Forecasting Post Fire Debris Flow Risk

Improving Predictions Through Geospatial Analysis and Machine Learning Algorithms

Presented by:

James Gumtau

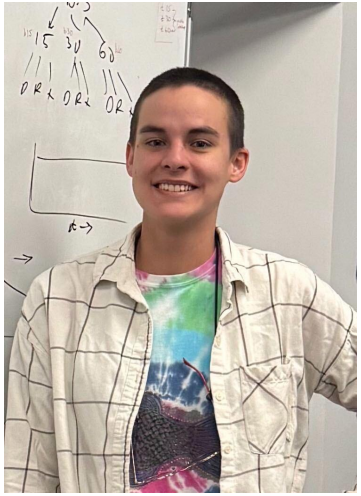
Computer Science
Engineering

Experience:
- Administrative Assistant

Wilking Chali Alphonse

B.A. Business Management

Experience:
- Project Management
- Startup Founder



Samantha Dicker

Mathematics

Experience:
- Healthcare .

Agenda

1. Libraries & Resources.
2. Data Overview.
3. Model Selection & Model Performance.
4. Explain key Predictors & Important Features.
5. Impact on communities and the environment.
6. Geospatial Analysis.
7. Conclusion.
8. Future Considerations

Libraries


jupyter

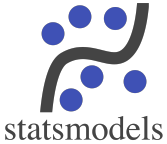

NumPy


python


scikit
learn
machine learning in Python


Pandas


simtable


statsmodels


plotly


tableau


pickleDB


USGS
data science

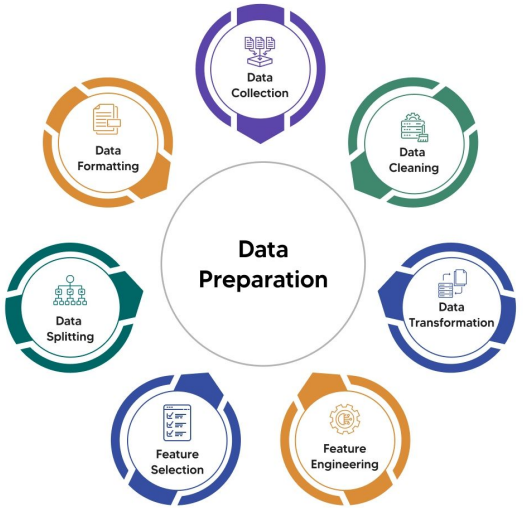

matplotlib

Data Overview

- Excel file, 202 kb.
- Data ranges from 2000 to 2012.
- 1550 Rows & 27 columns.
- States included (CA,UT,CO,AZ,MT,NM).
- 5 Identifiers.
- Data normalization (Drop columns, mean,Target Value).
- Data divided by rain intensity 15,30 & 60 minutes.

[USGS Data Source]

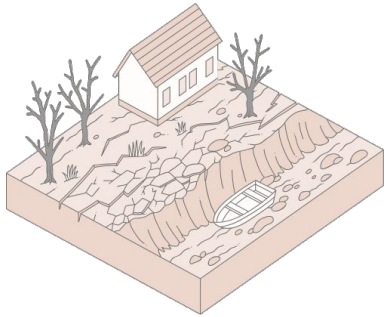
```python
def fill_nulls_with_mean(df):
    """
    Replaces null values in a DataFrame with the mean of each column.

    Args:
        df: The pandas DataFrame to process.

    Returns:
        A DataFrame with null values replaced by column means.
    """

    for column in df.columns:
        column_mean = df[column].mean()
        df[column].fillna(column_mean, inplace=True)

    return df
```

