International Journal of Research in Engineering, Science and Management
Volume 4, Issue 9, September 2021
https://www.ijresm.com | ISSN (Online): 2581-5792

Modeling Real Time Energy Utilization Patterns as a Scalable Sensing Systems Employing Social Media

Priyanka Kote[1*], M. Dhananjay[2]
[1]*M.Tech. Student, Department of Computer Science & Engineering, Guru Nanak Dev Engineering College, Bidar, India*
[2]*Associate Professor, Department of Computer Science & Engineering, Guru Nanak Dev Engineering College, Bidar, India*

***Abstract*: The theory of this undertaking is that subjects, communicated through huge scope web-based media organizations, rough power usage occasions (e.g., utilizing high force utilization gadgets like a dryer) with high precision. Customarily, scientists have proposed the utilization of keen meters to demonstrate gadget specific power usage designs. In any case, these methods experience the ill effects of versatility and cost difficulties. To alleviate these difficulties, we propose an online media network-driven model that uses huge scope text based and geospatial information to estimated power usage designs, without the requirement for actual equipment frameworks (e.g., like keen meters), therefore giving a promptly versatile wellspring of information. The technique is approved by considering the issue of power use disaggregation, where energy utilization rates from a nine-month duration in San Diego, combined with 1.8 million tweets from a similar area and interval of time, are used to consequently decide exercises that require huge or modest quantities of power to achieve. The framework decides 200 points on which to identify power related occasions and finds 38 of these to be substantial descriptors of energy usage. At long last, the generalizability of our model is contrasted and a climate based model, given by the U.S. Branch of Energy.**

***Keywords*: Social media networks, Smart meters, Weather based model, Big data, Machine Learning.**

1. Introduction

Social networking networks such as Twitter show Big Data" features which present threats and information discovery opportunities related to threat identification. For example, Twitter generates more than 500 million tweets per day (volume), corresponding to about one message per 173 microseconds (velocity). Those tweets contain a combination of text content, geo-details, video, etc. (Variety) [2].

Social Networking Network models can serve as an integrated, universal sensing device that provides physical sensor prediction with this additional benefit of 1) beings customizable 2) freely available 3) having lowers configuration and maintenances costs relative to certain physicals sensors (e.g. smarts meters or smart plug). Social networking sites such as Twitter, Google and Facebook process data every day from 12 terabytes (10^12) to the 20 petabyte (10^15), making then ideal for data mining and information exploration on a large scale.

The willingness of users to i) Recognize a phenomenon within a Social Media Network ii) Experience and perceive a phenomenon and iii) The timely and efficient dissemination of the effects of the Social Media Network phenomenon shows the ability of social media network to be used as large-scales sensors network.

Traditional wisdom has been that complex sensors (e.g. intelligent meters) are needed to understand, gather information, and make a conclusion in real time in order to better understand a complex object (e.g. energy use pattern) [1].

The main purpose of the future framework is to provide this mapping with application, objective assessment and analysis. The purpose of this paper is to question these traditional model of social media network and physical sensors by demonstrates the feasibility of social media network too be uses as interactive, universal sensor system to set up physical sensor system to accomplish similar objectives, delivering comparable levels of information and expertise [1].

2. Literature Survey

The different reviews and analysis carried out in the field interest and findings already reported, taking into account the different parameters of the project and the extent of the project, are seen in a literature review or a literature survey in project study.

The following is a literature survey that includes:

Current theories that are widely agreed about the subject.
Books written on the subject, generic as well as particular.
Study in the field is typically carried out in the order from oldest to newest.
Challenges that are faced and, if available, on-going work.

The literature survey explains the latest work on the project. It discusses the issue of the current system and also offers users a good idea of how to deal with the current problems and how to overcome the current problems.

*Corresponding author: priyanka7.kote@gmail.com

[1] Increasing Veracity of Detection on Social Media Networks by Using User Trust Modeling. Author: Todd Bodnar, Conrad Tucker, Kenneth Hopkinson, Sven G. Bil´en, 2014.

With large-scale success and ubiquity, social media networks are forced to determine the veracity of knowledge exchanged through them that informs people about current real-world events and trends.

For the dissemination of information on social media networks, we propose an authentication test model to detect text content generated by each user, including native language processing and machine learning algorithms. Major social networks (such as Twitter and Facebook) are considered to be network platforms, where information can be exchanged quickly and quickly, thus increasing access to information around the world. This paper analyzes four case studies that cover multiple areas, hazards and time periods to explain how real-world events affect how false information/information is exchanged and disseminated through a social network [2].

[2] The Trust-Aware User Recommendation Program made on social networks. Author: Magdalini Eirinaki, Malamati D. Louta [19], 2014.

