applied sciences

MDPI

Article

Do Spatial Trajectories of Social Media Users Imply the Credibility of the Users' Tweets During Earthquake Crisis Management?

Ayse Giz Gulnerman

Land Registry and Cadastre Department, Ankara Hacı Bayram Veli University, 06500 Ankara, Türkiye; ayse.gulnerman@hbv.edu.tr

Abstract: Earthquakes are sudden-onset disasters requiring rapid, accurate information for effective crisis response. Social media (SM) platforms provide abundant geospatial data but are often unstructured and produced by diverse users, posing challenges in filtering relevant content. Traditional content filtering methods rely on natural language processing (NLP), which underperforms with mixed-language posts or less widely spoken languages. Moreover, these approaches often neglect the spatial proximity of users to the event, a crucial factor in determining relevance during disasters. This study proposes an NLP-free model that assesses the spatial credibility of SM content by analysing users' spatial trajectories. Using earthquake-related tweets, we developed a machine learning-based classification model that categorises posts as directly relevant, indirectly relevant, or irrelevant. The Random Forest model achieved the highest overall classification accuracy of 89%, while the k-NN model performed best for detecting directly relevant content, with an accuracy of 63%. Although promising overall, the classification accuracy for the directly relevant category indicates room for improvement. Our findings highlight the value of spatial analysis in enhancing the reliability of SM data (SMD) during crisis events. By bypassing textual analysis, this framework supports relevance classification based solely on geospatial behaviour, offering a novel method for evaluating content trustworthiness. This spatial approach can complement existing crisis informatics tools and be extended to other disaster types and event-based applications.

Keywords: crisis informatics; spatial trajectory analysis; social media credibility; geospatial data mining; machine learning classification; earthquake response

check for updates

Academic Editor: Jose María Alvarez Rodríguez

Received: 9 May 2025
Revised: 5 June 2025
Accepted: 17 June 2025
Published: 18 June 2025

Citation: Gulnerman, A.G. Do Spatial Trajectories of Social Media Users Imply the Credibility of the Users' Tweets During Earthquake Crisis Management? *Appl. Sci.* **2025**, *15*, 6897. https://doi.org/10.3390/app15126897

Copyright: © 2025 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Earthquakes are severe disasters affecting wide areas that result in multidimensional hazards and losses in the short and long run. On-time management is the key to reducing the secondary impacts of these kinds of disasters on living beings, the environment, and the economy [1–4]. Geospatial data science immensely contributes to disaster management with various data sources (earth observatory satellites, aerial vehicles, and geospatial crowd sourcing) and information extraction algorithms [5,6]. SM became an actor of geospatial data sources since enabling location information features. This actor gives us the ability to see different dimensions of events apart from observed from the space, e.g., emergencies inside buildings, the status of living beings, and the sentiments [7,8].

There is a growing amount of research using SMD locations to identify urban dynamics [9–11], sentimental reactions [12,13], and changes in human mobility [14–16]. Most of these studies are largely concentrated on social media data intensity to answer coarse

grained questions, such as "Where are the most visited locations?", "How does sentiment sprawl after a major disaster?", and "When does human mobility vary?". What is not clear yet are the answers to the fine-grained questions on social media data, such as "Where are the victims' locations?", "What is the consistency between the content and the attached location for each post after an incident?", and "When did the aforementioned event in the post happen?". Current methods primarily focus on problems that can be addressed using density-based spatial algorithms or text frequency-based approaches [9–16]. As a result, these studies inherently rely on the clustering of multiple reports, which provides a form of internal validation. However, such methodologies tend to overlook individual emergency notifications and may also incorporate clusters formed as a result of misinformation. Neglecting individual emergency messages can lead to a scenario where only widely shared alerts gain visibility, while genuine victims in urgent need, who may have been able to send out only a single message, are left without assistance. Furthermore, the spread of misinformation can result in a high concentration of false reports in certain locations, leading to the misallocation of resources to areas where no real need exists. Therefore, social media data should be analysed at the individual level and supported with user-based credibility assessments. In this approach, each user is treated as an independent source of information, and the reliability of this source should be thoroughly examined.

There are also several studies carried out on SMD that contribute to disaster management in various cases [17]. Huang and Xiao (2015) [18] investigated Hurricane Sandy Twitter data by categorising them into topics regarding disaster management phases. Resch et al. (2018) [19] proposed a combined technique with semantic machine learning and spatiotemporal analysis for natural disaster monitoring and examined Napa Earthquake tweets. Most of these disaster domain studies are well documented and present significant results for coarse-grained problematics with a high intensity of data. However, these studies [17,18] do not assess the overall credibility of tweets; instead, they focus on classifying the relevance of content based on natural language processing, which may include second-hand information or misinformation. Additionally, mapping results using spatial clustering techniques based on high data intensity fails to address the extraction of individual earthquake-related tweets and, consequently, overlooks issues related to their reliability or credibility.

Yet, the credibility of SMD producers is a key issue in answering the fine-grained questions. So far, very little research has been carried out on social media data producers' credibility during events. Hecht et al. (2011) [20] was one of the first to examine spatial credibility and found that 34% of data producers on SM do not provide their correct location. To support this finding, the study compared user profiles with the location information mentioned in tweet content. Users' actual locations were determined based on their attendance at a concert event. Since the event was confined to a specific concert venue, the dataset was suitable for investigating the accuracy of user-reported locations. The study aimed to draw general conclusions about users' location-sharing behaviours and potential location manipulation. However, in large-scale disasters covering wide areas, it becomes challenging to verify the real-time accuracy of users' locations. Gupta and Kumaraguru (2012) [21] reported that solely 17% of gathered tweets provide credible situational awareness information. An algorithm utilising RankSVM and a relevance feedback approach was evaluated to rank tweets from 14 high-impact events based on the perceived credibility of the information in the tweet text (categorised as definitely credible, seems credible, definitely not credible, or undecided). This ranking algorithm focuses solely on the textual content and does not incorporate any spatial dimension of the tweets. Truelove et al. (2017) [22] presented a testing witnessing status of SM users based on evidence they share on their profiles. They present a framework for identifying

the potential social media users' witnessing to an event based on evidence (such as text and image content) in their SM accounts. Abbasi et al. (2013) [23] proposed the "credRank" algorithm to identify coordinated users (accounts acting together to dominate the results of something) based on the users' profile information such as tweeting timestamps and voting behaviour. Yang et al. (2019) [24] proposed a framework for SM data credibility during disasters and tested it with Hurricane Harvey Twitter data. The research is based on the calculation of event credibility with semantic and spatiotemporal aggregation. The main focus of the study is extracting location information from text and localising tweet content. The credibility rate of an event was calculated based on the supporting tweets and retweets counts. Neither of these studies are in the same domain, nor are the methods similar, yet all aim to reach credible first-hand information about an event from SM. Credibility is not only associated with text content but also with spatial credibility. Since textual content can be easily manipulated by social media users, the spatial dimension may, in the long run, offer a more reliable relevance signal for an event. Although spatial information can also be manipulated, it is less likely to be altered consistently and continuously. Therefore, our hypothesis in this study is that user trajectory data can provide meaningful relevance information for event verification. The motivation behind this study is to scrutinise second-hand or misleading information shared during earthquakes to distinguish between those who are physically present in the affected area, potential victims or eyewitnesses, and those who are not. To date, geo-social media (geo-SM) data producers' credibility determination has still not been extensively studied.

This study set out to identify the predictors for geo-SM users' credibility based on spatial trajectory data. Trajectory data mining was extensively reviewed by Mazimpaka and Timpf (2016) [25] and Zheng (2015) [26]. The pre-defined data processing steps vary based on the trajectory data sources, which are broadly classified as GPS, GSM, geo-SM, Wi-Fi, and RFID [26–29]. Trajectory data change with regard to the frequency of space and time based on these data sources' pre-defined or uncertain settings. There are also different types of moving objects (such as animals, vehicles, human, etc.) that shape the study methodology on the trajectory data mining.

Geo-SM is an unsystematic trajectory data source with its content at uncertain spatiotemporal granularity. Andrienko et al. (2013) [30] displayed a framework for the classification of trajectory data mining methods over georeferenced Twitter data. Chae et al. (2015) [31] demonstrated abnormal topics and movement patterns on Twitter during disasters by adopting Twitter data before, during, and after a disaster. Wang and Taylor (2018) [32] evaluate sentiment and human mobility together to understand the urban dynamics changes during earthquakes. The research illustrates the correlation between earthquake intensity and sentiment scores. Hübl et al. (2017) [33] adopt geo-tagged tweets for extracting refugee migration patterns with the V-Analytics Toolkit and visualise the pattern of migration from the Middle East to Europe. While determining the actual trajectory pertaining to refugees, they used both trajectory mining techniques combined with text content (e.g., language mix of Arabic and English and filtered refugee-related words). Ma et al. (2020) [34] structured the human activity scale of cities based on hot spots and street networks and discovered the changing activities before and after the Tokyo Earthquake. Kang et al. (2020) [35] worked to determine changing human mobility during COVID-19 through mobile phone trajectories used as the origin destination concept. The study infers dynamic population flows and compares the estimated results for each origin–destination pair with the gravity and radiation model.

