

main

1 branch

0 tags














Go to file

Add file

About

Code

Settings


	42abbe6	3 minutes ago	168 commits
	Data	4 of 4	yesterday
	ima...	done	20 minutes ago
	.giti...	Initial commit	5 days ago
	EDA...	Add files via u...	yesterday
	LICE...	Initial commit	5 days ago
	Mod...	done1	17 minutes ago
	REA...	Update READ...	3 minutes ago
	not...	done	20 minutes ago
	post...	Add files via u...	yesterday
	pre...	done	20 minutes ago
	upd...	fixed update_r...	18 hours ago
	web...	Add files via u...	yesterday

☰ README.md

ClassiFires

Wildfires

No description, website, or topics provided.

-  Readme
-  MIT license
-  Activity
-  0 stars
-  2 watching
-  0 forks

Releases

No releases published
[Create a new release](#)

Packages

No packages published
[Publish your first package](#)

Languages

 Jupyter Notebook 100.0%

Suggested Workflows

Based on your tech stack

Management Complexity Project



Project Overview

The goal of this project is to develop a classification model that can predict the fire management complexity level of a wildfire. Fire management complexity represents the highest management level utilized to manage a wildland fire. This target provides valuable insights into the resources needed and the potential scale, size and impact of a fire.



Actions Importer

Set up

Automatically convert CI/CD files to YAML for GitHub Actions.



Django

Configure

Build and Test a Django Project



Python application

Configure

Create and test a Python application.

[More workflows](#)

[Dismiss suggestions](#)

This classification model analyzes various features associated with the a wildfire incident, including meteorological data, bureaucratic data, and locational data. The developed classification model will enable fire management agencies to anticipate the resources required should a wildfire occur based on the location and current meteorological data. This model doesn't replace the realtime complex decision of determining fire management complexity level (like evaluating the risk to the firefighters). However this model can aid in helping predict if a new wildfire incident will be a large scale/impacting event based on the predicted fire management complexity.

Business Understanding

Wildfires pose significant risks to life, property, and the environment. Effective fire management is crucial for mitigating these risks and minimizing the impact of wildfires. Predicted fire management complexity level evaluates if a location is at risk of being resource intensive or potentially large threat to life and property should a wildfire occur. This can help identify regions that need to be on high alert and preparedness to minimize the impact of a wildfire.

The fire agencies administrator is responsible for setting the fire management complexity level. Their decision follows a set of standardized and subjective guidelines. Some of these guidelines are utilized as features in this project.

Fire management agencies, administrators, and other personnel responsible for allocating resources and planning fire response strategies would benefit from being able to accurately and efficiently predict the fire management complexity level of wildland fire incidents. This allows for them to anticipate the required resources, as well as assess the potential scale and impact of a fire. This information is crucial for fire management agencies to make informed decisions and ensured preparedness for faster fire response times.

Data Understanding

- [Target](#)
- [Data Sources](#)
- [Data Directory](#)
- [Key Features](#)

Target

- FireMgmtComplexity (Defined [here](#))

Factors contributing to the fire management complexity level:

- Area involved
- Threat to life and property

- Political sensitivity
- Organizational complexity
- Jurisdictional boundaries
- Values at risk
- Fire behavior
- Strategy and tactics
- Agency policy

Source:

https://gacc.nifc.gov/swcc/management_admin/Agency_Administrator/AA_Guidelines/pdf_files/ch5.pdf

FireMgmtComplexity Classes:

The levels of wildfire fire incidents range from Type 5 to Type 1. Each level represents a specific level of complexity

- Type 5:
 - lowest class
 - local resources
 - 2-6 firefighters
 - quickly contained
 - low impact risk
- Type 4
 - Local resources
 - low impact risk
 - slight increase in scale compared to Type 5
- Type 3
 - Mix of local and regional resources used
 - increased scale and risk
 - action plan created