Due to the emergence and proliferation of social media, such as blogs, social networking apps, microblogging, or customer review pages, social network analysis has recently received a lot of attention. The Trust becomes an important quality in the user experience in this setting, and suggestions for important content and trusted users are important to all members of the network. In this post, we propose a social media trust management system that focuses on the reputation process that takes expertise and clear relationships between network members, analyzes the semantics and complexity of these interactions, and provides network members with a user-focused response [4].

[3] Understanding Twitter Data with Tweet Xplorer. Author: Fred Morstatter, Shamanth Kumar, Huan Liu, Ross Maciejewski [15], 2013.

It is extremely hard for an analyst to derive useful knowledge from a sea of information in the age of big data. We present Tweet Xplorer, a method for analysts with little information on a case through the use of powerful visualization techniques to gain awareness. Using tweets gathered as an example during Hurricane Sandy, we will direct the reader through a job that shows the system's functionality.

[4] Social Network Modeling and Model-Based Imitation in Support of Crisis De-Escalation. Author: Michael J. Lanham, Geoffrey P. Morgan, and Kathleen M. Carleyn [6], 2014.

Decision-makers need the ability to model easily and reliably evaluate the consequences of actions and responses in crisis situations. Traditionally, there have been heavy processes of production resources and what-if the exercise of such models. This study demonstrates the quick and possible ways to build such models in functional environments.

Via the removal of the communication network from communication, in the analysis of networks to identify the main character, and then simulate exploring alternatives, advisors can encourage the practice and practice of declining levels of escalation. As part of the modeling process driven by the model, we explain how we used this process. We highlight the power of transition from experience to models and the beauty of data-driven simulation, which facilitates iterative purity. We conclude [16] by summarizing the shortcomings of this strategy and the expected future work.

[5] Low-Determination of Online Diagnosis via Twitter and Medical Records. Author: Todd Bodnar, Victoria C Barclay, Nilam Ram, Conrad S Tucker [11], 2014.

Social media is considered a source of disease monitoring data. Most studies, however, focus on models that emphasize close correlation with human disease levels beyond determining whether certain users are sick or not. Taking a different approach, we use a database of people diagnosed with clinics, creating a new diagnostic tool for social media.

In particular, we are building a framework for making a more accurate diagnosis of the flu based on the public Twitter profile. We find that on Twitter, about half (17 = 35 = 48:57 percent) of our sick sample users spoke directly about their illness. We are able to diagnose a person with more than 99 percent accuracy by creating a meta classifier that includes text analysis, inaccuracies, and social network analysis even if it does not take into account his or her health.

[6] Verification of diagnostic models Using Twitter. Author: Todd Bodnar, Marcel Salathé, 2013.

Mining data mining has become an important tool for monitoring infectious diseases. There are, however, major risks associated with predicting the wrong outbreak. In evaluating the results of the model, a large number of communication data combined with a limited number of real-world data and the general complexity of infectious diseases present some difficulty. We look at other methods in this paper that have been used to determine the spread of the flu through Twitter. With tests designed to prevent and detect problems with the standard process of verifying crossing, we will verify it.

We also find that small changes in the way data is categorized can have significant implications for the performance of the recorded model.

[7] Earthquake Shakes Twitter Users: Real-Time Event Discovery Discovery Through Public Sensors [20]. Author: Takeshi Sakaki, Makoto Okazaki, Yutaka Matsuo [14], 2010.

Twitter, a popular microblogging platform, has recently been given a lot of coverage. An important feature of Twitter is its real-time presence. For example, after an earthquake, people create multiple Twitter posts (tweets) linked to earthquake [9], making earthquake events more noticeable, simply by monitoring tweets. We explore real-time connections to events such as the earthquake on Twitter and suggest an algorithm for tracking tweets and finding targeted events, as described in this article. Depending on factors such as keywords in the tweet, the number of sentences, and the context [10] in which they find the intended case, we create the division of tweets. We then created a more sophisticated model of the case space, which could define the center and layout of the scene.

We view each Twitter user as sensors and use Kalman’s filters and particle filters, which are widely used on a ubiquitous/wide range of position ranking. In predicting

earthquake and hurricane centers, the particle filter performs better than other similar programs. We are developing an earthquake warning system in Japan as an app.

Due to the common earthquake and the large number of Twitter users nationwide (96 percent of earthquakes with a magnitude of 3 or more earthquakes detected by the Japan Meteorological Agency (JMA), we can expect earthquakes by following tweets with high probability Our system detects earthquakes there instantly also sends emails to registered users. Notifications are issued much faster than updates by the JMA.

3. Problem Statement

Researchers have historically recommended that uses of smart meters to copy particular electricity consumption habits of devices. These approaches, however, suffer from challenges of scalability and expense. For minimizing or predicting major spikes in electricity usage, the innovation presented by a smart grid is beneficial. For examples, organizing households not to simultaneously bring out high power consumption activity.