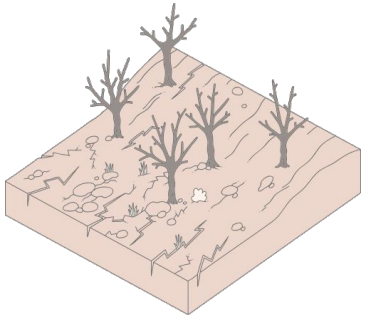


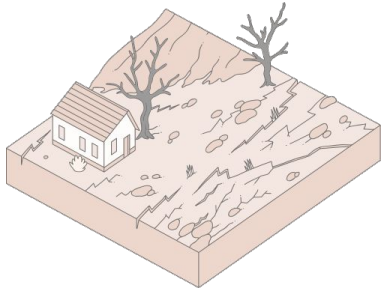
Model Selection

This is a supervised binary classification machine learning problem, where we are trying to predict whether there will be a post fire debris flow or not.

- **Logistic Regression:** simplicity, interpretability, and ease of implementation.
- **Gaussian Naive Bayes:** Determine the probability of each class occurring independently of any features.
- **Decision Trees / Random Forest / XGBoost:** Great for features selection, not scaling needed.

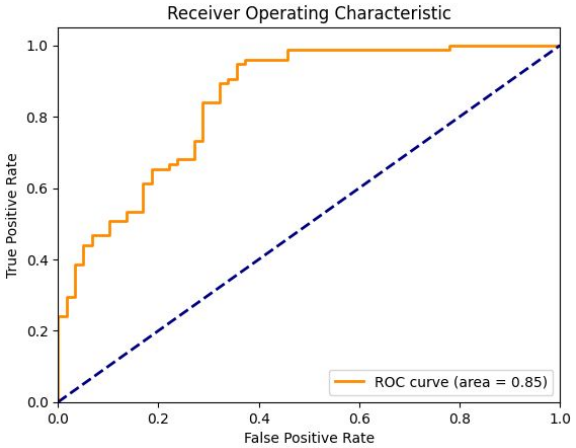






Model Performance

Logical Regression

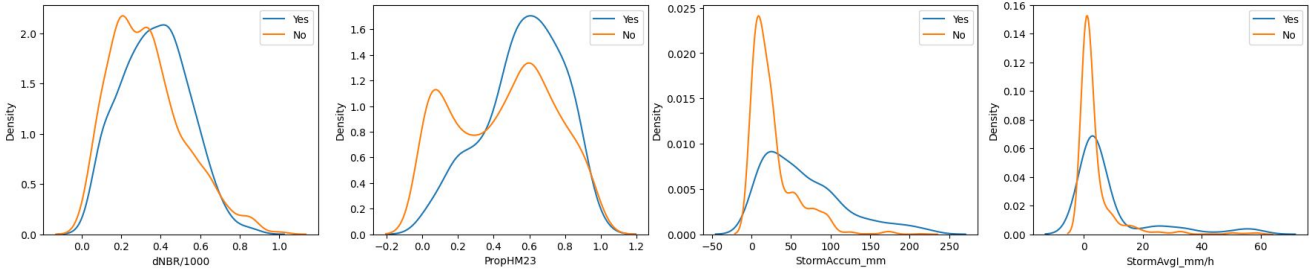


Rain Duration	Accuracy Score	F1 Score
15 minutes	74.6%	77.9%
30 minutes	83.6%	84.1%
60 minutes	82.8%	84.6%

The threshold value of burn severity above which the likelihood of debris flow significantly increases

Optimal threshold of 34.0%

```
Optimal dNBR/1000 threshold: 34.031815914731844
```



