Understanding and Modeling Human Mobility Response to California Wildfires

Justin Liu, Lyndsey Umsted, Piero Trujillo, Ellen Burrell

Sponsor: Somayeh Dodge | Mentors: Evgeny Noi, Laura Baracaldo Lancheros, Enbo Zhou

UC SANTA BARBARA | Data Science Initiative

Introduction

- **Goal:** Analyze changes in movement patterns during California wildfires, with a focus on the **Lake Fire** in Los Angeles County from August to September 2020.
- **First objective:** Investigate the suitability of large and multi-sourced **mobility data sets** for representing movement patterns in wildfire-impacted areas.
- **Second objective:** Apply machine learning techniques to identify and trace changes in **mobility time series** of wildfire events.

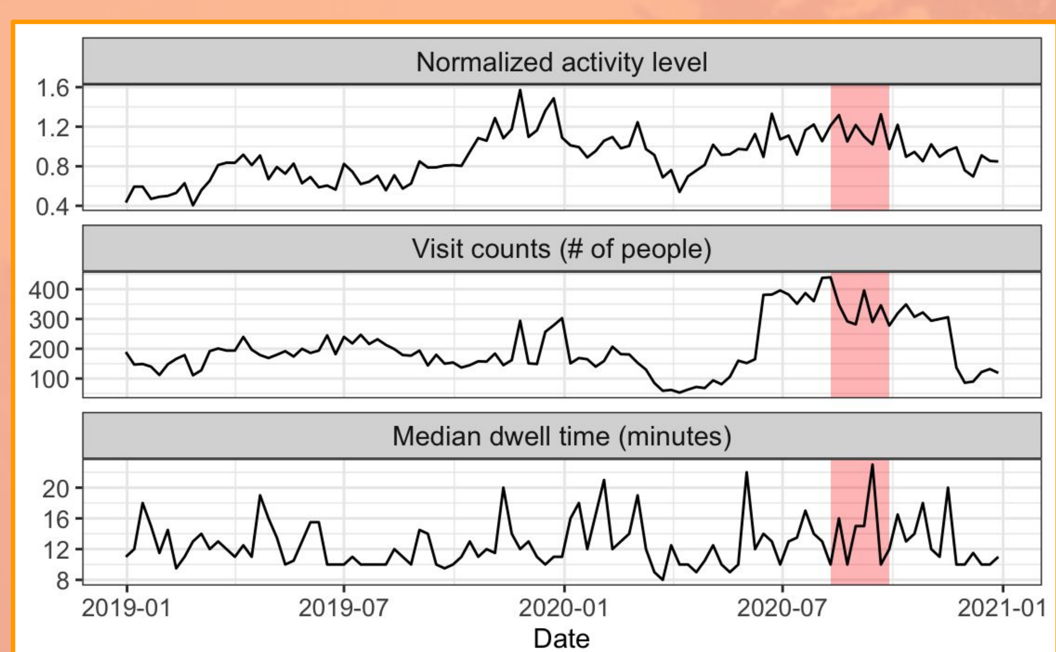

Figure 1: Time series plots of human mobility metrics for a Mobil gas station near Lake Fire (2020), aggregated by week. The red region represents the duration of the fire.

Data

- Utilized several human-mobility datasets sourced from **Mapbox** and **SafeGraph**.
- Mapbox and SafeGraph collect their data from users who have consented to location tracking through smartphone apps at **places of interest (POIs)**.
- Primary features: location (longitude and latitude), date, visit counts, dwell time, and activity levels.

