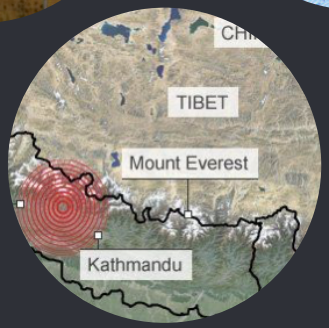
- Can social media predict natural catastrophes?

PROBLEM STATEMENT

Insurers are challenged to obtain real time, meaningful information about natural catastrophes



Effective disaster response helps insurers with:

- Claims Management
- Reinsurance Cover
- Reputation

INVESTIGATION GOALS

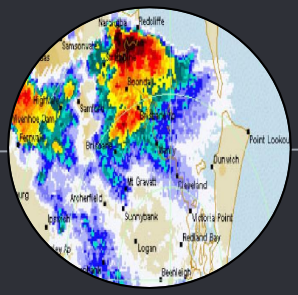
1. Can Twitter detect hail events from the language communicated in tweets?
2. Can Twitter identify hail events of differing severity?

DATA BEHIND THE SCIENCE

Brisbane

27th Nov 2014

$1,400m loss



Sydney

25th April 2015

$400m loss

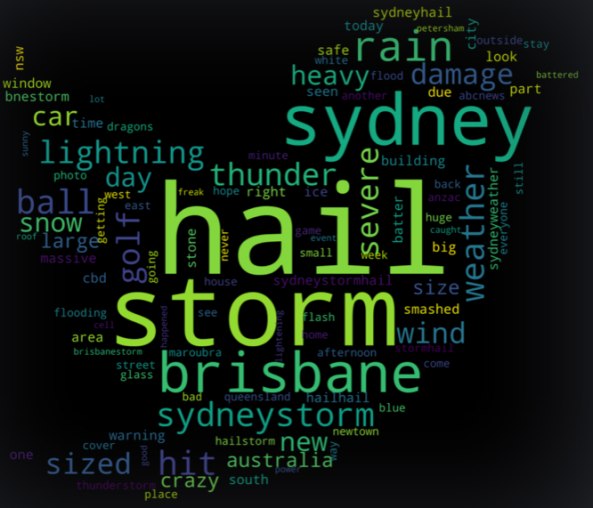
Tweet Text

User Information

Cat_ID	Text	favorites_count	followers_count	friends_count	statuses_count	influence_score	source
154	Love rain and thunder but that hail was too mu...	1966	367	576	4609	0.64	Twitter for iPhone
154	Full out snow storm	1119	410	495	2493	0.83	Twitter for iPhone
154	Thursday Thunder! https://t.co/ySwBPzDy30	728	123	258	2218	0.48	Twitter for Android
154	... I'll take your bad days with your good, Wa...	2541	2332	1624	61556	1.44	Twitter for Windows Phone
154	@The_Little_Sabo Tienes ya el war thunder para...	0	408	1	17147	408.00	Twitter Web Client

