# Event Prediction: Datasets

*2022-02-16 09:30AM (KST), 2022-02-15 07:30PM (EST)*
***Huigyu Yang***
*Sungkyunkwan University*
*huigyu@skku.edu*

# *Presentation Outline*

- Event Time Prediction
- Datasets in Papers
- Dataset in Kaggle

# Event Time Prediction

- Event time prediction can be categorized into three types:
- Event occurrence; simplest type of event time prediction task
  - This can be formulated as a binary classification problem whether the event will occur or not
- Discrete-time prediction
  - Approximate time prediction of future events through time-slot based data composition
- Continuous-time prediction
  - Continuous-time prediction extends discrete-time prediction by increasing granularity of discretization
  - Regression, Point processes

# Datasets in Papers (1/9)

- Event dataset from Social Media Platforms (SMP) [1]
  - Offline frameworks on event detection and predictions

| Author | Year | Application area | SMP | Techniques applied | Model/ Frameworks | Features | Gap/Future direction |
|---|---|---|---|---|---|---|---|
| Vanetik N et al. [39] | 2019 | Event detection | Twitter | Uses text and wavelet analysis | TWIST (Twitter event summarizer and Trend detector) | Solve challenges in processing the large volume of data. | – |
| Nolasco et al. [40] | 2019 | Event detection | Twitter | LDA, TF | Candidate selection algorithm, Score Ranking, Label Selection | Model identifies subevent, labels the subevent from collection, tags the subevent with suitable labels | Close representation od sub-event, addition of external resources to improve results |
| Fedoryszak et al. [41] | 2019 | Event detection | Full Twitter Firehose | Minimum similarity threshold, Louvain clustering resolution, Time window | Similarity based temporal event detection | Tracks important trends, achieves good performance with minimum similarity more than 0.4 | Many duplicates in subevent detection |
| Abbasi et al. [25] | 2019 | Adverse event prediction | Twitter | Signal detection based on Genetic algorithm, Sentiment analysis to reduce false positives | GASD framework for VOC (voice of customer) platform | Detection of events with accuracy more than 50–80% , | Timeliness, testing on multiple platforms, trying with different thresholds ,use of machine learning |
| Fan et al. [26] | 2019 | Disaster disruption detection on people and land | Twitter | Weighted graph, Time series analysis, cosine similarity | Graph based approach | Better results with graph based approach than clustering and topic modelling, Retweets does not add value to results | Improvements needed in algorithms on burst prediction and vectorising of tweets, need of new algorithm on identification of |

[1] *Poonam Tijare, Jhansi Rani Prathuri, "A Survey on Event Detection and Prediction Online and Offline Models using Social Media Platforms," Materials Today: Proceedings, 2021*

# Datasets in Papers (2/9)

- Event dataset from Social Media Platforms (SMP)
  - Offline frameworks on event detection and predictions

| Author | Year | Application area | SMP | Techniques applied | Model/ Frameworks | Features | Gap/Future direction |
|---|---|---|---|---|---|---|---|
| Yang et al. [27] | 2019 | Cross platform event detection | Flickr images from different sources | Shared data formats learning with dictionary learning | SMDR (Shared multi-view data representation model) | Model discovers heterogeneous events from crossdomains, incorporates shared dictionary and reconstruction error feature | Geolocalization can be added to improve quality, scaling up to big datasets, more platforms need to be explored |
| Saeed et al. [3] | 2019 | Event detection | Twitter | Graph analysis, KL-divergence | Enhanced Heartbeat Graph (EHG) | Low complexity approach, transforms the data into EHG and then detect events | Addition of temporal information, rank the detected events and application on live stream |
| Roy et al. [28] | 2019 | Natural disaster event | Twitter, New York taxi trip data | Location time series, normalized per userdisplacement | Quantifying mobility using resilience metrics by using geotagged information | Mobility information is studied based on disaster affected area | Better mobility data and infrastructure to conduct such experiments |