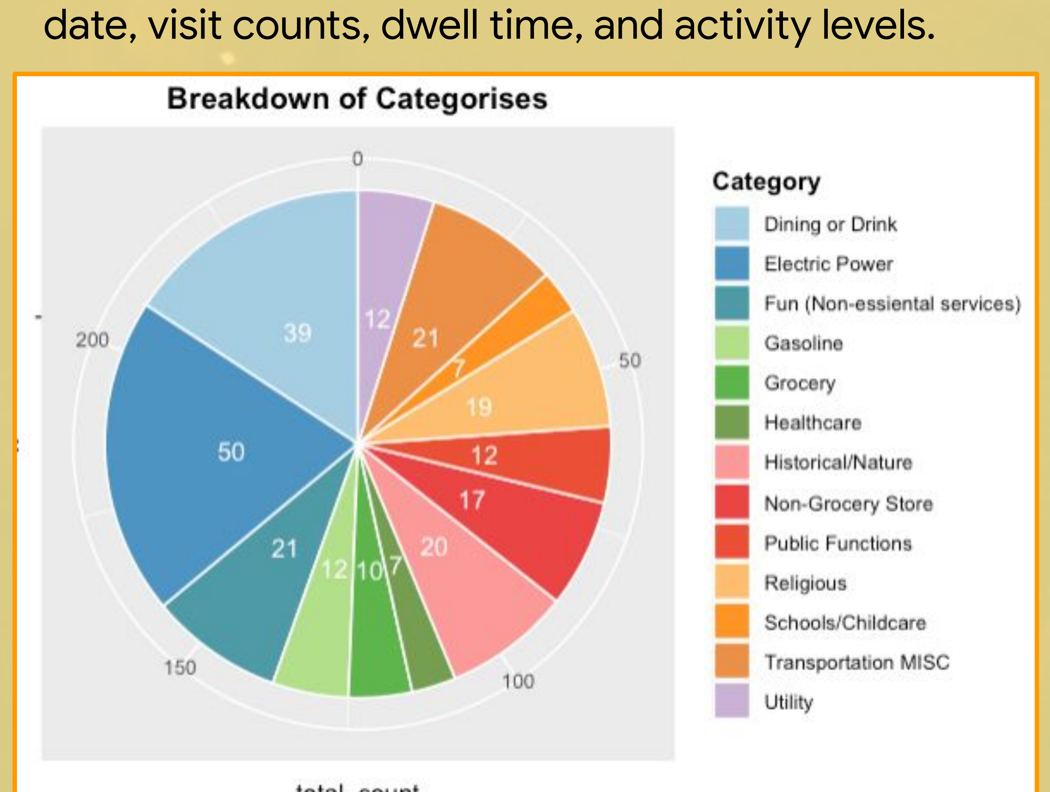

Figure 2: Pie chart breaking down the categories of our POIs.

Clustering Methods

- Explored several clustering methods to identify groups of POIs with similar characteristics.
- Used cluster labels to construct a classification model.
- **ST-DBSCAN** performs the best while also considering the **spatial and temporal aspects** of the data.

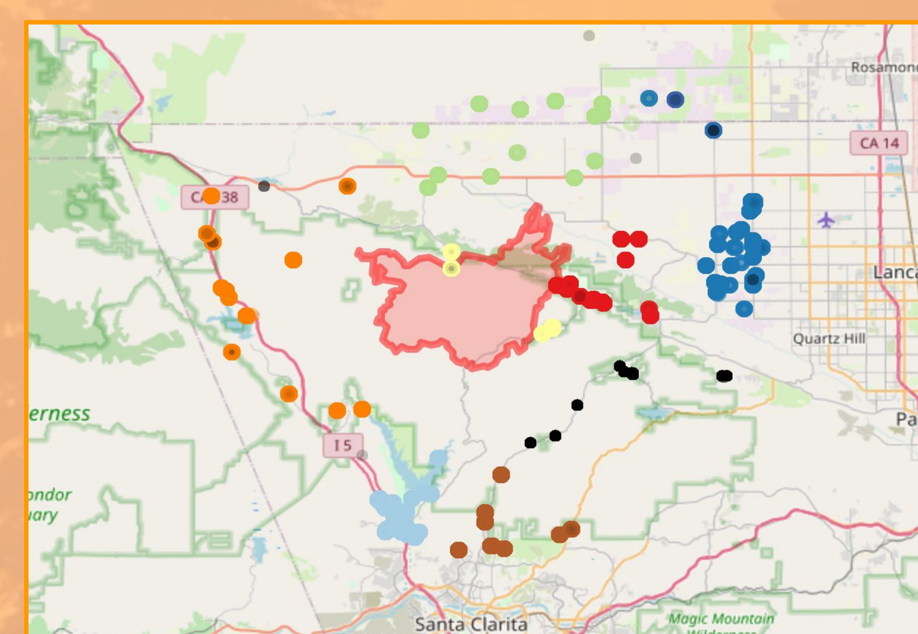

Figure 3: Spatial plot of our ST-DBSCAN clusters. Dots displayed in grey or with grey undertones represent noise.

Labeling Methods

- These methods label POIs due to missing truth labels (movement at POI impacted by wildfire) in our data.
- **Method I:** Clusters POIs based on **distance from fire** and **visit counts**, then applies **Binary Segmentation** to find changepoints. If a changepoint occurs during the fire, all POIs in that cluster are labeled as impacted.
- **Method II:** Analyzes individual POIs for anomalies during the 8-week period of the fire using **visit counts, median dwell time,** and **activity levels**. If the total number of anomalies meet the threshold (≥ 6 anomalies), the POI is labeled as impacted.