A. Existing System

The current system works on weather pattern measurement applications, disease diagnosis, earthquake monitoring, user recommendations, disaster action plans discovery, safety risk identification and obesity trends description. The relatively accessible and simple collecting of information, which unlike the conventional website is generated by greater community of user whose statistics are extra reflective of the common population, is part of the benefit of social media networks.

B. Disadvantages of Existing System

Doesn't provide scalability and not cost-efficient.

C. Proposed System

Proposed a framework that produces and checks relationship between Social media networks topics and the pattern of electricity use automatically. Those topics are next used to forecast the potential uses of energy or to test the causal connections between the subjects and the use of Granger. This Granger Causality is used for these connections to be confirmed. We considering a case study in which our methods implemented using Social Media Network data to disaggregate energy use.

That is, will our machine discover interesting social media network relationships that are trending with rates of electricity consumption? then comparing the topics identified by our systems to be accurate against actuals topics selected by energy field expert or against keyword directly extracted from dataset. We notice that our method replicate the topic selected by expert, in addition to other topics.

D. Advantages of Proposed System

Provides scalability, accuracy and efficiency.

4. Requirements Analysis

A. Software Requirement Specification

The Structure Requirement Specification (SRS) is a focal report that defines the basis of the headway handle for the object. It documents the specifications of a system and also has a description of its noteworthy highlight. A SRS is simply the seeing (in making) of an association of the edge work requirements and conditions of a customer or potential customer at a specific point in time (for the most part) before any actual design or shift work. It is a two-way insurance strategy that guarantees that at a given point in time, both the consumer and the affiliate recognize the needs of exchange from that viewpoint.

The programming need for detail synthesis decreases progress effort, as careful analysis of the report will expose oversights, mixed presumptions, and contradictions in front of the change cycle plan when these problems are less challenging to be correct. However, the SRS addresses the thing not the wander that made it, so the SRS fills in as a start for the finished thing to modify later.

The SRS may need to be updated, as it may provide a framework for proceeding with the development evaluation. The programming requires confirmation in clear words is the starting stage of the operation of the item update. The SRS involves unraveling the musings in the customers' brains-the data, into a structured chronicle-the production of the critical process. In this way, the performance of the stage is an arrangement of formally agreed requirements that are achieved and unrelenting in an ideal universe, whereas the data has that neither of these properties.

Table 1
Hardware requirement

Processor	intel i5 3.0 GHz
RAM	16 GB and above
System type	64-bits Operating system
Hard disk	500 GB

Table 2
Software requirement

Operating system	Windows 7/8/10
Programming Language	Python
Library	Matplotlib, numpy, Pandas, Sckit-learn, Tweepy, genism, nltk, PyLDAvis,
Simulation tool	Anaconda Navigator IDE 3.7.4, (jupyter Notebook).

B. Software Description

1) Jupyter Notebook

Fig. 1. Jupyter

Jupyter Notebook is an amazingly flexible platform for creating and displaying data science projects. This guide will show you how to set up and start using Jupyter Notes for data science projects on your local computer. However, first: what is a 'brochure'? In one article that includes visuals, narrative terminology, statistical simulations, and other rich media, the brochure covers the program and its output. This instructive

workflow promotes repetitive and incremental growth, making textbooks the center of current data processing, analytics, and science as a whole becoming increasingly common choices. Best of all as part of the open source community Jupyter, they are completely free [3].

The Jupyter project follows the previous IPthhton Notebook, first released in 2010 as an example. While various editing languages can be used within Jupyter textbooks, as the most commonly used scenario this article will focus on Python. (R Studio seems to be the most common method for R users).

For example, the Jupyter project follows the original IPython Notebook, which was first published in 2010. Although several different programming languages can be used within Jupyter Notebooks, this article will focus on Python since it is the most frequent use case. (One of the most common methods for R users seems to be R Studio) [3].

We will:

- Cover the fundamentals of downloading and building the first notebook with Jupyter.
- Delve deeper to grasp all the significant terms.
- Explore how to quickly share and publish notebooks online. This post is also the Jupyter Notebook! In the Jupyter Notebook environment, everything here was written, although you all viewing it in a read only form.

C. Functional Requirement

A Practical Necessity (FR) is a description of the service that the software needs to offer. A framework or part of software is specified. A functionality is nothing but inputs to the software framework, its behaviour, and outputs. It may be a measurement, data processing, business process, user interaction, or any other basic mechanism that decides what the intent of a device is likely to do. Often called design requirements are functional specifications. In terms of the whole mechanism of modules, one of the most critical aspects is the functional specifications of the project.