METHODOLOGY

DATA



- ‘USER’ DATA ANALYSIS

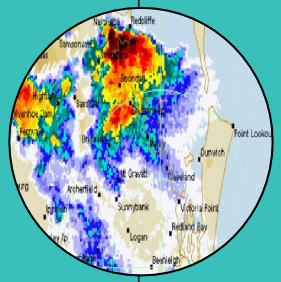
  cleaning to increase confidence

- ‘TWEET’ ANALYSIS

  natural language processing

- MODELLING

  machine learning algorithm investigation



DATA INSIGHTS

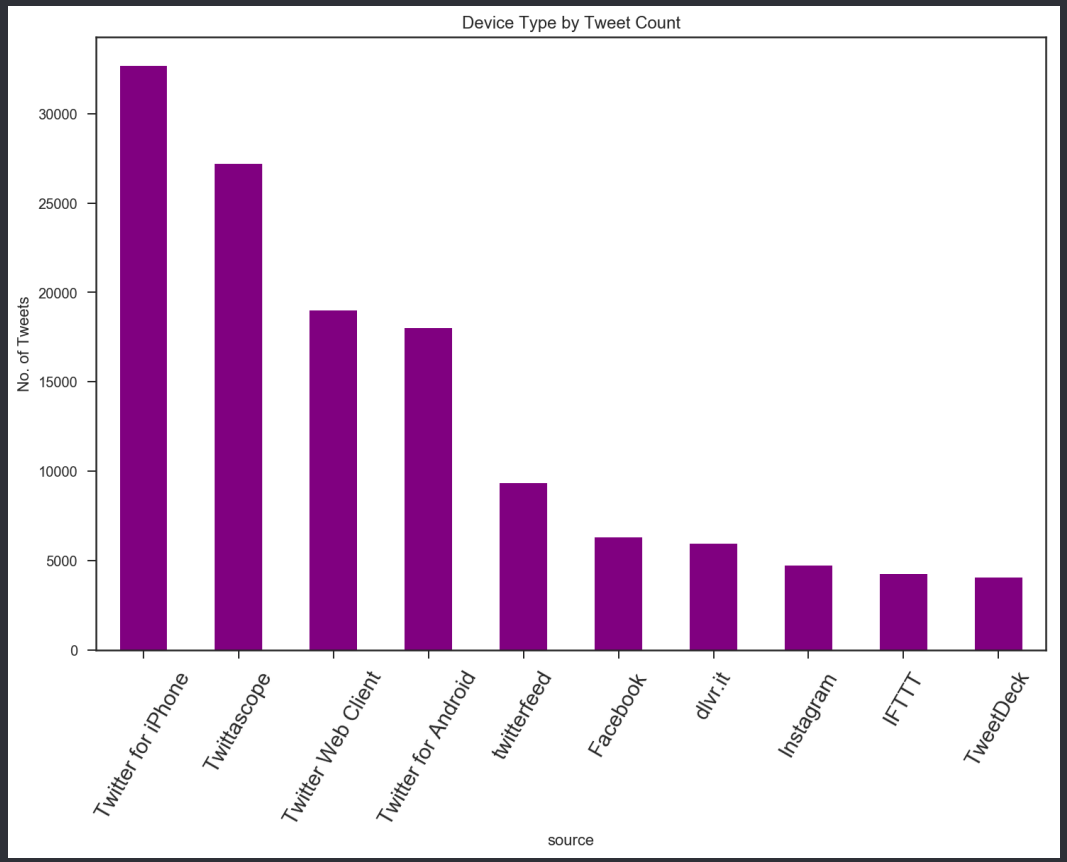
USER DATA ANALYSIS

Twittascope!?