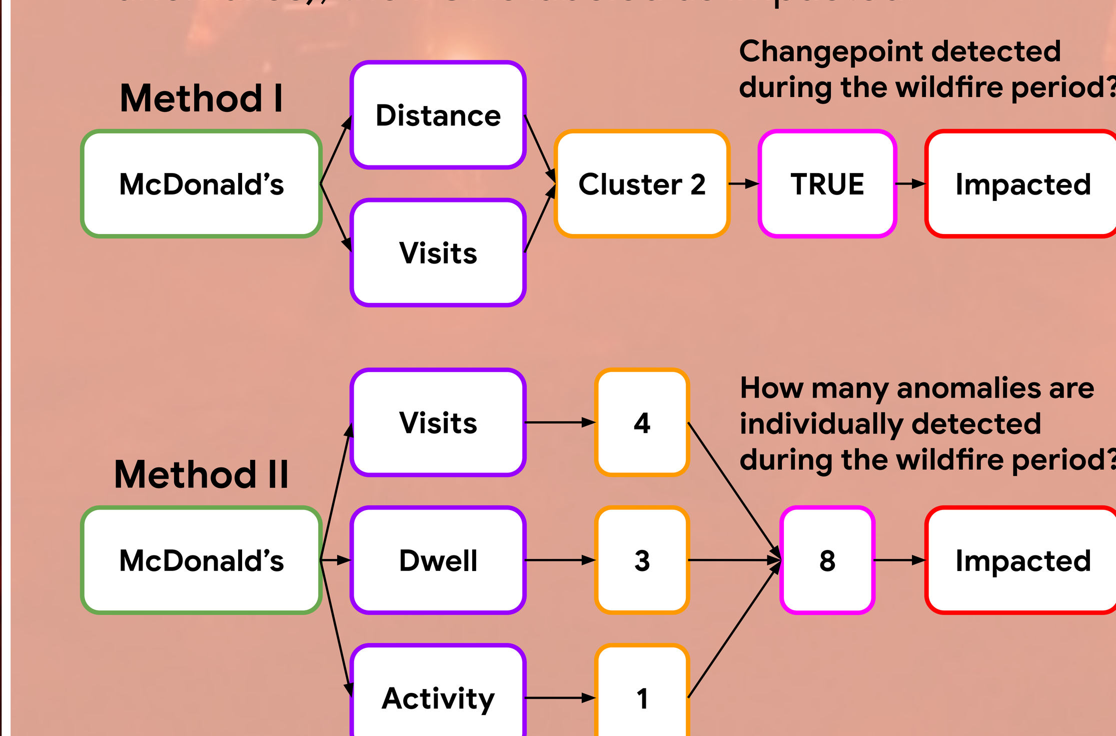

Figure 4: Methods to label POIs as being impacted by a wildfire. Method I (top) detects changepoints within clusters of POIs while Method II (bottom) detects anomalies at individual POIs.

Results

Method I: Changepoint Detection on Clusters

- **Logistic regression** binary classification model to predict whether each POI was impacted or not.
- Predictors: **distance** from fire edge (km), **angle** from fire direction (degrees), and **type of POI** (e.g., dining, grocery, utility).
- Performance: **test accuracy** of **0.75** and **ROC AUC** value of **0.80**.