Our study evaluates the trajectories of SMD producers to determine the relevance and event-related credibility of their content in relation to an earthquake. While previous studies on SMD content relevance have predominantly relied on fully or partially text-based

techniques [24,36], trajectory-based relevance assessment has not been extensively explored. Our research addresses this gap by extracting trajectory-based features and employing various machine learning techniques to assess the relevance and credibility of content in a language-independent manner.

The paper is structured as the following three sections: Section 2 presents the data and evaluation methodology, including data retrieval and filtering, data cleaning and tidying, data labelling, and processing parts. Section 3 illustrates the results of the processing steps and discusses the results and lists the limitations of the study. Section 4 concludes the research with the contribution to the scientific field and presents reproducible cases for its broader uses.

2. Materials and Methods

2.1. Data Retrieval and Filtering

Twitter (rebranded as X in July 2023, having had previously restricted free API access in February) provided a data stream for academic use through its API. This API returned a randomly sampled 1% of all data produced on the Twitter platform [37]. Using this methodology, we continuously harvested Twitter data for several years and inserted it into a PostgreSQL database. In order to retrieve the dataset for this study, we first queried earthquakes with a magnitude greater than 5 that occurred in Türkiye between the years 2017 and 2021. According to the Richter Scale, a magnitude >5 refers to moderate, strong, major, or great earthquakes [38]. This means that everyone felt the earthquake when it was of moderate scale or higher. Based on the Bogazici University Kandilli Observatory and Earthquake Research Institute-Regional Earthquake-Tsunami Monitoring Center (KOERI-RETMC) [39] database, there were 28 earthquakes (Table 1).

Table 1. Earthquakes with a magnitude of 5 or greater between 2017 and 2021 in Türkiye.

ID	Place Name	Date/Time	Latitude	Longitude	Magnitude	Tweet Count	Account Count
1	Ayvacik Aciklari-EGE DENIZI	6 February 2017 10:58:00	39.5014	26.0621	5.1	38	33
2	Ayvacik-Canakkale	7 February 2017 02:24:03	39.5164	26.0997	5.3	25	22
3	Ayvacik-Canakkale	12 February 2017 13:48:16	39.5048	26.1076	5	22	18
4	Samsat-Adiyaman	2 March 2017 11:07:27	37.5961	38.4724	5.2	27	26
5	Ula-Mugla	13 April 2017 16:22:16	37.1397	28.684	5	15	15
6	Saruhanli-Manisa	28 May 2017 11:04:57	38.7182	27.7715	5	23	23
7	Milas-Mugla	14 August 2017 02:43:47	37.1881	27.6585	5	13	13
8	Sehitkamil-Gaziantep	12 November 2017 18:19:58	37.1918	37.52	7.1	77	59
9	Koycegiz-Mugla	24 November 2017 21:49:14	37.0935	28.5999	6	29	29
10	Sivrice-Elazig	4 April 2019 17:31:11	38.3728	39.1712	5	60	58

Table 1. *Cont.*

ID	Place Name	Date/Time	Latitude	Longitude	Magnitude	Tweet Count	Account Count
11	Dazkiri-Afyonkarahisar	8 August 2019 11:25:30	37.9115	29.675	5.7	303	294
12	Cerkes-Cankiri	14 September 2019 06:03:11	40.7705	32.9597	5	23	21
13	Marmara Denizi (Orta)	26 September 2019 10:59:25	40.8678	28.164	5.7	350	334
14	Doganyol-Malatya	27 December 2019 07:02:28	38.2564	38.9967	5	5	5
15	Kirkagac-Manisa	22 January 2020 19:22:15	39.0656	27.8261	6	1870	932
16	Sivrice-Elazig	24 January 2020 17:55:16	38.3367	39.2637	6.4	767	675
17	Cihanbeyli-Konya	24 January 2020 17:56:25	38.3956	32.7884	5.3	767	675
18	Akhisar-Manisa	4 February 2020 17:55:23	39.0006	27.8441	5	42	37
19	Akhisar-Manisa	18 February 2020 16:09:22	39.0377	27.7558	5.2	22	21
20	Karliova-Bingol	14 June 2020 14:24:27	39.3081	40.8209	5.9	25	25
21	Karliova-Bingol	15 June 2020 06:51:31	39.3971	40.7076	5.4	12	12
22	Karayazi-Erzurum	16 June 2020 01:34:54	39.7849	42.0531	5.4	12	12
23	Dodecanese Islands	30 October 2020 11:51:26	37.8875	26.834	6.7	58	44
24	Puturge-Malatya	27 November 2020 08:27:55	38.2202	38.6871	5.2	4	3
25	Kurtalan-Siirt	3 December 2020 05:45:19	37.9463	41.6801	5.2	20	20
26	Antalya Korfezi-AKDENIZ	5 December 2020 12:44:40	35.9971	31.8088	5.5	19	19
27	Sivrice-Elazig	27 December 2020 06:37:34	38.4714	39.2593	5.3	28	28
28	Yayladere-Bingol	25 June 2021 18:28:38	39.2055	40.2233	5.4	14	14
					Total:	4670	3467

After an earthquake, Twitter data are generally queried using keywords similar to "earthquake". This method was replicated to determine our initial raw dataset for the 28 earthquakes listed in Table 1. While filtering the raw Twitter dataset for each earthquake in this study, three conditions were considered: the time interval, presence of the "earthquake" (e.g., "deprem", "sarsıntı", and "zelzele" in Turkish) keyword, and spatial boundary. The time interval was the first condition, starting with the time the earthquake occurred and ending 72 h afterward. The second condition was the presence of the word "earthquake" in the tweet content. The last condition was that the tweets had to be located within Türkiye's spatial boundary. When these three conditions were met for a tweet, the tweet was inserted into the earthquake raw dataset (Figure 1).

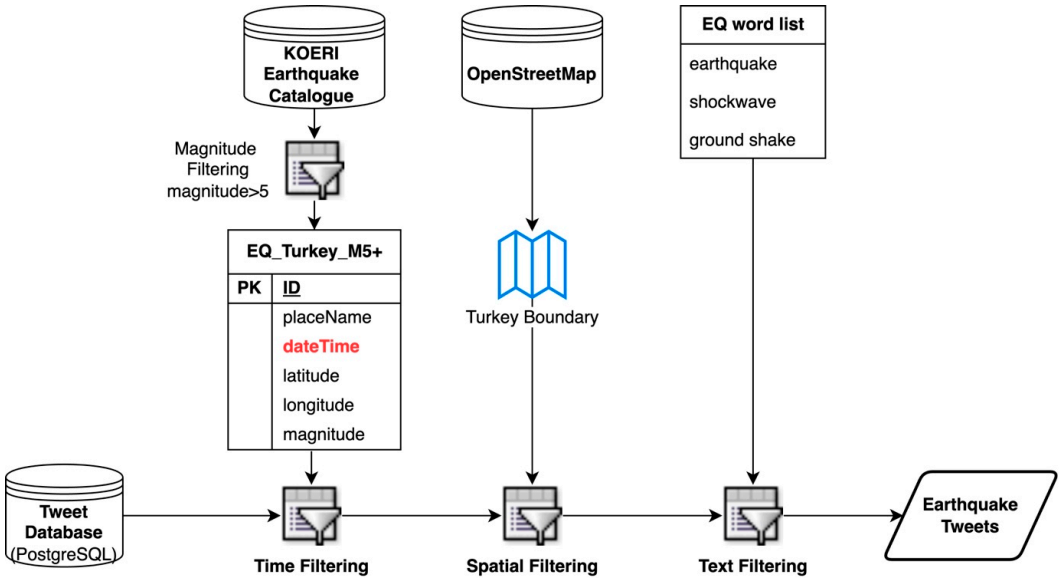


Figure 1. Earthquake tweet data retrieval flow.

2.2. Data Cleaning and Tidying

The retrieved earthquake tweet data are biased due to bot accounts that systematically report every earthquake worldwide, as well as tweets from newsfeed accounts. To mitigate this, we applied text-based filtering to tweets and accounts to exclude tweets from bot accounts in the dataset. We designed the data extraction process in three steps:

1. A word list of technical terms (e.g., "mww", "mwr", "mw", and "ml") was applied to tweet content to identify tweets from sensor-based bot accounts.
2. A word list of earthquake-related terms (e.g., "quake", "afad", "dask", "deprem", and "kandilli_info") was applied to account names to filter out tweets from earthquake institutions in Türkiye and worldwide.
3. A word list of news-related terms (e.g., "news", "haber", and "trendinalia") was applied to account names to exclude tweets from news-feeding bot accounts.

