

A DATA ANALYSIS OF THE CAMP FIRE

Anne Kerr
Michel Elias
Hugo Delgado
Stephen Godfrey

Problem Statement

Data

Text Analysis

Progress Models

Conclusions

Further Analysis

Problem Statement

Use data to **quantify** the Camp Fire to **prototype models** that could be employed during a wildfire response

A DATA ANALYSIS OF THE CAMP FIRE

Anne Kerr
Michel Elias
Hugo Delgado
Stephen Godfrey

Problem Statement

Data

Text Analysis

Progress Models

Conclusions

Further Analysis



Data

- News and Twitter Data for **Natural Language Processing**
- Cal-Fire Incident Data for analyzing **Progression, Magnitude, and Destruction**
- USGS Coordinate Data for **Mapping**
- FEMA Damage Data for Analyzing **Cost**

News

Twitter

Progress

Maps

News

News API

Time Frame:

November 8 to November 30th

Keywords for Search:

'wildfire', '"camp fire"', '"percent contained"', 'evacuate', 'flames', 'paradise', 'cal fire'.

Total Stories:

20,869 unique articles

Twitter

Time Frame:
Nov 8 through Nov 26

Keywords:
campfire

Language:
English

Profile Region:
California
Excluding retweets

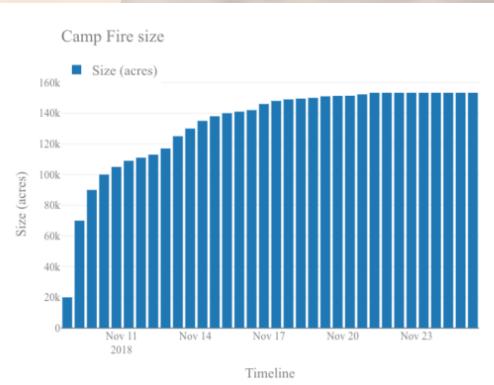
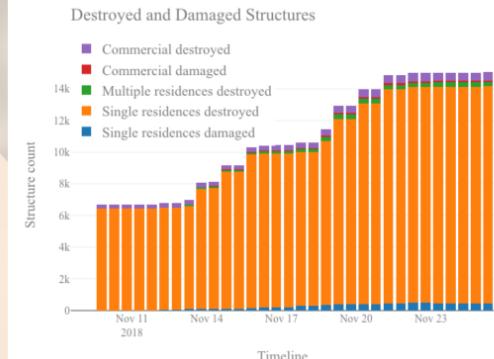
Total Tweets:
15,513 unique tweets



Progression

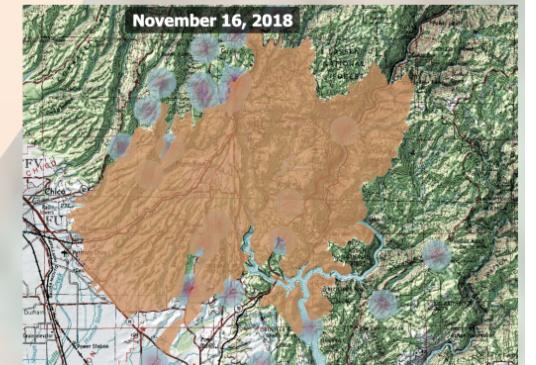
Incident Data

- **Acreage Burned**
- Fatalities
- Structures **Destroyed**
- **Personnel** and **Machinery** Deployed
- Evacuation Orders



Maps

- **Geospatial Multi-Agency Coordination** GeoMac
- **National Interagency Fire Center (NIFC)**
- **Cal Fire**
- **ArcGIS.com**



A DATA ANALYSIS OF THE CAMP FIRE

Anne Kerr
Michel Elias
Hugo Delgado
Stephen Godfrey

Problem Statement

Data

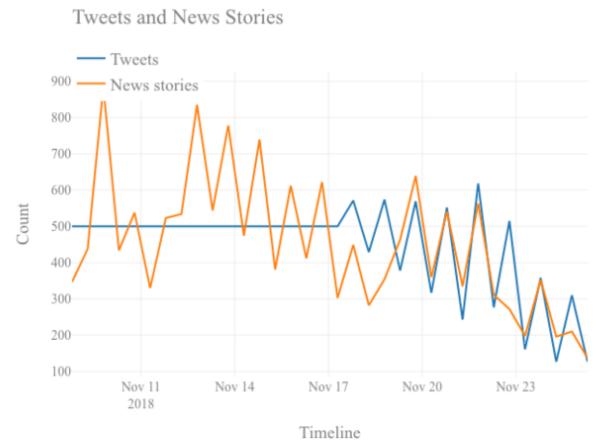
Text Analysis

Progress Models

Conclusions

Further Analysis

Text Analysis



Goal :