# Datasets in Papers (3/9)

- Event dataset from Social Media Platforms (SMP)
  - Online frameworks on event detection and predictions

| Author | Year | Application area | SMP | Techniques applied | Model/Framework | Features | Gap/Future direction |
|---|---|---|---|---|---|---|---|
| Hasan et al. [44] | 2019 | Event detection in real time | Twitter | TF-ID, Incremental clustering, cosine similarity | TwitterNews+ | Fast processing capability with constant process time, able to detect less bursty event | Temporal summarization of tweets in cluster can be done |
| Fedoryszak et al. [41] | 2019 | Event detection in real time | Full Twitter Firehose | Minimum similarity threshold, Louvain clustering resolution, Time window | Similarity based temporal event detection | Model experimented on both offline and online mode,tracks important trends | Model uses precomputed embeddings, dynamic embeddings are suitable |
| Mukhina et al. [47] | 2019 | Urban event prediction | Instagram | CNN | Proposed CNN based architecture | Predict user activity with average deviation of 1% compared to Ground Truth with 69.4% recall | Need to decrease in false positives, Need to build hybrid complex models for better accuracy |
| Hu et al [46] | 2020 | Event prediction | Twitter | Neural Network | Hierarchical attention-based model | Predicts the next possible event by forming word sequence that describes the event, uses joint learning | Accuracy improvement is needed |

- The datasets collected from SMP are huge and mostly given as natural language based data (*i.e.*, tweets)

# Datasets in Papers (4/9)

- Mobility related datasets: Next-location Prediction

| | Reference | Name | Year | DL Modules | Evaluation | Dataset | Code (https://bit.ly/) |
|---|---|---|---|---|---|---|---|
| Next-Location Prediction | Abideen et al. [1] | DWSTTN | 2021 | Encoder, Decoder, Attention, FC | Distance | [127] | - |
| | Tang et al. [186] | CLNN | 2021 | LSTM, Embedding, FC | Distance | [127] | - |
| | Bao et al. [10] | BiLSTM-CNN | 2020 | Embedding, BiLSTM, CNN | ACC@k | - | |
| | Chen et al. [36] | DeepJMT | 2020 | GRU, FC, Encoder | ACC@k | [218] | - |
| | Yang et al. [217] | Flashback | 2020 | Attention, RNN | ACC@k | [37] | Flashback-1 |
| | Ebel et al. [52] | - | 2020 | RNN, FC, Embedding | Distance | [127, 147] | - |
| | Rossi et al. [156] | - | 2019 | Attention, LSTM | Distance | [127, 147, 188] | - |
| | Gao et al. [67] | VANext | 2019 | CNN, GRU, Attention | ACC@k | [37] | - |
| | Kong et al. [103] | HST-LSTM | 2018 | LSTM | ACC | - | HST-LSTM |
| | Lv et al. [122] | T-CONV | 2018 | CNN, FC | Distance | [127] | T-CONV |
| | Feng et al. [57] | DeepMove | 2018 | Attention, GRU, FC | ACC | [57] | DeepMove |
| | Yao et al. [220] | SERM | 2017 | LSTM | ACC@k | - | SERM-Repo |
| | Liu et al. [118] | ST-RNN | 2016 | RNN | Rec@k, F1@k, MAPE, AUC | [37, 181] | STRNN |
| | De Brébisson et al. [47] | - | 2015 | FC | Distance | [127] | next-loc-1 |

- The data representations of mobility dataset are categorized into GPS trace and check-in based logs

# Datasets in Papers (5/9)

- Mobility related dataset lists [2]