The functional requirements here are:

1. Using recent most important dataset with Social Media Network Data collect from twitter API and SDGE dataset.
2. Creating evaluation metrics with word2Vec.
3. Check the Topics Score's based on machine learning algorithm.
4. Classify the topics data based on machine learning latent Dirichlet allocation(LDA), Forest (RF), Gaussian Nave Base (NB), Algorithms.
5. Applying Preprocessing Task.
6. Find out the best performance of various machine learning technique.

D. Non-Functional Requirement

1) Usability

The customer agrees that almost the buyer interfaces are traditional and dedicated to demanding ambush pressure in moving with another condition to a specific system.

2) Reliability

Both for the project pioneer and in addition to the test design, the improvements achieved by the programmers should be clear.

3) Security

Counting bugs following the system must have substantial protections and must avoid slamming of the whole process.

4) Performance

On a lone web server with a solitary database server out of reach, the system would be facilitated, so execution becomes a noteworthy concern.

5) Portability

This is needed because, because of a few complications, the web server that facilitates the system stalls out, which allows their framework to be taken to another framework.

6) Reusability

The structure can be split into components that could be used as part of another application without needing a lot of effort.

E. Technologies

1) Python

Python, developed by Guido Rossum in 1989, is an object-oriented programming language. Well designed to perform fast prototyping for complex applications. It has an interfaces and is extendable to C or C++ for several OS system calls and libraries. The Python programming language is used by several major establishments, including NASA, Google, Twitter, BitTorrent, etc. [5], [13].

In Artificial Intelligence, Natural Language Generation, Neural Networks and other advanced computer science fields, Python is widely used. Python has been very focused on reading the code and this section will teach you the basics of python [5].

Here python language used for IR sensor and HD camera.

Characteristics of Python:

- Provides richer types of data and easier to read syntax than any other programming language [5].
- Different written language on the platform with full access to the application API.
- It provides more versatility in run-time compared to other programming languages. Simple deceptive structures for Perl and Awk text are included.
- The module may have one or more categories and free Python functionality.
- Linux, Macintosh, and Windows cross-platforms are compliant with Python libraries.

F. Feasibility Study

In this phase, the project's feasibility is determined and a very traditional project plan and several cost figures are put out for the business proposal. During the program evaluation, a review of the availability of the proposed solution will be conducted. This is done to ensure that the proposed arrangement is not a risk to the company. To test the feasibility, some information from major device providers is required.

The three main factors included in the feasibility study are:

1. Economic Visibility
2. Technical Performance

3. Social Media Availability

1) Economic Visibility

This analysis is carried out in order to ensure that the approach can have an economic effect on the association. It is limited to the amount of funds that the organization will pour into the investigation and production of the process. It is necessary to justify the expenses. Thus, within the budget, the developed framework was also developed and this was done because most of the technology used are generously available. It was only appropriate to buy the changed products.

2) Technical Performance

This analysis is carried out in order to ensure that the approach can have an economic effect on the association. It is limited to the amount of funds that the organization will pour into the investigation and production of the process. It is necessary to justify the expenses. Thus, within the budget, the developed framework was also developed and this was done because most of the technology used are generously available. It was only appropriate to buy the changed products.

The method created must have a modest responsibility, since the application of this method needs only minimum or null modifications.

3) Social Media Availability

The study feature is to verify the user's level of process approval. This includes the process of teaching a person to use this technique effectively. Consumer demand does not feel threatened by the system, but rather we must accept it as necessary. The level of user acceptance depends entirely on the techniques used to educate the user about the process and to make it point. His level of assurance should be improved so that, as the end user of the process, he can also make constructive, well-received gratitude.

5. System Design

A. Introduction

A general overview of structure design is generated by system configuration planning. The programming diagram includes discussing the position of the item structure in a way that can be updated to at least one projected one. The key seen by the final client must be positioned in a organized way. The diagram is a creative system; an outstanding design is the best approach for a rational structure. The structure of the "Layout" is defined as "the methodology of applying distinctive frameworks and guidelines with the ultimate objective of defining a technique or system for a purpose sufficiently significant to allow its physical assertion." Afterwards, various concept segments are applied to the unit. The specifics of the design demonstrate the system elements, the parts or divisions of the structure, and their presentation to end-customers.

B. Design Consideration

The reason behind the strategy is to organize the course of action of the problem described in the Necessities Study. This stage is the underlying process of moving from the issue to the game plan's room. All things considered, start with what is needed; the diagram takes us to work on how to fulfill those needs. The system design is perhaps the most significant segment affecting the way of the item and the note dignifiedly affects the later stages, particularly testing and maintenance. All the vast data structure of the system, report game plan, yield and genuine modules are detailed in the system diagram and pick their specifications.

