Poisonous vs. Safe?

DS 4002 Case Study by Mark Haller

Link to GitHub Repo: github.com/Markmh03/DS-4002-CS3

Imagine you are hiking through the woods and you can only take one of two paths. Both of the paths require you to trudge through dense foliage. For path one you must go through the plant on the left, and for path two you must trudge through the plant on the left. Hopefully you selected the left path because if you picked the right, get ready for an itchy hike back to your car.

Poison ivy, oak and sumac are annoying, but a substantial portion of the American population are severely allegoric to these and other poisonous plants. None of us want to go through a patch of anything that makes us itch, but many people have to take strong precautions before hiking to make sure they do not encounter plants such as the above to avoid having to use an EpiPen or a trip to the ER. I can reasonably guess that your parents told you growing up that going outside was good for you and this is backed up by research. Studies show significant benefits for people going outside and enjoying green spaces.[1]

Now, you have recently been hired as a data scientist for a startup AI company. This company aims to create AI that gets people out into nature more to increase awareness of the great outdoors and just make us happier in general. This AI will be designed as a phone app so that the user can take photos of various plants and wildlife and have a system be able to accurately identify it.

Given this, you were staffed on the team dedicated to detecting poisonous vs non-poisonous plants so that anyone, and especially users with allergies, can avoid accordingly. Using image recognition, your goal is to train a model that can attain an accuracy of at least 70% for the product testing phase. You will use the dataset the company has provided from its contracted photographers so get a large enough sample for your results to be meaningful.

[1] pmc.ncbi.nlm.nih.gov/articles/PMC6562165/