Analyzing and Predicting Persistence of News Tweets

Maggie Liu[*1], **Jing Wang**[2], **Daniel Preoţiuc-Pietro**[2]
[1]University of British Columbia [2]Bloomberg
yitong.liu@stat.ubc.ca, jwang1621@bloomberg.net, dpreotiucpie@bloomberg.net

Abstract

News is read and consumed differently based on its topic and timeliness to the reader. Some stories attract readers immediately after they are published, while others capture readership consistently over multiple days after their publication, regardless of their overall popularity. This paper studies this less explored facet of news story consumption, which we name persistence, operationalized as the time for a story to reach a certain percent of its total interest. In particular, we study persistence though a novel, publicly available data set of news tweets from 353 news outlets. We perform an extensive linguistic analysis of persistence in social media to uncover the underlying topical and stylistic cues that impact short- or long-term interest in a story. We train several models for predicting persistence that achieve predictive performance of up to 0.353 Spearman correlation when extrapolating to tweets from days unseen in training and retain significant predictive performance even on tweets from accounts unseen in training. The ability to predict news persistence can be useful in several practical applications that drive news and social media consumption including alerting, search ranking or recommendations.

1 Introduction

The vast majority of news outlets follow a 24-hour news cycle to meet the demand for covering news consistently at every hour daily. Combined with the advent of online news and social media, users are thus now faced with a constant onslaught of news stories (Foundation, 2020). Faced with limited user time, past research focused on identifying important stories (Szabo and Huberman, 2010; Bandari et al., 2012) or recommending stories to specific users (Adomavicius and Tuzhilin, 2005; Borges and Lorena, 2010; Özgöbek et al., 2014).

News story *persistence* is a characteristic of news stories that is **distinct** from its overall popularity

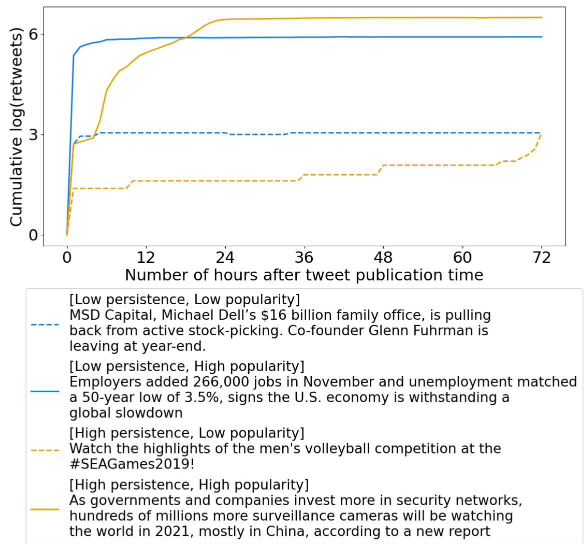


Figure 1: Examples of tweets of news stories of low and high persistence and popularity. Persistence is differentiated by color and popularity is differentiated by line type. These examples highlight that persistence is an orthogonal concept to popularity.

or its relevance to the reader. User interest is dependent on the content and type of the news story: some stories require the reader's immediate attention, such as breaking stories caused by political statements impacting day-to-day activities, financial markets or updates on a natural disaster; while other stories have interest more spread out over time, such as opinions or lifestyle content.

Figure 1 illustrates examples of stories of short-term (low persistence) and long-term (high persistence) interest expressed through news tweets. The short-term interest stories contain breaking news. These attracted immediate interest from users, with the majority of the total interest for the story amassed within 1 hour of their publication time, as the stories were actionable and, potentially, other stories or posts were written later to provide a more detailed analysis. One of the long-term interest stories provides a deeper analysis or report, with interest continuing even after a day of its pub-

[*] Work done during an internship at Bloomberg.

Proceedings of the 13th International Joint Conference on Natural Language Processing and the 3rd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics (Volume 1: Long Papers), pages 719–732
November 1–4, 2023.

lication time, as this content is not as time sensitive. The other long-term interest story contains information that is relevant well after publication (e.g., highlights of a competition), thus attracting sustained interest over time. We highlight that persistence is distinct from the overall popularity of a story, as both types of stories can either be of narrow interest attracting little total readership (e.g., volleyball competition) or can be of broad interest (e.g., security network report).

Contributions The goal of this paper is to study the persistence of news, as represented by news tweets, through their textual content alone at story publication time. The persistence of a news tweet is measured through the time in which it reaches a specific percent of its total interest. Our contributions include:

- A data set of news tweets associated with interest time series, extracted from a wide range of official news Twitter accounts.
- Correlation analysis showing that persistence is a distinct concept to popularity.
- An extensive quantitative analysis of the topics and linguistic features related to persistence.
- Experiments on predicting persistence at news tweet publication time from textual features that reach a predictive score of 0.445 Spearman correlation on held-out stories.
- Evaluation of persistence models on data from future time intervals which achieve 0.353 Spearman correlation. These show promise towards generalizability in a realistic scenario, when training is done frequently using the latest data and predictions happen on data from unseen future time periods.

Due to the nature of our problem framing, we only consider features available at story publication time, such as text. We consider outside the scope of this study the use non-textual information such as information about the identities and features of the authors, use of information from other modalities (e.g. images), as well as information about early interest in the story or cascades.

Applications Accurately predicting persistence of a story at publication time can lead to many direct applications such as:

- News prioritization – Identifying breaking news stories which require immediate attention is useful to prioritize news stories for readers such as investors interested in breaking stories that impact financial markets and trading or other journalists interested in identifying events that need more in-depth coverage.
- Search and Recommendation – Automatically identifying the persistence of stories can be used as a feature in ranking news search results, where long-term interest stories can be boosted in historical searches.
- Content promotion – Automatically flagging stories with high persistence that may benefit from additional promotion through advertisements or re-sharing through platforms available to publishers.
- Content resurfacing – Resurfacing persistent content for weekend editions or news recaps and summaries.

Understanding how news is consumed is a key research question in journalism. In addition, insights into persistence can be used by computational linguists to understand framing and style associated with different types of news and by social scientists to understand or suggest framing that can lead to different types of interest or sharing patterns.

2 Related Work

Ranking Tweets. Ranking tweets is a well studied problem (Nagmoti et al., 2010; Soboroff et al., 2012a,b) and relevance is often correlated with the overall popularity of the post as measured through existing retweet counts or the authoritativeness of the source (Duan et al., 2010). In addition, aging theory (Chen et al., 2003) was proposed for events, wherein a particular event receives boosts in interest when there are new stories about it, while also aging with a factor as time elapses. We expect that our predicted persistence factor will help better identify and surface relevant content, especially in domains such as finance, where timeliness and novelty are major characteristics of relevant news (Ceccarelli et al., 2016).

