Journal of Computational Social Science (2022) 5:1427–1465
https://doi.org/10.1007/s42001-022-00174-8

RESEARCH ARTICLE

Sentimental wildfire: a social-physics machine learning model for wildfire nowcasting

Jake Lever[1,2,3] · **Rossella Arcucci**[1,2,3]

Received: 6 January 2022 / Accepted: 30 June 2022 / Published online: 22 July 2022
© The Author(s) 2022

Abstract

The intensity of wildfires and wildfire season length is increasing due to climate change, causing a greater threat to the local population. Much of this population are increasingly adopting social media, and sites like Twitter are increasingly being used as a real-time human-sensor network during natural disasters; detecting, tracking and documenting events. The human-sensor concept is currently largely omitted by wildfire models, representing a potential loss of information. By including Twitter data as a source in our models, we aim to help disaster managers make more informed, socially driven decisions, by detecting and monitoring online social media sentiment over the course of a wildfire event. This paper implements machine learning in a wildfire prediction model, using social media and geophysical data sources with Sentiment Analysis to predict wildfire characteristics with high accuracy. We also use wildfire-specific attributes to predict online social dynamics, as this has been shown to be indicative of localised disaster severity. This may be useful for disaster management teams in identifying areas of immediate danger. We combine geophysical satellite data from the Global Fire Atlas with social data provided by Twitter. We perform data collection and subsequent analysis & visualisation, and compare regional differences in online social sentiment expression. Following this, we compare and contrast different machine learning models for predicting wildfire attributes. We demonstrate social media is a predictor of wildfire activity, and present models which accurately model wildfire attributes. This work develops the concept of the human sensor in the context of wildfires, using users' Tweets as noisy subjective sentimental accounts of current localised conditions. This work contributes to the development of more socially conscious wildfire models, by incorporating social media data into wildfire prediction and modelling.

Keywords Machine learning · Social media · Wildfires · Sentiment analysis · Twitter data · Satellite data

✉ Rossella Arcucci
r.arcucci@imperial.ac.uk

Extended author information available on the last page of the article

Abbreviations

MAE	Mean absolute error
SVR	Support vector regressor
RF	Random forest
GBRF	Gradient boosted random forest
MLP	Multi-layer perceptron regressor
NN	neural networks
ML	machine learning
RMSE	root mean square error
SA	Sentiment analysis
US	North American (United States and Canada)
AUS	Australian
GFA	Global fire atlas
DOY	Day of year

Introduction and motivation

In the advent of climate change, wildfire intensity and season length is increasing globally [1]. To control and mitigate the negative effects of wildfires, computational models exist which attempt to understand and predict the physical characteristics and evolution of these events [2–4]. To improve the accuracy of these models and make them more event oriented, we can include alternative, more diverse data sources [5, 6]. These models are increasing in sophistication but are still not fully effective for predictions in wildfire nowcasting systems [7–9] due to limited accuracy and geographic domain application.

Over the past years, the information era has led to an explosion of online activity. A large part of that activity is published on social media sites [10], representing a large amount of publicly available, unprocessed social data that is both opinionated and emotional.

Increasingly, studies are using social media as a data source [11] for investigating, modelling and mitigating natural disasters [12]. Due to its widespread use during events, this information has been developed for disaster management applications from a humanitarian perspective, which is now being employed by aid agencies [11, 13, 14]. In the context of wildfires, conventional sensing methods typically rely on more 'physical' means of sensing - i.e. remote or satellite sensing [15]. However, this is not the full picture. As with any natural or social disaster or event, people effected or witnessing these events post a subjective emotional account of this on social media. This account contains information about the users' experience, and represents an untapped data source in wildfire modelling. This information cannot be inferred using conventional sensing methods, and as a result using this data source adds previously unknown information to models. As with traditional sensing methods such as cameras [7] or satellites [16], this data can be used for early warning systems due to active monitoring capturing the real-time reactions to this event. This data is also easier and quicker to collect and process than data sources such as

satellites, and can therefore be used to complement or substitute these types of data sources when traditional sensing methods yield inaccurate or sparse data.

There has yet been comparatively little work undertaken into the social impacts and influences of wildfires. That is, understanding how people perceive different wildfire events, whether this perception is related to wildfire activity in local areas, and incorporating these into real-time wildfire models. This represents a shortcoming in current models [15]. This paper aims to understand how people perceive wildfires, and whether this can be measured and modelled to make predictions. The goal is combining social and physical data to develop nowcasting wildfire alert systems for modelling wildfire activity. In achieving this, we develop Sentimental Wildfires, a machine learning model trained on both social and physics wildfires data. We utilise Twitter as a data source for wildfire modelling, collecting large numbers of tweets associated with different wildfire events. Twitter's wide usage, and 280 character message length, make it an ideal platform for tracking disaster situations as they unfold [17]. Tweets represent noisy human sensor data, with an underlying flow of information surrounding an event that we aim to extract. By combining this data with historical satellite data, it is possible to create domain-specific Sentiment Analysis (SA) [18] of these wildfire events. We build up a socio-physical dataset of wildfire events by using satellite data from the Global Fire Atlas (GFA) 2016 Ignitions [19], and by collecting and sentimentally analysing social media data on these events. Following this, we investigate links between social wildfire data and physical wildfire characteristics. Our investigation shows that social sentiment can both be predicted by wildfire attributes, but also act as a predictor for these characteristics. This contributes to bridging the gap of understanding in the social impacts of wildfires, and uncovers relationships between social wildfire sentiment and physical fire attributes. The monitoring of the online social sentiment can be used to detect ignition geolocations. Given the presence of a wildfire in a specific area, we can use satellite data to model social sentiment over the course of the event. Using this satellite data to model social dynamics can be useful for capturing the effect wildfire activity has in the online space, and has been shown to be useful for quick disaster response [5].

The remainder of this paper is organised as follows. "Related works and contribution to present" examines current social media models in alternative applications and addresses the need for a social media based wildfire modelling system. "Sentimental wildfire: overview" outlines the system proposed by this paper for making wildfire predictions, and the contribution this system represents. The methodology of the system is then expanded on in "Sentimental wildfires: data collection", "Sentimental wildfires: social sentiment analysis" and "Sentimental wildfire: predictions". "Sentimental wildfires: data collection" describes the data collection and preprocessing methodology employed by the system proposed in this paper, which has been developed over multiple iterations. "Sentimental wildfires: social sentiment analysis" describes the use of sentimental analysis in the system, to convert qualatative textual data into quantitative numerical metrics, generating social sentiment variables for individual wildfire instances. We then introduce our prediction objectives for the system in "Sentimental wildfire: predictions", and describe the machine learning (ML) models used in the testing of the system. "Test cases and results"

outlines the test cases and results of our system. The system is tested in two separate domains; North America and Australia. We first compare and visualise the datasets, and then apply our system to this data using ML methods outlined in "Sentimental wildfire: predictions" to achieve our predictive goals defined here. Following testing and evaluation of the ML models and datasets, we present the final Sentimental Wildfire models which can be used to make wildfire instance predictions. The work concludes in Sect. 7 by summarising our results and pointing towards possible improvements to Sentimental Wildfires.

Related works and contribution to present work

Due to the abundance of social media data available today, there is increasing interest in the research community around leveraging this information to make predictions or understandings about social and physical events, both in real-time and retrospectively.

The human sensor

The real-time publishing of information on these platforms means that in disaster situations users act as human sensors, detecting and documenting events. Social Media has already been implemented to detect and predict aspects of influenza spread [13], social unrest [14, 20], polling and election election outcomes [21], mental illness [22] and addiction [23], as well as general natural disasters [11, 24] and finally wildfires [8, 25]. Studies related to disaster scenarios have revealed interesting results, building on the concept of the human sensor to inform decision making and recovery in these situations. Indeed, social media is playing an increasingly important role in disaster relief, helping fire and rescue services make bottom-up decisions as an event unfolds [26].

The study in [27] offers a comprehensive review on how social media can be used in disaster scenarios to extract both spatial and temporal information surrounding natural disasters. This highlights the potential uses of social media in facilitating disaster management and provides a motivation for implementing the system proposed by this paper. It has also been shown that during natural disasters Twitter is a good social media platform to use for collecting and distributing information. The study in [28] showed during the Tohoku earthquake in 2011 that 63.9% of people who used Twitter during this disaster that the platform was useful in providing information. This paper additionally emphasises the importance of public engagement for systems like this to be successful, and that these rates of engagement are rising year on year, representing a wider human sensor network available to researchers.

Two systems which have been recently developed in this area are the Artificial Intelligence for Disaster Response (AIDR) [29] and Europe Media Monitor (EMM) [30] systems, developed by the Qatar Computing Research Institute and the European Commission's Joint Research Centre respectively. Both these systems also utilise the human sensor, with the former offering advanced techniques for classifying

disaster-related tweets with up to 80% accuracy. The latter system monitors thousands of European news sources in over 70 languages, monitoring topics and news themes over the course of their development. These two systems are at the cutting edge of social media monitoring, the system proposed here aims to build on and combine these two methodologies to implement a wildfire monitoring & information collection system.

Modelling natural disasters using social media

This work combines social and physical data in a predictive model of wildfire activity based on Twitter data. A similar approach has been implemented in the context of flu spread [13]. This study predicts the number of hospital admissions based on the analysis of twitter data. Here, we apply this concept to the context of wildfires, and build on this methodology by attempting to model a number of wildfire characteristics. Additionally, wildfire modelling systems implemented previously have a number of restrictions which limit their application. The study in [8], like many retrospective wildfire analyses, yields findings that are highly geographically restricted. The study in [7, 9], uses cameras for an alert system for wildfires and requires large levels of expensive infrastructure. The system proposed by this paper avoids these geographic and economic restrictions by utilising the concept of the human sensor to use social media users for wildfire nowcasting. We use domain-specific language to isolate tweets associated with wildfire events, as in [11] and [15], using wildfire-specific keywords in our search queries. The wildfire keywords used are adapted from [15]. Additionally, there is evidence that during natural disasters, online sentiment in different geographic areas changes with the proximity to the disaster [17], and its perceived severity in that area [26]. To ensure this local sentimental reaction is captured, we generate queries to enforce a 'locality' on discussions in tweets analysed, by also using location keywords in search queries [18].

Sentiment analysis

Sentiment analysis (SA) has been used in the past to identify dynamic polarity of sentiments over the course of a disaster [26]. SA has been used in the past to gauge polarity in online reviews [31], and social media posts [32]. SA is also likely able to help evaluate levels of destruction in local areas, identify people or communities displaced or in need of rescue efforts, and improve overall disaster management and mitigation [33]. In the context of wildfires, this is of high importance, as there are often a multitude of time-sensitive factors involved, including; placement of fire fighters and rescue teams, alerts of blocked roads, power and telephone lines being cut, and the loss of infrastructure and life. This investigation uses SA on collected Twitter data in order to generate social sentiment variables and vectors for wildfires. This analysis of temporal data allows us to plot online sentiment over the course of a wildfire, introducing the notion of a sentimental arc of an event similar to those formulated in [17] and [8]. Previous studies have also used SA as a social analysis tool [11, 14] for natural disasters. These studies have shown that regional social

sentiment is correlated to perceived disaster severity in that area [11], and that social SA is a useful information extraction method in the context of social events.

To train our nowcasting models, we employ a ML model based on RF technology; namely the Gradient Boosted Random forest (GBRF) [34, 35] which has been shown to be effective even when the amount of the feature types is not very large [36]. Additionally, we compare our model with other different ML models: neural networks (NN) [37] and Support Vector Regressor (SVR) [38], and evaluate their potential suitability for this problem.

Sentimental wildfire: overview

An overview of the training of the system is shown in Fig. 1. The system consists of three parts: (A) data collection, (B) social/Sentimental analysis and (C) ML predictive models. These parts are better detailed in Sects. 3, 4 and 5 respectively.