| | Ref. | Name | Items | Time span | Area | Used By | Task | Link (https://bit.ly/) |
|---|---|---|---|---|---|---|---|---|
| GPS traces | [238] | GeoLife | 182 | 4.5 Years | Asia | [58] | Traj. Gen. | Geolife |
| | [235] | T-Drive | 10K | 1 Week | Beijing, China | [84] | Flow. Pred. | T-Drive-Data |
| | [86] | DeepCrowd | - | 4 months | Tokyo and Osaka, Japan | [86] | Crowd Flow Pred. | DeepCrowd |
| | [230] | ST-ResNet taxis | - | 4, 6 months | Beijing, China | [50, 91, 113, 117, 151, 184, 227, 230, 243] | Crowd Flow Pred. | ST-ResNet |
| | [230] | ST-ResNet bikes | - | 6 months | New York City, USA | [50, 91, 113, 117, 151, 184, 227, 230, 243] | Crowd Flow Pred. | ST-ResNet |
| | [147] | Taxi San Francisco | 500 | 30 days | San Francisco, USA | [52, 156, 224] | Next-Loc., Traj. Gen. | TaxiSF |
| | [127] | ECML-PKDD taxi | 441 | 9 months | Porto, Portugal | [47, 52, 122, 156] | Next-Loc. | TaxiPorto |
| | [188] | Taxi New York City | - | From 2009 | New York City | [156, 184, 187] | Next-Loc. | TaxiNYC-2 |
| | [110] | MDC | 185 | 2 years | Lausanne, Switzerland | [108, 133] | Traj. Gen | MDC-2 |
| | [95] | COVID 19 US Flows | - | From 2019 | United States | [169] | Flow Gen. | USFlows |
| check-ins | [37] | Gowalla | 196K | 20 months | California & Nevada, USA | [67, 118, 217] | Next-Loc. | GowallaData |
| | [37] | Brightkite | 58K | 30 months | - | [67, 118, 217] | Next-Loc. | Brightkite |
| | [219] | Foursquare | 800K | 10 months | NYC, Tokyo, World | public | Foursquare-Data | |
| | [217] | Foursquare | 46K | 1.5 Years | NYC | Flashback-1 | | |
| | [57] | DeepMove | 16K | 1 Year | New York City | [57] | Next-Loc. | DeepMove |
| | [229] | GMove | 1.4M | 4 Months | Los Angeles | [220] | Next-Loc. | SERM-Repo |
| | [16] | New York City bikes | - | from 2013 | New York City, USA | [116, 184, 187, 222] | Crowd Flow Pred. | BikeNYCData |
| | [17] | Washington DC bikes | - | from 2010 | Washington DC, USA | [184] | Crowd Flow Pred. | BikeWashington |

For each dataset, we provide a reference to the paper introducing it, the number of items (users or points) in the dataset (symbol "-" indicates that the dataset is aggregated, that the number is not available, or that the dataset is continuously updated), its time span, the geographic area covered, the list of selected papers that use it, the mobility tasks the dataset is used for, and the link to download it.

[2] *Massimiliano Luca, Gianni Barlacchi, Bruno Lepri, and Luca Pappalardo. "A Survey on Deep Learning for Human Mobility." ACM Computing Survey, Article 7. Nov. 2021*

# Datasets in Papers (6/9)

- Check-in based log dataset

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| [37] | Gowalla | 196K | 20 months | California & Nevada, USA | [67, 118, 217] | Next-Loc. | GowallaData |
| [37] | Brightkite | 58K | 30 months | - | [67, 118, 217] | Next-Loc. | Brightkite |

**Example of check-in information**

| [user] | [check-in time] | [latitude] | [longitude] | [location id] |
|---|---|---|---|---|
| 196514 | 2010-07-24T13:45:06Z | 53.3648119 | -2.2723465833 | 145064 |
| 196514 | 2010-07-24T13:44:58Z | 53.360511233 | -2.276369017 | 1275991 |
| 196514 | 2010-07-24T13:44:46Z | 53.3653895945 | -2.2754087046 | 376497 |
| 196514 | 2010-07-24T13:44:38Z | 53.3663709833 | -2.2700764333 | 98503 |
| 196514 | 2010-07-24T13:44:26Z | 53.3674087524 | -2.2783813477 | 1043431 |
| 196514 | 2010-07-24T13:44:08Z | 53.3675663377 | -2.278631763 | 881734 |
| 196514 | 2010-07-24T13:43:18Z | 53.3679640626 | -2.2792943689 | 207763 |
| 196514 | 2010-07-24T13:41:10Z | 53.364905 | -2.270824 | 1042822 |