Cascade Prediction. Cascade prediction aims to model future popularity and sharing patterns of a post based on sharing history up to a point-in-time. This is based on the observation that early rates of popularity are indicative of longer term trends (Szabo and Huberman, 2010). Models to predict cascades usually rely on features derived from past shares such as structural network features, resharer features (e.g., influence in the network) and temporal features (e.g., times between reshares) (Cheng et al., 2014), as well as features derived from the content of the post such as text (Hong et al., 2011) and poster information (Cheng et al., 2014). Several methods are proposed to model arrival times of retweets given past tweet history (Zaman et al., 2014; Zhao et al., 2015; Rizoiu et al., 2017), with time being modeled to account for the circadian nature of users and information decay (Kobayashi and Lambiotte, 2016). Work in this area is complementary to our approach which only uses the textual information available at story publication time. Comparing and combining they two types of approaches would be relevant future work.

Popularity Prediction. A popular related area of past research is popularity prediction. The goal is to predict the upcoming popularity of a news story or tweet measured through metrics such as total retweet counts, irrespective of when these were obtained compared to the posting time. The prediction is performed at posting time using features such as text (Petrovic et al., 2011), images (Wang et al., 2018), social relationships (Petrovic et al., 2011), time of posting (Yu et al., 2015) and account information (Bandari et al., 2012). The task is framed as either classification, where the goal is to predict if the target item will have popularity above a threshold (Jenders et al., 2013) or receives a response (Tsagkias et al., 2009; Petrovic et al., 2011; Artzi et al., 2012) or as regression or ordinal regression, with popularity intervals as the target value (Yu et al., 2015). A related thread is predicting if an article receives comments or responses (Yano et al., 2009; Yano and Smith, 2010), while also factoring in past account behavior and interests (Zhang et al., 2016). The impact of wording on tweet popularity was studied in (Tan et al., 2014) using tweets that have similar content, but were worded differently, uncovering significant predictive effects.

The only paper to study a related concept defined as persistence is (Wu et al., 2011), which studies the decay rate of URLs embedded in tweets by categorizing into two classes based on their shape after peak interest. Our paper presents a new more fine-grained operationalization of the concept as a regression task. We focus on a realistic setup that avoids feature leakage and emphasises model generalizability. We also introduce a more suitable *public* data set for the task by focusing on news tweets from a variety of sources and aim to avoid potential effects caused by measuring the network diffusion of content.

3 Task & Data

We define a news story's persistence as the time that elapses between its publication time and the time the collective interest reaches $T\%$ of its total interest. Note that our persistence metric only relates to the shape of the user interest in a story and is thus distinct to the story's popularity.

3.1 Data Collection

We create a new publicly available data set to study this task, as no other public data set is available. We release the persistence score with each tweet, as well as the full distribution of retweets across time used to compute the persistence score. We use Twitter as the source of our data since the vast majority of news outlets also maintain a Twitter account where they disseminate short news stories to users that use Twitter for reading news. Additionally, at least at the time of the data collection, tweets were mostly displayed to users in temporal order. Finally, Twitter maintains a public API which we can poll regularly to obtain a time series of the number of retweets over time.

We perform our analysis on a set of 353 major news accounts that publish news in English.[1] The list of accounts covers a variety of news sources including major local, national and international outlets across a variety of English-speaking locations.

We use the Twitter API[2] to regularly poll the timelines of all target accounts. Every time we poll, we obtain an updated retweet count and create a time series consisting of the number of retweets of a tweet over time. We use regular 15 minutes intervals to re-collect the same tweets due to Twitter API rate limit restrictions, which also impacts the number of tweets and accounts we can monitor. We argue this does not impact any of the core findings about the concept of persistence, as the content is from diverse accounts, or the text-based methods used to model this concept, as these can be readily applied if other data is available.

3.2 Interest Metric

Over 98% of retweets on Twitter occur within the 72 hours after the tweet was posted. Hence, we compute total interest within the first 72 hours after publication time.

[1]The full list is Table 10 in the Appendix.

[2]https://developer.twitter.com/en/docs/tweets/timelines/api-reference/get-statuses-user_timeline

Time spent	Percent of tweets
< 1 hr	**54.86%**
1-2 hrs	19.04%
2-5 hrs	5.07%
5-12 hrs	1.12%
12-24 hrs	0.21%
24-48 hrs	0.02%
48-72 hrs	0.00%

Table 1: Distribution of persistence score - the time until the retweet count reaches 60% of the total retweets obtained in the first 72 hours after publication.

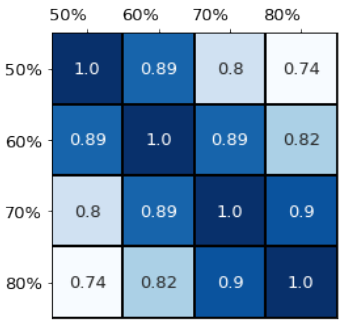


Figure 2: Correlations of persistence score distributions gaining $T\%$ of retweets in the first 72 hours after publication ($p < .0001$)

Each time we poll an account, we retrieve a time series of the number of retweets or likes of each tweet on that account. We found that aggregated retweet and favourite counts have a correlation of $r = 0.999$ ($p < 0.001$) [3], suggesting the two metrics are highly coupled. Hence, in the rest of the paper, we use the total number of retweets a news tweet received as a proxy for interest.

On average, interest in tweets accumulates very fast. As shown in Table 1, over half of the tweets in our dataset reached 60% of their total interest (retweets) within the first hour of publication. We also looked at the distribution of tweets at other $T\%$ thresholds, where $T \in [50, 60, 70, 80]$[4]. Figure 2 illustrates that the persistence scores based on each threshold value have high correlation with each other. When $T = 60$, the distribution has highest mean correlation with other thresholds with minimum deviation. Hence, in the rest of the paper, we define persistence as the amount of time it takes for a story to reach 60% of the total number of retweets it will obtain in the first 72 hours after publication. We also studied the shapes of the interest metric and noticed that most followed a similar pattern, with a burst in interest soon after publication, followed by a decay, with rarely any other significant burst during the decay stage. This evidence further supports our metric choice.

3.3 Data Processing

To assemble the final data set, we apply the following processing steps. First, we only keep tweets posted in English, as identified by the langid.py tool (Lui and Baldwin, 2012). To focus on original content posted by the account, we remove quoted tweets and retweets. If an account publishes a link to the same news story multiple times, we only keep the first tweet in chronological order. We discard the subsequent tweets and associated interest since it would lead to artificially increased interest. Finally, to remove time series having small counts with noisy persistence estimates, we drop all tweets with less than 20 likes and retweets combined.

We process the tweet texts by lower-casing, replacing all URLs and anonymizing all mentions of usernames with a placeholder token. We tokenize tweet text using the DLATK Twitter-aware tokenizer (Schwartz et al., 2017).

3.4 Data Set and Splits

Our final data set consists of 9104 unique tweets from 353 unique news Twitter accounts. In order to test that a predictive model of persistence is robust and able to extrapolate on unseen data, as in a real scenario, we create the following experimental setups:

Temporal The temporal setup reflects a realistic extrapolation scenario, where a model trained to predict persistence is used on unseen data posted after the model was trained. Thus, we split the data based on the time at which it was published, with the training data from the first three days (3 December 2019 – 5 December 2019) and the data held-out for testing authored on the next day (6 December 2019).