The data collection part of the investigation is represented by the top row of the diagram, with the lower section representing two types of models being trained, corresponding to the two prediction objectives which are outlined in the later Sect. 5 of this paper. These models will intake the newly collected socio-physical data. Following the running of this system the models will be trained on historical wildfire data and instances. We train the models on historical data and we test the models on unseen data. As we know this doesn't affect the generality of our study, we set up two test cases which covers wildfire activity for the whole of

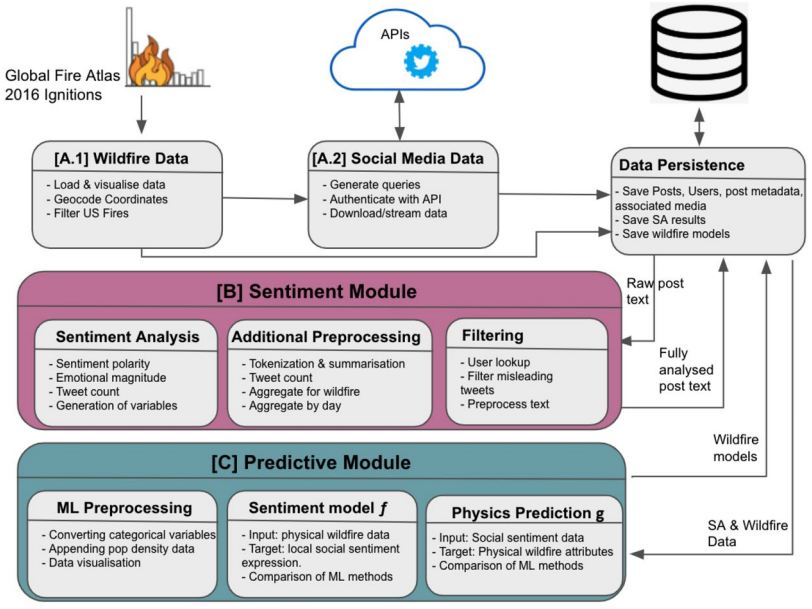


Fig. 1 Sentimental Wildfires: System Diagram for the **A** data collection, **B** social/sentimental analysis and **C** ML predictive models

North American (United States and Canada) (US) and Australian (AUS) regions in 2016. North America is often used as a case study in investigations involving Twitter data [15], due to its high usage of social media. Additionally, high levels of wildfire activity in the region ensure sufficient physical and social wildfire data to achieve the model goal of integration of these sources. Furthermore, Australia is also an area of frequent wildfire activity and is also English speaking, however, has less Twitter users [39]. This area is used as a second domain to compare and analyse generalisability of social wildfire models.

There are a number of aspects of this work that are novel and/or transferable to other domains:

(A) Firstly, we introduce a fast way to create a dataset for historical wildfire events and their associated sentimentally analysed tweets. As has been shown, twitter data is noisy [40], so our system creates a filtering system to ensure the data collected is relevant and provides an accurately trained system. The system is simple and easy to run and available at https://github.com/jakelever4/FSA. This is the first time that social media data has been collected for this dataset of historical wildfires. The resulting database is a novel collection of wildfires and their associated Tweets and media. The query and data collection methods used to isolate tweets are expansions on a number of methodologies [11, 14, 21, 26], optimised for this context.

(B) We introduce new datasets of sentimentally analysed wildfire events. Performing SA on Tweets associated with wildfires allows us to plot sentimental arcs for these events, as has been done with previous case studies [8]. In this work, we implement SA via ML approaches using Natural Language Processing [18, 41, 42]. This approach applies ML algorithms to the linguistic features of text data [43–45]. The result of creating social sentiment variables for these wildfires are new datasets of socially analysed fires, documenting local online social sentiment over the burn period of many wildfires. Databases of wildfire events including these generated values has not been previously created in this way. We show in Sect. 6.1 that these datasets capture wildfire regional seasons as well as individual events, and thus can be considered high-quality datasets.

(C) We train models which accurately predict attributes of the novel dataset. Accurate modelling social media sentiment during disasters may help disaster management and recovery crews make more informed, quick, data-driven decisions [11, 17, 24, 26, 33]. This is a key aspect of wildfire modelling which needs improvement, and could help fire suppression and reduce public danger. This is the first time these ML methods have been trained for the purpose of predicting these wildfire attributes from social media data. We compare differing ML for this task (GBRF [34, 36], SVR [38], NN [37]), and judge their compatibility with this sort of social ML problem, showing that GBRF are the best ML method for these models. This represents a previously unexplored avenue for wildfire prediction. Models resulting from this may be incorporated into existing wildfire frameworks [7, 15, 46], strengthening these models and adding a socially conscious dimension.

Sentimental wildfires: data collection

This section outlines the collection and preprocessing of the historical social media and wildfire data that the system is trained on, corresponding to part A of Fig. 1. For this study, data was collected for the year 2016. Both social and physical aspects of wildfire data need to be incorporated into the model during development. The two parts of this section correspond to the two stages of the development of the socio-physical dataset; fire data collection (represented by [A.1] in Fig. 1) and social media data collection ([A.2] in Fig. 1).

[A.1]: Physics wildfire data collection

The methodology of the data collection part of this investigation generates social sentiment variables for individual wildfire instances, based on supplied historical data. The system uses satellite data from the GFA 2016 ignitions data set, this data is publicly available and can be downloaded at https://www.globalfiredata.org/fireatlas.html. The physical wildfire characteristics taken from this data are: Latitude (∘), Longitude (∘), Size (km^2), Perimeter (Per) (km), Duration (d) (days), Speed (km/day), Expansion (Exp) (km^2/day), and Start (S_{DOY}) and End (E_{DOY}) DOY. The physics vector $\mathbf{x}$ is then expressed as:

$$\mathbf{x} = [\mathrm{Lat}, \mathrm{Lon}, \mathrm{Size}, \mathrm{Per}, d, \mathrm{Speed}, \mathrm{Exp}, S_{\mathrm{DOY}}, E_{\mathrm{DOY}}, \mathrm{Pop}_{\mathrm{Density}}] \quad (1)$$

where $Pop_{Density}$ denotes the population density which value is collected by UN WPP-Adjusted Population Density data [47].

These attributes represent the physical aspect of the system. A subset of GFA ignitions are considered, based on the input domain. The input domains are a bounding box of Longitudinal and Latitudinal coordinates which is the geographic scope of the system. This subset is achieved by geocoding fire coordinates using [48], to generate a string location from longitudinal and latitudinal coordinates (x, y), and then filtering fires located outside the input domain. Geocoding wildfire coordinates generated a corresponding string location with a granularity of 'administrative area level 2' [49]. This GFA subset is reflective of the input domain, and the geocoding data is appended and saved for use in "Social media wildfire data collection" to complete the instance in the GFA subset which will then become our dataset, and build social sentiment variables for each wildfire event in "Social media wildfire data collection".

[A.2]: Social media wildfire data collection

Having collected and filtered the physical wildfire data, social media data is collected and sentimentally analysed, to create social sentiment variables for the wildfire events. Algorithm 1 outlines the data collection algorithm implemented in this project for the historical 2016 wildfire data.

The algorithm requires as input the start and end DOY (namely S_{DOY} and E_{DOY}) and the latitude and longitude coordinates $\{(x_i, y_i\}$, where i denotes the row/instance

number of the wildfire event in the dataset, outlined in Sect. 3.1. Twitter data collection runs on the 2016 physical fire data, using dynamic query generation (in lines 2–3 of Algorithm 1), to search for tweets posted during the specific wildfire burn period, *and* satisfying the criteria for wildfire relevancy (described in Sect. 3.2). Once the social data is collected, filters are used (lines 4–7 of Algorithm 1, and outlined in Sect. 3.2) to remove misleading or irrelevant posts and reduce noise.

Algorithm 1 Twitter Data Collection Algorithm

Require: S_{DOY}, E_{DOY}, location(x_i, y_i)
Ensure: list of tweets for fire
1: **for** fire **in** Historical Wildfire Dataset **do**
2: query ← GenerateQuery(location(x_i, y_i))
3: list of tweets ← FullArchiveSearch(S_{DOY}, E_{DOY}, query)
4: **for** tweet **in** list of tweets **do**
5: **if** tweet **contains** [filter phrases] **then**
6: **delete** tweet
7: **end if**
8: **end for**
9: grouped tweets ← GroupTweetsByDay(list of tweets)
10: **return** grouped tweets
11: **end for**

In this paper, the collection of social media in Algorithm 1 is implemented using the Twitter V2 API with an academic research track key for a full archive search of Twitter's public data [50]. This endpoint takes a query formatted in guidelines with Twitters query syntax [51]. Queries specify keywords and date intervals to search on, and returns a list of tweets satisfying these constraints. The date interval refers to the burn period for a wildfire event, and the query is designed to isolate tweets specifically referring to this fire event.

Queries

Queries are a crucial part of the operation of the system. During training, the system searches Twitter's historical archives for tweets which are relevant to wildfire events, before sentimentally analysing them. Examples of queries have been implemented in another study [11], this system is highly complex and computationally expensive. This system [52] offers a way of geocoding tweets based on their text content alone, which would be extremely useful in this system. However, the geographical accuracy of this system in its current form would likely add more noise to the data. With further development in this area, methods like this could become viable.

In this paper, we propose a query generator based on location and also event-specific keywords. Due to the fact that only around 1–2% of tweets are published

with geocoding data [52], this paper employs a wider strategy for refining and collecting tweets which are relevant to wildfire events. Here, we include published tweets that mention the location of the burn area in the tweet text, and therefore consider these relevant to wildfires currently burning in that area. For each fire from "Physics wildfire data collection", a query is generated to target tweets referring to a particular fire. Queries target data relevant to its unique fire event, and ignore noise referring to other fire events/topics. This investigation has developed a criteria for generating queries dynamically which is designed to maximise relevancy to the wildfire event.

The following three criteria determine a tweet's relevancy:

(C1) The tweet must contain one location keyword: A location word at either the county, state or country level.
(C2) The tweet must also contain one fire keyword, defined as: Fires OR Wildfires OR Bushfires OR Landscape Burn OR "Wildland Burn".
(C3) Finally, the tweet is considered relevant if it contains one of a number of hashtags which are generated from the location and included in the query.

Geocoding wildfire coordinates in "Physics wildfire data collection" results in a string location generated in the form (County/district, State, Country). These terms are then used as location keywords to add locality to tweets, defined in Criteria (C1). Additionally, tweets need also to refer to landscape fire or burning to be considered relevant to a wildfire event. Thus, we also only return tweets which also contain a wildfire keyword, defined in Criteria (C2). An example of a dynamically generated query for an individual wildfire instance is shown in "Twitter data".

Finally, we utilise Twitter hashtags. Hashtags are a word or phrase preceded by a hash sign (#), used on social media websites and applications, to identify digital content 'tagged' to a specific topic. This allows users to search hashtags to find discussions on topics, and is an effective way of tying tweets to specific events. We use the location and wildfire keywords to generate hashtag search terms included in the query, outlined in Criteria (C3).

Filters

A large source of noise in the data is the collection of tweets that are not relevant to the specific fire in question/not related to wildfires in general. Thus, measures were taken to reduce noise and maximise relevancy to specific wildfires. Line 4–7 of Algorithm 1 outlines the filtering process during the data intake process.

- Significant noise came from tweets expressing political views and opinions rather than wildfire events. Filtering based on political keywords, such as 'corruption', 'Trump', 'Clinton' and 'feel the burn', reduced political noise in the data.

- Another source of noise are misuse phrases of the word 'wildfire'. An example would be the tweet "Corruption is spreading through the California construction industry like wildfire". Removing results containing the phrase 'like wildfire', reduced this noise.

Sentimental wildfires: social sentiment analysis

In this study, social SA is used to gauge online regional emotional expressions towards wildfire events, to investigate whether online social sentiment is reflective of local wildfire behaviour. Social SA converts qualitative social text data into quantitative metrics which can then be used as social characteristics for a dataset of wildfire events. SA is performed using Natural Language Processing [41, 42]. This approach applies ML algorithms to the linguistic features of text data [43–45]. If there is no labelled training data available, unsupervised methods need to be employed. However, in the presence of labelled training data, a number of different implementations, such as linear, decision tree, probabilistic and rule-based classifiers [53] can be used. This implementation is useful because it is a relatively simple yet effective way of quantifying peoples feelings towards a particular topic. Using Twitter data, we are able to take a wide range of these opinionated textual accounts of wildfire events, and build up a picture of the overall online feeling of the digital population towards the event. Here, by showing that social sentiment is a predictor of wildfire activity, we show that twitter data is useful in wildfire modelling, and subsequently open the door for more advanced textual information extraction methods. This study analyses the sentiment of tweets associated to wildfires, and uses these to generate social sentiment variables for individual wildfire events. Individual tweets are sentimentally analysed, and then these sentimental scores are aggregated to form overall social sentiment vectors and variables for historical wildfire events. Following the collection of the social media data for the physical wildfire dataset, we sentimentally analyse the tweets collected to generate social sentiment variables for the individual wildfire events. Tweets were collected for the wildfires and grouped by the date. URLs, punctuation, symbols, and repeated characters were removed from the text. Following this, the sentiment of each of the posts are analysed using Sentiment analysis, and then these values are aggregated by day. Sentiment is measured using two metrics which are defined as follows:

Definition *Sentiment Score* "Sentiment Score ranges between – 1.0 (negative) and 1.0 (positive) and corresponds to the overall emotional leaning of the text" [45].

Definition *Magnitude* "Magnitude indicates the overall strength of emotion (both positive and negative) within the given text, between 0.0 and +inf. Unlike score, magnitude is not normalized; each expression of emotion within the text (both positive and negative) contributes to the text's magnitude" [45].

From the above definition we defined two variables; S and M, which constitute our social sentiment variables. These variables observe the constraints $S \in \mathbb{R} : -1 \leq S \leq 1$ and $M \in \mathbb{R} : 0 \leq M$ respectively.