Categorize tweets and news stories into similar groups

Approaches:

- LDA
- SVDC
- Doc2vecC

LDA

SVDC

Doc2vecC

Latent Dirichlet Allocation (LDA)

Identify topics in a **collection of documents** by identifying **associated words**

LDA_c_0

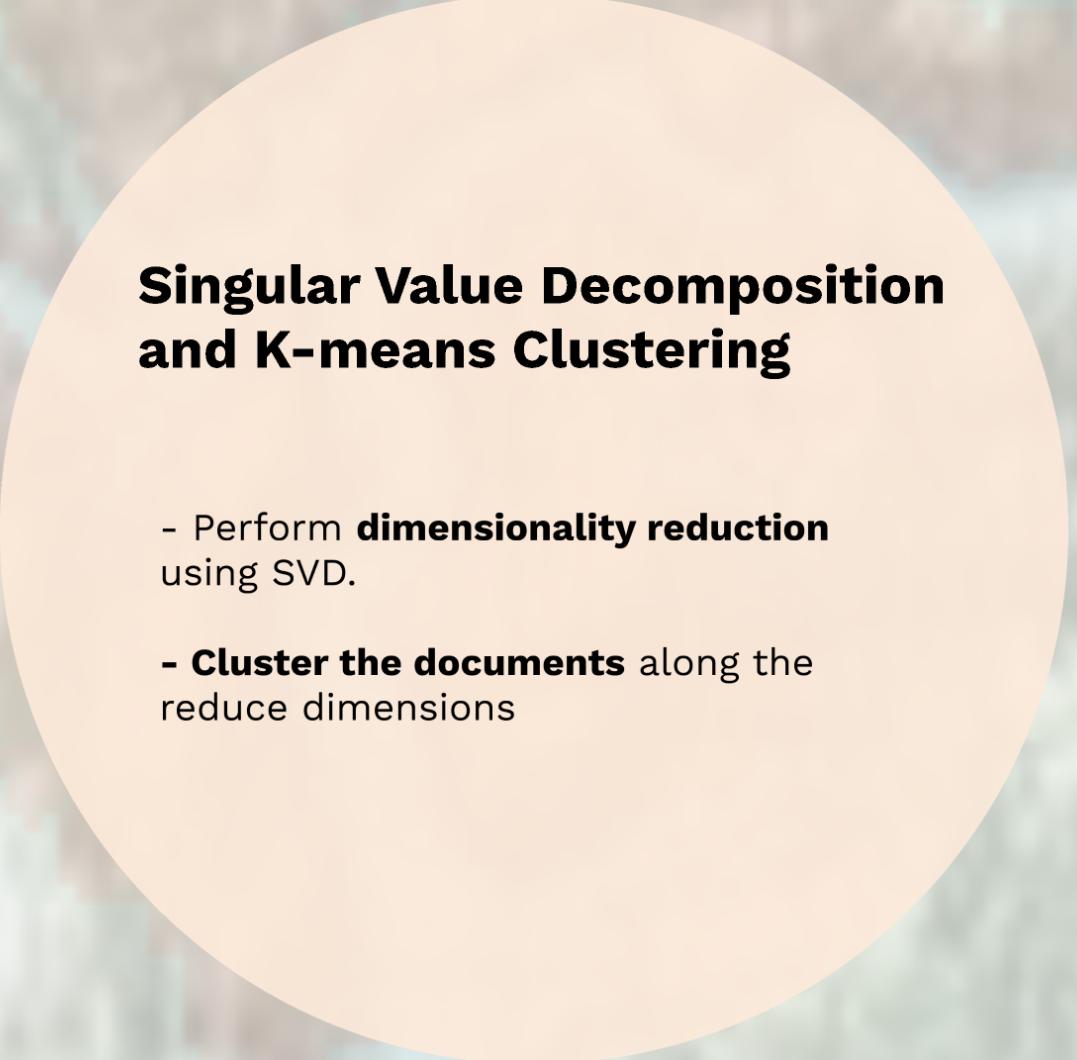
Walmart
California
People
Lot

LDA_c_1

Paradise
Lost
Home
Just

LDA_c_2

California
Smoke
Like
Fires



Singular Value Decomposition and K-means Clustering

- Perform **dimensionality reduction** using SVD.
- **Cluster the documents** along the reduce dimensions

Doc2Vec and K-Means Clustering

- **Vectorize** input documents
- **Cluster** the vectors using **K-Means**
Clustering with **cosine similarity**

A DATA ANALYSIS OF THE CAMP FIRE

Anne Kerr
Michel Elias
Hugo Delgado
Stephen Godfrey

Problem Statement

Data

Text Analysis

Progress Models

Conclusions

Further Analysis

Models

Our **Random Forest Regressor** provides a predicatively valid score indicating that, at least for the camp fire, we can **accurately predict** the size of the fire using components of news, twitter, and government reports