(dNBR) Differenced Normalized Burn Ratio

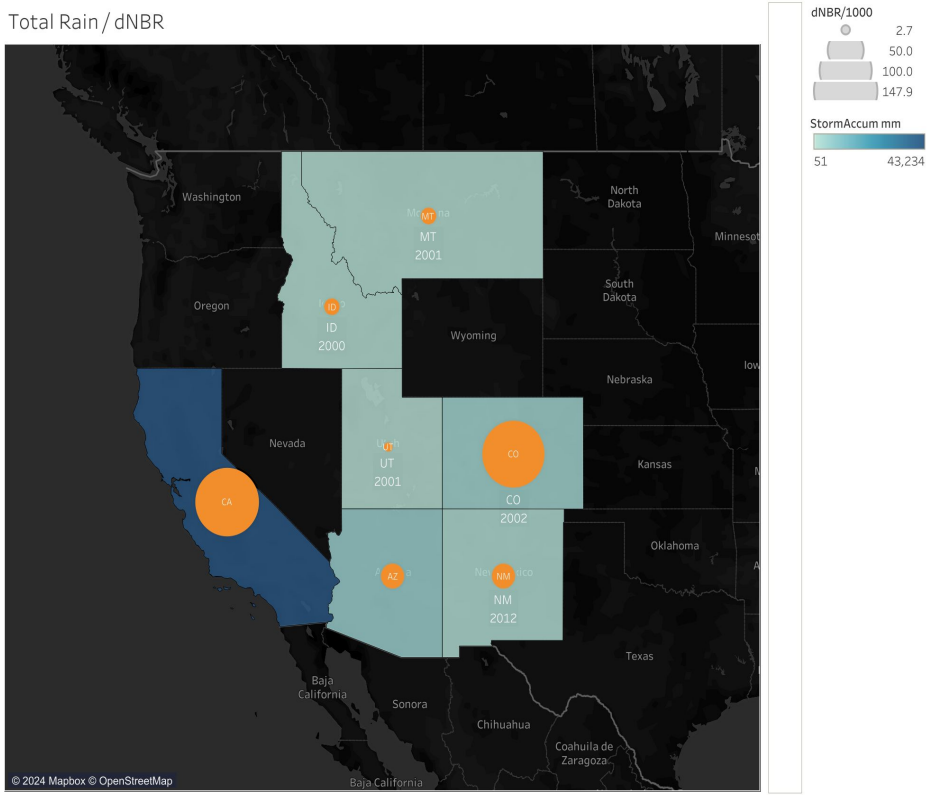
Wildfire severity indicates high levels of tree mortality, severe soil heating, and loss of soil organic matter. Such conditions increase the likelihood to erosion.

Steep Slopes (Gradients in excess of 23 degrees)

This metric identifies the portion of the watershed where both severe fire impacts and slopes occur at the same time.

Relationship between dNBR, Slope & Rain accumulation

High burn severity weaken vegetation and soil, reducing their ability to absorb and infiltrate water. When high storm accumulation occur, the runoff from these burned slopes can rapidly mobilize soil, vegetation debris and rocks, resulting in destructive debris flows.



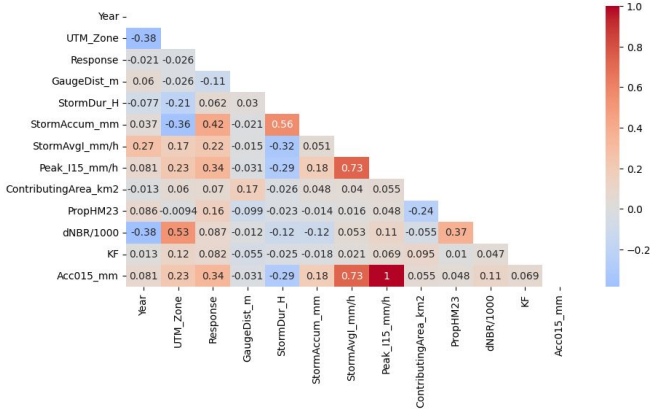
Naive Bayes

Analyze

- The longer a storm is happening(Storm_Dur), the more total rainfall, on average(Storm Accumulation)
- The more rainfall (Storm Accum_mm), the more likely there is to be debris flow (Response)
- dNBR correlation with location (UTM_zone) allows us to predict the severity of future fires in a given area of interest.

Evaluation

- When the storm lasts a long time, there is more rain. When there is more rain, there is more debris flow.
- The amount of times that someplace has burned in the past can determine the severity of future fires.