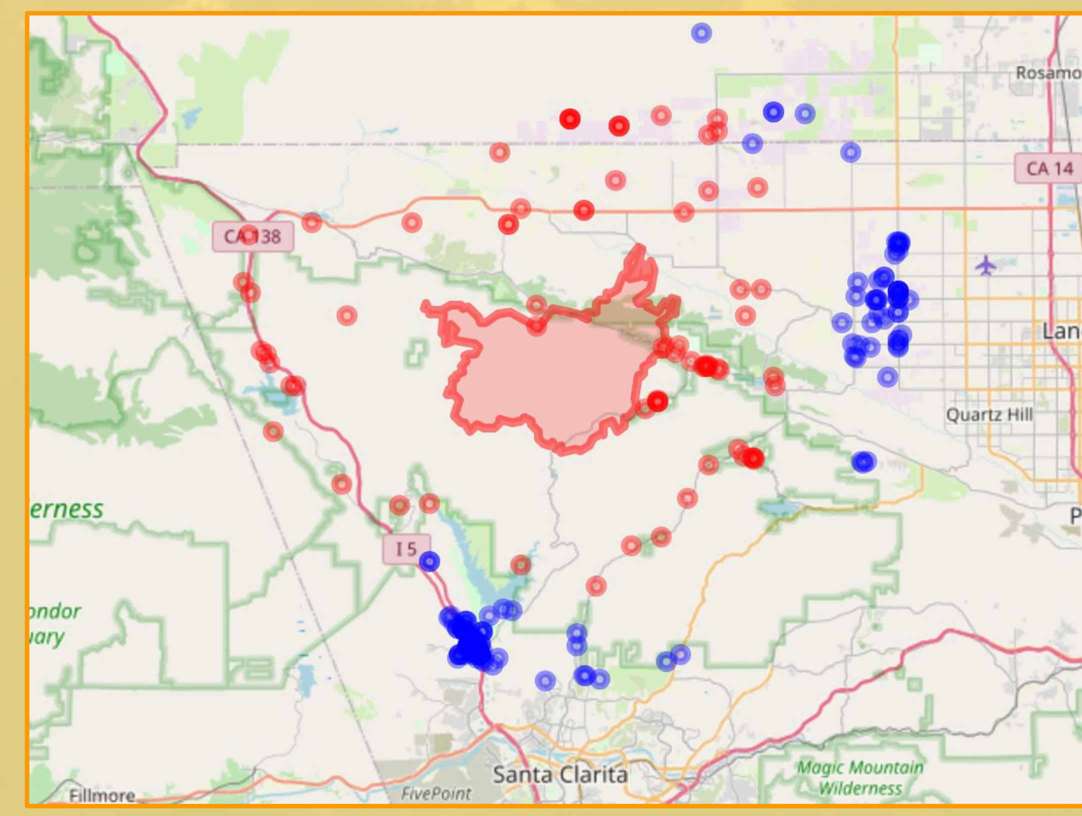

Figure 5: Spatial plot of the logistic regression model's predictions of impact levels using the changepoint detection labels. Impacted POIs (in red) tend to be closer to the fire edge compared to unimpacted POIs (in blue).

Method II: Anomaly Detection on Individual POIs

- Three classification models: logistic regression, random forest, and boosted trees.
- Predictors: binned **distance** from fire center (km), binned **angle** from fire (degrees), and **type of POI**.
- Best model: **random forest**, achieving a **test accuracy** of **0.71** and a **ROC AUC** value of **0.78**.