The score of RandomForest Regressor :

0.97

The test score of RandomForest Regressor is:

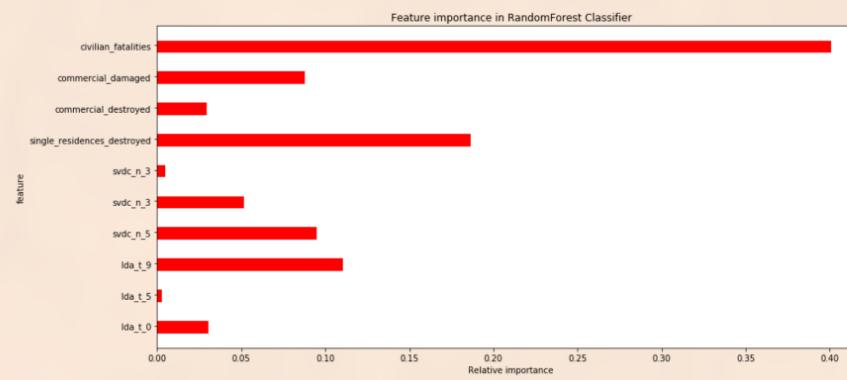
0.97

Features

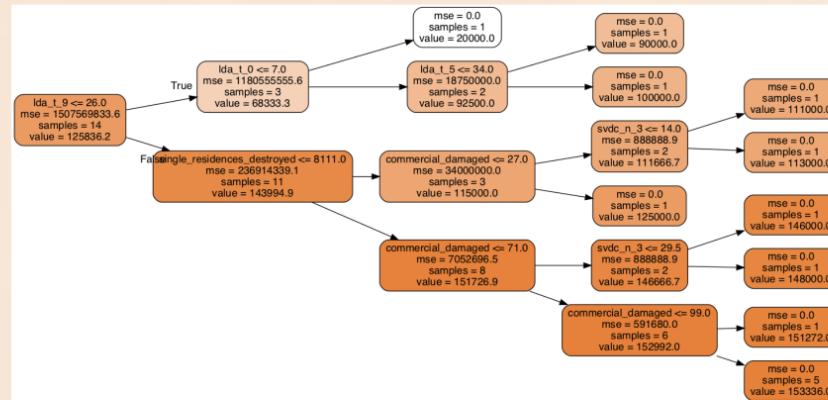
Model

Correlations

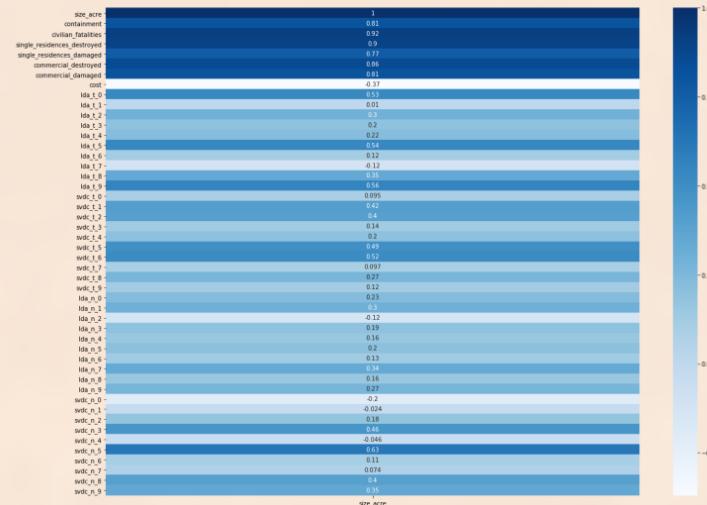
Features



Model



Correlations



A DATA ANALYSIS OF THE CAMP FIRE

Anne Kerr
Michel Elias
Hugo Delgado
Stephen Godfrey

Problem Statement

Data

Text Analysis

Progress Models

Conclusions

Further Analysis

Conclusions

- Combine **progression metrics** with **text data**
- **Categorize** text data into **small groups**
- **Geo-location** of tweets is a challenge
- **NLP** depends on the **text**
- Regression models rely on **several factors**

A DATA ANALYSIS OF THE CAMP FIRE

Anne Kerr
Michel Elias
Hugo Delgado
Stephen Godfrey

Problem Statement

Data

Text Analysis

Progress Models

Conclusions

Further Analysis

Further Analysis

- Build and test **more data**
- **Examine** other fires
- Gather more **granular data**
- Obtain data seen by emergency teams
- Employee a **wider range** of models

A DATA ANALYSIS OF THE CAMP FIRE

Anne Kerr
Michel Elias
Hugo Delgado
Stephen Godfrey

Problem Statement

Data

Text Analysis

Progress Models

Conclusions

Further Analysis