C. System Architecture

The method of architectural configuration is concerned with the design of a fundamental basic structure for a framework. It requires understanding the individual components of the system and the relationships between these segments. The starting configuration process of identifying these subsystems and building up a subsystem control and correspondence structure is called the outline of construction modeling, and the performance of this outline process is a reflection of the structural planning of the product. The suggested architecture is given below for this method. It illustrates the way this system is designed and the system functions briefly.

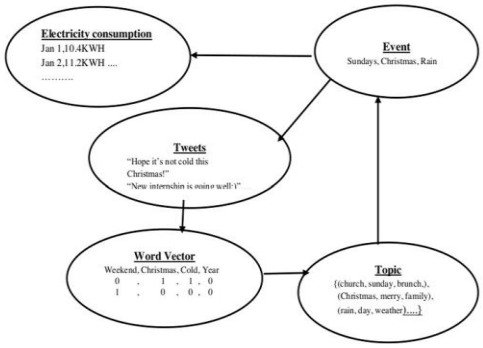


Fig. 2. Implementation of theoretical model with system architecture

D. Block Diagram

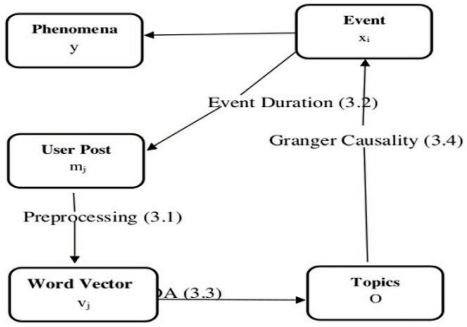


Fig. 3. Schematic view in High-level overview of the method to turn a stream of social media sites into theories about a real life event

E. Data Flow Diagram

DFD is a graphical representation of features or process that collect, manipulate, store and distribute data between a systems and its environment and between system components. It is a powerful means of communication between the user and the device designer due to the visual representation [12].

The DFD architecture makes it possible to start from a broader perspective and expands it into the category of complex graphics managers.

DFD has been used many times for the following reasons:

1. Logical flow of information system.

2. Determining the requirements for the structure of the body system. Ease of notification.
3. Establishment of manual requirements and automated programs.

DFD Components:

Using the following set of components, DFD will represent the source, destination, storage and flow of data [8]:

Entities: An external object is a person, agency, external organization, or other information system that provides system data or collects system results [7].

Process: Any system that alterations the data and generates an output. Based on business rules, it may perform computation or sort data based on the logic, or control the data flow.

Data Storage: Files or archives, such as table database or a memberships type, that contain information for later usage. A basic mark, such as "Orders", is obtained by each data store.

Data Flow: The path between the external bodies, systems and data storage the information requires.

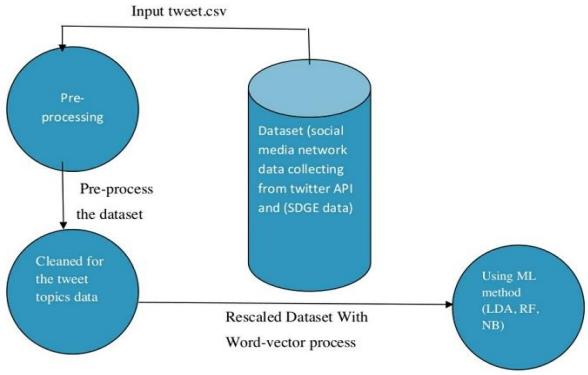


Fig. 4. DFD-L0

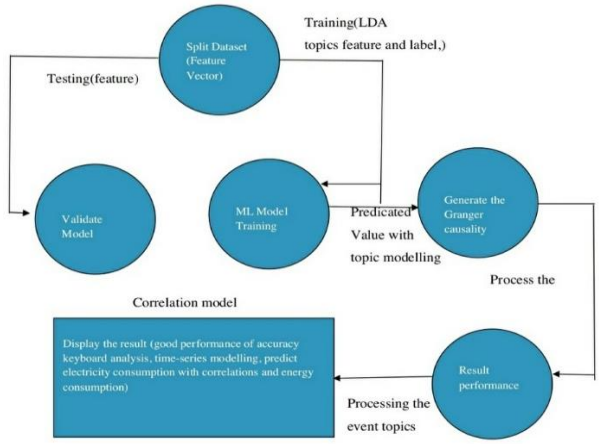


Fig. 5. DFD-L1

6. Implementation of Modules

A. Raw Social Media Network Data Cleaning

Information on social media networks is often defined as incredibly noisy, requires the social media networks stream intensively cleaned as a required first step. We do this one via the transformation of a character string into list of n- gram Pairs of contiguous words up to n. The n-grams calculated by tokenizing the strings on all non-alphabetical characters. Beginning with capitalized on social media networks can be erratic, the n-grams are then translated to lower- case values [1].