Algorithm 2 Tweet SA Algorithm

Require: list of tweets for fire
Ensure: Social Sentiment vectors S, M for fire
1: $S \leftarrow []$
2: $M \leftarrow []$
3: **for** fire **in** Historical Wildfire Dataset **do**
4: **for** tweets for day **in** list of tweets for fire **do**
5: $S \leftarrow$ Analyse(tweets for day).Sentiment
6: $M \leftarrow$ Analyse(tweets for day).Magnitude
7: **end for**
8: **return** S, M
9: **end for**
10: $S_{mean} \leftarrow AVG(S)$
11: $M_{mean} \leftarrow AVG(M)$
12: $M_{ovr} \leftarrow SUM(M)$
13: $S_{ovr} \leftarrow SUM(S)$
14: $tot_{Tweets} \leftarrow SUM(tweetsForFire)$

Algorithm 2 shows how tweets are analysed per day, for each fire. For each day, Sentiment *S* and Magnitude *M* values are calculated, then appended to sentimental vectors for the fire. *S* and *M* are computed as shown in lines 5 and 6 of Algorithm 2. The function Analyse analyses word frequency and polarity, assigning each word in the passage of text emotional valence and leaning scores. Each words emotional leaning score is positive or negative, depending on the overall positivity of the word and its context in the sentence. The scores for all of the words are aggregated by their frequency to give summed scores of emotional learning and strength, *S* and *M* respectively. *S* is then normalised to the bounds [– 1,1]. This gives the sentimental score *S* and magnitude *M* for an individual post. Line 10–14 shows these scores for individual posts are then aggregated to generate overall social sentiment variables for each wildfire event. Here, we generate a number of different variables from the social sentiment data. Each of these variables will be used to train an ML model in "Sentimental wildfire: predictions". "Test cases and results" then assesses the suitability of these variables and their corresponding models.

The results of SA are corresponding Sentiment and Magnitude values for each burn day, forming a vector over the burn period. The dimensionality of these vectors is $d + 2$ where d is the fire duration defined in "Physics wildfire data collection", accounting for the one day pre/post fire also considered in this analysis. Lines 10–14 of Algorithm 2 outlines the generation of the five social sentimental variables used in the prediction in Sect. "Test cases and results":

- S_{mean} : Average Daily Sentiment Score
- M_{mean} : Average Daily Magnitude Score
- S_{ovr}: Overall Sentiment Score
- M_{ovr}: Overall Magnitude Score
- Tot_{tweets}: Total number of Tweets for wildfire

These variables constitute the sentimental vector

$$\mathbf{y} = [S_{\text{mean}}, M_{\text{mean}}, S_{\text{ovr}}, M_{\text{ovr}}, Tot_{\text{tweets}}]. \tag{2}$$

Sentimental wildfire: predictions

We now present the part of the system which implements models to investigate whether social sentiment for wildfires is reflective of wildfire characteristics. An overview of the entire system is shown in Fig. 1. "Sentimental wildfires: data collection" and "Sentimental wildfires: social sentiment analysis" are represented by the top row of the diagram, with the lower modules outlined in this Section. Here we outline two types of models, corresponding to the two predictive objectives. These predictive objectives implement models which will intake the newly collected socio-physical data to predict different types of attributes.

The aim of these models is to investigate relationships in social sentiment towards fire events, and actual physical/geographic characteristics of the wildfire events themselves. Let $\mathbf{x}$ be a vector denoting the physical variables defined in (1) and let $\mathbf{y}$ be the vector denoting the sentiments as defined in (2), where S denotes the sentiment score and M denotes the magnitude. We make the following model objectives, the proving/disproving of which are the goals of the models:

1. *Sentiment prediction model*: there is some function $f : \mathbf{x} \rightarrow \mathbf{y}$, which predicts the social sentiment towards a wildfire event based on physical characteristics of the fire. This model uses the collected social media data as ground truth in training.
2. *Physics prediction model*: there is some other function $g : \mathbf{y} \rightarrow \mathbf{x}$, which predicts physics characteristics of a wildfire event based on it's social media sentiment. This model is trained on the analysed collected social media data to predict the observational wildfire data.

These goals are the basis of the investigation the resulting models form part of the socio-physical model shown in Fig. 1. Following the setting of the prediction goals, ML models were developed to use the social wildfire data to predict the physical wildfire variables, and vice versa. In this paper, we implement GBRF regressors, Multi Layer Perceptron Regressor NN and SVR models. For a comprehensive description of these ML methods, and their respective optimisable hyperparameters, please see Appendix A, B and C respectively.

Test cases and results

For testing our models, we chose two different test cases: North America (United States and Canada) and Australia over the duration of 2016. We use historical wildfire [19] and Twitter data and we investigate links in online social sentiment and wildfire evolution outlining a brief exploratory analysis of the datasets and a comparison of the AUS and US domains. The section ends with the training, testing and comparison of our ML prediction models.

Data collection

Physics wildfire data

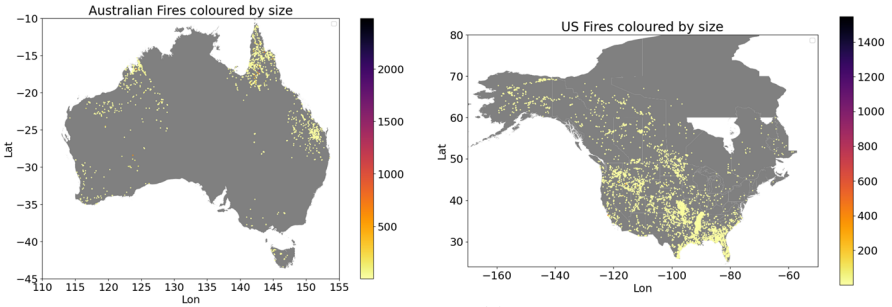
As mentioned in "Physics wildfire data collection", this report uses for the physical data a subset of the GFA 2016 ignitions data set—those in North America (comprising the US dataset), and also Australia (the AUS dataset).

The filtered fire data consisted of 2389 US fires 1365 and AUS wildfires. These are shown plotted in Fig. 2 for US (a) and AUS (b) wildfire events. The physical wildfire characteristics taken from this dataset were described in the vector **x** in Eq. (1).

Twitter data

Using Algorithm 1, social media data was collected about the wildfire events in the US and AUS in 2016. The queries described in Sect. "Queries" have been implemented for all fires in the US and AUS subsets of data. In this context, an example of the criteria implemented into a query for a given fire location is:

Example (Maui County, Hawaii, USA) → ("Maui County" OR USA OR Maui OR Hawaii) AND (Wildfire OR Wildfires OR "landscape burn" OR "wildland burn") OR #MauiCountyWildfires OR #MauiCountyFires OR #USAWildfires OR #USAFires OR #MauiWildfires OR #MauiFires OR #HawaiiWildfires OR #HawaiiFires'



(a) AUS Fire Data coloured by Size (km2)

(b) US Fire data coloured by Size (km2)

Fig. 2 Filtered AUS (**a**) and US (**b**) Fire data used in this investigation

For the 2389 US fires queried for social media data collection, 1,820,157 tweets were returned. For the 1364 AUS wildfire instances, 478,082 tweets were collected via the data collection process. This gives a combined database of 3753 wildfire events and 2,298,239 total tweets.

Having collected a large textual dataset of Tweets associated with both AUS and US geographic domains, we are now able to perform exploratory and compariative analysis on these datasets to assess their suitability. Figure 3 shows a Wordcloud for the full text of tweets collected for both the AUS (Fig. 3a) and US (Fig. 3b) wildfires. Here we can see the keywords 'wildfire' appearing frequently, however an interesting result is that 'bushfire' is an extremely prominent keyword used with high frequency in the AUS dataset, but is very rarely used in the US dataset. This differing terminology is certainly regional, and highlights the significance of localised language if this system is to be used in other domains or expanded globally.

We can see also in both wordclouds the geographic keywords such as 'qld', 'queensland', 'tasmania', 'south', 'california', 'oklahoma', 'kansas', 'alberta', 'fort', 'mcmurray' etc. and also fire service account names and hashtags such as 'qldfire', 'smokealert', 'tasfire' all play a prominant role in these word clouds. Finally we notice a large amount of wildfire specific language featuring frequently within the tweets for both datasets. Words such as 'smoke', 'vegetation', 'crew', 'help', 'massive', 'home', 'evacuation', 'burn', 'threaten' etc. are all highly domain specific language.

We can further observe the differences in vocabulary in Fig. 4. By tokenising the entire textual datasets and then performing frequency analysis on these tokens, we can show the top 20 tokens used in the tweet text for the two datasets. We can see here that the language used is quite varied between the domains, which was unexpected, as this may indicate that the different regions communicate about these events on social media in different ways.

It initally appears from Fig. 4 that the top tokens used in the US dataset are more negative than the AUS dataset. However, performing SA using Algorithm 2 on these 20 top tokens shows that this is not the case. Averaging the sentimental values for the top 20 tokens in the US and AUS dataset respectively gives values of 0.03 and 0.02, which shows very little difference despite the differing language. We therefore conclude that despite the regional differences in language in



Fig. 3 US (**a**) and AUS (**b**) Tweet text Wordcloud

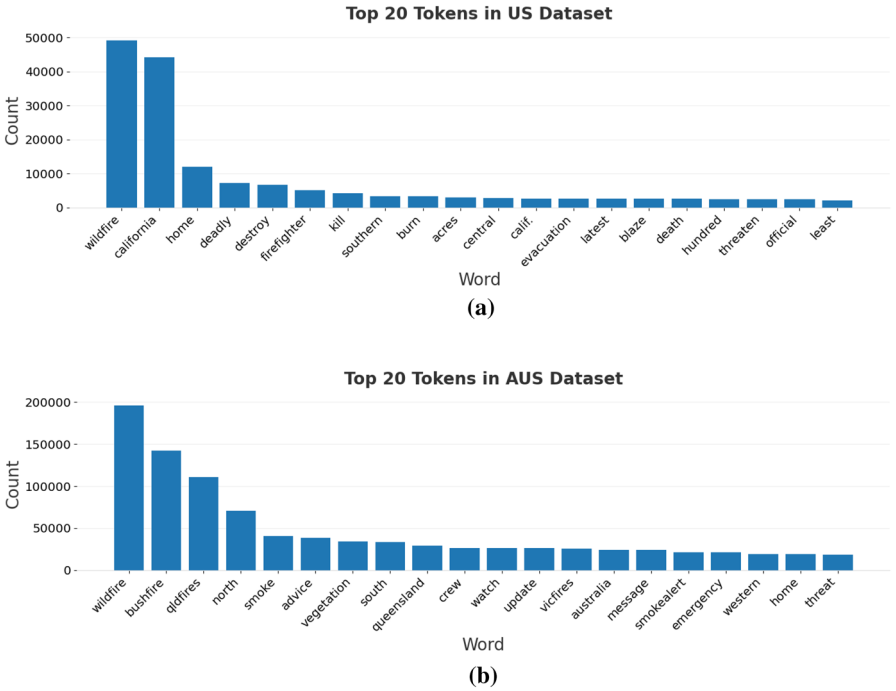
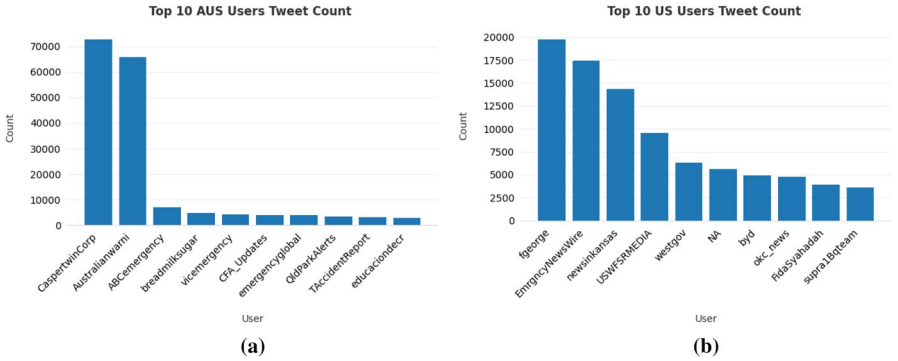


Fig. 4 Top 20 words used in Tweets in US dataset (**a**) and Top 20 words used in Tweets in AUS dataset (**b**)

these datasets, there is a similar expression of sentiment across the two datasets. It should also be noted that despite these tokens being used for textual data surrounding natural disaster events (which are inherently emotionally negative), the average sentiment of the top tokens of both datasets was still relatively neutral (close to 0). Both the similarity and the neutrality is an unexpected result here, given the difference in top tokens, and their emotionally negative domain application. This may be because both domains are English speaking, and there is a natural positive bias in the English language [54]. Had other domains, exhibiting other native languages been included, we may have had more comparative results, and this may be an important part of future work in this area.

