Georgia Tech

RED HOT

BY TEAM 152

Sierra Pitman, Christopher Holmes, Victor Tchervenobrejki and Gabriel Glazer

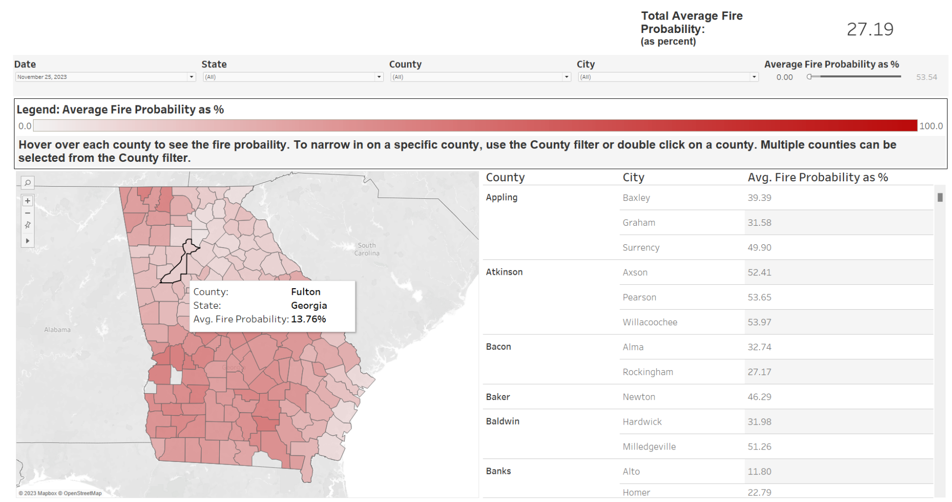
MOTIVATION/INTRODUCTION

WHAT IS THE PROBLEM:

- Goal: predict wildfires in Georgia based on weather forecasts and visualize results in Tableau
- How can we more quickly and accurately predict wildfire occurrences?
- How do we easily display the results to people who may not have a data science background?

WHY IS IT IMPORTANT:

- Wildfire damage accounted for $18B losses in 2022.
- $4.4B spent on wildfire suppression in 2021.
- Accurate predictions can guide efficient firefighting resource allocation.
- Communities can be alerted on the event of high risk.
- Potential wins: less casualties, less property damage, preserved ecosystem



OUR APPROACH

WHAT IS THE APPROACH AND HOW DOES IT WORK:

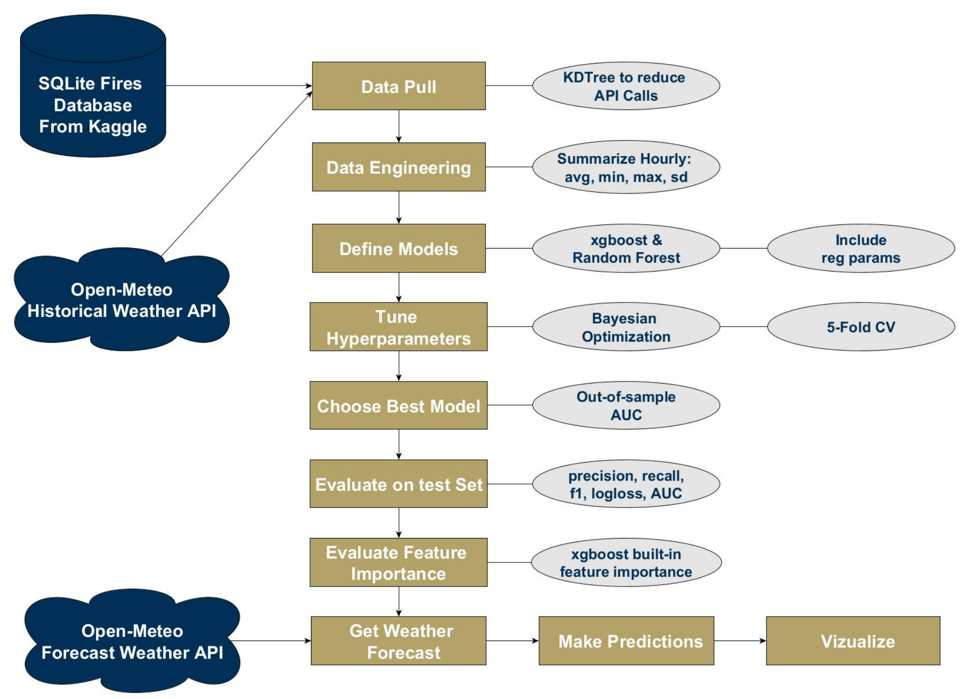
- Choose best between XGboost and Random Forest classification models.
- Elastic net regularization for variable selection.
- Hyperparameter tuning with Bayesian Optimization and k-fold cross-validation.
- Built-in feature importance of xgboost to assess factors of wildfire risk.
- Evaluation metrics on 20% test set: logloss, ROC AUC, precision, recall, f1.
- Visualize the predictions in Tableau.