Following the data extraction process, we summarised the tweet count by earthquakes and listed it in Table 1. The total tweet count for the 28 earthquakes is 4670, from 3467 data producers. While we test the hypothesis that a higher magnitude may result in a higher tweet count, the correlation test returns low significance for this hypothesis.

The allocation of the number of data producers in terms of the number of distinct earthquakes they tweet about and the number of tweets is summarised in Table 2. According to this summary, a great number of data producers (1011) tweeted one time related to a single earthquake. On the other hand, 1318 data producers tweeted only related to a single earthquake, while 138 and 607 data producers tweeted related to two and more than two earthquakes, respectively. The maximum number of tweets by a data producer and the maximum number of related distinct earthquakes are 18 and 8, respectively. The second largest group of data producers after the single-tweet accounts tweeted more than one and less than five times. Only five data producers tweeted 10 times or more. The maximum number of earthquakes a single data producer tweeted about, including repeated tweets due to overlapping time intervals of multiple earthquakes, is 8 out of the 28 earthquakes listed in Table 1.

Table 2. The allocation of data producers by the number of tweets and number of earthquakes they tweet about.

Number of Data Producers		Number of Earthquakes			Total
		1	2	>2 and <=8	
Number of tweets	1	1011	99	514	1624
	>1 and <5	179	33	81	293
	>=5 and <10	126	6	9	141
	>=10 and <=18	2	0	3	5
Total:		1318	138	607	2063

The second data collection stage of the study is to create a trajectory-based dataset by retrieving the previous tweets of each of the 2063 users. Before collecting the dataset, we first disambiguated tweets that referred to more than one earthquake event. We removed tweets based on their relative distance to the earthquake locations; specifically, we excluded those located farther from the disaster event and retained only the tweet that was closest to the corresponding event. After this process, we retained 3237 earthquake-related tweets.

To obtain trajectory data, we queried each user's tweets from up to three months prior to the timestamp of each earthquake-related tweet. These collected data were transformed into trajectory records for each earthquake event tweet, treating the earthquake tweet as the endpoint. In total, we compiled 3237 trajectories, each containing a varying number of trajectory points, as shown in Table 3. While nearly 10% of the trajectories consist of only the endpoint (i.e., the earthquake-related tweet itself), approximately 25% include between 77 and 1438 trajectory points.

Table 3. The number of trajectories by the number of trajectory points.

Number of Trajectory Points	Number of Trajectory (for Each Event Tweet)
1	317
2–6	540
6–24	768
25–76	804
77–1438	808
Total:	3237

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
1.00	6.00	24.00	57.35	76.00	1438.00

In the subsequent steps, aimed at determining earthquake relevancy based on user trajectories, the trajectory data were organised and feature engineered, as outlined in Section 2.4.

2.3. Data Labelling

The raw dataset was manually labelled in terms of its relevance to an earthquake. Table 4 displays relevance labels, from most relevant to least relevant, along with examples of tweet content. After labelling, the number of tweets in each category was 170, 1242, and 3258, respectively. However, some tweets were repeated due to overlapping time intervals across multiple earthquakes. As mentioned earlier, we removed duplicate tweets based on spatial proximity. Consequently, the number of distinct tweets in each category was reduced to 124, 655, and 2458, respectively.

Table 4. The frame of three categories of relevancy labels.

Relevancy Label	Explanation	Examples (Translated from Turkish)	Tweet Count	Distinct Tweet Count
1	Directly relevant: tweets directly from the field based on the experiences of tweet producers	Felt the earthquake. I am ok but we are not! #earthquake Oh my God, don't let those horror moments happen again.	170	124
2	Indirectly relevant: tweets wishing good recovery or spreading general information about earthquakes	We wish God's mercy on those who lost their lives in the earthquake that took place in Izmir, and a speedy recovery to the injured. May Allah help those who are under the rubble. Images of the building shaken like a cradle by the 5.8 magnitude earthquake.	1242	655
3	Irrelevant: mixed contents not directly meaning something	Earthquake scientist chicken feed #earthquakeVultures freight elevator legal person inflation Necromancer #KnowntheEarthquake be able to smack the skull ultimate enlightenment perfect	3258	2458
		Total:	4670	3237

2.4. Feature Extraction

To assess the relevance of each retrieved tweet (2.1), cleaned and processed tweet (2.2), and labelled tweet (2.3), we designed and extracted 126 features based on basic characteristics of earthquake-related (EQ) tweets and trajectory data. Previous tweets from each user were retrieved from the database, covering a period of up to 3 months before and 72 h after the event. Using this dataset, each earthquake-related tweet was incorporated into a trajectory. We leveraged trajectory data to extract features that help determine the relevance of tweets to the earthquake. Mainly, two calculations—spatial proximity and temporal distance—were performed for feature extraction. In the preprocessing of the trajectory data, the spatial proximity was calculated to measure the distance between the tweet location and other relevant locations (e.g., the earthquake event location or another tweet location) using the Haversine formula. Specifically, we computed the great-circle distance between each user's geotagged location (lon.x, lat.x) and the reference location (lon.y, lat.y). The distHaversine () function from the geosphere package [40] was used for this purpose, and the results were converted from metres to kilometres (1). In the preprocessing of the trajectory data, the temporal distance was calculated to measure the time difference between each user's tweet and another relevant timestamp (e.g., the earthquake event time or another tweet time). Specifically, we first ensured that the tweet timestamps (timestamp.x) were standardised to the UTC time zone using the force_tz () function. Then, we computed the time difference between each tweet and the reference timestamp (timestamp.y) using the difftime () function in R. The result, stored as the temporal distance, represents the time interval in hours (2).

Pseudocode for Calculating Spatial Proximity to Earthquake Epicentre

```
for each tweet in dataset:
    distance_meters = distHaversine(lon.x, lat.x, lon.eq, lat.eq)
    distance_km = distance_meters/1000
    append distance_km to dataset
```

(1)

Pseudocode for Calculating Temporal Distance to Earthquake Event

for each tweet in dataset:

convert timestamp.x and timestamp.y to POSIXct format
assign UTC timezone to timestamp.x and timestamp.y if not already set
time_difference_hours = difftime(timestamp.x, timestamp.y, units = "hours")
append time_difference_hours to dataset

(2)

The classification intervals were designed with two main considerations in mind: (1) the empirical characteristics of the dataset and (2) common practices in prior disaster-related spatiotemporal analyses. For example, magnitude classes were defined based on standard earthquake severity groupings used in seismology (e.g., magnitudes ≥5 indicating felt events and ≥7 indicating major quakes) [38,41]. Temporal and spatial proximity thresholds (e.g., ≤1 h, ≤100 km, etc.) were selected to reflect meaningful operational response windows and user behaviour patterns observed in disaster contexts, aligning with the approaches in other studies [42–44].

Figure 2 illustrates the complete feature design for our model dataset. We describe the feature design process in two main steps:

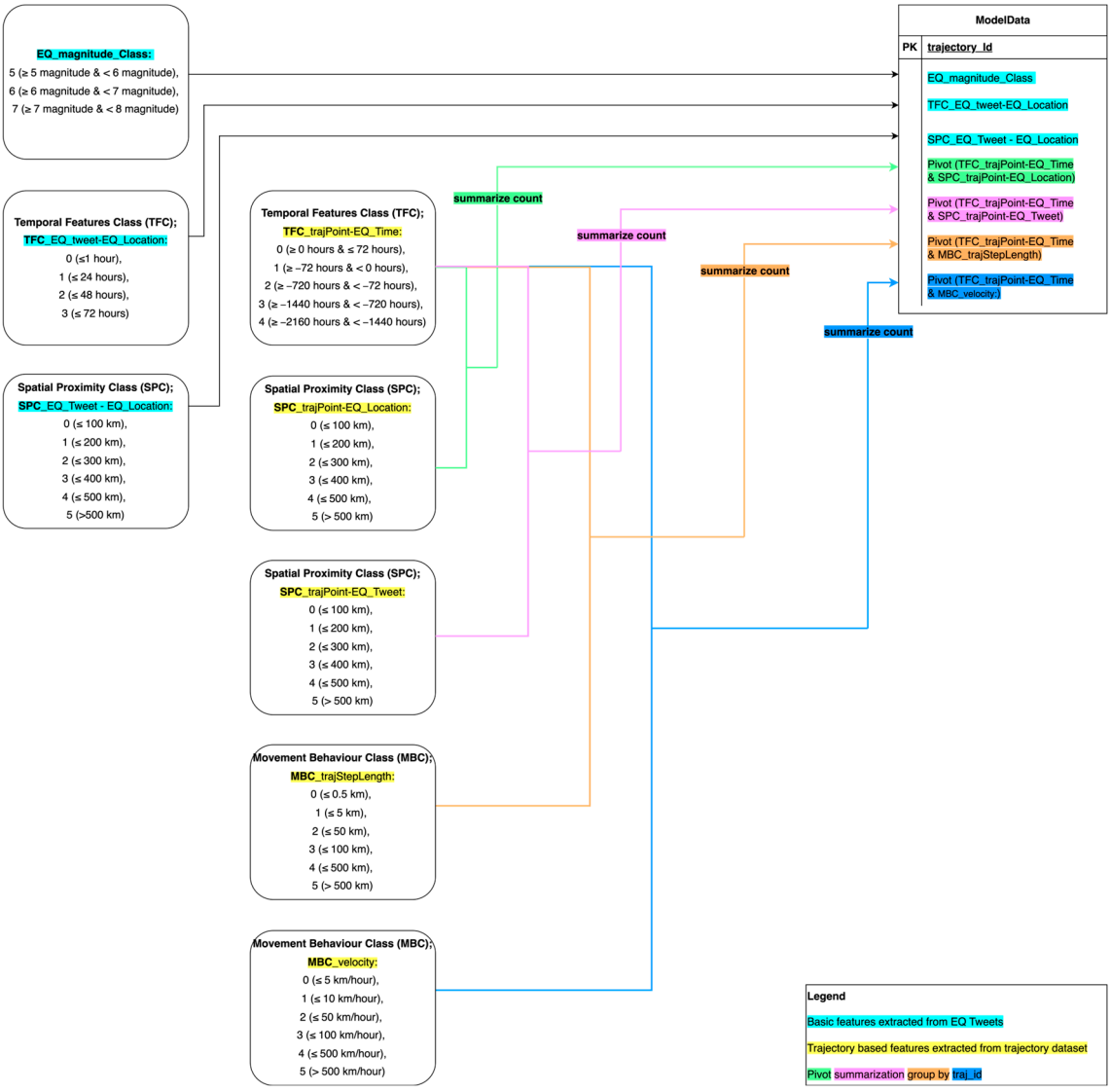


Figure 2. Trajectory-based feature design for the prediction model.

Step 1: Basic feature extraction from the EQ tweet dataset

- EQ_magnitude_Class: Categorised into three classes (5, 6, and 7), corresponding to EQs with magnitudes less than 6, 7, and 8, respectively. This feature is named EQ_magnitude_Class in the model dataset.
- Temporal Features Class (TFC): Categorised into four classes (0, 1, 2, and 3), representing the time difference between EQ occurrence and the corresponding tweet as follows: ≤1 h, ≤1 day, ≤2 days, and ≤3 days, respectively. This feature is named TFC_EQ_tweet-EQ_Location in the model dataset.
- Spatial Proximity Class (SPC): Categorised into six classes (0, 1, 2, 3, 4, and 5), representing the spatial distance between the EQ location and the location of the tweet as follows: ≤100 km, ≤200 km, ≤300 km, ≤400 km, ≤500 km, and >500 km, respectively. This feature is named SPC_EQ_Tweet-EQ_Location in the model dataset.

Step 2: Extraction of additional pivot features from trajectory data

- Temporal Features Class (TFC): Categorised into five classes (0, 1, 2, 3, and 4), representing the time difference between the EQ occurrence and the trajectory tweet as follows: within 3 days after the event, 3 days before, within 1 month before, within 2 months before, within 3 months before, respectively. This feature is named TFC_trajPoint-EQ_Time and is pivoted along with four other features—SPC_trajPoint-EQ_Location, SPC_ trajPoint-EQ_Tweet, MBC_trajStepLength, and MBC_velocity—in the model dataset.
- Spatial Proximity Class (SPC): Two SPC features are extracted under this title.
    - SPC_trajPoint-EQ_Location is categorised into six classes (0, 1, 2, 3, 4, and 5), representing the spatial distance between the EQ location and the location of the trajectory tweet as follows: ≤100 km, ≤200 km, ≤300 km, ≤400 km, ≤500 km, and >500 km, respectively.
    - SPC_ trajPoint-EQ_Tweet is categorised into six classes (0, 1, 2, 3, 4, and 5), representing the spatial distance between the EQ tweet location and the location of the trajectory tweet as follows: ≤100 km, ≤200 km, ≤300 km, ≤400 km, ≤500 km, and >500 km, respectively.
- Movement Behaviour Class (MBC): Two MBC features are extracted under this title.
    - MBC_trajStepLength is categorised into six classes (0, 1, 2, 3, 4, and 5), representing the step length between two trajectory tweets as follows: ≤0.5 km, ≤5 km, ≤50 km, ≤100 km, ≤500 km, and >500 km, respectively.
    - MBC_velocity is categorised into six classes (0, 1, 2, 3, 4, and 5), representing the velocity between two trajectory tweets as follows: ≤5 km/h, ≤10 km/h, ≤50 km/h, ≤100 km/h, ≤500 km/h, and >500 km/h, respectively.

2.5. Applied Models

The models employed in this study can be broadly categorised into three groups: 1. classical machine learning (ML) algorithms such as the Decision Tree (DT), Naïve Bayes (NB), Support Vector Machine (SVM), and k-Nearest Neighbours (k-NN); 2. ensemble-based methods exemplified by Random Forest (RF); and 3. deep learning (DL) approaches. Before applying any models, it is crucial to carefully determine appropriate model selection and assessment methods. Hastie et al. (2009) [45] suggest that when data are abundant, the optimal strategy is to randomly divide the dataset into three parts: training (50%) to fit the models, validation (25%) to estimate prediction error, and testing (25%) to evaluate the generalisation error of the final selected model. However, in many real-world problems, data are often limited, making it common to omit a separate validation set. In such cases,

k-fold cross-validation becomes essential, as it allows the model to be trained on a portion of the data while reserving a 1/k portion for validation in each fold.

As such, we considered the dataset size and class distribution. The dataset contains fewer than 5000 observations, with approximately 76%, 20%, and 4% of instances belonging to different classes. This indicates both a relatively small dataset and an imbalanced class distribution. To address these challenges, we implemented k-fold cross-validation techniques during model development. Additionally, we applied stratified sampling with an 80:20 ratio for the training and test sets to ensure a balanced representation of each class. During model training, we incorporated stratified 5-fold cross-validation and adjusted class sampling weights based on the distribution of each class to improve classification performance. Figure 3 illustrates the overall class distribution of the dataset, the stratified split of training and test sets by class, and the class distribution across the 5-fold stratified cross-validation within the training set.

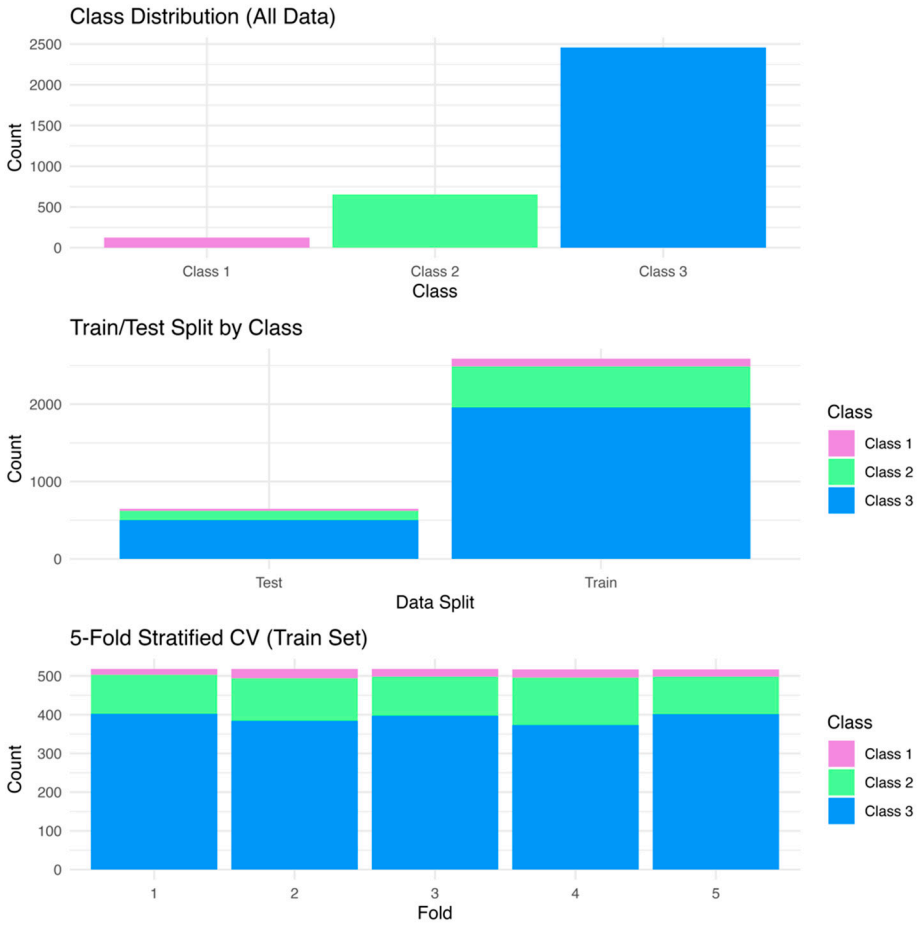


Figure 3. Overview of data partitioning and class distribution. The first panel shows the overall distribution of classes in the dataset. The second panel presents the stratified split of the dataset into training and test sets, preserving class proportions. The third panel illustrates the class distribution across the five folds used in stratified cross-validation within the training set.