We now compare the accounts which publish the tweets which comprise our datasets. There were 19,587 unique users stored in the AUS database. However, some users appear much more than others, and the top 20 users account for 190,167 tweets, and the top 10 users accounting for 171,891 tweets in 2016 alone, and these are shown in Fig. 5a. The top 2 accounts was the Australian Weather Fire Road and Police Warnings accounts @Australianwarni and @CaspertwinCorp, with 65,793 and 72,697 tweets, meaning that the top 2 accounts make up nearly 29% (138,490 of 478,082) of the entire dataset. The user with the third number of tweets was @ABCEmergency with 7,135 tweets, illustrating the



(a) (b)

Fig. 5 Top 10 Twitter accounts in the AUS Database (a) and in the US Database (b)

dominance of @Australianwarni and @CaspertwinCorp in this dataset. This was an unexpected result, and an aspect not shared by the US dataset.

Conversely, there were 67,879 unique users stored in the US database. The top 10 users accounted for 90,148 tweets and are shown in Fig. 5b. The top 20 users accounted for 122,582 tweets. The top user was the account of a journalist named @fgeorge, with 19,775 tweets in the dataset. We can see from Fig. 3 that the tweets are spread out over man more dominant accounts in the US dataset compared to the AUS dataset. This is an interesting result, and may be due to the decreased Twitter activity in the AUS domain in comparison to the US.

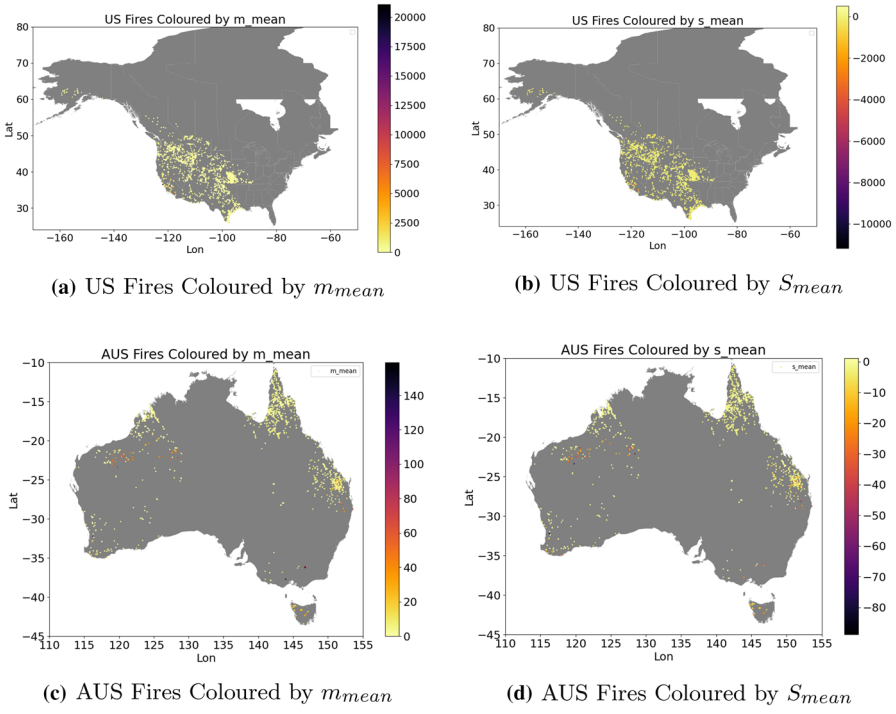
Sentiment analysis

Average daily sentiment and Magnitude scores are shown in Fig. 6, where the US and AUS wildfire events are shown plotted on a map, coloured by their average magnitude and sentiment scores respectively.

Finally, from this, the sentimental arcs of each individual wildfire could now be plotted over the course of the burn period. An example of one of these arcs is shown in Fig. 7, for the overall Sentiment and Magnitude, the average Sentiment, and finally the cumulative magntiude, representing the cumulative emotional expression over the course of the wildfire event.

The sentimental variables for the AUS dataset are shown plotted in Fig. 8. Figure 8a plots the Sentiment and Figure 8b shows the Magnitude respectively for the AUS dataset. Similarly, Fig. 8c plots the Sentiment and Fig. 8d shows the Magnitude respectively for the US dataset.

We can see from comparing Fig. 8a and b that the distributions for the Sentiment is very similar across the two datasets. Similarly, comparing Fig. 8c and d also shows that the Magnitude follows a similar distribution between the two domains. This indicates that our models should generalise reasonably well to predictions between these datasets.



(a) US Fires Coloured by m_{mean}

(b) US Fires Coloured by S_{mean}

(c) AUS Fires Coloured by m_{mean}

(d) AUS Fires Coloured by S_{mean}

Fig. 6 US and AUS Fires coloured by (average) magnitude and sentiment

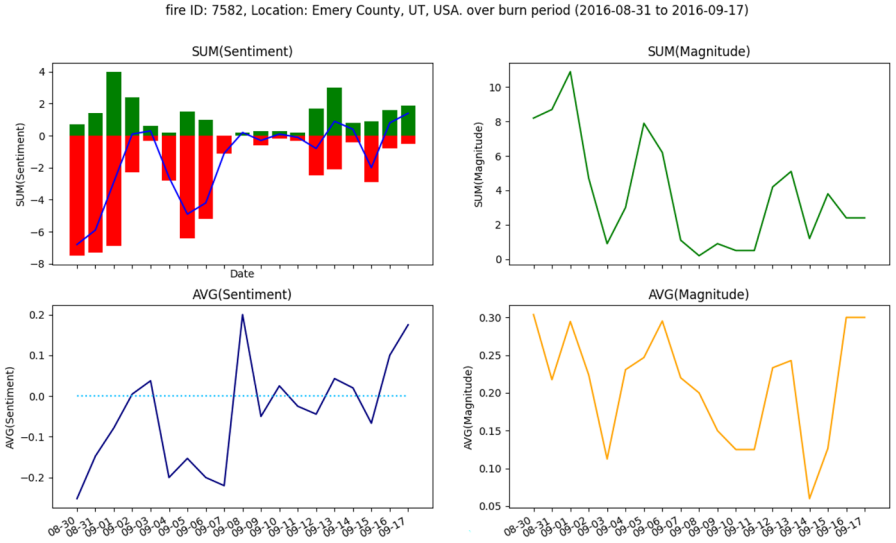
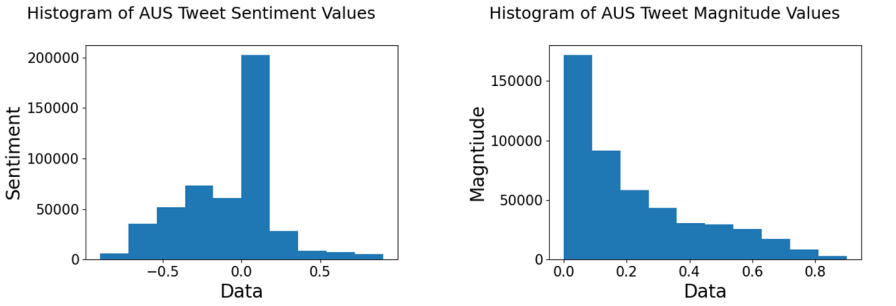
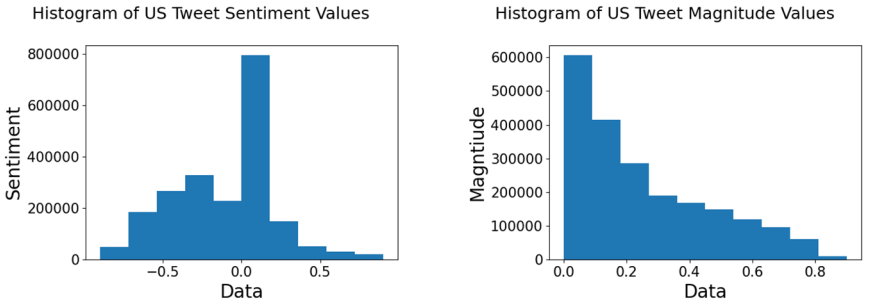


Fig. 7 Example sentimental Arc generated from SA



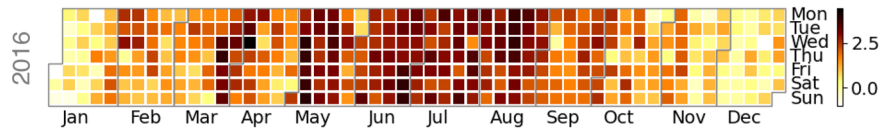
(a) Tweet Sentiment in AUS database **(b)** Tweet Magnitude in AUS database



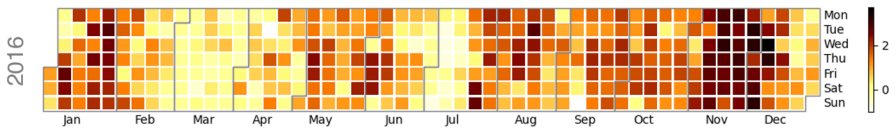
(c) Tweet Sentiment in US database **(d)** Tweet Magnitude in US database

Fig. 8 AUS (top) and US (bottom) sentiment (left) and magnitude (right)

By aggregating Sentimental values for the tweets for each day in 2016, we are able to visualise daily online wildfire social sentiment across the course of the year. These are shown in the form of sentimental heatmaps, for the US dataset in Fig. 9a, and additionally for the AUS dataset in Fig. 9b. Here we can see very clearly that online social sentiment correlates to regional seasonal wildfire

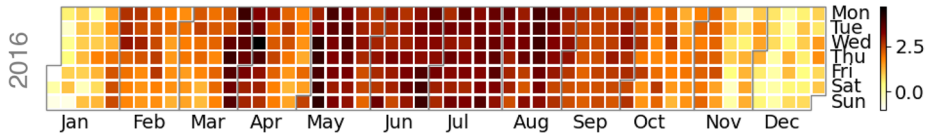


(a) Heatmap of US online social sentiment for 2016

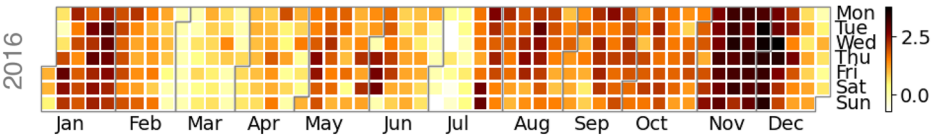


(b) Heatmap of AUS online social sentiment for 2016

Fig. 9 US (**a**) and AUS (**b**) online social sentiment heatmaps for 2016



(a) Heatmap of US online social Magnitude for 2016



(b) Heatmap of AUS online social Magnitude for 2016

Fig. 10 US (**a**) and AUS (**b**) online social Magnitude heatmaps for 2016

periods—with the US dataset showing hot-spots of activity from April to September, and the AUS dataset showing increased activity from October to Feburary. Again we can see this clearly in Fig. 10 which show the same plots but for the other variable, Magnitude.

This is a positive indicator for the quality of the data collected in Sect. 6.1, given that we are seeing here that the different regional wildfire seasons in the two domains is being accurately captured. Finally, we can also begin to see individual instances of wildfires manifesting themselves as dark patches here in the heatmaps, such as the Fort McMurray fire which is an event which features prominently at the start of May of this year in the heatmap.

This seasonal difference in domains further can also be observed in Fig. 11. The plot shows the summed daily magnitude (Absolute Log 10) over the course of 2016 for the US (Red) and AUS (Blue), datasets. Additionally, shown annotated are some of the biggest wildfire events to have occurred in US and AUS during their respective fire seasons. We can see that nearly all of the peaks in wildfire activity can be explained by one of these events. This is another good indicator of the quality of the data, as this shows that individual wildfire events, as well as regional seasons, are being captured by the data.

In summary, here we have visualised our textual datasets across different domains. We have shown that they exhibit different language patterns through the wordclouds in Fig. 3, and the top words in Fig. 4. However, we also showed that the distribution of the sentimental variables are actually very similar in Fig. 8. Furthering on this, Fig. 9 showed the regional seasonal differences between the domains, indicating that this element of wildfire activity was captured by the data. Finally, Fig. 11 showed that we may be able to pick up on activity spikes associated with individual wildfire instances, which would be highly beneficial from an analysis perspective, indicating that the data has a granularity of that level.

Prediction models

As already mentioned, this investigation utilises wildfire and social media data from North America and Australia in 2016. The system was tested on these two different domains; to assess differences in regional dialect, to investigate the generalisability of the individually trained models, and due to the abundance of wildfire and social media activity in both of these areas. In this section we employ the three ML models introduced in Sect. 5: GBRF, NN and SVR. These three models were trained on each of the three datasets: US, AUS and the combined dataset US+AUS. Then, to assess the generalisability of these models, we test them on the unseen set of data (i.e. models trained on US data are tested on AUS data and vice versa). This way we can evaluate how well the US model performs on the AUS data compared to the AUS model. We train the ML models on the 75% of data and we test them on the remaining 25% of unseen data.