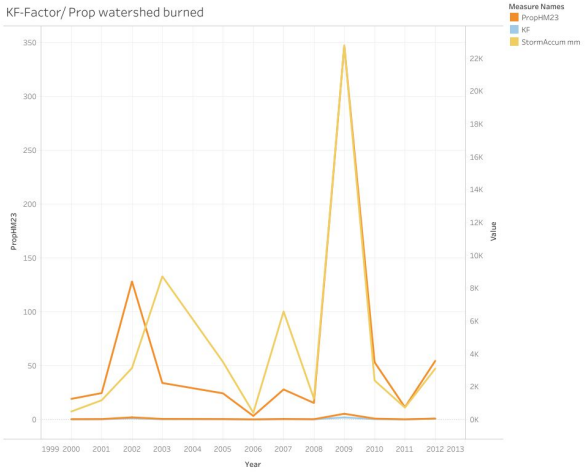


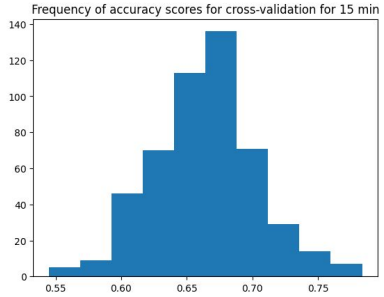
Rain Duration	Accuracy Score	F1 Score
15 minutes	65.7%	57.0%
30 minutes	64.2%	50.2%
60 minutes	70.1%	61.5%

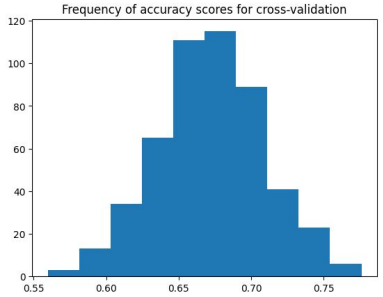
Comparing the GNB Models + KF-Factor

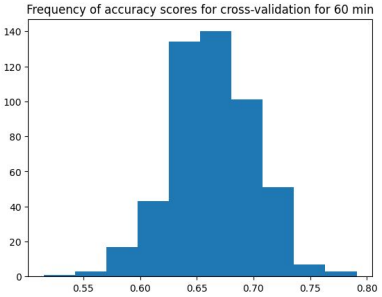
In the graph to the right, we can observe that the amount of rainfall (Storm_Accumulation mm) has close tracking with mild to severely burned watersheds(PropHM23). This tells us that a higher proportion of steeply sloped, severely burned areas might amplify the watershed's vulnerability to post-fire debris flows when storms bring significant rainfall.

The models below show us that the machine learning algorithm is consistent, informing that we could use the data we trained our model on to predict debris flows in the future.