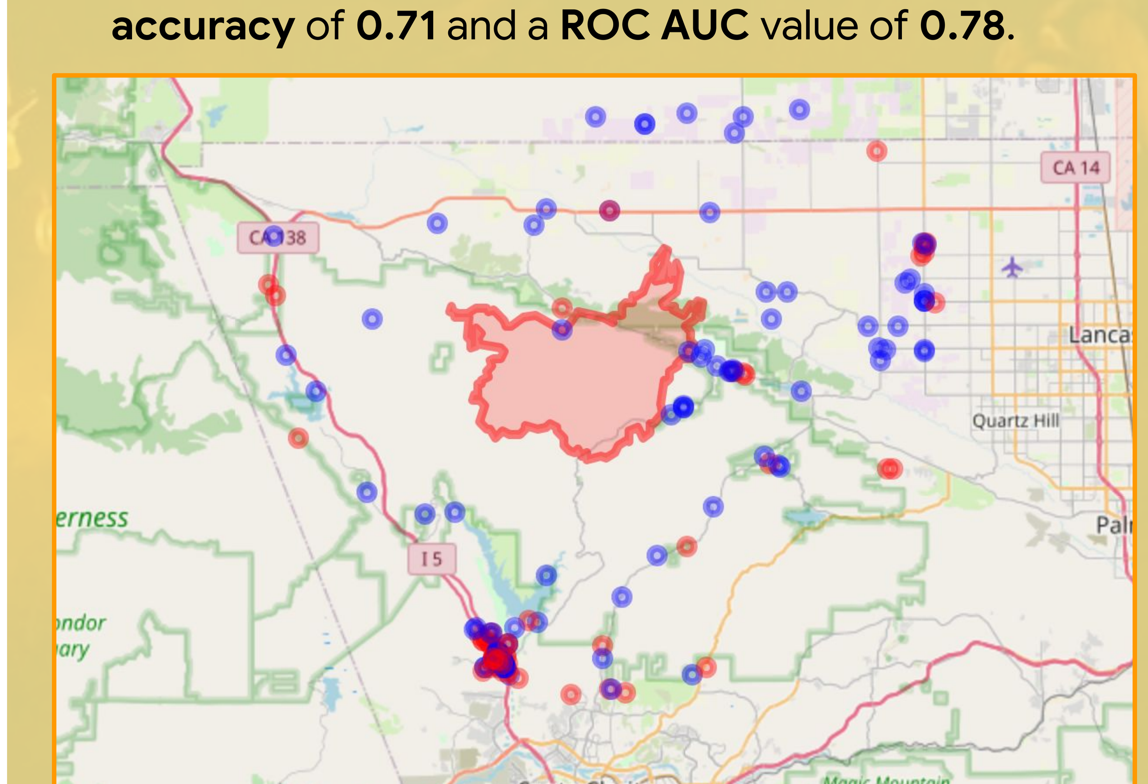

Figure 6: Spatial plot of the random forest model's predictions of impact levels using the anomaly detection labels. Both impacted and unimpacted POIs (in red and blue, respectively) are scattered.

Discussion & Findings

- Implemented ST-DBSCAN clustering to explore spatial-temporal clusters based on similar characteristics of space and mobility behavior.
- Self-labeled our data with anomaly and changepoint detection algorithms.
- Both models performed relatively well but **Method I: Changepoint Detection on Clusters, performed the best** based on accuracy and ROC AUC values.

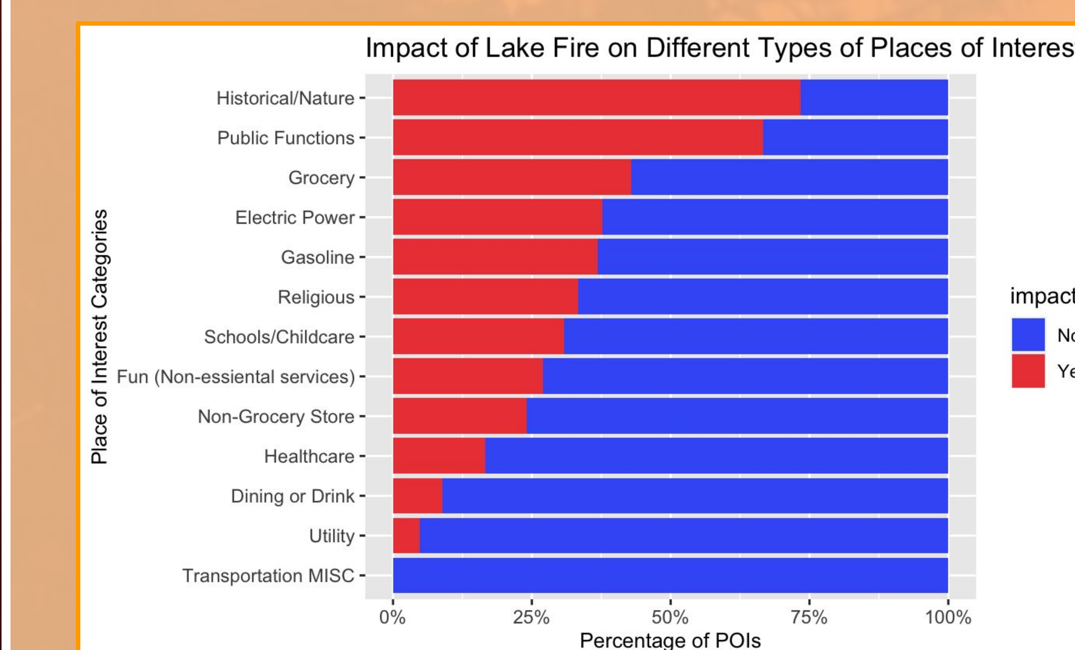

Figure 7: 100% stacked bar chart of types of POIs and their relative proportion of impacted and unimpacted POIs based on the model from Method I.

- Locations closer to the fire's edge and locations in the direction of the fire burn (W) are more heavily impacted by wildfires.
- Categories of locations experiencing a significant impact from wildfires are Historical/Nature, Public Functions, Grocery, Gasoline, Religious, and Childcare.

Future Work

- Evaluate our models on **additional wildfire datasets**.
- Incorporate **geographical features** (e.g., wind direction, elevation, air quality) to enhance predictions of wildfire-impacted areas.
- Account for the **impact of COVID-19** for wildfires that occurred during the pandemic.

References & Acknowledgments

[1] Sardá-Espinosa (2019). Comparing Time-Series Clustering Algorithms in R Using the dtwclust Package. https://cran.r-project.org/web/packages/dtwclust/vignettes/dtwclust.pdf

[2] Coburn (2022). *PSTAT 131: Machine Learning*. University of California, Santa Barbara. Department of Probability and Statistics.

This project is supported by NSF Award BCS #2043202. A special thanks to our sponsors and mentors at the MOVE Lab for their guidance and support.