Source We next split the data into training, development and test based on the source of the tweet. Thus, this setup reflects the scenario in which a model is trained on data from a set of sources and tested on tweets from unseen sources. We split the accounts randomly in a 60% train, 20% test, 20% development ratio, then assign all tweets from an account to either the train, development or test sets.

Random Finally, we also create a random split, where data is randomly split in training, development and testing using a 70% train, 10% development, 20% test proportion. We create this mostly to

[3] Figure 5 in the Appendix shows the aggregate time series plots of favorites and retweets.

[4] Detailed distribution of other thresholds are shown in Figure 6 in the Appendix.

Number of tweets	Number of twitter accounts
< 10	228
10-20	43
20-50	40
50-100	17
100-200	16
> 200	4

Table 2: Distribution of the number of tweets in our data set across accounts.

measure the gap in generalization induced by the previous more challenging, but realistic scenarios.

Table 1 shows the distribution of the persistence score we will aim to predict at news publication time. The mean persistence time is 2.227 hours, while the median time is 0.861 hours.

4 Analysis

This section details the analysis we performed on our dataset to discover and understand factors impacting persistence of news tweets.

4.1 Difference between Persistence and Popularity

We first aim to study the quantitative difference between persistence and popularity as they are defined as distinct concepts: popularity measures overall interest in a story, while persistence measures the temporal pattern of interest. In this setup, we can use the total number of retweets of a tweet as a proxy for its popularity. A correlation test uncovers a **small correlation between the overall interest** (total retweets) **and the persistence score** (time spent to reach 60% of total retweets) **with** $r = 0.131, p < .0001$. This illustrates that there is a relationship between popular tweets and tweets with higher persistence, albeit the effect size is small, proving these metrics operationalize distinct concepts.

4.2 Account Statistics

Next, we study the relationship between account identity and persistence. Table 2 displays the distribution of tweets in our data set across accounts. We see that only a couple of accounts have more than 250 tweets, with most accounts having between 10–50 tweets (>80%), with an average of 25.7 tweets per account. This shows that the data set is not dominated by content from a few prolific accounts and that our analysis may uncover more general patterns of persistence.

Table 3 presents lists of accounts that publish tweets with the lowest and highest persistence respectively. Only for this analysis, we kept accounts

Low Persistence		High Persistence	
Account Name	Average Persistence	Account Name	Average Persistence
BreakingNews	0.608	amazonnews	8.848
guardiannews	0.805	Echinanews	7.602
PTI_News	0.835	APTNNews	6.580
ELINTNews	0.852	UN_News_Centre	6.451
ReutersBiz	0.872	Consortiumnews	6.183
BNONews	0.875	9NewsMelb	5.854
NBCNightlyNews	0.919	TOLOnews	5.805
WSJ	0.938	IMFNews	5.761
NBCNews	0.976	NYPDnews	5.752
NewsBreaking	1.025	FAANews	5.731

Table 3: Top 10 Twitter accounts with lowest and highest average persistence in our data set. Persistence is measured as the average time (in hours) to reach 60% of the total retweets over 72 hours.

with ≥ 10 tweets in our data set in order to remove noise associated with this statistic when the average is computed over a small number of tweets.

We notice that accounts that post stories with low persistence include accounts that: a) specialize in breaking news (e.g. `NewsBreaking`, `BreakingNews`), which is natural as their goal is to publish stories that need to be read quickly after publication; b) specialize in finance news (e.g. `SEC_News`, `ReutersBiz`, `WSJ`), which is intuitive as financial stories move markets and quick reactions are common; c) are a popular publishing house (e.g. `guardiannews`, `NBCNews`, `PTI_News`).

On the opposite side of the spectrum, the main pattern emerging is regarding accounts of news from organizations or governmental authorities such as the Federal Aviation Administration (`FAANews`), the New York Police department (`NYPDnews`) or Amazon (`amazonnews`). Posts from these accounts are less time critical for their readers as, on a closer analysis, they are more likely to contain information about the organization or are advertising their operations.

We quantitatively study the relationship between the popularity of an account, measured through its number of followers, and news tweet persistence. A correlation test uncovers a statistically significant correlation between the number of followers and persistence ($r = -0.124$, $p < .0001$), highlighting that tweets from more followed accounts attract quick interest from users.

Finally, we also measure the relationship between the number of posts of an account and its average persistence score. A correlation test shows there is a statistically significant correlation between the two ($r = -0.205$, $p < .0002$), highlighting that the

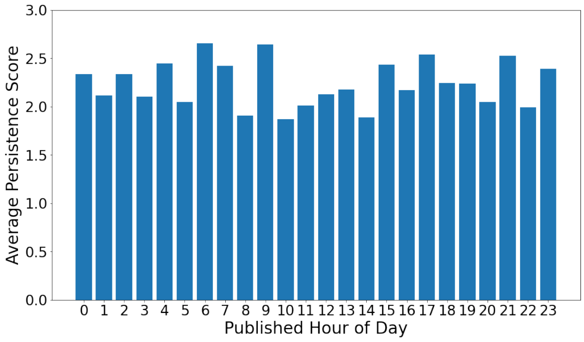


Figure 3: Distribution of the persistence score in our dataset based on time of tweet publication, in UTC.

accounts that publish more content have a lower average persistence score, which means more tweets attract quick interest from users. This could be an effect of tweets being harder to access in the interface by users given the larger volume of tweets from that account.

4.3 Time of Publication

As tweet publication and consumption varies across time (Statista, 2014), we study if there is any interplay between the persistence score and tweet publish time. Metrics of popularity were found to be predictable from tweets (Petrovic et al., 2011) or in the case of the Digg platform which consisted largely of news stories (Szabo and Huberman, 2010). Figure 3 shows the distribution of persistence score according to tweet publication hour of day in UTC. The histogram shows no consistent pattern between the two indicators, with a correlation value of -0.023 ($p = 0.915$). This highlights that publication time may not be a useful feature to model when trying to predict the persistence score.

4.4 Linguistic Analysis

Our final analysis studies the linguistic features associated with persistence in news tweets. We perform this analysis using the following feature sets:

Unigrams We use bag-of-words to represent each tweet as a normalised frequency distribution over the vocabulary consisting of all words used in at least 5 different tweets (5278 tokens)

GloVe Clusters An alternative to bag-of-word features is to represent each tweet as a distribution over group of words that are semantically and/or syntactically similar. The clusters help reducing the feature space and provides additional interpretability to the analysis (Preoţiuc-Pietro et al., 2015). Each tweet is thereby represented as a frequency distribution over these categories.

To conduct the analysis, we use the method from (Schwartz et al., 2013) and used in several other linguistic analyses of user traits or speech acts (Preoţiuc-Pietro et al., 2019). We rank the feature sets previously described using univariate Pearson correlation. Features are normalized to sum up to unit for each tweet. For each feature, we compute correlations independently between its distribution across posts and the persistence score of the tweet.