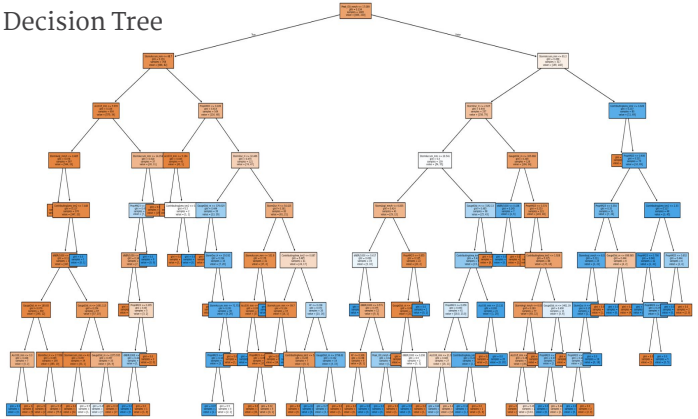








Decision Tree

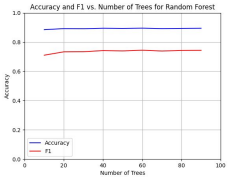


Tree Models

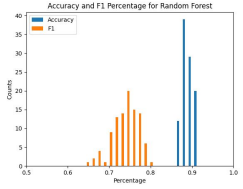
Random Forest & XGBoost

Model performance

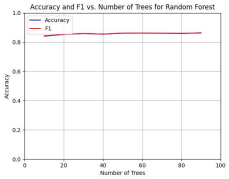
Skewed Data



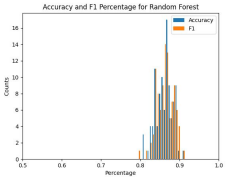
Response	
0	1216
1	334



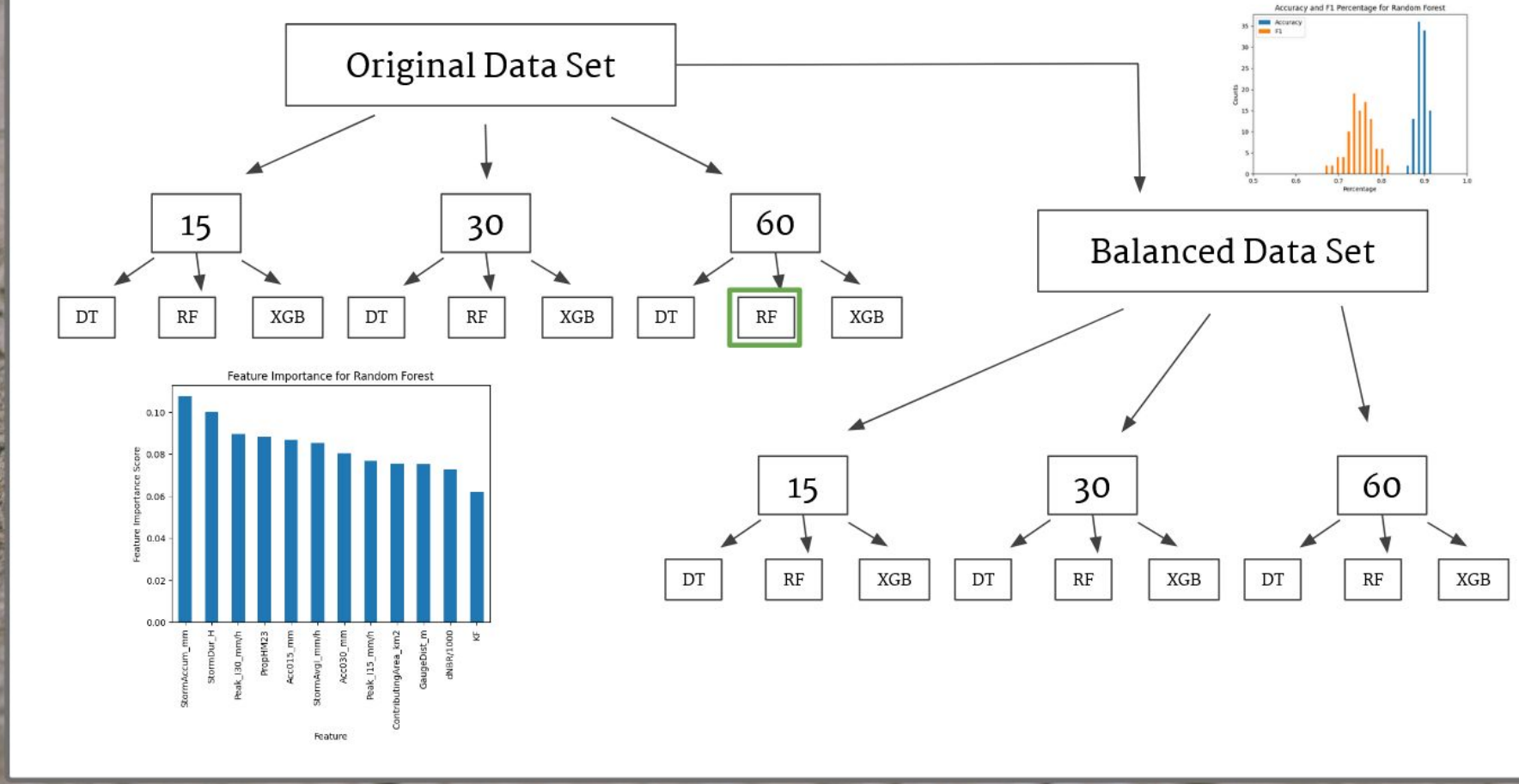
Balanced Data



Response	
0	334
1	334



Model	Skewed Performance	Balanced Performance
DT (15 min)	Accuracy: 0.8607956989247312 F1: 0.6738410295609579	Accuracy: 0.808258706467662 F1: 0.8063632329497099
RF (15 min)	Accuracy: 0.8936989247311828 F1: 0.7416156024925162	Accuracy: 0.8632835820895521 F1: 0.8660680169290758
XG (15 min)	Accuracy: 0.8878924731182795 F1: 0.7264997069781338	Accuracy: 0.8587562189054726 F1: 0.8595188660808282
DT (30 min)	Accuracy: 0.8593548387096774 F1: 0.6617485606828316	Accuracy: 0.8016915422885572 F1: 0.8032800855553732