**Example of check-in information**

| [user] | [check-in time] | [latitude] | [longitude] | [location id] |
|---|---|---|---|---|
| 58186 | 2008-12-03T21:09:14Z | 39.633321 | -105.317215 | ee8b88dea22411 |
| 58186 | 2008-11-30T22:30:12Z | 39.633321 | -105.317215 | ee8b88dea22411 |
| 58186 | 2008-11-28T17:55:04Z | -13.158333 | -72.531389 | e6e86be2a22411 |
| 58186 | 2008-11-26T17:08:25Z | 39.633321 | -105.317215 | ee8b88dea22411 |
| 58187 | 2008-08-14T21:23:55Z | 41.257924 | -95.938081 | 4c2af967eb5df8 |
| 58187 | 2008-08-14T07:09:38Z | 41.257924 | -95.938081 | 4c2af967eb5df8 |
| 58187 | 2008-08-14T07:08:59Z | 41.295474 | -95.999814 | f3bb9560a2532e |
| 58187 | 2008-08-14T06:54:21Z | 41.295474 | -95.999814 | f3bb9560a2532e |
| 58188 | 2010-04-06T06:45:19Z | 46.521389 | 14.854444 | ddaa40aaa22411 |
| 58188 | 2008-12-30T15:30:08Z | 46.522621 | 14.849618 | 58e12bc0d67e11 |
| 58189 | 2009-04-08T07:36:46Z | 46.554722 | 15.646667 | ddaf9c4ea22411 |
| 58190 | 2009-04-08T07:01:28Z | 46.421389 | 15.869722 | dd793f96a22411 |

# Datasets in Papers (7/9)

- Check-in based log dataset: Foursquare Dataset (by Dingqi YANG)

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| [219] | Foursquare | 800K | 10 months | NYC, Tokyo, World | public | Foursquare-Data | |
| [217] | Foursquare | 46K | 1.5 Years | NYC | Flashback-1 | | |
| [57] | DeepMove | 16K | 1 Year | New York City | [57] | Next-Loc. | DeepMove |

dataset_ubicomp2013_checkins.txt - 메모장

파일 편집 보기

```
35443 899
24973 42406
14860 177
222505	177
63524 609
51957 435580
7860	68829
8952	12790
42283 15071
14506 25864
121273	20624
25624 77986
31594 23124
23264 908
58758 43480
7128	34618
93642 12052
47596 12042
59796 14197
60062 381
```

줄 1, 열 1

dataset_ubicomp2013_tags.txt - 메모장

파일 편집 보기

```
15
20	andy cohen,bakery,bar,barbeque,bbq,bistro,boutique,bravo,bravoandy,celebrity sighting,coffee,cupcake,dessert,desserts,douchebag,douchebags,foodie,foody,french cuisine,gallery art,gossip girl,gossipgirl,home,manhattan,models,people hot spot,photo booth,restaurant,socalite,socialite,steak frites,theater,top chef,tourists,trendy,victoria's secret - bombshell,vs hotspot,west village,wine,zagat rated
25	colombian,cupcake,fish tacos,grilled corn,long wait,pork,rice and beans,sweet plantains,tinituan ma lala,trendy
26	brunch,happy hour,irish,pub
36	french,soho,zagat-rated,zagats
39	byob
40	+ jazz club,good food and jazz
42	bar,beer,brunch,burger,cozy,dinner,group,happy hour,lunch,nic
waitress,salad,sandwiches,sports,steak,tavern
46	bakery,bravo,cupcake,dessert,diner,frozen hot chocolate,froze
cream,socialite,top chef,tourists,upper east side
47	bagels,cafe,coffee,desserts,green tea cookies,irving farm,local
52	gossip girl
59	french restaurant,outdoor seating,wine bar
77	american,brunch,dinner,happy hour,jazz,live jazz,live music,pre-
```