Figure 4 presents the top unigrams and Table 4 presents the top clusters correlated with low and high persistence. The results uncover clear patterns regarding the type of content that attracts quick interest, thus resulting in low persistence: news about violent events (e.g. 'accused' - clusters denoted by their top word), political stories (dominated by US events in our data e.g. congress, president, trump, impeachment, ukraine) and stories about the economy. In addition, keywords such as 'breaking' and 'updates' are indicative of stories that are deemed breaking also by editors and stories that update ongoing breaking events.

On the opposite side of the spectrum, stories with high persistence show less distinctive patterns, albeit stories referring to people such as interviews or quotes ('his', 'saying') stand out. In addition, common words (e.g. 'and, 'the', 'our') hint that these stories are better edited and formed as compared to breaking stories, while words indicative of questions (e.g. 'can', 'will', 'how') show that analysis stories have higher persistence.

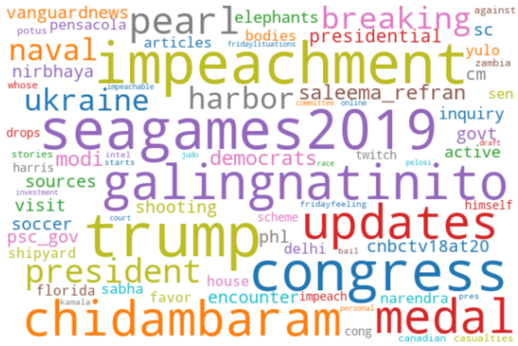
5 Predicting Persistence

Finally, we experiment with a series of approaches to predicting the persistence score of unseen news tweets including both linear models and models using pretrained contextual embedding models. We train regression models with three diverse scenarios of data splits as detailed in Section 3.4. For source split and random split, we train and evaluate performance over ten runs using different random splits for Twitter accounts and tweets respectively. For temporal split, we train and evaluate over ten runs using different random seeds.

5.1 Methods

Linear Model We use a linear regression model[5] with Elastic Net regularization tuned on the development set. We use the textual features from the tweet text as represented by the GloVe cluster to

[5]Implemented with SGDRegressor in scikit-learn



High Persistence Low Persistence

Figure 4: Unigrams with the highest Pearson correlations with high and low persistence shown as word clouds. The size of the unigram is scaled by its correlation with the persistence metric (statistically significant with $p < 0.01$).

High Persistence		**Low Persistence**	
r	**Words**	r	**Words**
.0492	infrastructure, operational, processing, irrigation, implementation, regulatory, technologies, improvements, automation, contractors	-.0406	accused, rape, abuse, arrested, allegedly, investigation, charged, murder, alleged, charges
.0490	his, after, says, more, been, their, than, into, years, being	-.0399	president, election, presidential, govt, minister, administration, voters, candidates, candidate, urges
.0409	saying, asked, asking, jail, sending, angry, mess, shut, happened, stopped	-.0297	medal, medals, swimming, meters, diver, medalist, hike, tennis, cycling, fitness
.0389	adoption, adopted, puppy, polar, puppies, lion, panda, pandas, unicorn, adopt	-.0270	economy, largest, economic, risk, increase, higher, concerns, measures, pollution, growth

Table 4: Pearson correlations between high and low persistence and text features. All correlations are significant at $p < 0.001$. Words in a category are sorted by frequency in our data set.

which they belong. We also create features from the Twitter account, as Section 4.2 showed there is a relationship between Twitter account and average persistence. Thus, similarly to the tweet text representation, we create an additional distinct bag-of-cluster representation from the Twitter account description and the number of followers associated with the account as an additional numerical feature.

In the learning stage, we fine-tune the linear regression model for 1000 epochs using the stopping criteria tol, regularization multiplier α, prediction accuracy threshold ϵ, and L1 ratio. We use dimension 200 text features created using GloVe 200 clusters based on preliminary results. When splitting the data by source, we use tol = 1e-3, α = 1e-5, ϵ= 0.5, and $L1\_ratio$ = 0.4. When splitting the data by publication time, we use tol = 1e-7, α = 1e-5, ϵ = 0.5, and $L1\_ratio$ = 1. When splitting the data randomly, we use tol = 1e-5, α = 1e-7, ϵ = 0.5, and $L1\_ratio$ = 0.

BERT In addition to the linear modes, we also utilize BERT (Devlin et al., 2019) as a representative for Transformer-based models. In our experiment, we fine-tune the model for persistence prediction with an output dense layer for regression. Similar to the Linear Regression model, we consider possible feature influence of textual inputs from tweet text and account description and follower numbers from tweet metadata. The numerical value of follower number is concatenated with pooled output vector of BERT before passing to the regression layer.

We fine-tuned the model for persistence prediction with pre-trained Bert-base-cased model (12-layer, 768-hidden, 12-heads, 110M parameter) for 3 epochs. The initial learning rate is set to 5e-5 and the batch size in training is set to 16. Additional details for facilitating reproducibility are presented in Section B in the Appendix.

5.2 Results

Table 5 shows the results of predictive models averaged across 10 runs. Results are measured using standard correlation metrics such as Spearman ρ, Pearson r and Root Mean Squared Error (RMSE). Models are tuned to obtain best Spearman ρ in the development set.

Model	Spearman ρ	Pearson r	RMSE
Temporal Split			
Random	0.019 ($\pm$0.020)	0.020 ($\pm$0.019)	6.659 ($\pm$0.061)
Linear Regression	0.285 ($\pm$0.005)	0.288 ($\pm$0.003)	1.081 ($\pm$0.029)
BERT	**0.353** ($\pm$0.009)	**0.321** ($\pm$0.009)	**0.977** ($\pm$0.002)
Source Split			
Random	0.014 ($\pm$0.020)	0.015 ($\pm$0.016)	6.875 ($\pm$0.437)
Linear Regression	0.111 ($\pm$0.103)	0.120 ($\pm$0.099)	1.121 ($\pm$0.042)
BERT	**0.167** ($\pm$0.042)	**0.167** ($\pm$0.033)	**1.110** ($\pm$0.037)
Random Split			
Random	-0.010 ($\pm$0.029)	-0.009 ($\pm$0.027)	7.240 ($\pm$0.138)
Linear Regression	0.323 ($\pm$0.009)	0.333 ($\pm$0.011)	1.078 ($\pm$0.011)
BERT	**0.445** ($\pm$0.013)	**0.437** ($\pm$0.020)	**1.014** ($\pm$0.012)

Table 5: Predictive results for persistence on the three different data splits. Except for the Random guess results, all correlations are statistical significant ($p < 0.001$).

Features	Spearman ρ
Temporal Split	
Tweet Text	0.227 ($\pm$0.011)
Tweet Text + Follower Number	0.212 ($\pm$0.015)
Tweet Text + Account Description	<u>0.353</u> ($\pm$0.009)
Tweet Text + Account Description + Follower Number	0.351 ($\pm$0.005)
Source Split	
Tweet Text	<u>0.167</u> ($\pm$0.042)
Tweet Text + Follower Number	0.159 ($\pm$0.056)
Tweet Text + Account Description	0.064 ($\pm$0.081)
Tweet Text + Account Description + Follower Number	0.063 ($\pm$0.096)
Random Split	
Tweet Text	0.290 ($\pm$0.026)
Tweet Text + Follower Number	0.307 ($\pm$0.016)
Tweet Text + Account Description	<u>0.445</u> ($\pm$0.013)
Tweet Text + Account Description + Follower Number	0.441 ($\pm$0.018)

Table 6: Predicted results of BERT model with different combinations of features ($p < 0.001$).