Since the aim of this step is to extract topic instead of keyword, using porter stemming, Stem each of the terms. This maps words to the same keyword with identical stems but with different suffixes. "Accept," "accept," and "acceptance," for instance, all is mapped to the same keyword.

A long-tail distribution is predicted to follow this list of n-grammer, results in the possibility that are some common or more common that it may assist in analysis. Popular terms such as "and," "is," "the" offer little or no information and can hide, have more details, less common words. Thus, common terms are omitted from the lists of n-grams, as described by Lewis et al.'s stop word list.

If a word is too rare, on the other hand, it does not occur sufficiently to be generalizable for any inferences about it. Since there is a long-tail distribution of n-grams, most terms would be too rare. Therefore, these very rare of n-grams have the potential to limit our ability to produce n- gram inferences [1].

Algorithm 1: Preprocessing Steps for Social Media Network Data

```
Data: Time tagged Messages M
Result: A set of aggregated and processed messages D
d_q = document of keywords at time q;
count_word = frequency of "word" in all documents;
W = set of all known stemmed words;
for m_j ∈ M do
    Break m_j into substrings on non-alphabetical
    characters ^[a ≠A ∉];
    j = hour m_j was posted;
    for non-empty Substring S in m_j do
        convert S to lowercase;
        stem S using porter stemmer;
        add S to W;
        push S onto d_j;
        count_S +;
end end
for word S in W do
    if count_S < δ_min then Remove
        S from each d_j; Remove
        S from W;
end end
```

B. Social Network Social Data and Real-World Comparisons

Social media networks records to be changed at a milli-2d degree however, it's miles in uncommon to document actual-international occasion at the sort of transient resolution. In addition, it's miles not possible that the unmarried social media community submit might include relevant, concrete expertise approximately the actual-international incident that we need to examine, or if it does, it's miles extraordinarily uncommon. We accurate this distinction with the aid of using normalizes the social media community records to the time scale of the actual-international records. That is, we don't forget a dj file to be the combination of all vj (as derived from mj) processed social media community messages that takes place in for the duration of the time frame among xq actual international occasion xq and the imminent occasion, xq+1. More technically,

$$dj = \{vj \mid time(xq) \leq time(mj) < time(xq+1) \quad (1)$$

C. Generating Topic Models

To generate topics via LDA, a given collection of documents

identified by the aggregation mentioned above can be used. To perform this analysis given that it containing a collection of n grams, LDA determining the likelihood of a document being about a subject. LDA first creates clusters of terms dependent on co- occurrences in document in order to do this. That is, word opportunities w appears as long as o (w) is the subject of the text. In order to describe these subjects in a human-readable way, we present the collection of words that are most likely to occur within the subject

Let's work on translating text data into a format that will serve as an addition to the LDA training model. We start by translating the documents into a simple presentation (BOW Word Wallet). First, we will convert the list of topics into a list of vectors, all of which are the same length in vocabulary. Depending on the outcome of these operations (document vectors list), we will then set ten most common goals.

Algorithm 2: LDA Algorithm in the Context of the Proposed Social Media Network Model

Data: set of **Documents** D, topics **O Result**:
a $|W| \times |O|$ matrix
for Document $d \in D$ **do**
  for *Word* $w \in d$ **do**
    w_{topic} = Random topic $\in \{0,\ldots, |O|\}$;
  en
end
for *Step in* $\{1,\ldots,$ *stop point*$\}$ **do**
  for Document $d \in D$ **do**
    for *word* $w \in d$ **do**
      for *topic* $o \in \{0,\ldots, |O|\}$ **do**

$$P(o|d) = \frac{|w \in d \text{ where } w_{topic} = o|}{|w \in d|} ; \quad \frac{|w \in D|}{|w \in D|} = o|$$

      end
      Assign w_{topic} based on $P(w\ o)$ $P(o\ d)$.
    end
  end
end

D. Determining Event-Phenomena Causality

The pattern of each of these event on more time. i.e. averaging all of an event's frequencies over time result in a time series to be contrasted with the phenomena of the natural world. Such topics will exhibit cyclical trends, such as Love, Electric, Thank You, Mondays, or Enjoying lunch, whereas others event, such a storm or festival, may being one's time, anomalous event.

By cross correlation, this timeline of events can be compared with the text of the time series' related to presenting real-world objects. This is explained by these positions in Pearson's Correlation, where a single point of drawing does not correspond to the frequency of real world events and conditions. This approach does not mediate positive or negative interactions: strong negative links between research and a real-world event cannot be as enthusiastic as positive. While all of these relationships can be strong, those who do not know the requirements dictate the causal relationship [1].