Following data partitioning and class distribution determination, we determined the used R libraries and algorithms parameters listed in Table 5.

Table 5. Machine learning models, R libraries, and parameters used.

Model	R Libraries	Method	Tuned Parameters/Settings	Cross-Validation
DT	caret, rpart	rpart	cp = {0.001, 0.011, 0.021, 0.031, 0.041, 0.051}, SMOTE	5-fold
NB	caret, klaR	nb	Default settings, SMOTE	5-fold
SVM	caret, kernlab	svmRadial	C = {0.25, 0.5, 1}, sigma = {0.01, 0.05, 0.1}, SMOTE	5-fold
k-NN	caret	knn	k = {3, 5, 7, 9}, SMOTE	5-fold
RF	caret, randomForest	rf	mtry = {2, 4, 6}, ntree = 100, SMOTE	5-fold
DL	h2o, caret	h2o.deeplearning	Hidden layers: {128, 64, 32, 16}, epochs = 200, early stopping (log-loss), balanced classes	5-fold

Following dataset partitioning and class distribution determination, we selected and tuned a variety of classical machine learning (ML), ensemble-based, and deep learning algorithms using the appropriate R libraries to ensure robust model performance.

Classical ML algorithms:

DT: We utilised the rpart library [46] to implement DT models, selecting the complexity parameter (cp) via grid search to prevent overfitting. SMOTE (Synthetic Minority Over-sampling Technique) was applied during training to address class imbalance. DTs are simple, interpretable models that are useful as a baseline classifier.

NB: Using the klaR and caret packages [47,48], we applied NB due to its efficiency and low computational requirements. It serves well in high-dimensional settings assuming conditional independence among features. SMOTE was applied during training to ensure balanced class distributions.

SVM: The kernlab and caret packages [48,49] were employed to fit SVM models with radial basis function kernels. An SVM is highly effective for handling non-linear decision boundaries. The models were tuned by optimising the regularisation parameter (C) and the kernel width (σ) to improve classification performance. SMOTE was used during training to mitigate bias toward majority classes.

k-NN: The caret package [48] was used to implement the k-NN algorithm. This method was selected for its simplicity and strong performance when sufficient local data patterns exist. The number of neighbours (k) was tuned through cross-validation, and SMOTE was applied during training to improve minority class representation.

Ensemble-based algorithm:

RF: The randomForest and caret packages [48,50] were utilised to implement RF classifiers, adjusting the number of variables randomly sampled at each split (mtry) and number of trees (ntree). SMOTE was used during training to correct class imbalance and improve generalisation. RF provides robustness to overfitting and improves generalisation by aggregating multiple trees.

Deep learning-based algorithm:

DL: We utilised the h2o library [51] to develop a DL model with multiple hidden layers (128-64-32-16 neurons) and Rectified Linear Unit (ReLU) activations. The model training incorporated stratified 5-fold cross-validation, early stopping based on log-loss minimisation, and adaptive learning rates. Unlike the classical models where SMOTE was used, class imbalance was addressed using balance_classes = TRUE along with customised

class_sampling_factors, which assigned higher weights to minority classes during training. This approach ensured better representation of rare classes without synthetic oversampling.

2.6. Applied Model Evaluation Metrics

To assess the performance of the models, we utilised a comprehensive set of evaluation metrics. Accuracy was calculated to measure the proportion of correct predictions, alongside its 95% Confidence Interval (CI) to estimate the reliability of the accuracy value. The No Information Rate (NIR) and corresponding *p*-value [Acc > NIR] were considered to evaluate whether the model significantly outperformed a random guess. Kappa statistics were employed to account for agreement occurring by chance. Additionally, the McNemar's Test *p*-value was used to statistically assess the differences in predicted versus actual classifications. For each class, detailed per-class metrics were extracted, including Sensitivity (True Positive Rate), Specificity (True Negative Rate), Positive Predictive Value (PPV), Negative Predictive Value (NPV), Prevalence, Detection Rate, and Detection Prevalence. Balanced Accuracy was calculated to equally weigh Sensitivity and Specificity in cases of class imbalance. Finally, the F1 Score was reported as a harmonic mean of the Precision and Recall, providing a robust indicator of classification performance especially in imbalanced datasets.

3. Results and Discussion

The performances of six classification models—DT, NB, SVM, k-NN, RF, and DL—were evaluated using several key metrics. The metric results are shown in Table 6.

Table 6. The frame of three levels of relevancy labels.

Model	DT			NB			SVM			k-NN			RF			DL		
Accuracy	0.85			0.77			0.86			0.87			0.89			0.85		
95% CI	(0.82, 0.87)			(0.74, 0.81)			(0.82, 0.88)			(0.84, 0.90)			(0.86, 0.90)			(0.82, 0.87)		
NIR *	0.76			0.76			0.76			0.76			0.76			0.76		
p-value [Acc > NIR]	4.14×10^{-8}			0.22			0.00			2.12×10^{-12}			5.59×10^{-16}			7.25×10^{-8}		
Kappa	0.63			0.19			0.66			0.6807			0.7153			0.6215		
Mcnemar's Test *p*-value	0.0006			$<2 \times 10^{-16}$			3.583×10^{-10}			0.0007			0.0051			0.0025		
Classes (Cl)	Cl:1	Cl:2	Cl:3	Cl:1	Cl:2	Cl:3	Cl:1	Cl:2	Cl:3	Cl:1	Cl:2	Cl:3	Cl:1	Cl:2	Cl:3	Cl:1	Cl:2	Cl:3
Sensitivity	0.42	0.72	0.90	0.29	0.05	0.99	0.58	0.84	0.87	0.63	0.76	0.91	0.46	0.82	0.92	0.46	0.73	0.89
Specificity	0.94	0.93	0.86	0.97	1.00	0.21	0.92	0.92	0.94	0.95	0.94	0.88	0.96	0.94	0.90	0.94	0.92	0.84
Pos Pred Value	0.20	0.71	0.95	0.25	0.75	0.80	0.23	0.73	0.98	0.32	0.75	0.96	0.29	0.77	0.97	0.23	0.71	0.95
Neg Pred Value	0.98	0.93	0.74	0.97	0.80	0.89	0.98	0.96	0.69	0.98	0.94	0.76	0.98	0.95	0.79	0.98	0.93	0.71
Prevalence	0.04	0.20	0.76	0.04	0.20	0.76	0.04	0.20	0.76	0.04	0.20	0.76	0.04	0.20	0.76	0.04	0.2	0.76
Detection Rate	0.02	0.15	0.69	0.01	0.01	0.75	0.02	0.17	0.66	0.02	0.15	0.69	0.02	0.17	0.70	0.02	0.15	0.68
Detection Prevalence	0.08	0.20	0.72	0.04	0.01	0.94	0.09	0.23	0.67	0.07	0.21	0.72	0.06	0.22	0.73	0.07	0.21	0.72
Balanced Accuracy	0.68	0.82	0.88	0.63	0.52	0.60	0.75	0.88	0.90	0.79	0.85	0.89	0.71	0.88	0.91	0.70	0.83	0.87
F1 Score	0.27	0.71	0.93	0.27	0.09	0.88	0.33	0.78	0.92	0.42	0.76	0.93	0.35	0.79	0.95	0.31	0.72	0.92

* NIR = No Information Rate.

The RF model achieved the highest accuracy at 89% (95% CI: 86–90%), followed closely by k-NN (87%) and SVM (86%). DL and DT yielded comparable performance (85% and 85%, respectively), while NB had the lowest accuracy at 77%. Comparisons against the No Information Rate (NIR = 76%) showed statistically significant improvement for all models ($p < 0.001$) except NB ($p = 0.22$), indicating that NB did not significantly outperform random classification.