We investigate the difference in training using different models and regional datasets of socially analysed wildfires, which are generated in Sect. 6.1. We created three of these datasets; one for the US, AUS, and one combined dataset for all 2016 ignitions, and their associated tweets.

We then trained three different ML models on these datasets. A GBRF model has been implemented using the XGBoost package in python [35], this model and it's optimisable hyperparamters are described in Appendix A. Additionally, a NN (Multi-layer Perceptron regressor) model in SKLearn is also implemented. A description of the model hyperparameters is given in Appendix B. Finally, an SVR implementation is presented, also in SKLearn. It's corresponding hyperparameters is outlined in Appendix C. The models were trained on the data using a wide grid search and a training set size of 75%—around 5000 wildfire instances. $\mathbf{x}$ denotes the 10 physical attribute values defined in (1) for the wildfire gathered from the GFA

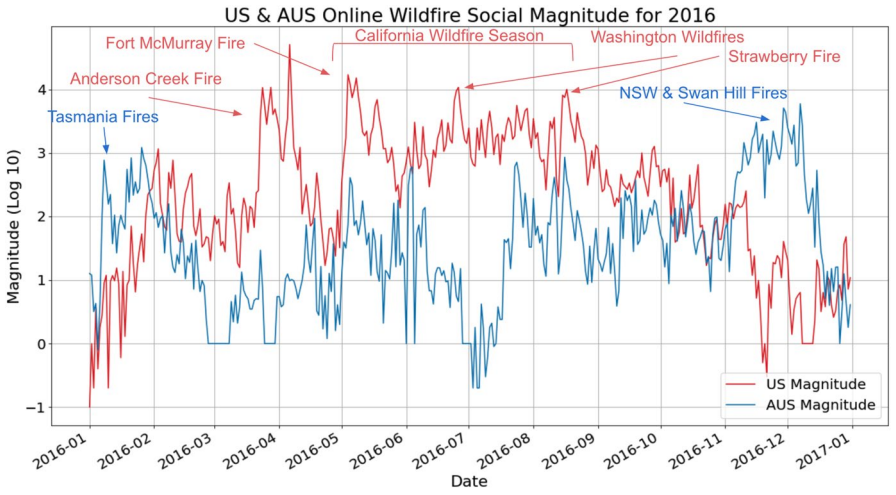


Fig. 11 Absolute magnitude (Log 10) plotted for both AUS (blue) [55] and US (red) datasets over the course of 2016, annotated with significant events during this period

dataset, and **y** denotes the 5 social sentiment variables defined in (2), generated from the data collection stage were used as the target. For the Physical Prediction Model, the same models were used, utilising the GBRF, NN and SVR, however this time using **y**, i.e. the five social variables, to predict **x**, i.e. the 10 physical characteristics. We use the RMSE as the metric of comparisons.

Prediction model *f*

Table 1 shows the differing results gathered across the three datasets. Tables 2 and 3 show results of the execution time for training and testing, respectively. For the GBRF models, in Table 1, we see that there is a reasonably good generalisation between the different domain datasets, and also that there is the best RMSE values across a number of different variables for each dataset. Overall the GBRF models perform well, both in terms of generalisation and in accuracy across the different datasets. For the NN models, we see that these models perform well in their native datasets, for the domain specific models. i.e. the US model performs well on the US data and the AUS model performs well on the AUS data. However, we see a big gap in results in terms of generalisation. US models do not perform well on the AUS data and AUS models perform equally badly on US data. It would seem that in this case NN do not generalise well to our geographic variances in online expression. Finally, the SVR shows the worst RMSEs for most of the social sentiment variables modelled here. In summary, we observe here that GBRF models show good RMSEs and reasonable generalisation ability to new and unseen domains. Additionally, NN models give low RMSEs on native datasets, however do not cope well in new domains. Finally, SVR models did not perform well in terms of RMSEs. We can also observe here that in general, AUS models generalise worse to US data than US models do to AUS data. This may be because of the reduced data variance in the AUS dataset. Models fit on this data could be vulnerable to overfitting. Models trained on the combined dataset showed good results, and are shown to be the most generalised model. For the subsequent Sect. 6.3, we use the combined dataset in our models to maximise this generalisability/accuracy.

Following this, we investigate using different ML models in the modelling prediction part of the system. We test three models here; GBRF, NN, and SVR models, and compare their training and also exceution times on the test set. The training set used here is the combined dataset, split 75% into training data and 25% into testing data. The exceution time on the test data is the 'test time'.

The SVR model showed consistently the lowest training times across the three models, however Table 1 shows that this was at the cost of accuracy in comparison to the GBRF and NN models. The GBRF and NN implementations both show similar results, in both training time and RMSE values. The RMSEs are sufficiently low, and show good results for these models. The training times for these models are higher, however, given that we are training this model on retrospective historical datasets, a high training time is of less detriment to this specific application. Following the (offline) training of the system, we can then use this in an online implementation, making real time predictions on data where accuracy is of greatest importance.

Table 1 Prediction model f ML method comparison results: using GBRF, NN, and SVR models for predicting social sentiment variables from physical wildfire characteristics. Included are results from US, AUS and combined datasets of wildfires and tweets

Training data	US model		AUS model		Combined model
Testing data	US data	AUS data	US data	AUS data	Combined data
Variable	GBRF model				
S_{mean}	32.52	30.55	44.39	16.83	24.56
M_{mean}	50.09	58.65	67.45	19.80	50.49
M_{ovr}	314.75	565.24	648.5	95.2	498.5
S_{ovr}	202.25	383.85	357.30	44.19	167.65
tot_{tweets}	2343.5	2281.1	3701.70	408.5	2635.5
NN Model					
S_{mean}	28.98	9.23	407.60	4.52	21.10
M_{mean}	62.88	18.75	65.20	9.84	44.92
M_{ovr}	1017.73	149.27	646.52	110.88	435.26
S_{ovr}	238.27	115.45	353.23	77.42	244.54
tot_{tweets}	2160.65	22994.96	3641.59	674.20	2610.37
SVR Model					
S_{mean}	39.59	13.51	43.63	6.09	29.61
M_{mean}	66.56	21.24	63.85	11.91	54.22
M_{ovr}	523.92	4837.70	2211.65	139.33	466.76
S_{ovr}	252.42	100.04	351.59	96.70	275.48
tot_{tweets}	2461.10	41719.41	14814.51	637.95	2625.48

Metric used here is RMSE

Table 2 Comparison of training times in seconds for the three ML models for prediction model f

Variable	GBRF (secs)	SVR (secs)	NN (secs)
S_{mean}	349	108	293
M_{mean}	366	66	504
M_{ovr}	334	27	658
S_{ovr}	340	44	449
tot_{tweets}	330	21	348

Table 3 Comparison of physics-based prediction model f execution times on the test set in milliseconds for the three ML models

Variable	GBRF (msecs)	SVR (msecs)	NN (msecs)
S_{mean}	2.94	43.99	1.92
M_{mean}	5.75	54.11	0.09
M_{ovr}	3.37	45.78	0.24
S_{ovr}	2.28	41.24	0.65
tot_{tweets}	2.94	47.21	0.29

Multiple models can be trained offline and substituted into the online system due to its modular structure. For this reason, we prioritise a low RMSE in these results, and the GBRF and NN models are clearly the best models for this application. For the test times, a low execution time on new data is important for making new predictions on live data. we find that the SVR models consistently showed the highest execution times on the test set. This is not desirable, and the GBRF and specifically the NN models perform considerably better here.

It has been shown that NN models give less interpretability [56] and are less generalisable to real-world domains. Interpretability is highly desirable in this context given the socially conscious angle needed to successfully implement a system such as this. GBRF models are beneficial for this because they have built in feature importance [57], which gives a clear justification for the predictions made. Conversely, NNs are somewhat of a black box in this regard. This is a problem given the application case of this system, where we need to know how any why our model has made the prediction it has - "The model must also explain how it came to the prediction (the why) because a correct prediction only partially solves the original problem." [58]. Additionally, NNs need very large amounts of training data in comparison to the other ML models in order to achieve good accuracy [59]. Finally, we have seen that the NN models do not generalise well to our differing geographic domains. This is is a key drawback in this ML method for modelling. Here, despite good performance metrics on most test sets, the low interpreability and poor generalisation of the NN method means that we conclude at this point this method is not suitable for this type of problem.

As a result, we will be using GBRF models in the final Sect. 6.3 of the Sentimental Prediction model. Their interpretability, generalisability and good scoring metrics on the test set make this the optimal ML method this prediction problem.

Final sentimental prediction model *f*

Following testing in the previous sections, we present our final models for wildfire social sentiment prediction. This model uses GBRF models and is trained on a combined US and AUS training dataset.

Table 4 shows the results from the models trained on physical data with a social sentiment variable target. Results show that a number of social sentiment values can be accurately predicted by modelling physical wildfire variables. Models for predicting average Sentiment and Magnitude (S_{mean} and M_{mean}) show good results when attempting to predict these variables, with low MAEs of 6–10. Overall Sentiment and Magnitude (S_{ovr} and M_{ovr}) models also preformed well, with satisfactorily low MAEs. tot_{tweets} did not show good results.

Additionally, an interesting result here is the feature importance of the well-performing GBRF models. Figure 12 show the feature importances for the Average Magnitude and Sentiment variables respectively. The feature importance is calculated by the F-score - the frequency each attribute is used to make a split in the tree across the GBRF model. These results show that the models with the greatest social sentiment prediction power placed most importance on the attributes corresponding to S_{DOY} and E_{DOY}, Lon, Lat, and $\text{Pop}_{\text{Density}}$. These attributes were always the features

with the highest F-score (not necessarily in that order) for both models across 3 independent runs of training. These attributes are also significant because they are wildfire independent—that is, they are variables concerned with the time and place of the fire, rather than the fire itself. This result implies that important factors in predicting the online sentiment of a wildfire event is the time and location of the burn.

This could be because of regional variances in online social expression towards wildfires. The seasonal nature of wildfires and the differing of these seasons across North America and Australia may also be a contributing factor in this result.

In summary, these results indicate that we can indeed predict our target social sentiment variables by using physical fire characteristics as predictors. Additionally, the most important features were found to be the time and location of the wildfire burn. Both sentiment and magnitude of social emotion have been shown here to be accurately predicted for US and AUS 2016 ignitions, using only physical wildfire data as an input. This leads us to make the conclusion that the sentiment depends, and can be predicted, by the physical attributes of the wildfire by $f(\mathbf{x})$.

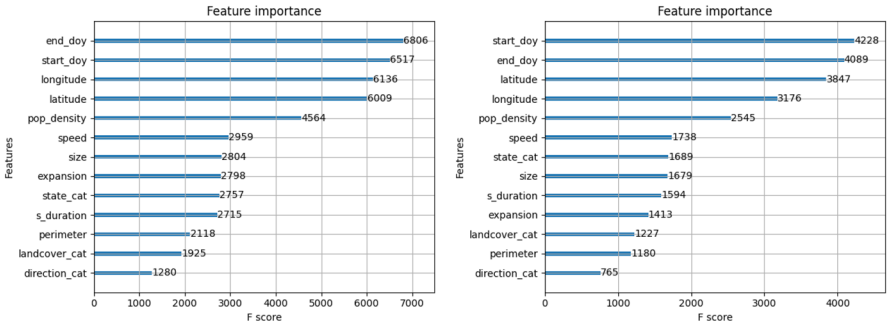
Physics prediction model *g*

This Section outlines the development of the physics-based wildfire prediction model g. As defined in "Sentimental wildfire: predictions", conversely to our development of the sentimental prediction function f, for this model we use the physical wildfire data as the ground truth, and train is our g models on social sentiment wildfire data, which was used as the target in "Prediction model f". As in "Prediction model f", we start our development of this model with an assessment of the datasets and comparative generalisability of the models trained on each regional dataset. In achieving this, we again train nine models for each variable, each one employing one of the three ML models; GBRF, NN or SVR, and trained on one of the domain-specific datasets; US, AUS or combined data. Following this, we can assess which variables are best modelled by which ML models, which variables and ML models generalise well for this problem, and conversely which ML models and datasets are not suitable for this type of problem or variable. Finally, we present our final model for our physical wildfire characteristics based on the social sentiment data collected for these wildfire events using the datasets and ML models explored. All results shown in this section are the average of three independent runs of training.

Table 4 Sentimental prediction model f results from predicting social sentiment variables from physical wildfire characteristics

Variable	MAE	RMSE	Gini	$Gini_N$	Score (%)
S_{mean}	6.34	17.83	0.367	0.846	38.7
M_{mean}	10.88	40.85	0.316	0.822	38.3
M_{ovr}	107.27	407.36	0.338	0.872	24.4
S_{ovr}	54.75	364.76	0.381	0.865	15.3
$\text{tot}_{\text{tweets}}$	458.927	2777.28	0.323	0.863	10.6

Included are results from US, AUS and combined datasets of wildfires and tweets



(a) Feature Importance of M_{mean} model. **(b)** Feature Importance of S_{mean} model.