There are two benefits to this duals time series analysis approach: it quantifies the meaningful duration of a lag and determining which one sampled topic are important. This Granger's causal test will measure the causals relationship between a phenomenon's (a changes in powers using) and the occurrence (a defined by these one or more's topics in socials media). This causalities measurements are the primary's method of evaluating's causality applied in this approach. Socials networking post can't be processed into topics in advance, and within new posts, these topics to be found in linear timing and also enables these causals relationships will be transformed online. If the effectiveness of this prediction of their causal relationships undermines new samples they may be drawing and re-calculating (see Alg 3). This one allow us, in setting of relayed exclusively on old's data, to change and use new data.

Algorithm 3: Computational Complexity of This Methodology

input : Social Media Posts
output: Predictions
Social Media Posts arrive: $O(1)$;
Preprocessing: $O(m)$ where m = number of posts;
topics ← Generate Topics (LDA): $O(Nm^2)$ (see alg 2);
CausalTopics ← Granger(topics) $O(Len($topics$))$;

Dataset:

1. Rate _ Of _topics

MOST_LIKELY_WORDS_IN_THE_TOPIC	R-RATE
xe2 x80 xa6	79.07047448
x80 xa6 https	71.50566982
xa6 https co	71.50566982
this full message	48.4225151
fernand50948390 si si	36
trying this full	35.18262169
x94 x8d xfxfxxd	34.61698727
xf0 x9f x94	34.61698727
x9f x94 x8d	34.61698727
xe2 x80 x99t	27.06903143
love electricity heyso	23.59920072
thanks spreading love	22.01354701
spreading love electricity	22.01354701
can xe2 x80	21.36633991
knock knock thing	19.96920706
knock thing find	19.96920706
isrecruit xf0 x9f	19.93108488
kamungamukanya xf0 x9f	19.93108488
x80 xa6 rt	19.21075844

2. Rate _ Of _Topics _ Correction

DATE	MOST_LIKELY_WORDS_IN_THE_TOPIC	TOPICS	R-RATE
11/1/2019	xe2 x80 xa6	Electricity	69.56267828
12/1/2019	xa6 https co	Electricity	63.22294244
13-01-2019	x80 xa6 https	Electricity	63.22294244
14-01-2019	this full message	Electricity	42.97605153
15-01-2019	fernand50948390 si si	Electricity	32
16-01-2019	trying this full	Electricity	31.24175643
17-01-2019	x9f x94 x8d	Electricity	30.76946856
18-01-2019	xf0 x9f x94	Electricity	30.76946856
19-01-2019	x94 x8d xfxfxxd	Electricity	30.76946856

3. Clean _Tweets

Tweets	Clean_tweet
b'@maunyk Hey @maunyk, more of an RSVP. Love, Electricity.'	b' hey rsvp love electricity'
b'@A_CHEEKYBOY Cheekyboy indeed. Love, Electricity.'	b' cheekyboy indeed love electricity'
b"Half an hour... I'm a bit nervous. #ElectricityTalks"	bhalf hour i'm bit nervous electricity talk
b"@KungFuCutBug @Porsche Fair point. But maybe there's another wheel I can take."	fair point but mayb there' another wheel can take
b'@PompeyCal @sophiepb_03 Hey @PompeyCal , I love providing tweets with power, but spreading my message in morse felt\xe2\x80\xa6 https://t.co/b5EBSHMZdQ'	b' hey i love providing tweet power spread message morse felt xexxa httpstcobebshmzdq'

4. SDGE-ELEC-2020-Q1

Zip Code	Month	Year	Customer Class	Combined	Total Customers	Total (kWh)	Average (kWh)
91901	1	2020	A	Y	0	0	0
91901	1	2020	C	Y	0	0	0
91901	1	2020	I	Y	0	0	0
91901	1	2020	R	Y	8729	5338772	612
91902	1	2020	A	Y	0	0	0
91902	1	2020	C	N	327	1266079	3872
91902	1	2020	I	Y	0	0	0
91902	1	2020	R	N	6476	3126529	483
91905	1	2020	A	Y	0	0	0

5. Tweets

b'@captainbrown @andy_park @AustralianArmy @Ausgrid Wait, does this mean that ALL of the people affected by fires sin\xe2\x80\xa6 https://t.co/3OswZT6Mah'
b"RT @graham_copp: 425,500 workers went on strike in America last year. The most since 2019.\nI'm ecstatic that nearly half a million people e\xe2\x80\xa6"
b"RT @graham_copp: 425,500 workers went on strike in America last year. The most since 2019.\nI'm ecstatic that nearly half a million people e\xe2\x80\xa6"
b"425,500 workers went on strike in America last year. The most since 2019.\nI'm ecstatic that nearly half a million p\xe2\x80\xa6 https://t.co/MNoul3apdn"
b'@UmemeLtd please try to connect me power I applied since 2019 on account p10020313053412'