Kappa statistics, indicating agreement beyond chance, were highest for RF (0.72); moderate for k-NN (0.68), SVM (0.66), DT (0.63), and DL (0.62); and poor for NB (0.19).

The Sensitivity and Specificity varied across classes (Cl:1, Cl:2, and Cl:3). For Cl:3 (the most prevalent class, 76%), all models showed high Sensitivity (>87%) and strong Specificity (mostly above 0.85), with RF and k-NN performing particularly well with Cl:3 sensitivities of 0.92 and 0.91, respectively. However, the Sensitivity for the minority classes (Cl:1 and Cl:2) was considerably lower, especially for NB (Cl:1 = 0.29; Cl:2 = 0.05) although RF and DL offered better performances across these classes.

Balanced Accuracy, which accounts for imbalanced class distributions, was highest for DL (Cl:1: 0.82, Cl:2: 0.86, and Cl:3: 0.92) and RF (Cl:1: 0.71, Cl:2: 0.88, and Cl:3: 0.91), highlighting their robustness across classes.

The F1 Scores, representing a balance between Precision and Recall, were highest in RF (0.35–0.95 across classes) and DL (0.41–0.94), emphasising their ability to maintain high predictive performance even on minority classes.

McNemar's Test p-values further suggested that for most models, there were significant disagreements between predicted and true classes ($p < 0.05$), indicating room for improvement in model consistency, particularly for NB ($p < 2 \times 10^{-16}$) and SVM ($p = 3.58 \times 10^{-10}$).

Overall, RF provided the best overall balance between accuracy, class-wise performance, and reliability. K-NN was a close second, especially effective in handling class imbalance and delivering high accuracy and F1 Scores across all classes. Both outperform traditional models like DT, NB, and even SVM in multi-class classification tasks.

Feature importance plots were generated for the DL, DT, and RF models, as these algorithms provide interpretable importance scores based on internal mechanisms. For DT and RF, feature importance is derived from the frequency and quality of splits involving each feature, while in DL models, importance is computed using methods such as variable permutation or network weights analysis. In contrast, models like NB, SVM, and k-NN do not inherently offer meaningful feature importance scores. NB assumes feature independence and relies on conditional probabilities, SVM focuses on support vectors without ranking features globally, and k-NN is instance based and lacks an internal model structure to evaluate feature contribution. Therefore, direct feature importance visualisation is not applicable for these models without additional model-agnostic techniques.

The feature importance plot in Figure 4 illustrates the relative significance of the top 20 predictors across three different models—DL, RF, and DT—with normalised importance scores enabling direct comparison. Several features, such as Class_distanceToEvent (SPC_EQ_Tweet-EQ_Location) and Cl_td_sd_toEvent1_0 (Pivot: TFC_trajPoint-EQ_Time and SPC_trajPoint-EQ_Location: Class: 1 and Class: 0), consistently exhibit high importance across all three models, indicating their strong and stable predictive power regardless of the underlying algorithm. In contrast, features like Cl_tD_movBx_x (Pivot: TFC_trajPoint-EQ_Time and MBC_trajStepLength: Class: x and Class: x) demonstrate model-specific relevance, generally being highly influential in the DL and RF models, respectively, while receiving low importance scores in DT. This suggests that different models may capture distinct aspects of the data structure. Moreover, features such as EQ_magnitude_Class and Cl_tD_velocity2_0 and 0_0 (Pivot: TFC_trajPoint-EQ_Time and MBC_velocity: Class: 2,0

and Class: 0,0) show moderate to high importance in specific models, potentially reflecting algorithmic sensitivity to particular variable patterns or interactions. Overall, the variability in feature importance highlights both the consensus among models on core predictive features and the unique perspectives each model brings to the learning task.

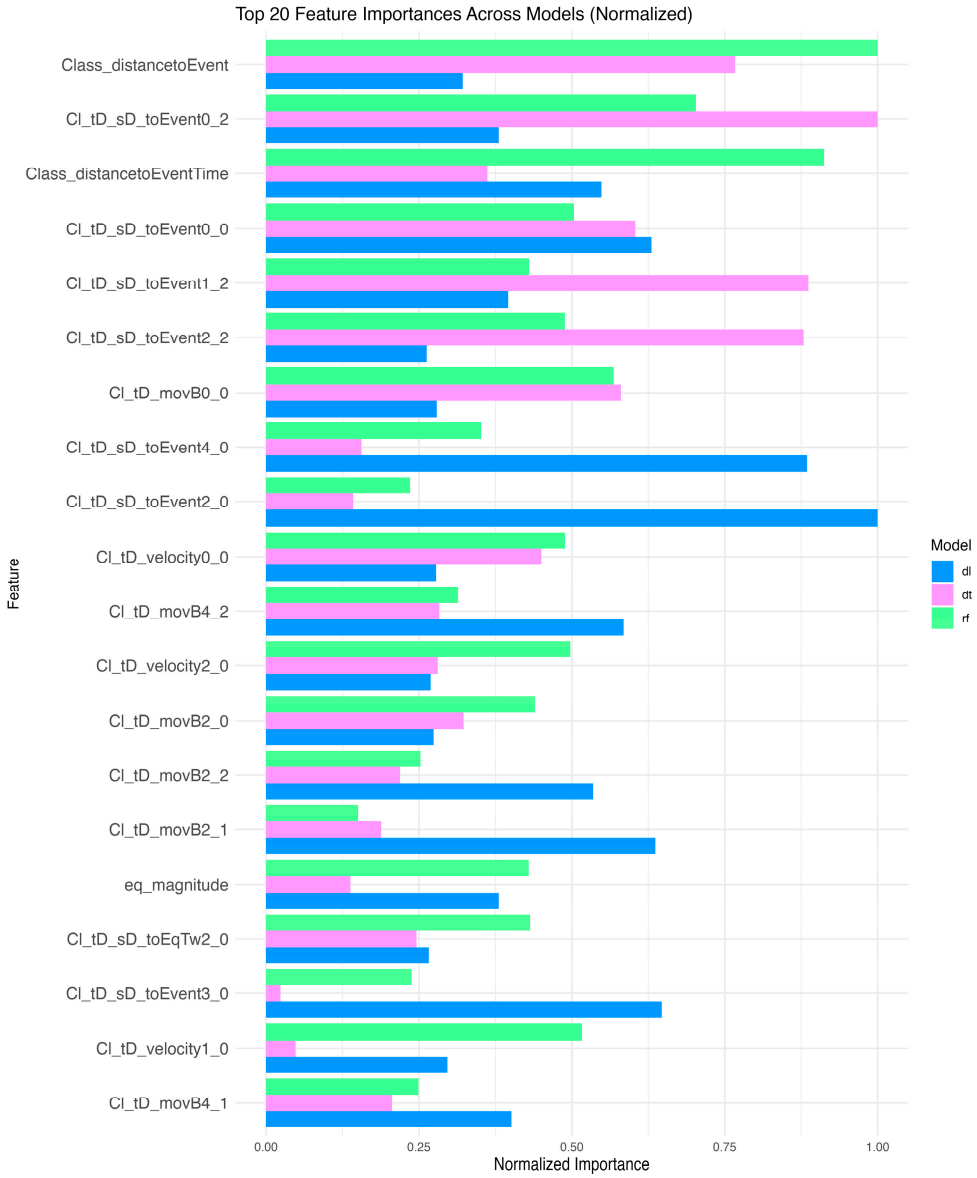


Figure 4. Top 20 normalised feature importances across DT, RF, and DL. The bar lengths represent the relative contribution of each feature to the model's predictive performance, scaled between 0 and 1 for comparability. Variations in importance across models highlight both commonly influential features and model-specific sensitivities.

This study demonstrates that trajectory-based and context-specific features, when combined with machine learning (ML) techniques applied to social media data (SMD), can effectively assess the relevance and event-related credibility of earthquake-related content. Six ML techniques were employed to assess the relevance of individual earthquake-related tweets, achieving an overall accuracy of 89%. In general, class prediction accuracy tends to increase from the most relevant class to the least. Confusion matrices reveal a high level of misclassification between Class 1 and Class 2, likely due to the inherent semantic overlap between them. Specifically, Class 1 includes content directly related to the earthquake, while Class 2 represents indirectly relevant information. From a trajectory

perspective, it is reasonable that individuals might mention an earthquake both directly and indirectly in nearby locations. This overlap introduces ambiguity and represents a limitation of the study. Nevertheless, the classification of data at this level of detail constitutes a valuable contribution.

4. Conclusions

In this study, we examined the spatial trajectories of social media (SM) users to infer the relevance of their content to earthquake events. Social media content is inherently uncertain, unstructured, and heterogeneous.