Fig. 12 Feature importance of M_{mean} (**a**) and S_{mean} (**b**) models

Table 5 shows the comparison of datasets across all of the physical variables which are modelled in the scope of this paper. Here we can see for Lat and Lon variables, the native models do not generalise well to the unseen domains. NN models specifically show very poor results here. In general, the NN models also genralise poorly again, as in Sect. 6.3. However, they perform well on their native datasets and also on the combined dataset. Furthermore, the result that the Lat and Lon variables did not generalise well is likely to be expected given the wide geographic divide between the two domains of US and AUS. Conversely, models trained on the physical characteristics Exp, Speed, d, Per and Size all show considerably better results in terms of generalisability across the different datasets.

The GBRF models here show reasonably good RMSE values for both the native and unseen domains for most of the physical variables modelled in this section. This method has similar RMSE values to the NN models in native datasets, however NNs have the disadvantage of poor generalisability to unseen domains. Finally, SVR models again do not show particulary good results either in terms of RMSE values or generalisability.

S_{DOY} and E_{DOY} also showed interesting results here, as the two native models do not generalise well, however, the combined model still shows reasonably good comparative results. This is likely due to wildfire seasons in the two domains (US and AUS) being at different times of years. This was somewhat expected, and likely a result of what we have observed in Fig. 9. This plot shows how the fire seasons of the two domains are almost the complete inverse of one another, and a model trained on one of the individual native datasets would not have the training data to account for this. Furthermore, this was seemingly largely accounted for in the combined model. This is a good indication that the quality of the data has allowed these models to generalise better and account for these regional differences. These results indicate that the use of the combined dataset will make the physical prediction models more general except for the case where location is the target.

In summary, Table 5 shows that ML models such as GBRFs and NN can accurately model physical wildfire variables such as d, Speed, Exp, as well as Lat and

Table 5 Sentimental wildfire prediction model g: results from predicting physical variables from sentimental wildfire data

Training data	US model		AUS model		Combined model
Testing data	US data	AUS data	US data	AUS data	Combined data
Variable	GBRF model				
Lat	5.51	60.09	67.49	4.41	14.62
Lon	9.7	244.4	243.7	6.00	52.94
Size	37.42	106.41	41.1	41.55	80.08
Per	14.34	42.84	20.34	40.93	25.66
d	1.786	4.88	3.79	1.95	2.12
Speed	0.83	1.47	1.07	1.38	1.05
Exp	1.20	5.14	2.45	4.43	4.40
$\text{Pop}_{\text{Density}}$	643.33	378.14	775.26	290.90	733.76
S_{DOY}	58.53	95.76	103.91	50.81	59.50
E_{DOY}	62.88	99.77	96.61	52.44	59.10
NN model					
Lat	5.85	1893.56	1.01e+04	5.26	19.18
Lon	11.52	239.08	243.80	7.62	72.38
Size	53.73	1580.05	216.19	73.40	84.05
Per	19.58	45.23	1.92 e+04	47.93	20.33
d	0.36	16.75	8169.72	2.46	2.48
Speed	1.18	1.44	1.83	1.38	0.94
Exp	1.16	732.05	57965.15	4.00	4.20
$\text{Pop}_{\text{Density}}$	819.26	1478.88	770.06	28.76	567.03
S_{DOY}	178.61	6810.54	10331.92	84.14	149.71
E_{DOY}	82.13	24112.91	2.00 e+05	79.76	84.32
SVR model					
Lat	6.30	58.99	6.53	5.90	27.81
Lon	11.73	252.98	12.89	7.77	130.75
Size	22.12	8.29 e+05	6.36 e+09	60.40	48.60
Per	9.59	42.12	23.92	54.86	35.05
d	1.10	8.08	6.85	3.09	3.25
Speed	0.85	1.42	1.31	1.58	1.26
Exp	2.14	5.12	2.36	5.43	3.58
$\text{Pop}_{\text{Density}}$	917.85	384.22	771.0	28.13	569.27
S_{DOY}	76.97	110.92	83.08	78.65	91.46
E_{DOY}	76.04	109.03	83.43	80.94	88.63

Included are results from US, AUS and combined datasets of wildfires and tweets. Metric used here is RMSE

Lon and S_{DOY} for native datasets. Variables Size, Per and $\text{Pop}_{\text{Density}}$ all showed satisfactory early-stage results. We have shown that these geographic (Lat and Lon) and temporal (S_{DOY} and E_{DOY}) variables do not generalise well to unseen domains due to differing fire seasons and locations. However, this information is somewhat arbitrary

for this application, as these values can be inferred from the tweet data itself. Additionally, we found that NN models too do not generalise to unseen domains. Overall, most of the physical variables show good results in at least one of the datasets at this early stage of model development. We conclude these are good results, and further investigate different ML models for the Physical prediction model in Sect. 6.3.

Again we test the same three models; GBRF, NN, and SVR, and assess the training time on the training set and the execution time on the test set. The results from recording the training time and the test times on their respective sets are shown in Tables 6 and 7 respectively.

Similarly to "Prediction model f", we see that SVRs perform poorly in terms of a large execution time on the test set. SVRs do however, have a low training time, but as discussed previously in "Prediction model f", this is of little relevance to this application. Conversely, again we observe that NN models report a much larger training time, but have a very short execution time on the test set. This is a good combination of results for our given application, as discussed previously. Finally, GBRF models give reasonable training times given the application, and also good execution times for operation. We can summarise from Tables 6 and 7 that SVR models perform poorly in terms of training time, and that NN and GBRF models perform well in terms of operational speed.

In summary, we have explored how the different ML models perform on the different datasets, evaluating them in terms of operational time and error metrics. Table 5 shows that d, Speed and Exp call all be modelled to a good degree of accuracy. Additionally, the variables which are more domain specific, i.e. S_{DOY} and E_{DOY} , Lat and Lon, do not generalise well and perform better on their native datasets. The variables were genrally modelled best using the GBRF and NN models, which both give similarly low RMSE values. We also see, similarly to Table 1 that the NN models have worse generalisability across these physical variables also. Additionally, as outlined in "Prediction model f", GBRF models give greater interpretability and transparency, which is a key aspect for this application. As a result, as in "Prediction model f" we will prefer to use the GBRF models over the NN moethod for the final models in the Physical prediction model which will be outlined in the following "Prediction model f".

Final physics prediction model *g*

The final model uses GBRF models on the combined dataset to achieve the best results. The average of these results are shown in Table 8, showing the models trained on the social sentiment data for the physical target variables. Results were generated by averaging three independent runs of training.

The most notable result shown in Table 8 is the model predicting fire Duration (d). This model showed very positive results, with an MAE of 0.84. This shows that the resulting model was able to predict wildfire duration to within one day from social sentiment values/social media data alone. Additionally, Speed and Exp both performed well. Size and Per did not perform as well as hoped, which was disappointing as it is an important wildfire attribute. Other methods may need to be incorporated to further predict these attributes. Lon, Lat, S_{DOY} and

Table 6 Sentimental prediction model g: comparison of training times in seconds for the three ML models

Variable	GBRF (secs)	SVR (secs)	NN (secs)
Lat	276.73	20.23	919
Lon	279.43	20.1	949.06
Size	305.77	32.66	360.83
Per	299.33	27.33	742.33
d	298.33	39.8	1446.87
Speed	286.33	68	663.66
Exp	286.66	62.33	253.33
$\text{Pop}_{\text{Density}}$	311.33	32	70
S_{DOY}	282.63	17.79	2953
E_{DOY}	261.50	16.62	2359

Table 7 Sentimental wildfire prediction model g: comparison of execution times on the test set in milliseconds for the three ML models

Variable	GBRF (msecs)	SVR (msecs)	NN (msecs)
Lat	4.90	23.74	0.03
Lon	6.11	33.20	0.06
Size	1.95	34.39	0.77
Per	3.69	45.95	0.054
d	4.22	28.36	0.53
Speed	2.32	23.24	0.47
Exp	2.09	26.88	0.47
$\text{Pop}_{\text{Density}}$	2.22	27.18	1.95
S_{DOY}	2.83	32.86	0.45
E_{DOY}	5.48	34.13	3.46

E_{DOY} all showed relatively good MAEs, despite being trained on the combined dataset. These are additionally interesting results because they were also the most important features for predicting social sentiment in the Sentimental Prediction Model described in Sect. 6.3. However, while these results indicate there is predictive power in these modes, MAEs for these results are not low enough for use to conclude that these are highly predictable attributes.

In summary, the results show that wildfire d, Speed and Exp can be predicted from its social sentiment variables. Also, the location (Lat and Lon) and time of year (S_{DOY} and E_{DOY}) showed promising preliminary results, however these results still show error in predictions. More data sources may be needed to more accurately predict these values. It is reasonable to conclude that these results indicate that physical wildfire characteristics can also be predicted using only social sentiment data, by $g(\mathbf{y})$.

Conclusion and future work

In this paper, we proposed a new predictive model for wildfires which uses social media and geophysical data sources with sentiment analysis to predict wildfire characteristics. We combined geophysical satellite data from the Global Fire Atlas with social data provided by Twitter.

In the paper we provided both formulations of the model and the code. We tested the new model with two different test cases: US and AUS dataset domains over the duration of 2016. The results show that social media is a predictor of wildfire activity, and present models which accurately model wildfire attributes. The algorithm and numerical methods proposed in this work can be applied to other natural events involving other equations and/or state variables.

The results in "Prediction model f" and "Prediction model f" indicate that both Sentimental and Physical prediction objectives outlined in "indicate that both Sentimental and" were achieved by the system described in "Sentimental wildfire: overview". The models trained were able to predict both social sentiment variables; Sentiment score and Magnitude, which had been generated from the data collection and analysis outlined in "Physics wildfire data collection" and "Social media wildfire data collection" of this paper. Additionally, models trained on the social sentiment data were able to predict wildfire duration d, Speed and Exp accurately, as well as the wildfire location (Lat and Lon) and S_{DOY} and E_{DOY} to less accuracy.

The initial insight to be gained by this paper is the link between physical wildfire activity and online social media expression. Implementing this system is a proof of concept model incorporating the concept of the human sensor, combined with satellite data, to predict wildfire attributes. Demonstrating a relationship between online social sentiment and wildfire activity indicates additional capabilities of social media for use in wildfire prediction. Further analysis of this data would reveal social sentiment dynamics at a finer level, which could allow the identification of areas of extreme danger to the public. This could additionally be used to automate allocation of emergency services, by monitoring local social media channels for local users documenting events.

Table 8 Sentimental wildfire prediction model g: results from predicting physical variables from social sentiment data

Variable	MAE	RMSE	Gini	$Gini_N$	Score
Lat	3.72	5.2	0.344	0.868	0.75
Lon	6.185	7.40	1.73	0.929	0.79
Size	9.66	33.89	0.354	0.835	0.14
Per	7.22	15.89	0.227	0.764	0.24
d	0.84	2.01	0.207	0.961	0.87
Speed	0.533	0.86	0.143	0.659	0.25
Exp	0.88	3.24	0.226	0.652	0.16
$\mathrm{Pop}_{\mathrm{Density}}$	69.05	575.18	0.397	0.424	0.02
S_{DOY}	40.73	60.90	0.100	0.736	0.56
E_{DOY}	39.62	61.15	0.095	0.734	0.56

In terms of improving the system, there is evidence from [13] that including more languages in the input can improve results when using SA to predict events. This could easily be incorporated into the future work by using a translation system. Additionally, we will explore the generation of either additional, reduced or different social sentiment variables. It is likely that additional analysis of the social media data will allow us to generate social sentiment variables with more physical predictive power for wildfires, while also being better reflective of online social expression. This could be achieved through context analysis and entity extraction, and will be generally facilitated by the extended linguistic analysis of the social media posts.

There are also results that indicate differences in regional expressions of online social sentiment at both a national and international level [60]. "Prediction model f" and "Final sentimental prediction model f" both show improved generalisation with combined/more geographic domains. Domains wider than those used in "Test cases and results" would likely improve model generalisation and results, and this could be achieved by additional rounds of data collection, focused on other regions.

One dimension which was limited in the data collection was the geotagging information tying a tweet to a location, as only 1–2% of Tweets contain geo-tagged data [52]. However, if only these Tweets were considered, it would considerably limit the data. "Queries" partially solves this by implementing a query algorithm to search on tweets mentioning specific locations, however, this is an imperfect solution. Models have been implemented to address this problem, geolocating Tweets automatically [52] when geolocation data is not available. However, these implementations are still relatively inaccurate, and incorporating them into this model would likely introduce additional distortion in already noisy Tweet text data.