- An automated horoscope posting

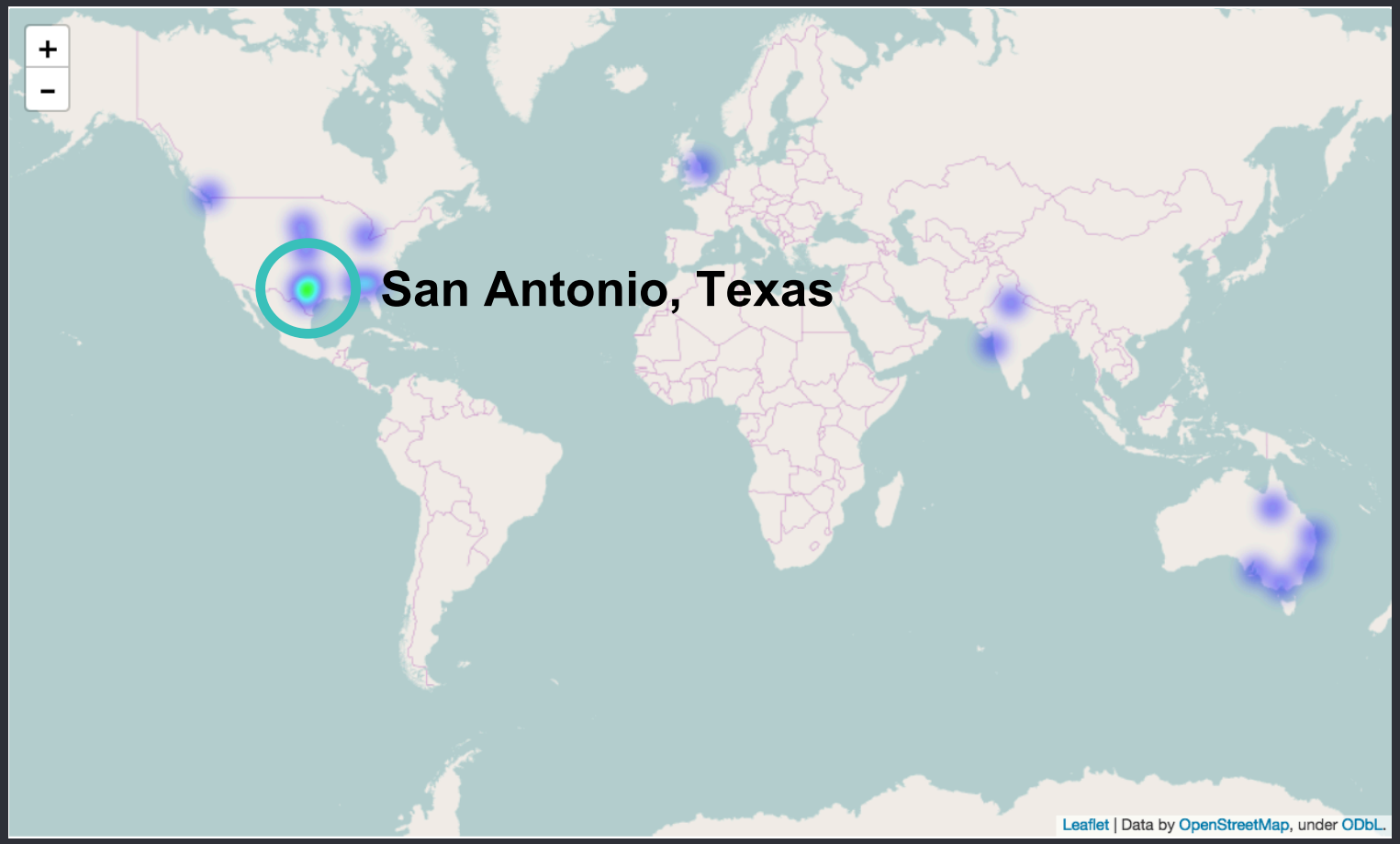
Cleaning the data increases confidence of tweets by:

- excluding anomalous data sources
- removing users with extreme number of friends, followers…
- replacing null values



A THIRD HAIL EVENT?

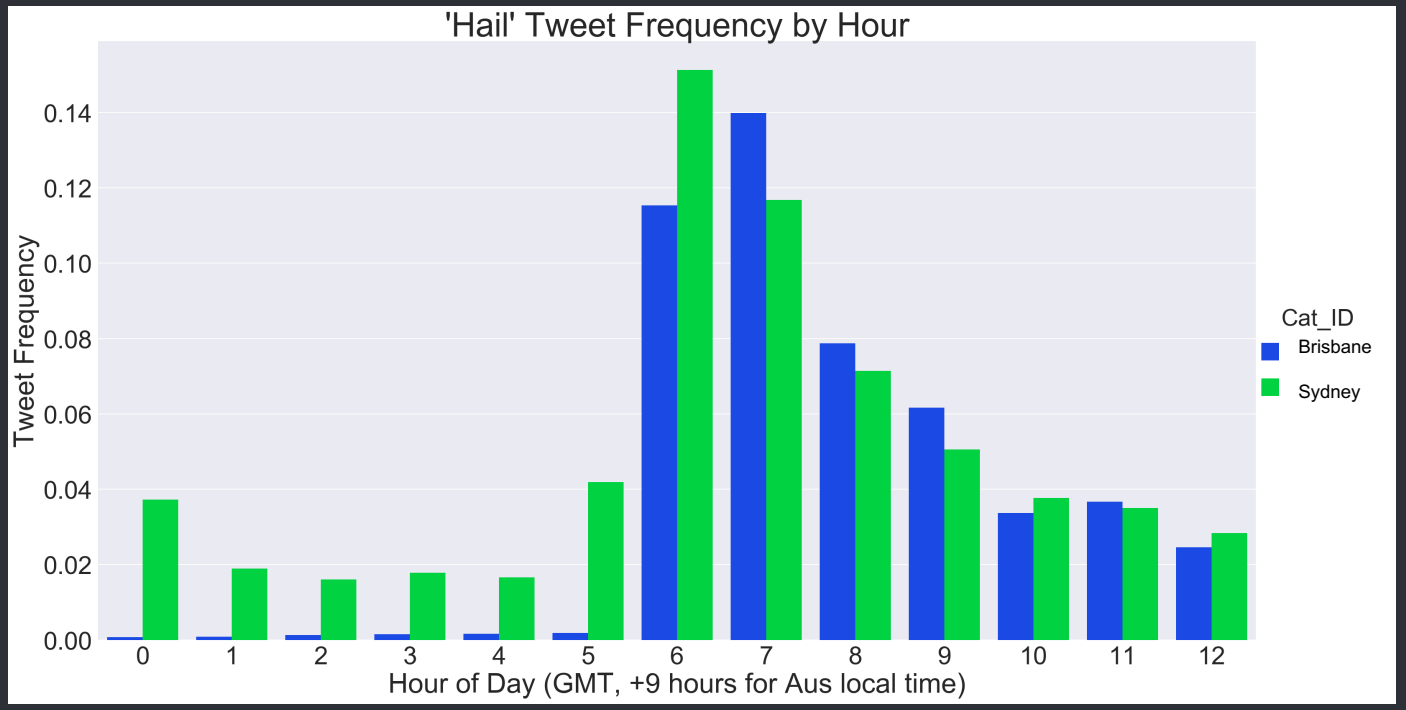
Geographic analysis on 25TH April 2015 identified