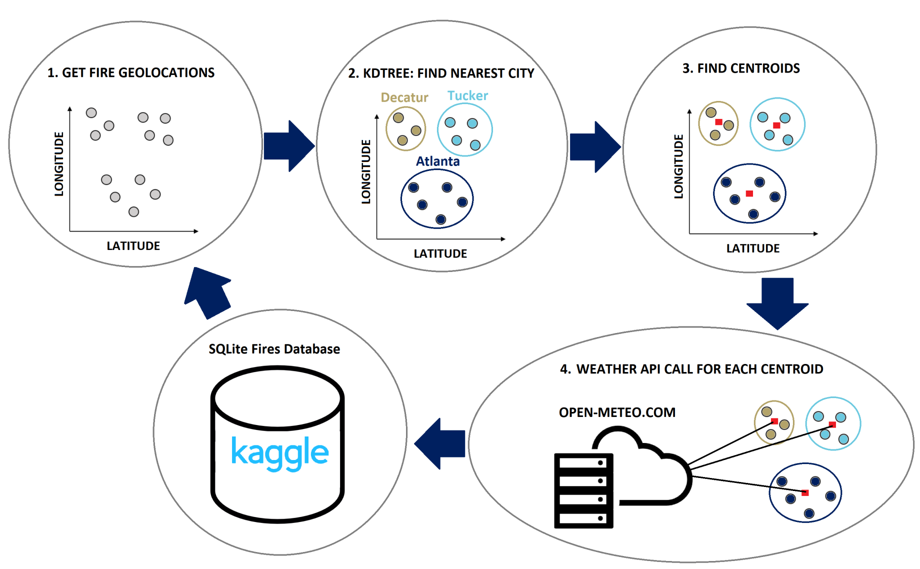
WHY DOES IT SOLVE THE PROBLEM:

- Computational efficiency of XGBoost and Bayesian Optimization.
- Simplicity and interpretability of Random Forest.
- Cross-validation ensures that model is robust and not prone to overfitting.
- Holistic set of evaluation metrics that indicate the model's real-world relevance.
- Leveraging visualization power of Tableau.

WHAT IS NEW IN OUR APPROACH:

- Robust variable selection through elastic net.
- Capturing non-linear functional forms through ensemble algorithms.
- Enhanced data collection strategy using KDTree.





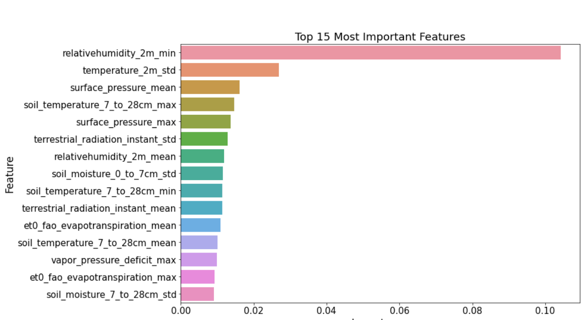
DATA

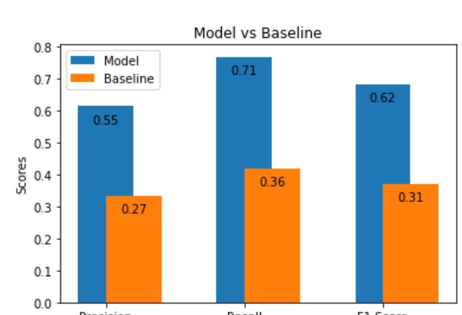
HOW DATA WAS COLLECTED:

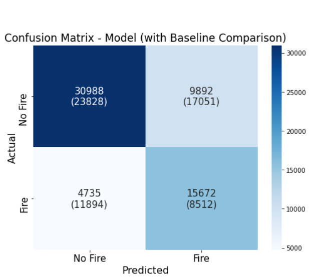
1. Get dataset from Kaggle that contains 1.8 million fire occurrences.
2. Identify the nearest city of each fire geo-location using KDTree mapping.
3. Reduce geolocations, to a single central location for each city.
4. Perform API pulls from open-meteo.com for each of the 660 cities in Georgia.
5. Join fire data with weather data and rebalance the dataset.

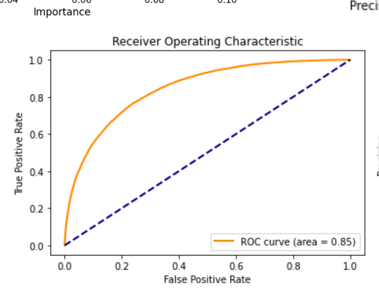
WHAT ARE THE DATA CHARACTERISTICS:

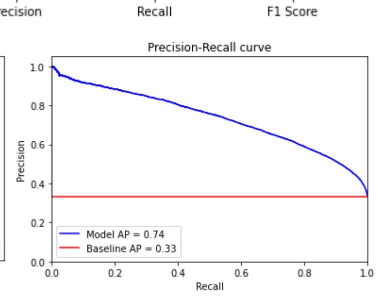
- Fire data includes: location and time of fire occurrences
- Weather data includes 178 weather features relating to: temperature, humidity, dewpoint, precipitation, snowfall, cloud cover, visibility, evapotranspiration, wind metrics, soil conditions, UV index, and solar radiation.











EXPERIMENTS AND RESULTS

HOW THE PROJECT EVALUATED:

- Model selection based on out-of-sample ROC AUC
- Model accuracy based on 20% test-set: logloss, ROC AUC, Precision, Recall, F1 Score
- Built-in feature importance of xgboost

WHAT ARE THE RESULTS:

- Model is twice better than baseline.
- Precision: 0.55 versus 0.27 baseline
- Recall: 0.71 versus 0.36 baseline.
- F1 Score: 0.62 versus 0.31 baseline.
- AUC: 0.85 versus 0.5 baseline.
- Logloss: 0.64 versus 0.46 baseline.
- Most important features: relative humidity 2 meters above the ground, temperature and soil-temperature related features, surface pressure, evapo-transpiration and terrestrial radiation

HOW OUR METHOD COMPARES TO OTHER METHODS?

Our method is unique in the choice of algorithms, data engineering and data fetching approaches.