- Type 2
 - large scale 200+ firefighters
 - Many units required
 - regular planning and briefing
- Type 1
 - highest class
 - Same characteristics of type 2 incident
 - 500+ firefighters
 - aircraft and aviation is used
 - Greater access to resources
 - larger scale and impact

Data Sources

The data used in this project comes from the following sources below:

- Wildfire Occurrences
 - <https://data-nifc.opendata.arcgis.com/datasets/nifc::wildland-fire-incident-locations/about>
 - This dataset gets updated daily and contains data going back to roughly 2014
- Live RAWS Data (Remote Access Weather Stations)
 - <https://data-nifc.opendata.arcgis.com/datasets/nifc::public-view-interagency-remote-automatic-weather-stations-raws/about>
 - Dataset of live RAWS data

- Historical RAWS Data
 - <https://raws.dri.edu/>
 - Contains historical data for around 3k RAWS
- Elevation data:
 - open elevation api

Data Directory

Data
station_list.csv
threshold_year.pickle
nessid.csv
RAWS_Historical_Full
RAWS.csv
stations_dates.csv.zip
RAWS_stations.csv.zip

Wildland_Fire_Incident_Locations.c

clean_fire_data.csv.zip

fire_elevation.csv

fire_model_data.csv

Key Features

Below are the key features used in this project. Several features in the dataset have corresponding features that contained the same or similar data. These features were utilized to fill in missing values whenever possible. There are many more features than what is listed here, refer to source websites for an indepth overview.

Fire Incidents:

Definitions provided by source

- **FireMgmtComplexity:** The highest management level utilized to manage a wildland fire

- **FinalAcres:** Final burn acres, nulls filled in with IncidentSize
- **site:** Created in Modeling.ipynb, closest RAWS that has at least 50% data coverage over the duration of a fire incident.
 - It is used as a reference point for analyzing weather conditions during the fire event.
- **DispatchCenterID:** A unique identifier for a dispatch center responsible for supporting the incident. Nulls filled in with POODispatchCenterID
- **POODispatchCenterID:** A unique identifier for the dispatch center that intersects with the incident point of origin (point where fire incident occurred)
- **POOJurisdictionalAgency:** The agency having land and resource management responsibility for a fire incident as provided by federal, state or local law
- **POOFips:** Code identifies counties and county equivalents. The first two digits are the FIPS State code and the last three are the county code within the state.

- **FireDiscoveryDateTime:** The date and time a fire was reported as discovered or confirmed to exist
- **FireOutDateTime:** The date and time when a fire is declared out
- **OBJECTID:** Incident ID for dataset
- **EstimatedFinalCost:** Nulls filled in with EstimatedCostToDate
- **elevation:** Elevation of fire incident (meters)

RAWS data:

For each fire incident, all meteorological metrics were computed as averages of the fire duration.

- **NESSID:** NESS ID for identifying RAWS
- **X:** Longitude
- **Y:** Latitude
- **date:** date when data was collected, if null then no data collected on that day
- **total_solar_radiation_ly:** Solar radiation
- **ave_mean_wind_speed_mph:** Average wind speed (mph)

- **ave_mean_wind_direction_deg**
: Average wind direction
(degree)
- **max_maximum_wind_gust_mph**
h: Maximum wind gust (mph)
- **ave_average_air_temperature_deg_f**: Average air temperature
(°F)
- **ave_average_relative_humidity**
: Average relative humidity
- **total_precipitation_in**: Total
precipitation (inches)

Data Set length: over 250K and the final model dataset has a length of 7731
RAWS: There are roughly 2252 RAWS sites with usable data, around 3k in total

Project Rundown

- All needed data is saved under data folder refer to each notebook to import those datasets
 - refer to [Data Directory](#) for dataset source/curation and where its used.
1. Run model [Modeling notebook](#) up until Thiessen Polygon, this provides a general overview of the fire incident occurrences and features