Finally, if this system were to employ a domain-specific language model, similar to one used in [11], it is likely the sentiment analysis methodology outlined in "Sentiment analysis" would yield significantly more accurate results. To achieve this advanced level of analysis, language models implementing rule-based and probabilistic approaches may be implemented. Packages exist to facilitate this, such as the VADER (Valence Aware Dictionary and sEntiment Reasoner) package in Python [61]. This would also allow for contextual & topic analysis to be implemented, to extract concern themes over the course of a wildfire event. Additionally, entity extraction would allow for the tracking of named entities and their corresponding topics in relation to wildfires. This would also allow for a much more complex information extraction process from the tweets, which is key for a fully integrated model.

The findings in "Test cases and results" act as proof of concept for the sociophysical model proposed in "Sentimental wildfire: overview" and shown in Fig. 1. This paper shows that social sentiment and wildfire activity are linked, a relationship which will be explored further in subsequent work. We can expand on this methodology by investigating both temporal and spatial online social dynamics over the course of a wildfire. This will allow the development of a real-time social sentimental wildfire system, using linguistic analysis of social media to predict areas of interest to disaster management teams and danger to the public. This system could

be incorporated to create more socially conscious, real-time wildfire models which provide actionable insights for disaster management teams.

Appendix A GBRF Models

Random Forests are an ensemble method extension of Classification and Regression Trees (CART) [62]. Figure 13 shows the architecture of the GBRF model. For the training phase, these methods start with a simple model, often a single tree (or weak learner), and then additive training occurs where trees are generated and added to build up a forest of trees which is used for the final prediction. For new data, predictions are then made by averaging the majority vote of all trees in the forest. The GBRFs are built and evaluated using the MAE, and the Gini coefficient [38]. The Gini coefficient is a scoring metric which is a measure of the degree to which a particular element is wrongly classified when randomly chosen and it is expressed by the formula:

$$\sum_{i=1}^{n} p_i(1-p_i) \tag{A1}$$

where p_i denotes that the physical event i happens (or that the sentiment $i \in S$ is evaluated) and n is the number of possible events (or possible sentiments). The Gini coefficient is used in the tree-building stage to chose which variables to split on in the decision tree models [38]. When splitting at a node in the decision tree, each variable is considered by calculating the new Gini coefficient if the node is split on this variable. The variable which gives the greatest Gini gain is the optimal variable to perform the node split on. Gradient boosting RF algorithms are particular RFs which begin with a base (weak) learner (tree), and consecutively add more weak learners to the ensemble with the goal of minimising a loss function [63]. Given a training sample $\{\mathbf{x}_i, \mathbf{y}_i\}_1^n$ (or $\{\mathbf{y}_i, \mathbf{x}_i\}_1^n$), the goal is to find function of $F'(\mathbf{x})$ (or $G'(\mathbf{y})$) such that the expected value of the loss function $\psi(\mathbf{y}, F(\mathbf{x}))$ (or $\phi(\mathbf{x}, G(\mathbf{y}))$) is minimised [38]. The gradient boosting problem is then expressed as:

$$F'(\mathbf{x}) = \arg\min_{F(\mathbf{x})} \mathbb{E}_{\mathbf{x},\mathbf{y}} \psi(\mathbf{y}, F(\mathbf{x})) \tag{A2}$$

or as:

$$G'(\mathbf{y}) = \arg\min_{G(\mathbf{y})} \mathbb{E}_{\mathbf{y},\mathbf{x}} \phi(\mathbf{x}, G(\mathbf{y})) \tag{A3}$$

where $\mathbb{E}$ denotes the expected value. The boosting algorithm then approximates $F'(\mathbf{x})$ (or $G'(\mathbf{y})$) by additive expansion, which can be summarised as:

$$F(\mathbf{x}) = \sum_{m=0}^{s} \beta_m h(\mathbf{x}, a_m) \tag{A4}$$

or

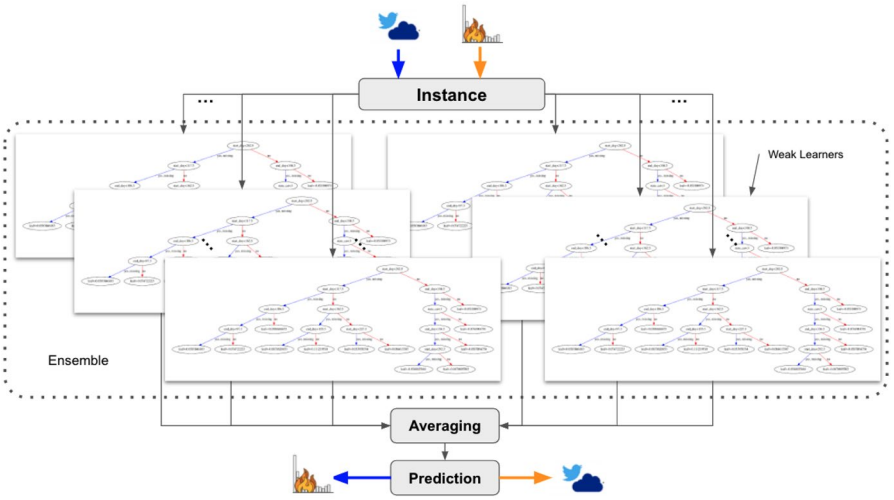


Fig. 13 General diagram of the GBRF architecture employed in this work for training the function G (blues arrows) and F (orange arrows)

$$G(\mathbf{y}) = \sum_{m=0}^{s} \alpha_m k(\mathbf{y}, b_m) \tag{A5}$$

where the functions $h(\mathbf{x}, a)$ (or $k(\mathbf{y}, b)$) are known as the 'weak learners', represented by the trees in Fig. 13.

Over a series of s steps, the weak learners are sequentially added, and the expansion coefficients $a = \{a_1, a_2, ...\}$ (or $b = \{b_1, b_2, ...\}$) and $\{\beta\}_0^s$ (or $\{\alpha\}_0^s$) are jointly fit to the current models pseudo-residuals. For $m = 1, 2, \ldots, s$, this gives

$$(\beta_m, a_m) = \arg\min_{\beta,a} \sum_{i=1}^{n} \psi(\mathbf{y}_i, F_{m-1}(\mathbf{x}_i) + \beta h(\mathbf{x}_i; a)) \tag{A6}$$

or

$$(\alpha_m, b_m) = \arg\min_{\alpha,b} \sum_{i=1}^{n} \phi(\mathbf{x}_i, G_{m-1}(\mathbf{y}_i) + \alpha k(\mathbf{y}_i; b)), \tag{A7}$$

and

$$F_m(\mathbf{x}) = F_{m-1}(\mathbf{x}) + \beta_m h(\mathbf{x}; a_m) \tag{A8}$$

or

$$G_m(\mathbf{y}) = G_{m-1}(\mathbf{y}) + \alpha_m k(\mathbf{y}; b_m) \tag{A9}$$

which shows the step-wise optimisation of F (or G). Equations (A6) and (A7) are then optimised to solve the loss function ψ and ϕ by fitting least squares on $h(\mathbf{x}, a)$

and $k(\mathbf{y}, b)$ respectively. This replaces the optimisation problem presented in Eqs. (A2) and (A3) with one based on reducing least squares in Eqs. (A6) and (A7).

These models were trained on a wide grid search covering the hyper-parameters denoted by:

- ETA (Alias—learning rate): Step size shrinkage used in update to prevents overfitting [64].
- Max depth: Maximum depth of a tree.
- Min child weight: Minimum sum of instance weight (hessian) needed in a child.
- Learning rate: Learning rate schedule for weight updates
- Subsample: Subsample ratio of the training instances.
- Col sample by tree: Subsample ratio of columns when constructing each tree.
- N estimators: Number of trees in the ensemble.

Appendix B NN methods

A Multi-layer Perceptron Regressor (MLP) is feed-forward artificial NN model, which maps a set of input MLP consists of multiple layers of nodes in a directed graph, with each layer fully connected to the next one and at least one hidden layer [56]. The architecture of this type of network is shown in Fig. 14.

Each neuron in this network includes a nonlinear activation function that is differentiable, and this network is trained via a method called back propagation, in two stages [37]: In the forward phase, the input signal is propagated through the network until it reaches the output layer and the network weights remain fixed. Then in the backward phase, an error signal is produced by comparing the network output with the true target. The error signal is propagated backwards through each layer of the network and successive adjustments of the weights between neurons are adjusted to reduce the error signal with each round of back propagation [56].

The hyper-parameters optimised during the training of this model are denoted as:

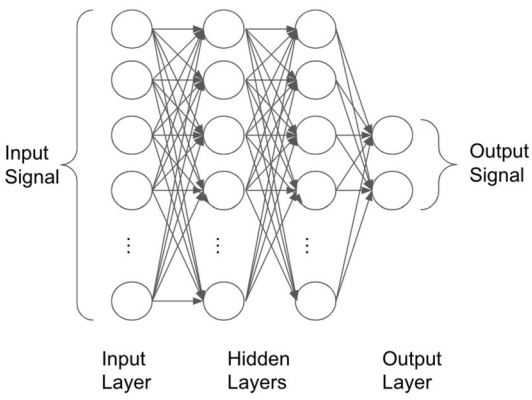


Fig. 14 Architectural graph of a multi-layer perceptron with two hidden layers

- Solver: Solver method for weight optimization.
- Activation: Activation function for the hidden layer.
- Alpha α: L2 penalty (regularization term) parameter.
- Learning Rate: Learning rate schedule for weight updates.
- Hidden Layer Sizes: Network architecture.

Appendix C SVR methods

An SVR method is also implemented [38]. SVRs are supervised learning method which employ the concept of the support vector machine [65]. The training of these models is achieved by plotting all of the N training data points in an d dimensional space, where d is the dimensionality of the training data. The algorithm then aims to find the 'best fit line' - a (d dimensional) regression hyperplane in this space which best models the N training data points.

Figure 15 shows the parameters for the SVR. The dotted lines are known as the 'soft margins' and form an 'ϵ-insensitive' region or ϵ-tube, which allow for more leniency in the errors. This region is bounded by the 'support vectors', and these are chosen to minimise the errors on the training data. Points classified outside of the margin have errors associated to them by assigning slack variables, proportional to the distance from the ϵ-insensitive region/tube shown in Fig. 15. Once the hyperplane is generated in the d dimensional space, kernel methods are used to project unseen data into this space for prediction.

We implement a grid search on the hyper-parameters [66] denoted by:

- Kernel: kernel type to be used in the algorithm.
- Degree: Degree of the polynomial kernel function.
- Gamma γ: Kernel coefficient for 'rbf', 'poly' and 'sigmoid'.
- Tolerance (tol): Tolerance for stopping criterion.
- Epsilon ϵ: Epsilon in the epsilon-SVR model.
- C: Regularization parameter.

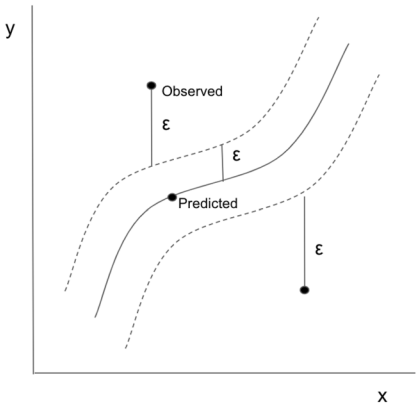


Fig. 15 Included parameters of the support vector regression

Funding This research is funded by the Leverhulme Centre for Wildfires and Imperial College London Data Science Institute.

Availability of data and materials Data from this project is available by request to the author.

Code availability The code used in this project is avaiable in Github at https://github.com/jakelever4/FSA.

Declaration

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval Data collected as part of this research is publicly available, anonymous and authorised by Twitter's Terms of Service agreement.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

1. World Health Organisation: Wildfires. (2021). https://www.who.int/health-topics/wildfires/.
2. Weise, D. R., & Biging, G. S. (1997). A qualitative comparison of fire spread models incorporating wind and slope effects. *Forest Science, 43*(2), 170–180.
3. Silvani, X., Morandini, F., & Dupuy, J.-L. (2012). Effects of slope on fire spread observed through video images and multiple-point thermal measurements. *Experimental Thermal and Fluid Science, 41*, 99–111.
4. Just, M. G., Hohmann, M. G., & Hoffmann, W. A. (2016). Where fire stops: vegetation structure and microclimate influence fire spread along an ecotonal gradient. *Plant Ecology, 217*(6), 631–644.
5. Owen, G., McLeod, J. D., Kolden, C. A., Ferguson, D. B., & Brown, T. J. (2012). Wildfire management and forecasting fire potential: the roles of climate information and social networks in the southwest united states. *Weather, Climate, and Society, 4*(2), 90–102.
6. Rodriguez-Aseretto, D., De Rigo, D., Di Leo, M., Cortés, A., & San-Miguel-Ayanz, J. (2013). A data-driven model for large wildfire behaviour prediction in Europe. *Procedia Computer Science, 18,* 1861–1870.
7. Aniedu, A., Chukwuneke, C., & Asogwa, D. S. N. (2016). Real-time wildfire monitoring and alert system using GSM technology. *IOSR Journal of Mobile Computing & Application (IOSR-JMCA), 3,* 1–10. https://doi.org/10.9790/0050-03040110.
8. Du, H., Nguyen, L., Yang, Z., Abu-Gellban, H., Zhou, X., Xing, W., Cao, G., & Jin, F. (2019). Twitter vs news: concern analysis of the 2018 California wildfire event. In: 2019 IEEE 43rd Annual Computer Software and Applications Conference (COMPSAC), vol. 2, pp. 207–212 . https://doi.org/10.1109/COMPSAC.2019.10208.
9. Govil, K., Welch, M. L., Ball, J. T., & Pennypacker, C. R. (2020). Preliminary results from a wildfire detection system using deep learning on remote camera images. *Remote Sensing*. https://doi.org/10.3390/rs12010166.
10. Esteban Ortiz-Ospina: The rise of social media. (2019). https://ourworldindata.org/rise-of-social-media.