RF (30 min)	Accuracy: 0.8924946236559141 F1: 0.7374975247615316	Accuracy: 0.8606467661691544 F1: 0.8627147838115127
XG (30 min)	Accuracy: 0.8907956989247313 F1: 0.7332846715670152	Accuracy: 0.8544776119402985 F1: 0.8570675563847203
DT (60 min)	Accuracy: 0.8652258064516128 F1: 0.6789314522512142	Accuracy: 0.7985074626865672 F1: 0.7985314065033233
RF (60 min)	Accuracy: 0.8961075268817205 F1: 0.7476221083505411	Accuracy: 0.8510447761194029 F1: 0.8543690279883429
XG (60 min)	Accuracy: 0.8955913978494624 F1: 0.7463907352985568	Accuracy: 0.8518905472636819 F1: 0.8545725069845822


Original Data Set
15
30
60
DT
RF
XGB
DT
RF
XGB
DT
RF
XGB
Balanced Data Set
15
30
60
DT
RF
XGB
DT
RF
XGB
DT
RF
XGB
Accuracy and F1 Percentage for Random Forest
Accuracy
F1
Counts
Percentage
Feature Importance for Random Forest
Feature Importance Score
0.00
0.02
0.04
0.06
0.08
0.10
StormAccum_mm
StormDur_H
Peak_i30_mm/h
PropHM23
Acc015_mm
StormAvgi_mm/h
Acc030_mm
Peak_i15_mm/h
ContributingArea_km2
GaugeDist_m
dNBR/1000
KF
Feature