2. Webscrape urls, nessids, RAWs code, and more [webscrape notebook](#)
3. Mimics post request to pull RAWs historial data from 2014-2023 [post request notebook](#)
4. Load in RAWs json files, creates 2 dataframes for further EDA and cleaning, a simple data cleaning is performed to reduce datasize [EDA1 notebook I](#)
5. Continue with [Modeling notebook](#) Further EDA, generates meteorological attributes for each fire, modeling, and evalutaion, This is the main notebook

Data Preprocessing

EDA - Historical RAWs Data

This invloved changing -9999 values to nulls, dropping rows that are beyond each stations data collecting time period, creating dataframe where eaach column is a RAWs and row inputs are days, if data doesn't exist for that date then a null is placed,

- changing from the 4 digit code

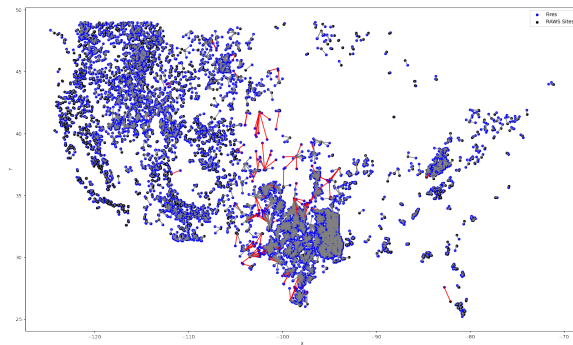
used with post request to the
NESSID

- fix datetimes columns

EDA - Fire Occurrences Data

- Dropping duplicates
- remove days where end date was before the start date and other irregularities, set to only use points within the contiguous USA
- fix datetimes columns

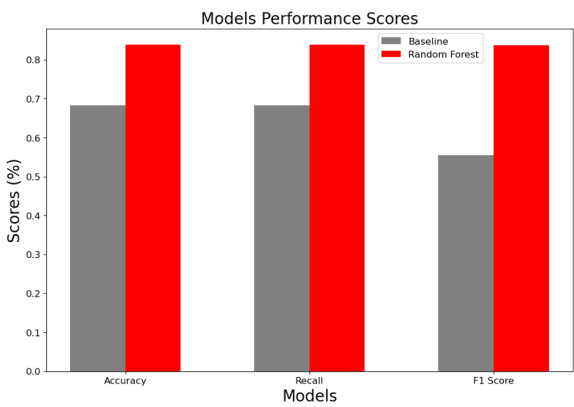
EDA - Final Dataset



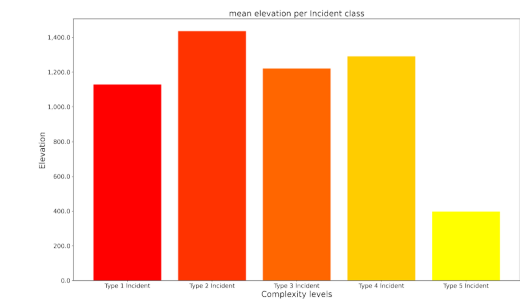
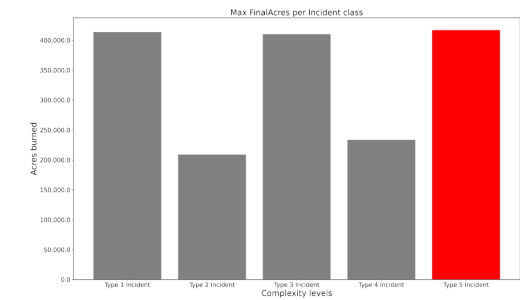
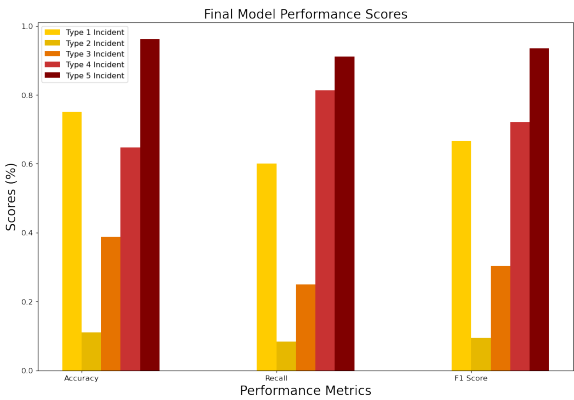
- generate the mean meteorological metrics for each fire
- drop nulls
- fix datetimes columns