11. Yao, F., & Wang, Y. (2020). Domain-specific sentiment analysis for tweets during hurricanes (dssa-h): A domain-adversarial neural-network-based approach. *Computers, Environment and Urban Systems, 83*, 101522. https://doi.org/10.1016/j.compenvurbsys.2020.101522
12. Ahmed, W. (2019) Using Twitter as a data source: an overview of social media research tools (2019). https://blogs.lse.ac.uk/impactofsocialsciences/2019/06/18/using-twitter-as-a-data-source-an-overview-of-social-media-research-tools-2019/.
13. Alkouz, B., Aghbari, Z. A., & Abawajy, J. H. (2019). Tweetluenza: Predicting flu trends from twitter data. *Big Data Mining and Analytics, 2*(4), 273–287. https://doi.org/10.26599/BDMA.2019.9020012
14. Gallagher, R. J., Reagan, A. J., Danforth, C. M., & Dodds, P. S. (2018). Divergent discourse between protests and counter-protests: #blacklivesmatter and #alllivesmatter. *PLoS One, 13,* 1–23. https://doi.org/10.1371/journal.pone.0195644.
15. Boulton, C., Shotton, H., & Williams, H.T.P. (2016). Using social media to detect and locate wildfires. In: EcoMo@ICWSM.
16. Koltunov, A., Ustin, S. L., & Prins, E. M. (2012). On timeliness and accuracy of wildfire detection by the goes wf-abba algorithm over California during the 2006 fire season. *Remote Sensing of Environment, 127,* 194–209. https://doi.org/10.1016/j.rse.2012.09.001.
17. Kryvasheyeu, Y., Chen, H., Obradovich, N., Moro, E., Van Hentenryck, P., Fowler, J., & Cebrian, M. (2016). Rapid assessment of disaster damage using social media activity. Science Advances **2**(3) https://advances.sciencemag.org/content/2/3/e1500779.full.pdf. https://doi.org/10.1126/sciadv.1500779.
18. Feldman, R. (2013). Techniques and applications for sentiment analysis. *Communications of the ACM, 56*(4), 82–89.
19. Andela, N., Morton, D. C., Giglio, L., Paugam, R., Chen, Y., Hantson, S., van der Werf, G. R., & Randerson, J. T. (2019). The global fire atlas of individual fire size, duration, speed and direction. *Earth System Science Data, 11*(2), 529–552. https://doi.org/10.5194/essd-11-529-2019
20. Compton, R., Lee, C., Lu, T.-C., De Silva, L., & Macy, M. (2013). Detecting future social unrest in unprocessed twitter data: "emerging phenomena and big data". In: 2013 IEEE International Conference on Intelligence and Security Informatics, pp. 56–60 . https://doi.org/10.1109/ISI.2013.6578786.
21. Hendrickx, R. (2020). Correcting public opinion trends through machine learning and data assimilation.
22. Reece, A. G., Reagan, A. J., Lix, K. L. M., Dodds, P. S., Danforth, C. M., & Langer, E. J. (2017). Forecasting the onset and course of mental illness with twitter data. *Scientific Reports, 7*(1), 13006. https://doi.org/10.1038/s41598-017-12961-9
23. Yang, Z., Nguyen, L.H., & Jin, F. (2019). Opioid relapse prediction with gan. In: 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), pp. 560–567 . https://doi.org/10.1145/3341161.3342951.
24. Young, C., Kuligowski, E., & Pradhan, A. (2020). A review of social media use during disaster response and recovery phases. Technical Note (NIST TN). National Institute of Standards and Technology. https://doi.org/10.6028/NIST.TN.2086.
25. Mustaqim, T., Umam, K., & Muslim, M. A. (2020). Twitter text mining for sentiment analysis on government's response to forest fires with Vader lexicon polarity detection and k-nearest neighbor algorithm. *Journal of Physics: Conference Series, 1567,* 032024. https://doi.org/10.1088/1742-6596/1567/3/032024.
26. Wang, Y., & Taylor, J. E. (2018). Coupling sentiment and human mobility in natural disasters: a twitter-based study of the 2014 south Napa earthquake. *Natural Hazards, 92*(2), 907–925. https://doi.org/10.1007/s11069-018-3231-1.
27. Phengsuwan, J., Shah, T., Thekkummal, N. B., Wen, Z., Sun, R., Pullarkatt, D., Thirugnanam, H., Ramesh, M. V., Morgan, G., James, P., et al. (2021). Use of social media data in disaster management: a survey. *Future Internet, 13*(2), 46.
28. Mavrodieva, A.V., & Shaw, R. (2021). In: Shaw, R., Kakuchi, S., Yamaji, M. (eds.) Social Media in Disaster Management, pp. 55–73. Springer. https://doi.org/10.1007/978-981-16-0285-6_4.
29. Imran, M., Castillo, C., Lucas, J., Meier, P., & Vieweg, S. (2014). Aidr: Artificial intelligence for disaster response. In: Proceedings of the 23rd International Conference on World Wide Web. WWW '14 Companion, pp. 159–162. Association for Computing Machinery, New York, NY, USA . https://doi.org/10.1145/2567948.2577034.
30. Commission, E. (2022). Europe Media Monitor—Top Stories . https://emm.newsbrief.eu/NewsBrief/clusteredition/en/latest.html.

31. Jain, P.K., Saravanan, V., & Pamula, R. (2021). A hybrid CNN-lstm: A deep learning approach for consumer sentiment analysis using qualitative user-generated contents **20**(5) . https://doi.org/10.1145/3457206.
32. Employing bert-dcnn with sentic knowledge base for social media sentiment analysis. Journal of Ambient Intelligence and Humanized Computing (2022). https://doi.org/10.1007/s12652-022-03698-z.
33. Beigi, G., Hu, X., Maciejewski, R., & Liu, H. (2016). An overview of sentiment analysis in social media and its applications in disaster relief. Springer. https://doi.org/10.1007/978-3-319-30319-2_13$.
34. Friedman, J.H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of Statistics*, 1189–1232.
35. XGBoost Developers: DART Booster. (2020). https://xgboost.readthedocs.io/en/latest/tutorials/dart.html.
36. Cui, H., Huang, D., Fang, Y., Liu, L., & Huang, C. (2018). Webshell detection based on random forest–gradient boosting decision tree algorithm. In: 2018 IEEE Third International Conference on Data Science in Cyberspace (DSC), pp. 153–160 . IEEE.
37. Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.
38. Bishop, C. M. (2006). *Pattern recognition and machine learning*. Berlin: Springer.
39. Twitter: Most users by country. Published by Statista Research Department, 19 Nov (2021). https://www.statista.com/statistics/242606/number-of-active-twitter-users-in-selected-countries/.
40. Godfrey, D., Johns, C., Meyer, C., Race, S., & Sadek, C. (2014). A case study in text mining: interpreting twitter data from world cup tweets.
41. Chong, W.Y., Selvaretnam, B., Soon, & L.-K. (2014). Natural language processing for sentiment analysis: an exploratory analysis on tweets. In: 2014 4th International Conference on Artificial Intelligence with Applications in Engineering and Technology, pp. 212–217 . IEEE.
42. Nasukawa, T., & Yi, J. (2003). Sentiment analysis: Capturing favorability using natural language processing. In: Proceedings of the 2nd International Conference on Knowledge Capture, pp. 70–77.
43. Yi, J., Nasukawa, T., Bunescu, R., & Niblack, W. (2003). Sentiment analyzer: Extracting sentiments about a given topic using natural language processing techniques. In: Third IEEE International Conference on Data Mining, pp. 427–434 . IEEE.
44. Google sentiment analysis documentation: analyzing sentiment (2021). https://cloud.google.com/natural-language/docs/analyzing-sentiment.
45. Google Sentiment Analysis Documentation: Interpreting sentiment analysis values (2021). https://cloud.google.com/natural-language/docs/basics#interpreting_sentiment_analysis_values.
46. Beneduce, R., Hill, R., & Schelle, C. Alert Wildfire (Group 6). https://www.alertwildfire.org/.
47. Center for International Earth Science Information Network—CIESIN—Columbia University: Gridded population of the world, version 4 (gpwv4): Population density adjusted to match 2015 revision un WPP country totals, revision 11 (2018).
48. Google: Geocoding Overview. (2021). https://developers.google.com/maps/documentation/geocoding/overview.
49. Google: Geocoding Service Documentation (2021). https://developers.google.com/maps/documentation/javascript/geocoding.
50. Developers, T. Twitter API V2. https://developer.twitter.com/en/docs/twitter-api.
51. Twitter: Building queries for Search Tweets. (2021). https://developer.twitter.com/en/docs/twitter-api/tweets/search/integrate/build-a-query.
52. Schlosser, S., Toninelli, D., & Cameletti, M. (2021). Comparing methods to collect and geolocate tweets in great Britain. *Journal of Open Innovation: Technology, Market, and Complexity, 7*(1), 44.
53. Dhaoui, C., Webster, C., & Tan, L. (2017). Social media sentiment analysis: lexicon versus machine learning. *Journal of Consumer Marketing, 34*, 00–00. https://doi.org/10.1108/JCM-03-2017-2141
54. Kloumann, I. M., Danforth, C. M., Harris, K. D., Bliss, C. A., & Dodds, P. S. (2012). Positivity of the English language. *PLoS One, 7*(1), 1–7. https://doi.org/10.1371/journal.pone.0029484.
55. Annual Australian climate statement 2016. Australian Government—Bureau of Meteorology. http://www.bom.gov.au/climate/current/annual/aus/2016/#tabs=Events.
56. Elogeel's, A. (2010). Multilayer Perceptron. https://elogeel.wordpress.com/2010/05/05/multilayer-perceptron/.
57. Denisko, D., & Hoffman, M.M. (2018). Classification and interaction in random forests. Proceedings of the National Academy of Sciences 115(8), 1690–1692 https://www.pnas.org/content/115/8/1690.full.pdf. https://doi.org/10.1073/pnas.1800256115.
58. Carvalho, D. V., Pereira, E. M., & Cardoso, J. S. (2019). Machine learning interpretability: A survey on methods and metrics. *Electronics*. https://doi.org/10.3390/electronics8080832.

59. Dehghani, M., Severyn, A., Rothe, S., & Kamps, J. (2017). Avoiding your teacher's mistakes: Training neural networks with controlled weak supervision. CoRR **abs/1711.00313**arXiv:1711.00313**.**
60. Mitchell, L., Frank, M. R., Harris, K. D., Dodds, P. S., & Danforth, C. M. (2013). The geography of happiness: Connecting twitter sentiment and expression, demographics, and objective characteristics of place. *PloS One, 8*(5), 64417.
61. Hutto, C., & Gilbert, E. (2014). Vader: A parsimonious rule-based model for sentiment analysis of social media text. In: Proceedings of the International AAAI Conference on Web and Social Media, vol. 8, pp. 216–225.
62. Denison, D. G., Mallick, B. K., & Smith, A. F. (1998). A Bayesian cart algorithm. *Biometrika, 85*(2), 363–377.
63. Si, S., Zhang, H., Keerthi, S.S., Mahajan, D., Dhillon, I.S., & Hsieh, C.-J. (2017). Gradient boosted decision trees for high dimensional sparse output. PMLR . https://proceedings.mlr.press/v70/si17a.html.
64. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. Sklearn.neuralNetwork.Mlpregressor. https://scikit-learn.org/stable/modules/generated/sklearn.neural_network.MLPRegressor.html.
65. Awad, M., & Khanna, R. (2015). Support Vector Regression, pp. 67–80. Apress. https://doi.org/10.1007/978-1-4302-5990-9_4.
66. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. sklearn.svm.SVR Documentation. https://scikit-learn.org/stable/modules/generated/sklearn.neural_network.MLPRegressor.html.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Jake Lever[1,2,3] · Rossella Arcucci[1,2,3]

[1] Department of Earth Science and Engineering, Imperial College London, London, UK

[2] Data Science Institute, Imperial College London, London, UK

[3] Leverhulme Centre for Wildfires, Environment and Society, Imperial College London, London, UK