

# The Power of Data to Predict Forest Fires

**Date :** September 29, 2020

Fire season on the West Coast of the U.S. has been nothing short of disastrous in 2020. The numbers are staggering. More than [5 million acres](#) combined have burned in California, Oregon and Washington so far in some of the largest fires ever recorded. By all accounts, the experts say that since the summer, 2020 is the most active fire year on record for the West Coast – ever.

And the flames rage on.

“The El Dorado Fire, which was caused by [pyrotechnics] from a gender-reveal party, is 15 miles from my house in California,” says Emily Fu, an undergraduate in business analytics at the Wharton School who is now back in Philadelphia for her senior year. “For us, it has become such a normal thing. Every couple of years the mountain catches fire and if it’s really bad, we’ll evacuate. I call home pretty often and my family didn’t even mention it. I asked my dad, ‘How close is the fire?’ He sent me a video and said, ‘Oh, you can see the flames.’”

## A Little Modern Data Mining

While the wildfires of recent years may no longer stoke Emily’s anxiety, the numbers rising out of the smoke and ash have ignited her curiosity.

Emily and her classmates Zhun Yan Chang and Melisa Lee, also Wharton seniors, began digging deeply into wildfire data last year as students in [Prof. Linda Zhao](#)’s Modern Data Mining class at Wharton. Data mining is the practice of examining large databases – for instance “Wildfires in the U.S. for the Past 30 Years” – in order to generate new insights.

“Social impact should be at the center of your data project.” — Melisa Lee, Wharton Student

Inspired by Emily’s personal connection to the fires and their collective fascination with data analytics, the team set out to answer a specific question: “Given its features, can we predict the size of a fire?”

The team ultimately created a data model, a descriptive diagram of relationships between various types of information, for their class project that may actually help prevent forest fires in California. Their work earned them a spot among the industry professionals, Ph.D students and professors presenting at the Women in Data Science Conference at the University of Pennsylvania in February 2020.

They were excited to tell their fire-research story, inspired by weeks of data collecting and analysis. “We found out that a very small percentage of fires, less than 1%, cause 80% of the destruction. That told us if we can prevent that less than 1% and focus on stopping those fires, we can cut down on a big portion of the destruction,” notes Emily. “So, what’s causing those fires? Most of these are fires from lightning striking a tree or forest...When we ran the model, it spit out some specific vegetation types that were predictors in how big a fire was — like timber litter or dead branches that fall on the ground and are dry. They are causing these large fires. Our storyline: These lightning fires are the most destructive, they spike up in the summer, and are also linked with vegetation.”

Plans to share their findings with the California government have been delayed by the pandemic. Still, the three data scientists-in-training were eager to pass along some takeaways from their research process when Wharton Global Youth

---

checked in with them last week.

## **Fruit Trees or Lightning?**

First, data science demands you to extract the most compelling story behind the numbers. You can build models and encode things, and suddenly you're left with 30 variables that, for example, influence fire size. Finding the commonalities between all these things? That's up to you. And arriving at the best conclusions will take time. "Once you get the proper data, you shouldn't straight out build a model," says Zhun, a finance major from Malaysia. "You should look at the data, figure out how it is going to fit into the big picture and the impacts of each variable. Use data analysis, simple charts and visualizations to have an idea of what your final results should be."

Then you begin to shape the story that you want to tell. "Our team could have told many stories with our data," notes Melisa, who is pursuing a career in marketing and business analytics. "An alternate story was that the vegetation type around orchards is also very flammable because of the fruits. But that story would not have driven as much impact or incited people to act as much as our lightning story. Social impact should be at the center of your data project. Building a model should be for some ultimate purpose to let people know that something is happening and you have the data to back it up."

Emily, Melisa and Zhun are convinced that data has immense power (did you know it has replaced oil as the world's most valuable resource?). They plan to spend part of their senior year working to get their project into the hands of decision makers who are confronting the worsening West Coast wildfires. "Our model highlighted a lot of high-risk areas that we can target.

We can target these by using prescribed burns, which are smaller controlled fires that will burn up all the timber litter by summer so that once fire season comes along, there's not much to catch on fire," says Emily. "We want our model to motivate the California State Department to focus on having more prescribed burns, but also to make sure the ones we do have are well targeted so we're preventing the larger fires going forward."

## Do You Data?

Emily Fu, Melisa Lee and Zhun Yan Chang share some moments of inspiration on their paths to data domination:

**ASK QUESTIONS.** Don't be afraid to raise your hand because you think you're the only one who doesn't get it. You're not! Be shameless in asking questions.

**LEVERAGE YOUR SKILLS.** A technical model that nobody understands has little value. "I bring myself back to that when I feel like a male classmate has more technical skills," says Melisa. "If you can't communicate it to people, then it won't drive impact. Our team uses our strength in communication to our advantage."

**PUT IT IN CONTEXT.** Even in a high school statistics class, apply data to your own life. "Over the summer, I was playing around and made this Excel spread sheet of my Pokemon battle strategy," says Emily. "Relate what you're learning to an interest outside of class."

**SHOW UP.** Don't be intimidated by data. Both women and men need to have a voice in this space.