Machine learning is traditionally created and used by data scientists, so this is an excellent opportunity to further your knowledge as a data analyst and get experience with more complex projects.

You'll first have to determine which areas can use machine learning. Next, you'll look at data that already includes answers so that machine learning can be trained to recognize similar answers (supervised learning). Then you'll clearly communicate what you found (your results) and how they'll affect ClimateWins' strategy.

After you've identified the best models to use in Achievement 1, you'll put them to use in Achievement 2 as you work with ClimateWins' data and start making predictions.

Project Background

ClimateWins wants you to assess the tools available to categorize and predict the weather in mainland Europe. It's concerned with the increase in extreme weather events, especially in the

past 10 to 20 years. However, it thinks that even weather extremes could be predicted and planned for using advanced tools such as machine learning. With data from the past century, it hopes to create a model for what the future will hold.

Here are the questions it wants to answer first:

- How is machine learning used? Is it applicable to weather data?
- ClimateWins has heard of ethical concerns surrounding machine learning and AI. Are there any concerns specific to this project?
- Historically, what have the maximums and minimums in temperature been?
- Can machine learning be used to predict whether weather conditions will be favorable on a certain day? (If so, it could also be possible to predict danger.)

Context

Python is one of the most-used programming languages when it comes to working with machine learning. There are also many different types of machine learning available, each of which is suitable for different subjects and classifications. One of the largest divisions exists between supervised and unsupervised learning models. Both can be used to predict new data—just in different ways. Throughout this Achievement, you'll learn the building blocks of optimization and machine learning so that you can assess which types of models would be most helpful for predicting weather data. By the time you finish, you should be able to prepare data for machine learning analysis, perform the analysis with both simple and complex algorithms, and derive information from the models' results.

Ethical Considerations

AI and machine learning are hot topics, and both disciplines are developing rapidly. As with any new technology or theory, there is a period of time before regulatory bodies catch up. Think of seatbelts, for example, which were not required in cars until the 1960s, and not made mandatory to wear until a few decades later! Machine learning elements have just as much capability to change the world as cars had and will require more regulation in the future. For now, it's important to be aware of what ethical concerns might arise.

Data Sets

You'll be using a data set based on weather observations from 18 different weather stations across Europe, which contain data ranging from the late 1800s to 2022. Recordings exist for

almost every day with values such as temperature, wind speed, snow, global radiation, and more. This data is collected by the European Climate Assessment & Data Set project.

- Download the temperature data set (.csv, 16.6MB).
- Create a “Data Sets” directory where you can download raw data.
- Create folders labeled “Supervised” and “Unsupervised.”

Project Deliverables

Throughout this Achievement, you’ll be working from Exercise to Exercise to train on the workings of machine learning while assessing the correct tools for your eventual predictive analysis.

For the task in each Exercise, you’ll submit a deliverable that directly contributes to the final product—in this case, a proposal for the best way to predict weather patterns. Your mentor will review the deliverables for Exercises 1.1 through 1.5 and your final proposal. They’ll also give you feedback on how your project relates to the work of a data analyst.

Below is a breakdown of your course project deliverables by Exercise:

Exercise 1.1 Machine Learning History and Tools

- Describe the difference between machine learning and AI
- Explain where machine learning is used in industry
- Identify the difference between linear and nonlinear goals
- Identify tools for machine learning

Exercise 1.2 Ethics and Direction of Machine Learning Programs

- Explain why there are ethical issues in machine learning and automation
- Describe the data landscape for your project
- Explain the difference between supervised and unsupervised learning
- Prepare your data for analysis

Exercise 1.3 Optimization in Relation to Problem-Solving

- Identify whether your data is linear or nonlinear
- Understand how weights, loss, backpropagation, and learning rates are used
- Describe the different types of optimization algorithms and how they work.
- Identify local and global maximums and minimums in your data

Exercise 1.4 Supervised Learning Algorithms Part 1

- Identify what supervised machine learning can and can’t be used for
- Describe inductive problem solving
- Explain the terminology behind supervised learning and how classification is used
- Use supervised learning to predict if a day will be pleasant or not

Exercise 1.5 Supervised Learning Algorithms Part 2

- Identify what supervised machine learning can and can’t be used for
- Explain the uses for regression and classification in supervised learning
- Use supervised learning to predict if a day will be pleasant or not

Exercise 1.6 Presenting Machine Learning Results

- Explain why nearly all data analysts are well-positioned to advise on strategic issues
- Identify the dangers in cognitive bias
- Describe why humans are still needed to assess machine learning results
- Recognize how to lead a presentation to ensure valuable information is retained
- Create and give a presentation of your assessment of machine learning tools for ClimateWins