TIME ANALYSIS

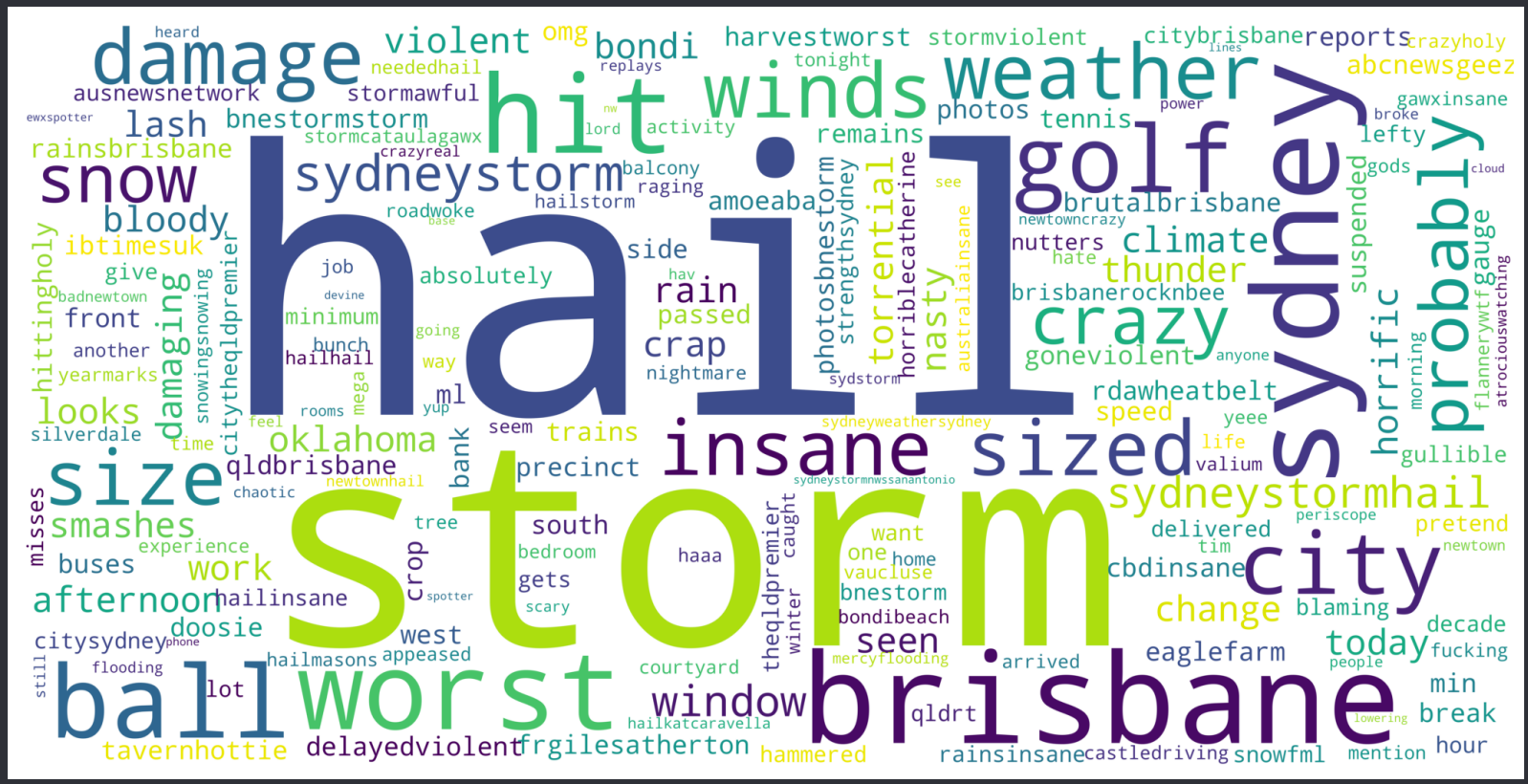
What can we infer from the time data?