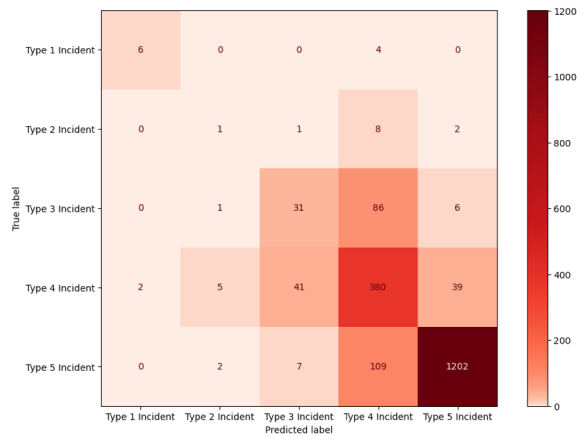
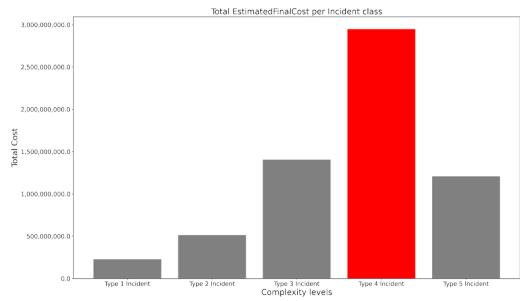
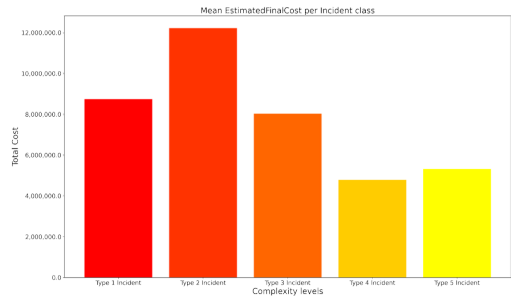
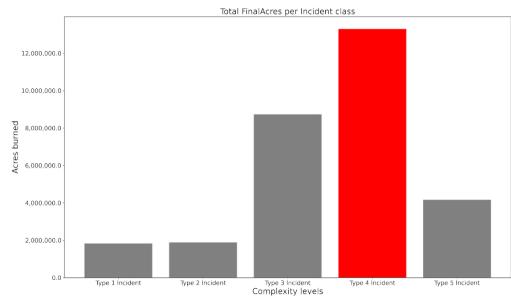
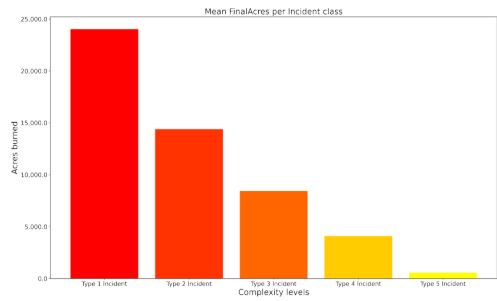
Modeling

