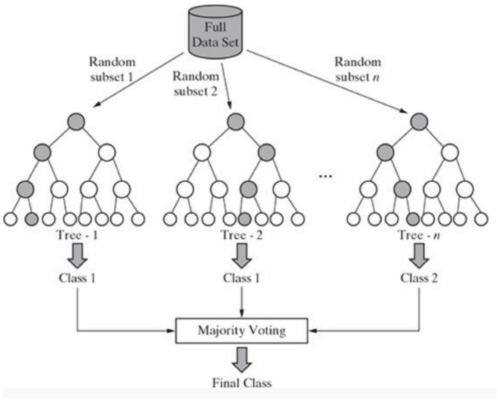
- https://today.usc.edu/using-ai-to-predict-wildfires/
    - Used a cWGAN, which stands for Conditional Wasserstein Generative Adversarial Network
        - “operates like a standard GAN, but with the added ability to "condition" the generated output based on extra information (like class labels), by feeding this conditioning data to both the generator and discriminator, essentially guiding the generator to produce outputs that align with the specified conditions, while still using the Wasserstein distance metric to measure the quality of generated data, leading to more stable training compared to traditional GANs”
        - Using a GAN makes sense because the use cases of GANs is image generation. There are two parts to a GAN — the discriminator and the generator. The goal of the generator is to generate images that mimic the training distribution and fool the discriminator into thinking that the image actually came from the training set. Once the discriminator has reached a certain threshold of confusion, then the training stop. By training the generator to fool the discriminator, the generator learns features and behaviors in the data set. This way, you could use cWGANs to simulate real wildfires.
- https://www.sciencedirect.com/science/article/pii/S2666719324000244
    - Actually does a survey about different methods to predict wildfires and finds that random forest performed the best *for predicting wildfires in central valley.*
    - *“Utilizing the dataset in the following two variations were a random split and a chronological split of training and testing sets. The best-performing algorithm using this dataset was Decision Trees at 550 maximum splits with an F1-Score of 0.689.”*
    - *“the specific variables used within many of the ML models were wind speed, temperature, soil moisture, soil depth, precipitation, population, Normalized Difference Vegetation Index (NDVI), land cover, global human modification, elevation, drought, and topography.”*
    - 

    - Based on results I can play around with Decision Tree and Random forest and compare results to qwen.

- https://essopenarchive.org/users/713243/articles/697596-a-global-probability-of-fire-pof-forecast
    - “XGBoost improves the performance of decision trees by sequentially adding 86 models to correct errors made by previous ones and uses regularisation techniques to prevent overfitting. The 87 optimisation of XGBoost makes it computationally efficient when compared to RF. “

- **https://fireecology.springeropen.com/articles/10.1186/s42408-024-00335-2**
    - Also agreed that in general, Random Forest/Decision trees have the best performance
    - Past results they talk about:
        - What is interesting is they talk about an assembled model of SVM and Random Forest
        - Also, a specific time of Decision tree was mentioned: logit boost group-based decision tree (LBGb-DT)
    - They test LR and Random Forest
    - Results are lower than I like:
        - “To better assess the skill of the LR and RF in wildfire prediction, the models are trained using data from 2001 to 2009 and then validated with 2010–2018 datasets in both case studies. Results show that the performance rates of the LR for the validation period are **54.6%** and **69.8%** in the Okanogan and the Jamésie regions, respectively. Similarly, the performance rates of the RF are **55.5% and 75.5%**”
- https://ral.ucar.edu/solutions/benefits/improving-wildfire-prediction-cutting-edge-satellite-imagery-and-ai
    - Our scientists fed the satellite data into a machine learning (AI) model known as a “random forest” survey that estimates tree mortality. In fact, the AI model was able to accurately update the LANDFIRE fuel data, defining the majority of the fuels previously categorized as “timber litter” or “timber understory” to “slash blowdown,” the term used for heavy tree mortality.

NEXT STEPS:

1. Try the model second paper claimed works the best: Decision Trees at 550 maximum splits with an F1-Score of 0.689.
2. Try random forest + XGBoost or with ADA Boost
3. Try ensemble Random Forest and SVM
4. Try Qwen.

It may be better to use DT or Random Forest as base model and LLM that communicates and talks with the users.