While text-based filtering and classification techniques are widely used, only a limited number of studies have employed spatial trajectories to classify content relevance and assess user credibility. Ballatore (2018, 2020) [52,53], for example, highlights the influence of geodemographics on SM posts, showing how they shape representations of urban places in Greater London and Los Angeles. From a reverse-engineering perspective, spatial representations can also reveal underlying geodemographic patterns. Our findings demonstrate that trajectory data provide valuable insights for assessing the relevance of SM content to earthquake events.

We applied machine learning (ML) models to trajectory-based features extracted from a relatively small dataset with imbalanced class distributions. In configuring model parameters, we accounted for dataset size and class imbalance, as emphasised by Hastie et al. (2009) [45], to reduce bias and variance. Despite these constraints, the models achieved high overall accuracy. However, studies of this nature require larger datasets to increase impact and improve generalisability. To support future research, our data and code are available upon reasonable request.

A key limitation of this study is the reliance on labelled data, which may introduce bias or limit scalability. Future work could address this through semi-supervised learning or crowdsourced annotation to improve both the quality and quantity of training data. Additionally, while our models performed well on the current dataset, methodological robustness could be strengthened through spatial or temporal cross-validation to evaluate performance across different geographic or temporal conditions.

The true merit of modern mapping technologies lies not only in the existence of crowdsourced platforms but in their purposeful application to generate actionable value. In this study, we focused on leveraging such platforms for life-saving efforts in earthquake crisis management. In earthquake-prone countries such as Türkiye, SM plays a vital role in incident mapping, particularly given the vulnerabilities and infrastructural damages that follow major disasters.

On 6 February 2023, two devastating earthquakes with magnitudes of 7.7 and 7.6 struck 10 cities in Türkiye, resulting in over 50,000 deaths and millions of injuries. Telecommunications infrastructure failed in many areas, and Internet bandwidth was restricted by the government to suppress the spread of misinformation. Despite this, demand for emergency assistance was high, and the country was largely unprepared for a disaster of this scale. Social media remained operational and could have played a pivotal role in emergency mapping, yet it was underutilised. Although some online communities mobilised to support crisis response, SM was overwhelmed by a flood of posts, many of which included false information, rumours, or irrelevant content, obstructing effective incident mapping.

Unlike prior studies that focus solely on textual or network-based features, our approach integrates spatiotemporal trajectory patterns with content-level analysis, offering a more grounded and language-independent method of inferring tweet relevance. By relying on users' geolocations and temporal behaviours instead of linguistic cues, the method can

generalise across different languages and regions. The fine-grained classification of tweets into directly relevant, indirectly relevant, and unrelated categories further enhances the interpretability and usefulness of the results. To validate the scalability of our approach, future work may apply the methodology to other seismic events using localised SM datasets. Moreover, the approach is adaptable to other types of crises (e.g., floods and wildfires), where spatiotemporal behaviour similarly reflects situational relevance.

In this regard, although the current study focuses specifically on assessing the credibility of tweets shared after medium- to high-intensity earthquake events, the underlying trajectory-based framework shows promise for broader applicability. Disasters with similar characteristics—those that are sudden, short term, and geographically widespread—may also benefit from this approach. For example, if a tornado affects a region within a specific time window, that period could serve as a meaningful basis for designing temporal feature classes. Likewise, spatial proximity-based feature classes could be developed by examining the size and centre of the tornado's impact area. Therefore, the tweet credibility assessment approach proposed in this study may contribute to evaluating information reliability in other types of disasters that are destructive within a defined timeframe.

Our study highlights the potential of trajectory-based classification algorithms to support timely and accurate incident mapping during crises. Such systems are crucial for filtering out irrelevant or misleading content, mitigating the spread of misinformation, and improving the quality of information used in emergency response. As false content often spreads rapidly, especially from users not directly affected by a disaster, robust classification strategies are essential to prevent false-positive mappings.

Systems like the one proposed here can improve disaster resilience, help save lives, and reduce secondary fatalities. We also stress the importance of developing national-level social media infrastructures that enable citizens to communicate freely and contribute to emergency mapping in real time. While our research focused on earthquake response, the methodology can be extended to other types of disasters, provided the feature extraction process is adapted to the unique spatial and temporal dynamics of each context.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are unavailable due to privacy or ethical restrictions. Data generated in this study will be made available upon reasonable request.

Acknowledgments: The author would like to thank all social media users who shared their data publicly and Twitter for being responsible previously by sharing free data for academic purposes. During the preparation of this work, the author used ChatGPT (GPT-4-turbo & o3) for proofreading. After using this tool, the author reviewed and edited the content as needed and takes full responsibility for the content of the publication.

Conflicts of Interest: The author declares no conflicts of interest.

References

1. Feng, Y.; Cui, S. A review of emergency response in disasters: Present and future perspectives. *Nat. Hazards* **2021**, *105*, 1109–1138. [CrossRef]
2. Lu, Y.; Xu, J. The progress of emergency response and rescue in China: A comparative analysis of Wenchuan and Lushan earthquakes. *Nat. Hazards* **2014**, *74*, 421–444. [CrossRef]
3. Shigihara, Y.; Ariga, E.; Fukutani, Y.; Tada, T. Assessing the rescue capabilities of administrative agencies and estimating rescue team dispatch for tsunami disasters. *Int. J. Disaster Risk Reduct.* **2024**, *115*, 105030. [CrossRef]

4. United Nations Office for the Coordination of Humanitarian Affairs (OCHA). 5 Essentials for the First 72 Hours of Disaster Response. 2017. Available online: https://www.unocha.org/publications/report/world/5-essentials-first-72-hours-disaster-response (accessed on 1 January 2025).
5. Saroj, A.; Pal, S. Use of social media in crisis management: A survey. *Int. J. Disaster Risk Reduct.* **2020**, *48*, 101584. [CrossRef]
6. Yu, M.; Yang, C.; Li, Y. Big data in natural disaster management: A review. *Geosciences* **2018**, *8*, 165. [CrossRef]
7. Neppalli, V.K.; Caragea, C.; Squicciarini, A.; Tapia, A.; Stehle, S. Sentiment analysis during Hurricane Sandy in emergency response. *Int. J. Disaster Risk Reduct.* **2017**, *21*, 213–222. [CrossRef]
8. Mihunov, V.V.; Lam, N.S.N.; Zou, L.; Wang, Z.; Wang, K. Use of Twitter in disaster rescue: Lessons learned from Hurricane Harvey. *Int. J. Digit. Earth* **2020**, *13*, 1454–1466. [CrossRef]
9. Zhou, X.; Xu, C. Tracing the spatial-temporal evolution of events based on social media data. *ISPRS Int. J. Geo Inf.* **2017**, *6*, 88. [CrossRef]
10. Chen, M.; Arribas-Bel, D.; Singleton, A. Quantifying the characteristics of the local urban environment through geotagged Flickr photographs and image recognition. *ISPRS Int. J. Geo Inf.* **2020**, *9*, 264. [CrossRef]
11. Gulnerman, A.G.; Karaman, H.; Pekaslan, D.; Bilgi, S. Citizens' spatial footprint on Twitter—Anomaly, trend and bias investigation in Istanbul. *ISPRS Int. J. Geo Inf.* **2020**, *9*, 222. [CrossRef]
12. Beigi, G.; Hu, X.; Maciejewski, R.; Liu, H. An overview of sentiment analysis in social media and its applications in disaster relief. In *Sentiment Analysis and Ontology Engineering*; Springer: Cham, Switzerland, 2016; pp. 313–340.
13. Chapman, L.; Resch, B.; Sadler, J.; Zimmer, S.; Roberts, H.; Petutschnig, A. Investigating the emotional responses of individuals to urban green space using Twitter data: A critical comparison of three different methods of sentiment analysis. *Urban Plan.* **2018**, *3*, 21–33.
14. Calafiore, A.; Palmer, G.; Comber, S.; Arribas-Bel, D.; Singleton, A. A geographic data science framework for the functional and contextual analysis of human dynamics within global cities. *Comput. Environ. Urban Syst.* **2021**, *85*, 101539. [CrossRef]
15. Ahmouda, A.; Hochmair, H.H.; Cvetojevic, S. Using Twitter to analyze the effect of hurricanes on human mobility patterns. *Urban Sci.* **2019**, *3*, 87. [CrossRef]
16. Gulnerman, A.G. Changing pattern of human movements in Istanbul during COVID-19. In Proceedings of the International Conference on Computational Science and Its Applications, Cagliari, Italy, 13–16 September 2021; Springer: Berlin/Heidelberg, Germany, 2021; pp. 220–230.
17. Houston, J.B.; Hawthorne, J.; Perreault, M.F.; Park, E.H.; Goldstein Hode, M.; Halliwell, M.R.; McGowen, S.E.T.; Davis, R.; Vaid, S.; Mcelderry, J.A.; et al. Social media and disasters: A functional framework for social media use in disaster planning, response, and research. *Disasters* **2015**, *39*, 1–22. [CrossRef]
18. Huang, Q.; Xiao, Y. Geographic situational awareness: Mining tweets for disaster preparedness, emergency response, impact, and recovery. *ISPRS Int. J. Geo Inf.* **2015**, *4*, 1549–1568. [CrossRef]
19. Resch, B.; Uslander, F.; Havas, C. Combining machine-learning topic models and spatiotemporal analysis of social media data for disaster footprint and damage assessment. *Cartogr. Geogr. Inf. Sci.* **2018**, *45*, 362–376. [CrossRef]
20. Hecht, B.; Hong, L.; Suh, B.; Chi, E.H. Tweets from Justin Bieber's heart: The dynamics of the location field in user profiles. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Vancouver, BC, Canada, 7–12 May 2011; ACM: New York, NY, USA, 2011; pp. 237–246.
21. Gupta, A.; Kumaraguru, P. Credibility ranking of tweets during high impact events. In Proceedings of the 1st Workshop on Privacy and Security in Online Social Media, Lyon, France, 17 April 2012; pp. 2–8.
22. Truelove, M.; Vasardani, M.; Winter, S. Testing the event witnessing status of micro-bloggers from evidence in their micro-blogs. *PLoS ONE* **2017**, *12*, e0189378. [CrossRef]
23. Abbasi, M.-A.; Liu, H. Measuring user credibility in social media. In Proceedings of the International Conference on Social Computing, Behavioral-Cultural Modeling, and Prediction, Washington, DC, USA, 2–5 April 2013; Springer: Berlin/Heidelberg, Germany, 2013; pp. 441–448.
24. Yang, J.; Yu, M.; Qin, H.; Lu, M.; Yang, C. A Twitter data credibility framework—Hurricane Harvey as a use case. *ISPRS Int. J. Geo Inf.* **2019**, *8*, 111. [CrossRef]
25. Mazimpaka, J.D.; Timpf, S. Trajectory data mining: A review of methods and applications. *J. Spat. Inf. Sci.* **2016**, *13*, 61–99. [CrossRef]
26. Zheng, Y. Trajectory data mining: An overview. *ACM Trans. Intell. Syst. Technol.* **2015**, *6*, 1–41. [CrossRef]
27. Spinsanti, L.; Berlingerio, M.; Pappalardo, L. Mobility and geo-social networks. In *Mobility Data: Modeling, Management, and Understanding*; Cambridge Press: Cambridge, UK, 2013.
28. Pelekis, N.; Theodoridis, Y. *Mobility Data Management and Exploration*; Springer: New York, NY, USA, 2014.
29. Jurdak, R.; Zhao, K.; Liu, J.; AbouJaoude, M.; Cameron, M.; Newth, D. Understanding human mobility from Twitter. *PLoS ONE* **2015**, *10*, e0131469. [CrossRef] [PubMed]