A Random forest classifier, decision tree and dummy model were used

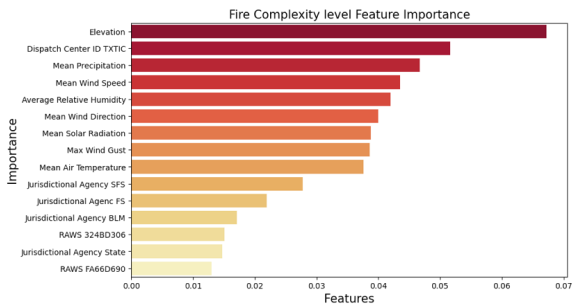


Performance





Feature Importance



Conclusion

The final model scores:

Complexity level	precision	recall
Type 1 Incident	0.75	0.60
Type 2 Incident	0.11	0.08
Type 3 Incident	0.39	0.25
Type 4 Incident	0.65	0.81
Type 5 Incident	0.96	0.91
accuracy		
macro avg	0.57	0.53

weighted avg	0.84	0.84
-----------------	------	------

The final model performs best at predicting type 5 incidents, even though I used smote, the majority of wildfires occur at the type 5 incident. This means that most fires are put out within a few days and or only require a few firefighters. Type 4 incident is one level up and type 1 incidents have the next best performance. With type 2 and 3 performing poorly. Looking at the usability of this model it is more significant to be able to predict both extremes well. If a fire incident is 1 day old it is likely still at type 5, this model will be able to use current and forecasted meteorological data, and bureaucratic features such as agency and dispatch center to predict the fire incidents fire complexity level. Further evaluation shows that the highest mean acres burned and economic cost correlate with type 1 incidents, This for one confirms that fire complexity levels do correlate with fire scale and impact. However, this is not absolute, when evaluating the max acres burned for each level types 1, 3, and 5 all share close max acres burned. This could be an error in the data or Possibly more underlying factors influencing the fire complexity level. One speculation is that large fires occurring in heavily

remote regions are less of a risk to people and communities. Further analysis also shows that type 4 incidents have the largest cumulative acres burned and economic costs. This is likely due to just the class imbalance as the mean shows that type 1 incidents are significantly higher in both features.

Next Steps

Looking at the next steps, I am looking to further improve model performance by adding additional features such as calculating drought data, remoteness index, and improving RAWs site selection. After this, I am looking to build a streamlit deployment of the model. This will involve setting up APIs and pulling current RAWs data, and potentially forecasted Meteorological data.

Contact Information

Email:

ginell_k1@denison.edu

Github:

Linkedin:

<https://www.linkedin.com/in/keanan-ginell>

Repo Structure

```
|— .gitignore
|— Data
|   |— RAWS.csv
|   |— RAWS_Historical_Full
|   |   |— set1.zip
|   |   |— set2.zip
|   |   |— set3.zip
|   |   |— set4.zip
|   |—
|   |— RAWS_stations1.csv.zip
|   |—
|   |— RAWS_stations2.csv.zip
|   |—
|   |— RAWS_stations3.csv.zip
|   |—
|   |— Wildland_Fire_Incident_Location
|   |—
|   |— clean_fire_data.csv.zip
|   |   |— fire_elevation.csv
|   |   |— fire_model_data.csv
|   |   |— nessid.csv
|   |   |— station_list.csv
|   |   |—
|   |— stations_dates.csv.zip
|   |—
|   |— threshold_year.pickle
|   |— EDA1.ipynb
|   |— LICENSE
|   |— Modeling.ipynb
|   |— README.md
```

```
|— images
|   |— Incident_scores.png
|   |— Target_Acres_max.png
|   |—
|Target_Elevation_mean.png
|   |—
|Target_acres_mean.png
|   |— Target_acres_sum.png
|   |— Target_cost_mean.png
|   |— Target_cost_sum.png
|   |—
|feature_importance.png
|   |— fire_site_map.png
|   |— matrix.png
|   |— matrix_red.png
|   |— mike-newbry-
|_AwSiaesk40-unsplash.jpg
|   |— scores.png
|— notebook.pdf
|— post_request.ipynb
|— presentation.pdf
|— update_readme.py
|— web_scraper.ipynb
```