Jesus Figueroa
FLOOD EMERGENCY IN EFFECT
EVACUATIONS IN RUIDOSO
BREAKING
NEWS
KOAT

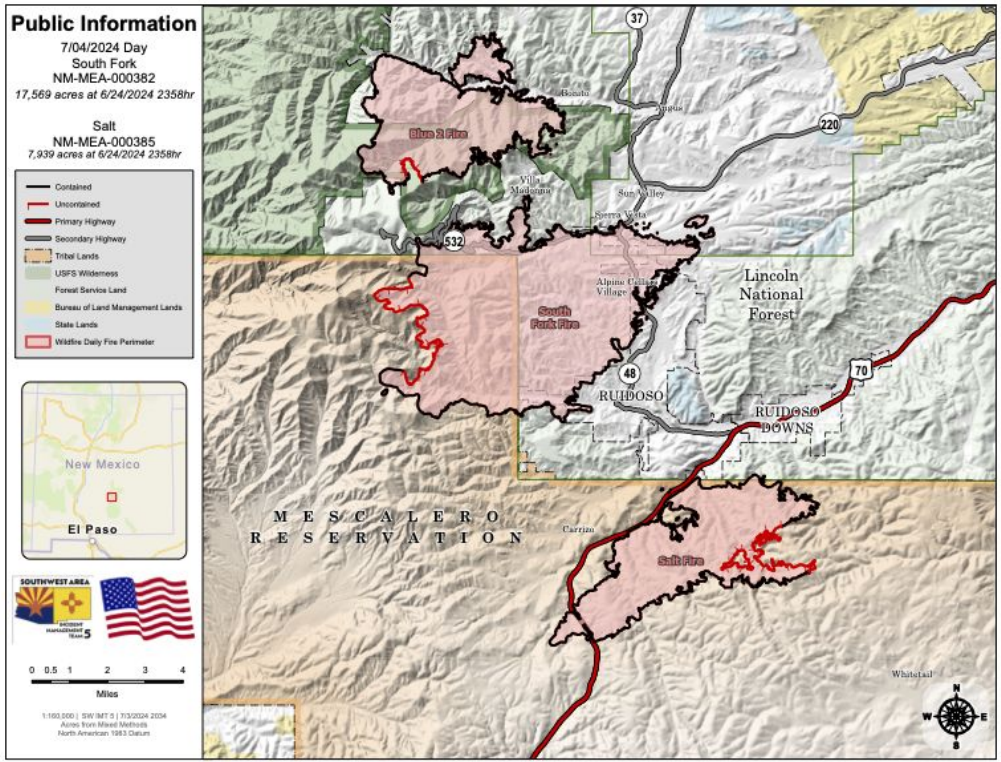

Mark Stambaugh

Ruidoso, NM

Geospatial Analysis

Ruidoso recent fires

- Blue 2 fire (7,532 acres)
  May/16/2024 to June 12, 2024
- South Fork Fire (17,569 acres)
  June/17/2024 to July 15, 2024
- Salt Fire (7,939 acres)
  June/17/2024 to July 7, 2024
- Flash floods
  Sunday, June 30, 2024



Socioeconomic Consequences

Livelihood Disruption

Local communities that depend on forests for their livelihoods are often the hardest hit.

Population Displacement

Local communities that depend on forests for their livelihoods are often the hardest hit.

Water Scarcity

Reduced forest cover and increased drought lead to lower water availability.

Health Impacts

Increased healthcare costs burden both individuals and healthcare systems.

Conclusion

- Areas with both high burn severity and steep terrain exhibited the highest likelihood of debris flow events.
- Rainfall characteristics were crucial factors in triggering debris flows.
- The model demonstrated reasonable predictive accuracy in identifying areas at high risk of post-fire debris flow. However, further model refinement and validation are necessary.
- The findings emphasize the need for targeted mitigation efforts in areas with high dNBR/1000 values and steep slopes. This may include slope stabilization measures, improved drainage, and early warning systems.
- Further research is needed to investigate the influence of other factors, such as soil type, vegetation cover, and antecedent moisture conditions, on post-fire debris flow occurrence.

Future considerations

- Investigate how the risk of debris flow changes over time after a fire, accounting for factors such as vegetation regrowth and soil stabilization.
- Validate model predictions through field surveys and observations to assess their accuracy and reliability.
- Integrate the model with real-time weather data and monitoring systems to provide timely warnings of potential debris flow events.
- Regularly update the model with new data and refine it as new information becomes available.

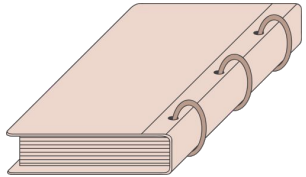
Thanks!
Gracias!

Questions?

Acknowledgments

Subject Matter Expert

Stephen Guerin
CEO, Founder
https://simtable.com/



Data Science Instructor

Robert Citek

Java & Android Instructor

Nick Bennett

Deep Dive by CNM Management Staff

Devonna James
Nora Meah
Esteban Martinez
Sue Andres