- Hail events occurred at similar times of day (late afternoon)
- Frequency of tweets declined at same rate following each event
- Hail warnings for Sydney but not for Brisbane?



SENTIMENT = SEVERITY?

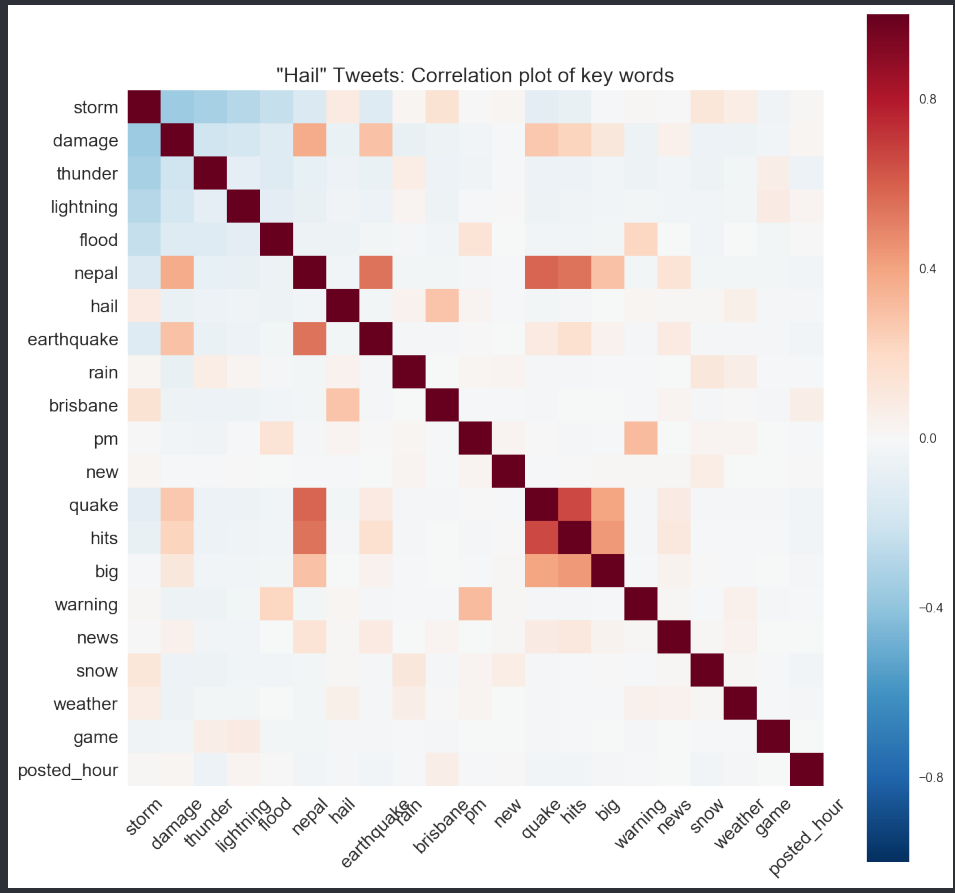
- High frequency of **negative** tweets relative to the total sample
- Strong indicators of damage in the most negative tweets