In all three experimental setups, the results show that both linear regression and BERT models predict persistence consistently above random chance, even on data from unseen time intervals and from unseen sources.

The results show that the model with pre-trained contextual embeddings outperforms linear regression model with handcrafted features, with an average Spearman correlation increase of 0.05 on the source split, 0.07 on the temporal split and 0.12 on the random split.

Results from the temporal split show that the model for predicting persistence can generalize on future data than that used in training. The models achieve a Spearman correlation of 0.353 (BERT) and of 0.285 (Linear Regression). This setup is the most realistic for real-world applications, where data from sources of interest are available for training, but the predictions are on future data from a time range not observed in training. This performance is naturally lower than on the random split setup where there is a temporal and source overlap that the model can leverage (Huang and Paul, 2018). Our results highlight the importance of model retraining using the most recent data.

The effect of source drift introduces the largest impact on model performance with Spearman correlation dropping to 0.167 from 0.445 (BERT) and to 0.111 from 0.323 (Linear Regression) when compared with the random split setup. The large standard deviation of the source split compared to the others demonstrates that the performance is impacted by the topic and style differences in posts by accounts in training data and unseen test data.

Overall, we note that the task of predicting persistence is challenging as we used reasonable NLP methods with a vast range of feature types in our experiments for a best effort modelling approach. We note that the models can predict better than chance in all setups and show good generalization in out-of-time experiments, which prove generalizability and robustness in a real-world scenario. The raw Spearman correlation values are in a similar range as those for related tasks such as predicting popularity (retweets) using content alone e.g. 0.229 – 0.358 ρ with textual, social and visual features (Wang et al., 2018).

Table 6 shows the results (Spearman correlation) of the BERT model with various combinations of features. The best results on both the temporal and random splits are obtained with the combined textual features from tweet text body and account description. However, the best performance on the source split is obtained only using tweet text body. Adding account description features significantly dropped the performance, showing that the account features do not extrapolate well for predicting persistence of other accounts.

6 Discussion and Future Work

The key assumption for the experimental setup used in the paper is that persistence of news can be predicted primarily through textual content, and other limited information information that is available at story publication time (follower count, account description). This enables applications such as news prioritization, which aims to identify a priori and more prominently feature stories that are likely to be breaking and attract a high interest immediately after their publication.

The experimental design around data used in the paper, including ranges and splits, also aims to mirror the approach that is used in a system designed for news prioritization. Experiments with the temporal split of the data emulates a setup in which the model can be retrained regularly using recent stories and their persistence-related observations. Predictions can then be made on new data from a recent day unseen in training, which can see a drift in topics or events. The temporal split assumes that the news sources remain constant across training and prediction within a given date range. The source split experiments aim to show the generalizability of the model, and establish its baseline performance on data from a new data source. Once this new data source is included for an extended period of time, retraining with that data would bring performance close to that of the temporal split experiments.

Other applications that can use a news persistence measure include search and recommendation, content promotion, and content resurfacing. For most of these applications, the prediction of persistence does not need to be made at story publication time and can benefit from measures of initial interest in the story. We consider this information to be complementary to that available through the story content. Combining our methods with those from the cascade prediction literature would be valuable future work for understanding persistence and further improving the predictive performance.

Additional structured information about the account or publisher, such as network information, can also be used to further improve predictive performance.

Separately, while we focused on written news stories, the concept of persistence can be relevant to other settings and types of content, such as images, and information or transcripts from videos or audio.

7 Conclusions

This paper presented a computational linguistic study of the concept of persistence. We operationalized this concept by measuring the time in which tweets from news accounts obtain 60% of their total interest, as defined through counting the number of retweets and showed this is different than news popularity or importance. We performed several analyses to uncover patterns in our data and trained machine learning models that are able to predict at tweet publication time its persistence score significantly above chance, even for accounts unseen in training, demonstrating the robustness and ability to generalize. We expect this task to be useful to downstream applications such as search, recommendations and identifying content to promote. Future work will study additional modelling approaches for this challenging task, use social factors like cascades and information about users engaging with the content to study and predict persistence and the impact of the persistence score in predicting cascades.

8 Limitations

We only studied English-language news tweets published on a given week. Working with more data would only improve the prediction results. We acknowledge the feature analysis and magnitude of results may not hold to other languages, types of news or different time ranges, but our methods for analysis are easily applicable and transferable to any new domain, language or time range.

Acknowledgements

We would like to thank Umut Topkara, Diego Ceccarelli, Jonny Marks and other members of the Bloomberg AI Group for valuable discussions and feedback on the paper.

Ethics Statement

Our work complies with Twitter's data policy for research.[6]

References

Gediminas Adomavicius and Alexander Tuzhilin. 2005. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE transactions on knowledge and data engineering*, 17(6):734–749.

Yoav Artzi, Patrick Pantel, and Michael Gamon. 2012. Predicting responses to microblog posts. In *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 602–606, Montréal, Canada. Association for Computational Linguistics.

Roja Bandari, Sitaram Asur, and Bernardo Huberman. 2012. The pulse of news in social media: Forecasting popularity. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 6, pages 26–33.

Hugo L Borges and Ana C Lorena. 2010. A survey on recommender systems for news data. *Smart Information and Knowledge Management: Advances, Challenges, and Critical Issues*, pages 129–151.

[6]https://developer.twitter.com/en/developer-terms/agreement-and-policy

Diego Ceccarelli, Francesco Nidito, and Miles Osborne. 2016. Ranking financial tweets. In *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval*, pages 527–528.

Chien Chin Chen, Yao-Tsung Chen, Yeali Sun, and Meng Chang Chen. 2003. Life cycle modeling of news events using aging theory. In *European Conference on Machine Learning*, pages 47–59. Springer.

Justin Cheng, Lada Adamic, P Alex Dow, Jon Michael Kleinberg, and Jure Leskovec. 2014. Can cascades be predicted? In *Proceedings of the 23rd international conference on World wide web*, pages 925–936.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Yajuan Duan, Long Jiang, Tao Qin, Ming Zhou, and Heung-Yeung Shum. 2010. An empirical study on learning to rank of tweets. In *Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010)*, pages 295–303, Beijing, China. Coling 2010 Organizing Committee.

Knight Foundation. 2020. American views 2020: Trust, media and democracy. https://knightfoundation.org/wp-content/uploads/2020/08/American-Views-2020-Trust-Media-and-Democracy.pdf. Accessed: 5-17-2023.

Liangjie Hong, Ovidiu Dan, and Brian D Davison. 2011. Predicting popular messages in twitter. In *Proceedings of the 20th international conference companion on World wide web*, pages 57–58.

Xiaolei Huang and Michael J. Paul. 2018. Examining temporality in document classification. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 694–699, Melbourne, Australia. Association for Computational Linguistics.