줄 3, 열 115 100%

dataset_ubicomp2013_tips.txt - 메모장

파일 편집 보기

```
24436 15	make your own sandwich: tuna salad on country white bread, cheddar, lettuce, tomato, cucumber, mayo. yuuuum.
8550	20	The calamari as an appetizer and the quail. Pretty good, if a bit rich. The berry and creme desert thing was really, really good!
1537	20	Be careful with the napkins on your pants. They leave remnants behind. As in lint.
59283 20	try the steak frites, it's the best - believe me
713	20	go visit google across the street. see the big coffee robot
40518 20	Great drinks & good appetizers. Love the dessert menu.
33560 20	Try the English breakfast
17435 20	total madhouse and that's because the french onion soup keeps everyone coming back.. oh, and celebrity sightings too. leave room for dessert, my 2 faves? Tarte Aux Pommes & Cr?es Suzette.
59010 20	And by the way if u are looking for a Job on advertising sales and travelling around the world check that after the grattin dauphinois: http:₩/₩/bit.ly₩/aQiWtc
59010 20	Go for the grattin dauphinois!
37602 20	When here you have to order the Mac & Cheese. Amazing
68863 20	People watching + Cheesecake + Sex & The City
118428	20	para brunchear
105277	20	love eating outside here :)
211609	20	If you want an authentic experience, think again if you are pondering a full, sit-down meal here-stick to a pre-dinner cocktail or a night cap₩u2013 always reliable and fun. Check out the rest
```

줄 12, 열 41 100% Unix (LF) ANSI

# Datasets in Papers (8/9)

- GPS and Check-in based log dataset: Citi Bike

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| [16] | New York City bikes | - | from 2013 | New York City, USA | [116, 184, 187, 222] | Crowd Flow Pred. | BikeNYCData |
| [17] | Washington DC bikes | - | from 2010 | Washington DC, USA | [184] | Crowd Flow Pred. | BikeWashington |

## Citi Bike Trip Histories

We publish downloadable files of Citi Bike trip data. The data includes:

- Ride ID
- Rideable type
- Started at
- Ended at
- Start station name
- Start station ID
- End station name
- End station ID
- Start latitude
- Start longitude
- End latitude
- End Longitude
- Member or casual ride

# Datasets in Papers (9/9)

■ GPS and Check-in based log dataset: Citi Bike

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| [16] | New York City bikes | - | from 2013 | New York City, USA | [116, 184, 187, 222] | Crowd Flow Pred. | BikeNYCData |
| [17] | Washington DC bikes | - | from 2010 | Washington DC, USA | [184] | Crowd Flow Pred. | BikeWashington |