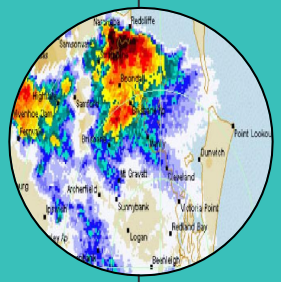


A CATASTROPHIC DISCOVERY

A serendipitous discovery of a second natural catastrophe…

- Nepal Earthquake on 25th April 2015
- $6 billion economic loss
- Earthquake detected during word frequency analysis and topic modelling





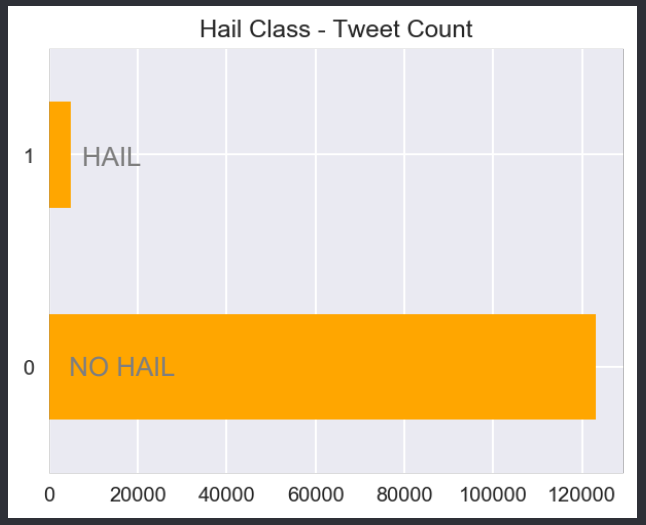
MODELLING RESULTS

HAIL IDENTIFICATION MODELLING

Can twitter detect hail events using language in tweets?

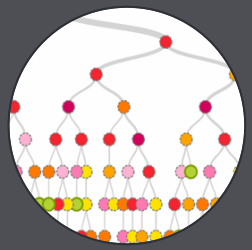
Predictors: Tweet Text

Target: Hail Class



Optimal Model:

- Random Forest Classifier



	Predict Hail	Predict No Hail
Hail	**83%**	17%
No Hail	0%	100%

CATASTROPHE MODELLING

Can twitter classify different types of natural catastrophes?

hail vs. earthquake

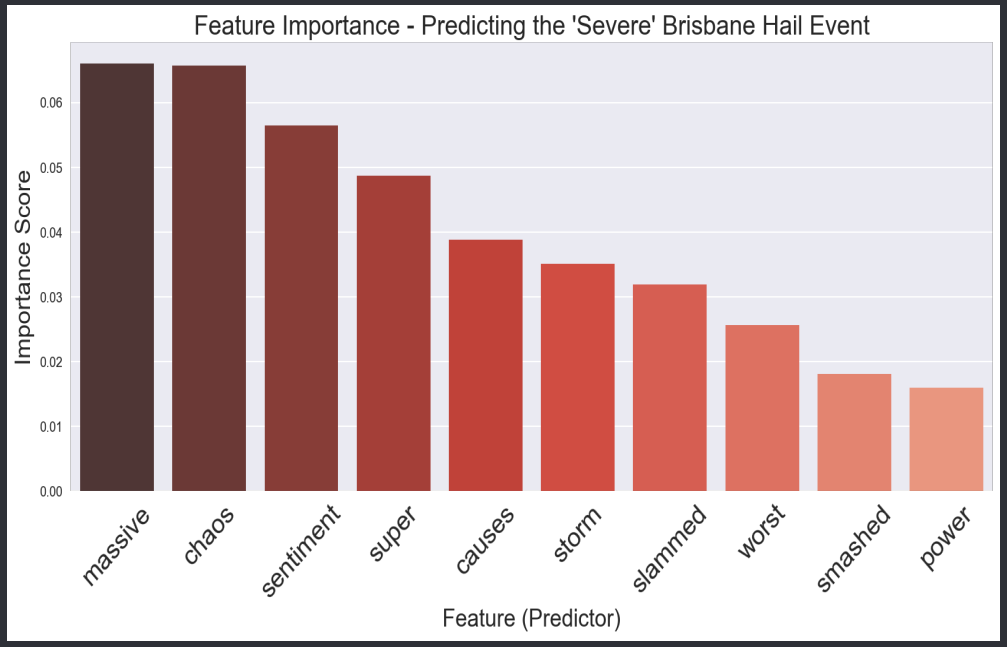
- predicts earthquake class *moderately* well
- combined model reduces hail predictability

	predict earthquake	predict hail	predict no catastrophe
earthquake	**70%**	0%	30%
hail	0%	**29%**	71%
no catastrophe	1%	0%	99%