Maximilian Jenders, Gjergji Kasneci, and Felix Naumann. 2013. Analyzing and predicting viral tweets. In *Proceedings of the 22nd international conference on world wide web*, pages 657–664.

Ryota Kobayashi and Renaud Lambiotte. 2016. Tideh: Time-dependent hawkes process for predicting retweet dynamics. In *Tenth International AAAI Conference on Web and Social Media*.

Marco Lui and Timothy Baldwin. 2012. langid.py: An off-the-shelf language identification tool. In *Proceedings of the ACL 2012 System Demonstrations*, pages 25–30, Jeju Island, Korea. Association for Computational Linguistics.

Rinkesh Nagmoti, Ankur Teredesai, and Martine De Cock. 2010. Ranking approaches for microblog search. In *2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*, volume 1, pages 153–157. IEEE.

Özlem Özgöbek, Jon Atle Gulla, and Riza Cenk Erdur. 2014. A survey on challenges and methods in news recommendation. In *WEBIST (2)*, pages 278–285.

Sasa Petrovic, Miles Osborne, and Victor Lavrenko. 2011. Rt to win! predicting message propagation in twitter. In *Fifth International AAAI Conference on Weblogs and Social Media*.

Daniel Preoţiuc-Pietro, Mihaela Gaman, and Nikolaos Aletras. 2019. Automatically identifying complaints in social media. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 5008–5019, Florence, Italy. Association for Computational Linguistics.

Daniel Preoţiuc-Pietro, Vasileios Lampos, and Nikolaos Aletras. 2015. An analysis of the user occupational class through Twitter content. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 1754–1764, Beijing, China. Association for Computational Linguistics.

Marian-Andrei Rizoiu, Lexing Xie, Scott Sanner, Manuel Cebrian, Honglin Yu, and Pascal Van Hentenryck. 2017. Expecting to be hip: Hawkes intensity processes for social media popularity. In *Proceedings of the 26th International Conference on World Wide Web*, WWW '17, page 735–744.

H Andrew Schwartz, Johannes C Eichstaedt, Margaret L Kern, Lukasz Dziurzynski, Stephanie M Ramones, Megha Agrawal, Achal Shah, Michal Kosinski, David Stillwell, Martin EP Seligman, et al. 2013. Personality, gender, and age in the language of social media: The open-vocabulary approach. *PloS one*, 8(9):e73791.

H. Andrew Schwartz, Salvatore Giorgi, Maarten Sap, Patrick Crutchley, Lyle Ungar, and Johannes Eichstaedt. 2017. DLATK: Differential language analysis ToolKit. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 55–60, Copenhagen, Denmark. Association for Computational Linguistics.

Ian Soboroff, Dean McCullough, Jimmy Lin, Craig Macdonald, Iadh Ounis, and Richard McCreadie. 2012a. Evaluating real-time search over tweets. *Proc. of ICWSM*.

Ian Soboroff, Iadh Ounis, Craig Macdonald, and Jimmy J Lin. 2012b. Overview of the trec-2012 microblog track. In *TREC*, volume 2012, page 20.

Statista. 2014. *Distribution of daily Twitter activity worldwide as of April 2014, by hour of day*. https://www.statista.com/statistics/

188741/global-twitter-activity-by-hour-of-day/ (accessed May 23, 2020).

Gabor Szabo and Bernardo A Huberman. 2010. Predicting the popularity of online content. *Communications of the ACM*, 53(8):80–88.

Chenhao Tan, Lillian Lee, and Bo Pang. 2014. The effect of wording on message propagation: Topic- and author-controlled natural experiments on Twitter. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 175–185, Baltimore, Maryland. Association for Computational Linguistics.

Manos Tsagkias, Wouter Weerkamp, and Maarten de Rijke. 2009. Predicting the volume of comments on online news stories. In *Proceedings of the 18th ACM Conference on Information and Knowledge Management*, CIKM '09, page 1765–1768.

Ke Wang, Mohit Bansal, and Jan-Michael Frahm. 2018. Retweet wars: Tweet popularity prediction via dynamic multimodal regression. In *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 1842–1851. IEEE.

Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Rémi Louf, Morgan Funtowicz, et al. 2019. Huggingface's transformers: State-of-the-art natural language processing. *arXiv preprint arXiv:1910.03771*.

Shaomei Wu, Chenhao Tan, Jon M Kleinberg, and Michael W Macy. 2011. Does bad news go away faster? *Proc. of ICWSM*.

Tae Yano, William Cohen, and Noah A Smith. 2009. Predicting response to political blog posts with topic models. In *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 477–485.

Tae Yano and Noah A Smith. 2010. What's worthy of comment? content and comment volume in political blogs. In *Fourth international AAAI conference on weblogs and social media*.

Haihao Yu, Xu Feng Bai, ChengZhe Huang, and Haoliang Qi. 2015. Prediction of users retweet times in social network. *International Journal of Multimedia and Ubiquitous Engineering*, 10(5):315–322.

Tauhid Zaman, Emily B Fox, Eric T Bradlow, et al. 2014. A bayesian approach for predicting the popularity of tweets. *The Annals of Applied Statistics*, 8(3):1583–1611.

Qi Zhang, Yeyun Gong, Jindou Wu, Haoran Huang, and Xuanjing Huang. 2016. Retweet prediction with attention-based deep neural network. In *Proceedings of the 25th ACM international on conference on information and knowledge management*, pages 75–84.

Qingyuan Zhao, Murat A. Erdogdu, Hera Y. He, Anand Rajaraman, and Jure Leskovec. 2015. Seismic: A self-exciting point process model for predicting tweet popularity. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '15, page 1513–1522.

A Choice of Interest Metric

Figure 5 shows the aggregated cumulative distributions of retweet counts and like counts over time as a fraction of the total number of retweets within the first 72 hours of publication. The Pearson correlation between the aggregated number of retweets and aggregated number of favorites at each hour after publication, up until 72 hours, across all tweets in our dataset, is 0.999, significant at $p < 0.001$. Additionally, the Pearson correlation between the total number of retweets and total number of favorites within the first 72 hours after publication, is 0.814, significant at $p < 0.01$. The figure, along with these correlations show that retweets and favorites are congruent proxies for reader interest.

Figure 6 shows histograms of the number of hours it took for tweets in our dataset to reach $T\%$ of their total retweets within 72 hours of publication, at thresholds $T \in \{50,60,70,80\}$.

B Hyperparameters

We trained our model with HuggingFace Transformers (Wolf et al., 2019). In this section we elaborate on the hyperparameters and training details involved in persistence predictions for individual split scenarios in order to facilitate reproducibility.

B.1 Linear Models

In the learning stage, we tune the linear regression model for m epochs, where $m \in \{1000, 5000\}$ using stopping criteria *tol* $\in$ {1e-3, 1e-5, 1e-7, 1e-9}, regularization multiplier $\alpha \in$ {1e-3, 1e-5, 1e-7}, prediction accuracy threshold $\epsilon \in$ {0.5, 1e-1, 1e-3, 1e-5}, and L1 ratio $\in$ {0, 0.2, 0.4, 0.6, 0.8, 1} where an L1 ratio of 0 corresponds to L2 loss and an L1 ratio of 1 corresponds to L1 loss. We use GloVe 200 clusters as features. The final best parameters are determined by validation set performance based on Spearman Correlation as shown in Table 7.