| 1 | ride_id | rideable_type | started_at | ended_at | start_station_name | start_station_id | end_station_name | end_station_id | start_lat | start_lng | end_lat | end_lng | member_casual |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2 | 1589851B36BB0B5C | classic_bike | 2022-01-07 12:56 | 2022-01-07 13:01 | Adam Clayton Powell Blvd & W 126 St | 7738.04 | Frederick Douglass Blvd & W 139 St | 7876.07 | 40.80949535 | -73.94776493 | 40.81900582 | -73.94476891 | member |
| 3 | 4C0BB6BD8AFCA917 | classic_bike | 2022-01-06 16:01 | 2022-01-06 16:06 | Adam Clayton Powell Blvd & W 126 St | 7738.04 | Frederick Douglass Blvd & W 139 St | 7876.07 | 40.80949535 | -73.94776493 | 40.81900582 | -73.94476891 | member |
| 4 | 765572ACD0D65972 | classic_bike | 2022-01-31 15:31 | 2022-01-31 15:36 | E 56 St & Madison Ave | 6732.01 | E 48 St & 5 Ave | 6626.01 | 40.761573 | -73.972628 | 40.75724568 | -73.97805914 | member |
| 5 | 86E8E7C4791EA81D | classic_bike | 2022-01-21 17:38 | 2022-01-21 17:46 | 3 Ave & E 100 St | 7414.17 | E 85 St & 3 Ave | 7212.05 | 40.7877214 | -73.94728331 | 40.77801203 | -73.95407149 | member |
| 6 | D3B80E976AC4DBCF | classic_bike | 2022-01-23 18:37 | 2022-01-23 18:43 | E 88 St & Park Ave | 7293.1 | E 85 St & 3 Ave | 7212.05 | 40.7814107 | -73.95595908 | 40.77801203 | -73.95407149 | member |
| 7 | 283D619884B13025 | classic_bike | 2022-01-13 12:17 | 2022-01-13 12:20 | E 85 St & York Ave | 7146.04 | E 85 St & 3 Ave | 7212.05 | 40.77536905 | -73.94803392 | 40.77801203 | -73.95407149 | member |
| 8 | CC02A6C3FA1F2083 | classic_bike | 2022-01-11 9:09 | 2022-01-11 9:20 | Broadway & Madison St | 4483.1 | Suydam St & Broadway | 4689.03 | 40.68822 | -73.91966 | 40.69544 | -73.93223 | member |
| 9 | 968470449EEB57C0 | classic_bike | 2022-01-09 12:33 | 2022-01-09 12:50 | Broadway & W 58 St | 6948.1 | E 85 St & 3 Ave | 7212.05 | 40.76695317 | -73.98169333 | 40.77801203 | -73.95407149 | member |
| 10 | 16C601B498A35DBF | classic_bike | 2022-01-14 8:28 | 2022-01-14 8:33 | Lenox Ave & W 111 St | 7602.05 | E 102 St & Park Ave | 7488.24 | 40.7987859 | -73.9523 | 40.7904828 | -73.95033068 | member |
| 11 | 7E477AA47C849CE5 | classic_bike | 2022-01-19 13:27 | 2022-01-19 13:40 | Hancock St & Bedford Ave | 4255.05 | Suydam St & Broadway | 4689.03 | 40.68216564 | -73.95399026 | 40.69544 | -73.93223 | member |
| 12 | 4DC494327327DD7E | classic_bike | 2022-01-31 13:17 | 2022-01-31 13:28 | E 97 St & Madison Ave | 7393.09 | E 85 St & 3 Ave | 7212.05 | 40.787801 | -73.953559 | 40.77801203 | -73.95407149 | member |
| 13 | 2FF6406688A54185 | classic_bike | 2022-01-27 12:45 | 2022-01-27 12:53 | E 97 St & Madison Ave | 7393.09 | E 85 St & 3 Ave | 7212.05 | 40.787801 | -73.953559 | 40.77801203 | -73.95407149 | member |