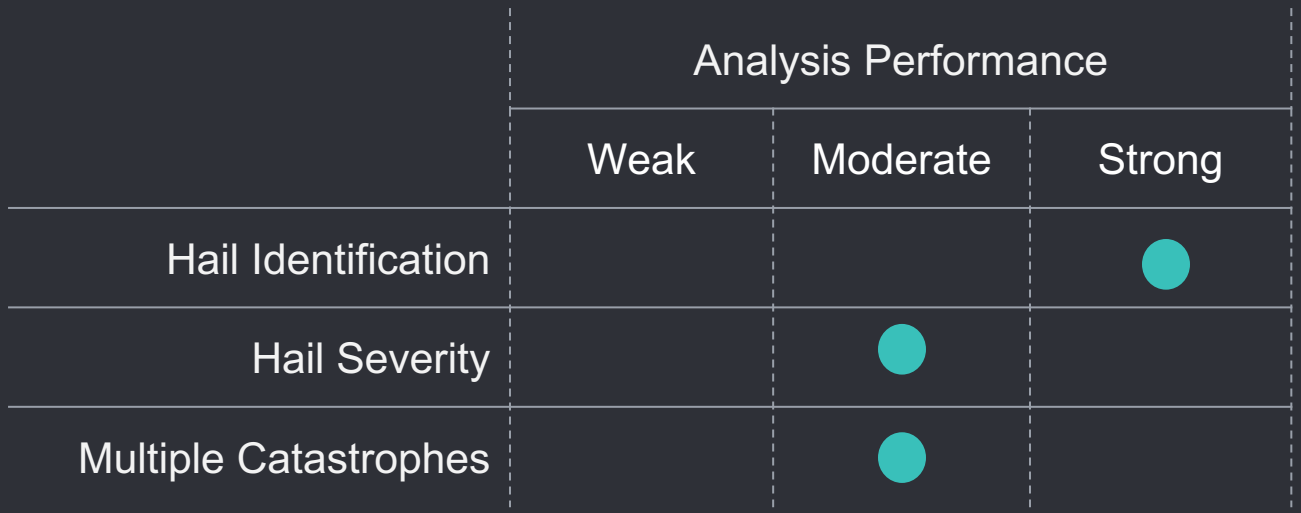
SEVERITY MODELLING

Can twitter identify hail events of differing severity?

	Predict Severe	Predict Moderate
Severe	**63%**	37%
Moderate	8%	92%



RESULTS + RECOMMENDATIONS



Recommendations:

- Train models on further historical and future data
- Develop algorithms independently for each catastrophe type
- Investigate location: tweets and clustering
- Bushfires, floods, cyclones....

MODEL DEPLOYMENT

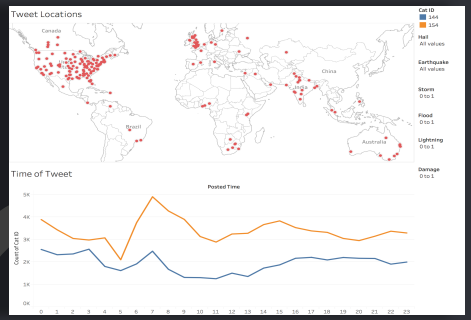
Twitter API

Event search criteria
"hail" or "damage" or "storm" or "flood"

User Data Parsing

Tweet Parsing

Tweet Event Monitoring Dashboard



Vectorise Dataset

Random Forest Classifier

Train: 80%

Test 20%

Class: **Hail**

Severity: **0.57**

size: 0.34
falling: 0.13
1cm: 0.20

Tweet Classification

Severity Classification

Event Severity Score

Thanks!

ANY QUESTIONS?

You can find me at:

@tomcroshaw

tcroshaw.github.io