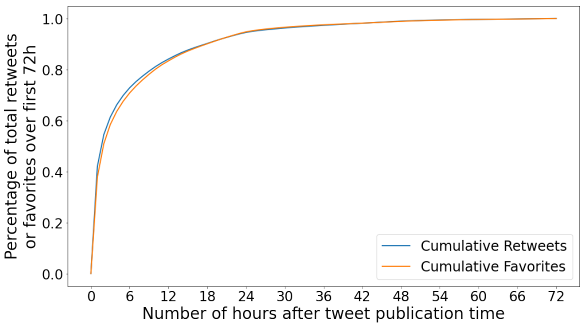


Figure 5: Cumulative distribution of retweet counts and favorite counts over time as a fraction of the total number of retweets over 72 hours.

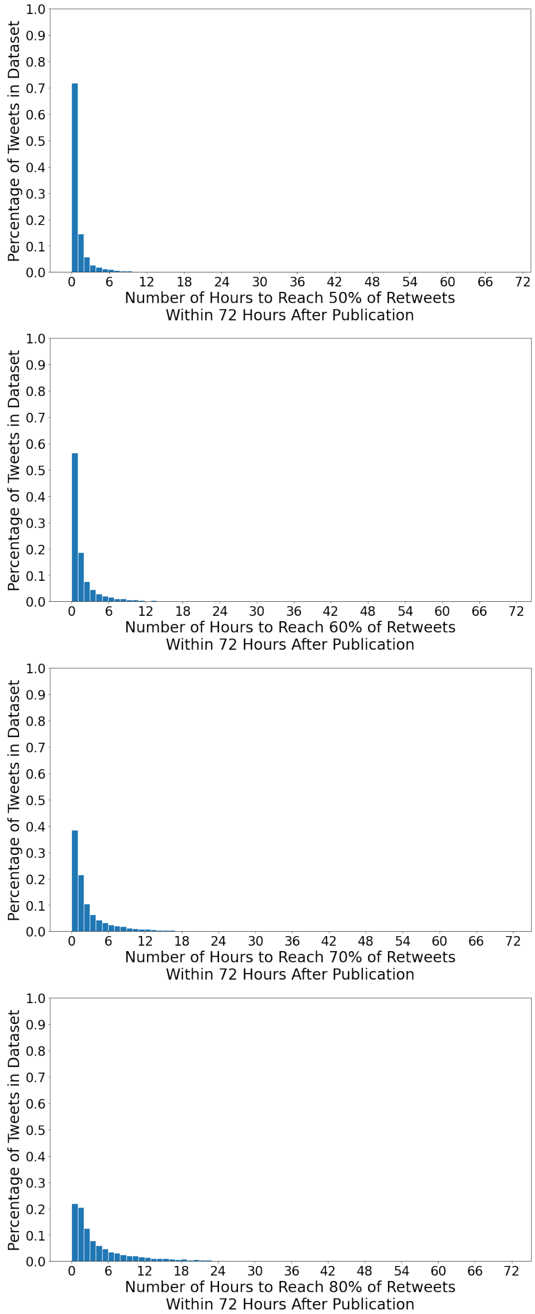


Figure 6: Histogram of time to reach $T\%$ of total retweets within the first 72 hours of publication.

B.2 BERT

We fine-tune the transformer network with pre-trained Bert-base-cased model for m epochs, where $m \in \{1, 3\}$. We experiment with learning rate $lr \in$ {1e-4, 1e-5, 5e-5}, weight decay in {1e-6, 1e-8}, and training batch size varies in $\{16, 32\}$. We also hypertune warm-up ratio in values of $\{0.1, 0.01\}$. Evaluation batch size is set to 64. The final best parameters in Table 8 are determined by the Spearman Correlation values on development set across

Linear Model					
Scenario	epochs	α	tol	ϵ	$L1\_ratio$
Source Split	1000	1e-5	1e-3	0.5	0.4
Temporal Split	1000	1e-5	1e-7	0.5	1.0
Random Split	1000	1e-7	1e-5	0.5	0.0

Table 7: Linear model parameters

BERT					
Scenario	epochs	lr	decay	warmup	batch
Source Split	3	5e-5	1e-6	0.1	16
Temporal Split	3	5e-5	1e-8	0.1	16
Random Split	3	5e-5	1e-8	0.1	16

Table 8: BERT model parameters

Model	**Spearman ρ**		**Pearson r**		**RMSE**	
	Dev	**Test**	**Dev**	**Test**	**Dev**	**Test**
Source Split						
Random	-0.002 ($\pm$0.018)	0.014 ($\pm$0.020)	-0.004 ($\pm$0.019)	0.015 ($\pm$0.016)	6.892 ($\pm$0.561)	6.875 ($\pm$0.437)
Linear Regression	0.200 ($\pm$0.087)	0.111 ($\pm$0.103)	0.214 ($\pm$0.087)	0.120 ($\pm$0.099)	1.104 ($\pm$0.030)	1.121 ($\pm$0.042)
BERT	**0.242** ($\pm$0.035)	**0.157** ($\pm$0.044)	**0.247** ($\pm$0.038)	**0.167** ($\pm$0.033)	**1.102** ($\pm$0.031)	**1.110** ($\pm$0.037)
Temporal Split						
Random	0.012 ($\pm$0.034)	0.019 ($\pm$0.020)	0.011 ($\pm$0.030)	0.020 ($\pm$0.019)	7.411 ($\pm$0.219)	6.659 ($\pm$0.061)
Linear Regression	0.297 ($\pm$0.003)	0.285 ($\pm$0.005)	0.309 ($\pm$0.002)	0.288 ($\pm$0.003)	1.056 ($\pm$0.035)	1.081 ($\pm$0.029)
BERT	**0.457** ($\pm$0.013)	**0.357** ($\pm$0.009)	**0.469** ($\pm$0.008)	**0.321** ($\pm$0.009)	**1.071** ($\pm$0.017)	**0.977** ($\pm$0.002)
Random Split						
Random	-0.003 ($\pm$0.045)	-0.010 ($\pm$0.029)	-0.002 ($\pm$0.041)	-0.009 ($\pm$0.027)	7.150 ($\pm$0.337)	7.240 ($\pm$0.138)
Linear Regression	0.314 ($\pm$0.026)	0.323 ($\pm$0.009)	0.327 ($\pm$0.022)	0.333 ($\pm$0.011)	1.073 ($\pm$0.018)	1.078 ($\pm$0.011)
BERT	**0.463** ($\pm$0.041)	**0.444** ($\pm$0.016)	**0.467** ($\pm$0.036)	**0.437** ($\pm$0.020)	**1.014** ($\pm$0.010)	**1.014** ($\pm$0.012)

Table 9: Comparison of predictive results for persistence on the three different data splits in development and test stages. Except for the Random guess results, all correlations are statistical significant ($p < 0.001$).

ten trials for each split scenario.