30. Andrienko, G.; Andrienko, N.; Bosch, H.; Ertl, T.; Fuchs, G.; Jankowski, P.; Thom, D. Thematic patterns in georeferenced tweets through space-time visual analytics. *Comput. Sci. Eng.* **2013**, *15*, 72–82. [CrossRef]
31. Chae, J.; Cui, Y.; Jang, Y.; Wang, G.; Malik, A.; Ebert, D.S. Trajectory-based visual analytics for anomalous human movement analysis using social media. In *EuroVA@EuroVis*; The Eurographics Association: Eindhoven, the Netherlands, 2015; pp. 43–47.
32. Wang, Y.; Taylor, J.E. Coupling sentiment and human mobility in natural disasters: A Twitter-based study of the 2014 South Napa earthquake. *Nat. Hazards* **2018**, *92*, 907–925. [CrossRef]
33. Hübl, F.; Cvetojevic, S.; Hochmair, H.; Paulus, G. Analyzing refugee migration patterns using geo-tagged tweets. *ISPRS Int. J. Geo Inf.* **2017**, *6*, 302. [CrossRef]
34. Ma, D.; Osaragi, T.; Oki, T.; Jiang, B. Exploring the heterogeneity of human urban movements using geo-tagged tweets. *Int. J. Geogr. Inf. Sci.* **2020**, *34*, 2475–2496. [CrossRef]
35. Kang, Y.; Gao, S.; Liang, Y.; Li, M.; Rao, J.; Kruse, J. Multiscale dynamic human mobility flow dataset in the US during the COVID-19 epidemic. *Sci. Data* **2020**, *7*, 390. [CrossRef]
36. Hanny, D.; Resch, B. Multimodal GeoAI: An integrated spatio-temporal topic-sentiment model for the analysis of geo-social media posts for disaster management. *Int. J. Appl. Earth Obs. Geoinf.* **2025**, *139*, 104540. [CrossRef]
37. Xdevelopers. [Tweet]. X. Available online: https://x.com/XDevelopers/status/1621026986784337922 (accessed on 2 February 2023).
38. Richter, C.F. An instrumental earthquake magnitude scale. *Bull. Seismol. Soc. Am.* **1935**, *25*, 1–32. [CrossRef]
39. Boğaziçi University Kandilli Observatory and Earthquake Research Institute (KOERI). Earthquake Catalog Search System. 2025. Available online: http://udim.koeri.boun.edu.tr/zeqmap/hgmmapen.asp (accessed on 26 August 2024).
40. Hijmans, R. Geosphere: Spherical Trigonometry. R Package Version 1.5-20 2024. Available online: https://CRAN.R-project.org/package=geosphere (accessed on 1 January 2025).
41. USGS. Earthquake Magnitude, Energy Release, and Shaking Intensity. Available online: https://www.usgs.gov/programs/earthquake-hazards/earthquake-magnitude-energy-release-and-shaking-intensity (accessed on 1 June 2025).
42. Crooks, A.; Croitoru, A.; Stefanidis, A.; Radzikowski, J. #Earthquake: Twitter as a distributed sensor system. *Trans. GIS* **2013**, *17*, 124–147.
43. Zheng, Y.; Zhang, L.; Xie, X.; Ma, W.Y. Mining interesting locations and travel sequences from GPS trajectories. In Proceedings of the 18th International Conference on World Wide Web, New York, NY, USA, 20–24 April 2009; pp. 791–800.
44. Feng, C.M.; Wang, T.C. Highway emergency rehabilitation scheduling in post-earthquake 72 hours. *J. 5th East. Asia Soc. Transp. Stud.* **2003**, *5*, 3276–3285.
45. Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning*; Springer: New York, NY, USA, 2009; p. 222.
46. Therneau, T.; Atkinson, B. Rpart: Recursive Partitioning and Regression Trees (Version 4.1.24). 2025. Available online: https://CRAN.R-project.org/package=rpart (accessed on 12 April 2025).
47. Weihs, C.; Ligges, U.; Luebke, K.; Raabe, N. klaR: Analyzing German business cycles. In *Data Analysis and Decision Support*; Baier, D., Decker, R., Schmidt-Thieme, L., Eds.; Springer: Berlin/Heidelberg, Germany, 2005; pp. 335–343.
48. Kuhn, M. Building predictive models in R using the caret package. *J. Stat. Softw.* **2008**, *28*, 1–26. [CrossRef]
49. Karatzoglou, A.; Smola, A.; Hornik, K. Kernlab: Kernel-Based Machine Learning Lab (Version 0.9-33). 2024. Available online: https://CRAN.R-project.org/package=kernlab (accessed on 12 April 2025).
50. Liaw, A.; Wiener, M. Classification and regression by randomForest. *R News* **2002**, *2*, 18–22. Available online: https://CRAN.R-project.org/doc/Rnews/ (accessed on 26 August 2024).
51. Fryda, T.; LeDell, E.; Gill, N.; Aiello, S.; Fu, A.; Candel, A.; Click, C.; Kraljevic, T.; Nykodym, T.; Aboyoun, P.; et al. H2o: R Interface for the 'H2O' Scalable Machine Learning Platform (Version 3.44.0.3). 2024. Available online: https://CRAN.R-project.org/package=h2o (accessed on 12 January 2025).
52. Ballatore, A.; De Sabbata, S. Charting the geographies of crowdsourced information in Greater London. In Proceedings of the Annual International Conference on Geographic Information Science, Lund, Sweden, 12–15 June 2018; Springer: Cham, Switzerland, 2018; pp. 149–168.
53. Ballatore, A.; De Sabbata, S. Los Angeles as a digital place: The geographies of user-generated content. *Trans. GIS* **2020**, *24*, 880–902. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.