| 14 | 7BA60209DC4F5607 | classic_bike | 2022-01-04 6:54 | 2022-01-04 6:58 | 5 Ave & E 135 St | 7769.06 | Park Ave & E 124 St | 7682.01 | 40.812191 | -73.937838 | 40.8045555 | -73.9396861 | member |
| 15 | 4E4ADEA1887ACAC3 | classic_bike | 2022-01-04 18:38 | 2022-01-04 18:48 | E 91 St & 2 Ave | 7286.01 | Park Ave & E 124 St | 7682.01 | 40.78115276 | -73.94963041 | 40.8045555 | -73.9396861 | member |
| 16 | 2DB47657BEC9445B | classic_bike | 2022-01-09 14:11 | 2022-01-09 14:16 | E 91 St & 2 Ave | 7286.01 | E 85 St & 3 Ave | 7212.05 | 40.78115276 | -73.94963041 | 40.77801203 | -73.95407149 | member |
| 17 | B0536F12DE7E5311 | classic_bike | 2022-01-19 8:44 | 2022-01-19 8:52 | 5 Ave & E 135 St | 7769.06 | Park Ave & E 124 St | 7682.01 | 40.812191 | -73.937838 | 40.8045555 | -73.9396861 | member |
| 18 | 6D9D2C2222D7E33A | classic_bike | 2022-01-04 13:59 | 2022-01-04 14:05 | 5 Ave & E 135 St | 7769.06 | Park Ave & E 124 St | 7682.01 | 40.812191 | -73.937838 | 40.8045555 | -73.9396861 | member |
| 19 | CD49F9CD64CD6D89 | classic_bike | 2022-01-23 12:24 | 2022-01-23 12:28 | E 91 St & 2 Ave | 7286.01 | E 85 St & 3 Ave | 7212.05 | 40.78115276 | -73.94963041 | 40.77801203 | -73.95407149 | member |
| 20 | 280A334FE7464EB4 | classic_bike | 2022-01-20 18:29 | 2022-01-20 19:20 | E 91 St & 2 Ave | 7286.01 | E 102 St & Park Ave | 7488.24 | 40.78115276 | -73.94963041 | 40.7904828 | -73.95033068 | member |
| 21 | 3C96570C77108EA3 | classic_bike | 2022-01-28 6:28 | 2022-01-28 6:33 | W 41 St & 8 Ave | 6602.03 | E 48 St & 5 Ave | 6626.01 | 40.75640548 | -73.9900262 | 40.75724568 | -73.97805914 | member |
| 22 | D95AD5F2DAD2C0A6 | classic_bike | 2022-01-25 8:21 | 2022-01-25 8:40 | Central Park W & W 91 St | 7453.01 | E 48 St & 5 Ave | 6626.01 | 40.78866499 | -73.96680057 | 40.75724568 | -73.97805914 | member |
| 23 | 9C086BDA0F37CB0E | classic_bike | 2022-01-02 8:49 | 2022-01-02 8:52 | E 82 St & East End Ave | 7049.04 | E 85 St & 3 Ave | 7212.05 | 40.7724607 | -73.9468208 | 40.77801203 | -73.95407149 | member |
| 24 | 8DCFD9368A79B545 | classic_bike | 2022-01-03 8:14 | 2022-01-03 8:18 | E 82 St & East End Ave | 7049.04 | E 85 St & 3 Ave | 7212.05 | 40.7724607 | -73.9468208 | 40.77801203 | -73.95407149 | member |
| 25 | 7B7B5A23EEF32F3E | classic_bike | 2022-01-06 8:40 | 2022-01-06 8:50 | Old Broadway & W 133 St | 7881.09 | Park Ave & E 124 St | 7682.01 | 40.818212 | -73.955277 | 40.8045555 | -73.9396861 | member |
| 26 | 2DAEA88A2670038E | classic_bike | 2022-01-24 15:41 | 2022-01-24 15:49 | 3 Ave & E 100 St | 7414.17 | E 85 St & 3 Ave | 7212.05 | 40.7877214 | -73.94728331 | 40.77801203 | -73.95407149 | member |
| 27 | 12BAFFB86A19D6DA | classic_bike | 2022-01-23 17:13 | 2022-01-23 17:20 | 3 Ave & E 100 St | 7414.17 | E 85 St & 3 Ave | 7212.05 | 40.7877214 | -73.94728331 | 40.77801203 | -73.95407149 | member |