C Development Data Set Performance

Using the parameters tuned in Appendix B, we make a through comparison of performance between development and test datasets across all models, all split scenarios. The full results are shown in Table 9.

D Runtime

Our experiments on linear models were conducted on CPU with 8 cores. Averaged training time for one combination of hyperparameters is 0.672 seconds. Averaged inference time for each single data point is 0.082 milliseconds.

For BERT, the averaged training time is approximately 2 minutes per epoch on one GPU with 6 cores. Averaged inference time for each single data point is 2.207 milliseconds.

E Twitter Accounts

Table 10 lists the 353 accounts from which we retrieved the tweets for our final dataset.

10News	11AliveNews	16WAPTNews	1TVNewsAF	4029news	41actionnews	69News
6News	7News	7NewsMelbourne	7NewsSydney	8NEWS	8NewsNow	9NEWS
9NewsAUS	9NewsGoldCoast	9NewsMelb	9NewsSyd	ABBgroupnews	ABC13News	ABC7News
ABCNewsLive	ABCNewsPR	ABCWorldNews	ABSCBNNews	AFnewsroom	AHRQNews	AJENews
APTNNews	ARYNEWSOFFICIAL	AbacusNews	AbbottNews	AccesswireNews	AmericaNewsroom	ArianaNews_
Automotive_News	BBCLondonNews	BBCNews	BBCNewsAsia	BBCNewsNI	BBCNewsnight	BBCScienceNews
BBCScotlandNews	BBYNews	BNONews	BangkokPostNews	BofA_News	BreakingNews	BreakingNewsKE
BreakingNewsUK	Breakingviews	BreitbartNews	BuzzFeedNews	CBCNews	CBKKenya	CBNNews
CBS21NEWS	CBS7News	CBSEveningNews	CBSNews	CBSNewsPress	CNBCTV18News	CNBCnow
CNETNews	CTVBarrieNews	CTVNews	CWUnews	CeresNews	Channel4News	ChannelNewsAsia
CityNews	Consortiumnews	CordCuttersNews	Cyclingnewsfeed	DDNewsLive	DMAnews1	DPRK_News
DeltaNewsHub	Diageo_News	DuPont_News	EBA_News	EDARNEWS	EDNYnews	EDVAnews
ELINTNews	EMA_News	EUCouncilTVNews	EYnews	Echinanews	EnergyVoiceNews	FAANews
FAOnews	FATFNews	FOX21News	FOX2News	FOX46News	FOX61News	FTMarkets
FarmsNews	FoxNewsSunday	GroundUp_News	HPE_News	HRBlockNews	HawaiiNewsNow	HealthITNews
HighNorthNews	HuaweiEUNews	IAFFNewsDesk	ICOnews	IIGCCnews	IMFNews	IRSnews
InstaNewsAlerts	IrrawaddyNews	JNJNews	JTAnews	JovemPanNews	KAKEnews	KARK4News
KATUNews	KATVNews	KBTXNews	KCCINews	KGETnews	KGWNews	KHNews
KHONnews	KHQLocalNews	KRIS6News	KSNNews	KSNTNews	KTREnews	KTULNews
KUTV2News	KXAN_News	KitcoNewsNOW	LEX18News	LSEnews	LeedsNews	LivEchonews
Live5News	Local4News	MENnewsdesk	MFB_NEWS	MPRnews	MSFTnews	MastercardNews
MercoPressNews	MetroUKNews	MylanNews	NAR	NBCNews	NBCNewsBusiness	NBCNewsPR
NBCNewsWorld	NBCNightlyNews	NDILnews	NDOHnews	NEWSTALK1010	NOLAnews	NOPDNews
NTSB_Newsroom	NTVNEWS	NYDailyNews	NYPDnews	Nairobi_News	NatureNews	NewStatesman
News12CT	News3LV	NewsBFM	NewsBreaking	NewsDayZimbabwe	NewsHour	News_8
News_Executive	Newsbreak_Lotus	NewsfromScience	NewshubBreaking	NewshubNZ	NewsroomNZ	NewstalkFM
OPP_News	PGPDNews	PIX11News	PMDNewsGov	PNCNews	POWER987News	PRI_News
PTI_News	PTVNewsOfficial	PennDOTNews	PhilstarNews	RI_News_Alert	RealHKNews	Recode
Region8News	ReutersBiz	ReutersUS	SABCNewsOnline	SABreakingNews	SAPoliceNews	SAfmnews
SCMPNews	SDCAnews	SDNYnews	SDOHnews	SEC_News	SPOTNEWSonIG	STVNews
SkyNews	SkyNewsAust	SkyNewsBreak	SkyNewsPolitics	SkySportsNews	SneakerNews	SpaceNews_Inc
Starbucksnews	StockNewsNow	TOIIndiaNews	TOLOnews	TVAnews	TargetNews	TeamNews24
TelegraphNews	TfLTrafficNews	TheBuffaloNews	TheCitizen_News	TheCompoundNews	TheHackersNews	TheNationNews
TheRealNews	TheStar_news	The_Japan_News	TransferNewsCen	UN_News_Centre	UPS_News	UrgentNews911
VOANews	VerizonNews	WAVY_News	WBBJ7News	WBRCnews	WBTV_News	WCBINEWS
WDKYnews	WDPAnews	WDRBNews	WFSBnews	WGNNews	WHSVnews	WISCTV_News3
WISN12News	WMCActionNews5	WNEMTV5news	WOWK13News	WOWT6News	WSJ	WSJecon
WTAJnews	WWLP22News	WXXINews	WrestlingNewsCo	XHNews	YahooNews	YonhapNews
YourAnonNews	ZBCNewsonline	ZeeNews	abc7newsbayarea	abcnews	airnewsalerts	amazonnews
amna_newseng	bmsnews	catalannews	cbcnewsbc	cleveland19news	crikey_news	crypt0snews
dailynewstz	dailystarnews	dallasnews	detroitnews	dwnews	eha_news	enews
euronews	fbnewsroom	fcpsnews	firstdraftnews	fox32news	fox8news	foxcarolinanews
geonews_english	globalnews	globalnewsto	globaltimesnews	gmanews	goodnewsfinland	guardiannews
gulf_news	haveigotnews	hongkong_news	i24NEWS_EN	irish_news	itvnews	kcranews
kfdmnews	koconews	komonews	kron4news	ksatnews	ksdknews	ksfynews
kxly4news	ladailynews	lcdnews	lufthansaNews	mcpnews	mercnews	natnewswatch
news24tvchannel	news5wcyb	news6wkmg	news8news	newsbusters	newschambers	newschannelnine
newscientist	newscomauHQ	newsobserver	nknewsorg	nowthisnews	rnz_news	rtenews
rthk_enews	standardnews	statnews	supchinanews	thenews_intl	thenewsminute	usnews
vanguardngrnews	vicenews	vicsesnews	vmwarenews	wgbhnews	witfnews	wsfa12news
wwmtnews	wyffnews4	ylenews				

Table 10: List of Twitter accounts involved in the data analysis and experiments