# Dataset in Kaggle

■ Event log of hospital

| | A | B | C | D |
|---|---|---|---|---|
| 1 | patient | action | org:resource | DateTime |
| 2 | patient 0 | First consult | Dr. Anna | 2017-01-02 11:40:11 |
| 3 | patient 0 | Blood test | Lab | 2017-01-02 12:47:33 |
| 4 | patient 0 | Physical test | Nurse Jesse | 2017-01-02 12:53:50 |
| 5 | patient 0 | Second consult | Dr. Anna | 2017-01-02 16:21:06 |
| 6 | patient 0 | Surgery | Dr. Charlie | 2017-01-05 13:23:09 |
| 7 | patient 0 | Final consult | Dr. Ben | 2017-01-09 08:29:28 |
| 8 | patient 1 | First consult | Dr. Anna | 2017-01-02 12:50:35 |
| 9 | patient 1 | Physical test | Nurse Jesse | 2017-01-02 13:59:14 |
| 10 | patient 1 | Blood test | Lab | 2017-01-02 14:20:19 |
| 11 | patient 1 | X-ray scan | Team 1 | 2017-01-06 09:13:40 |
| 12 | patient 1 | Second consult | Dr. Anna | 2017-01-06 10:38:04 |
| 13 | patient 1 | Medicine | Pharmacy | 2017-01-06 11:47:36 |
| 14 | patient 1 | Final consult | Dr. Anna | 2017-01-06 16:49:21 |
| 15 | patient 2 | First consult | Dr. Anna | 2017-01-04 10:02:49 |
| 16 | patient 2 | Physical test | Nurse Jesse | 2017-01-06 09:05:01 |
| 17 | patient 2 | X-ray scan | Team 2 | 2017-01-10 08:05:47 |
| 18 | patient 2 | Blood test | Lab | 2017-01-10 09:12:09 |
| 19 | patient 2 | Second consult | Dr. Anna | 2017-01-12 16:32:00 |
| 20 | patient 2 | Medicine | Pharmacy | 2017-01-13 10:32:51 |
| 21 | patient 2 | Final consult | Dr. Ben | 2017-01-17 11:54:48 |
| 22 | patient 3 | First consult | Dr. Anna | 2017-01-05 16:10:58 |
| 23 | patient 3 | Blood test | Lab | 2017-01-06 08:15:00 |
| 24 | patient 3 | X-ray scan | Team 2 | 2017-01-11 11:36:39 |
| 25 | patient 3 | Physical test | Nurse Corey | 2017-01-11 12:38:58 |
| 26 | patient 3 | Second consult | Dr. Anna | 2017-01-13 10:48:35 |
| 27 | patient 3 | Medicine | Pharmacy | 2017-01-13 13:51:21 |

# Datasets in Papers

- Event dataset from Social Media Platforms (SMP)

| Author | Year | Dataset | Is dataset Public? |
|---|---|---|---|
| Offline models | | | |
| Hu et al. [29] | 2017 | Twitter, Instagram, Newyork taxi trip data | No |
| Sangameshwar et al. [30] | 2017 | Twitter tweets on Natural disaster with lattitude and longitude data | No |
| Chong et al. [31] | 2017 | Avenue, Subway and UCSD | Yes |
| Acharya et al. [32] | 2017 | Surge in power system blackouts and views on web page | No |
| Jin et al. [33] | 2017 | data of Google search, news and twitter from January 1, 2012 to July 31, 2013 | No |
| Rule et al. [34] | 2018 | Charlottesville rally and torch light march past | No |
| Garg et al. [35] | 2018 | First story detection | Yes |
| Repp et al. [36] | 2018 | Event 2012 | Yes |
| Feng et al. [37] | 2018 | ACE 2005 | yes |
| Gupta et al. [38] | 2018 | 6.5 million Tweets on MTV Lady Gaga 2017 event and ICC Championship Trophy 2017 | No |
| Nolasco et al. [40] | 2019 | Brazil's political protest tweets 2013, Zika virus tweets 2015 | No |
| Fedoryszak et al. [41] | 2019 | Full Twitter Firehose, English tweets from United states | No |
| Abbasi et al. [25] | 2019 | Tweets, forums postings, and search query logs on automotive and pharmaceutical industries. | No |
| Fan et al. [26] | 2019 | Hurricane Harvey data in 2017 in Housten | No |
| Yang et al. [27] | 2019 | Multi-domain and Multi-modality Event Detection dataset, and MediaEval SED 2014 | Yes |
| Saeed et al. [3] | 2019 | Football association challenge cup final, US election 2012, Super Tuesday | Yes |
| Roy et al. [28] | 2019 | Hurricane Sandy, New York taxi trip data | Yes |
| Online Models | | | |
| Hasan et al. [47] | 2016 | Event 2012 corpus | Yes |
| Ranganath S et al. [42] | 2016 | Nigerian Election | No |
| Ngyyen et al. [40] | 2017 | FA cup and Super Tuesday | Yes |
| Singh et al. [48] | 2017 | Twitter disaster tweets stream on floods in eastern and southern states, India | No |
| Hasan et al. [44] | 2019 | Event 2012 corpus | Yes |
| Fedoryszak et al. [41] | 2019 | Full Twitter Firehose | No |
| Mukhina et al. [45] | 2019 | Instagram posts on New York | No |
| Hu et al. [46] | 2020 | Real world dataset | Yes |