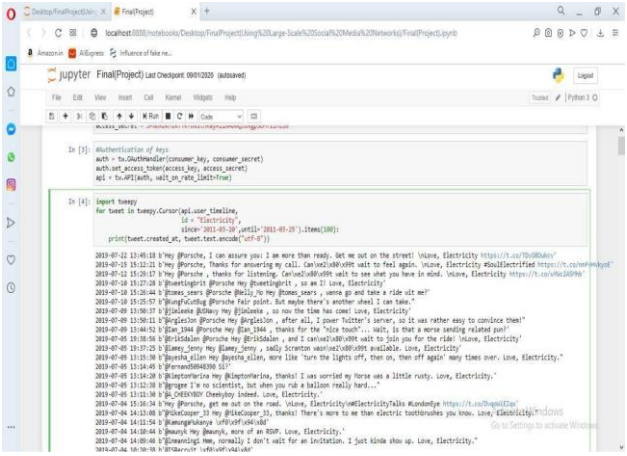
7. Results and Discussion

Although our system enables information exploration with a minimal need for expert knowledge, it is not useful if it does not work well. To explain the presence of our system, we equate the outcomes of our system to topics popular in the literature on power consumption. In addition, we conduct keyword mining to identify terms that are related to electricity usage instead of topics [1].

If we only consider some topics in recently and local papers that appear more than once" (temperature," "jobs," "electricity price," "air conditioner, "and" heater), "then we can detect two clusters of topics informally: 1)" climate control "and 2)" economic causes". Through our automated system important measurements of electricity consumption were also found to be both of these two topics. Twenty topics related to electricity use have been found in our system. Twenty subjects linked to the use of energy were also included in our literature review.

In struggling with periodic grouping, testing keywords instead of topics contributed to certain contrasts. 2 Our keyword assessment, however, makes a range of measures equal to the scale of the corpus, which is impossible to explicitly equate with 20 topics being tested. As we just considering the top 20 keywords, we find out keywords with this highest positives correlation to be “don” with value r=0.344 and keyword with the highest negatives correlations to be with r=-0.475. Our methodology detects incidents where 0.458 is the highest positive correlation and the highest negative correlation is -0.523, a collective increase of 16.4 percent and 8.06 percent.

Output screenshot 1:



Collecting tweets using twitter API

Output screenshot 2:

Pre-processing and LDA algorithm

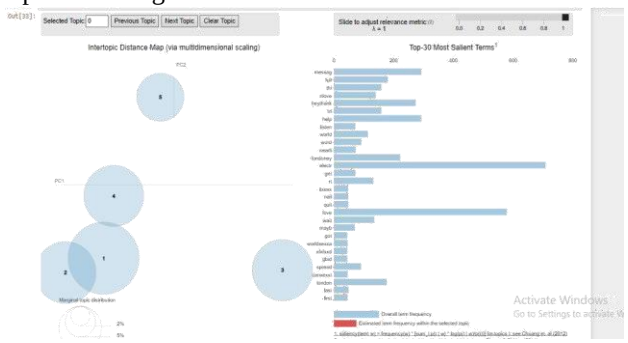
```
In [67]: from gensim import models
         #Latent Dirichlet Allocation (LDA)
         #num_topics=3
         #model = models.ldamodel.LdaModel(corpus, num_topics=3, id2word=dictionary, passes=15)
         #num_topics=20
         model = models.ldamodel.LdaModel(corpus, num_topics=20, id2word=dictionary, passes=15)
         print(model)
         topics = model.print_topics(num_words=3)
         # print(topics)
         for topic in topics:
             print(topic)

LdaModel(num_terms=262, num_topics=20, decay=0.5, chunksize=2000)
(0, '0.104*"electr" + 0.053*"b" + 0.053*"love"')
(1, '0.093*"electr" + 0.093*"love" + 0.070*"b"')
(2, '0.083*"b" + 0.078*"hey" + 0.052*"electr"')
(3, '0.082*"electr" + 0.082*"love" + 0.082*"b"')
(4, '0.126*"b" + 0.048*"electr" + 0.048*"hey"')
(5, '0.085*"b" + 0.054*"rt" + 0.048*"help"')
(6, '0.087*"world" + 0.072*"messag" + 0.071*"full"')
(7, '0.059*"electr" + 0.059*"love" + 0.059*"happi"')
(8, '0.121*"b" + 0.117*"thi" + 0.116*"help"')
(9, '0.096*"b" + 0.064*"listen" + 0.064*"itxexx"')
(10, '0.091*"b" + 0.060*"londoney" + 0.060*"london"')
(11, '0.086*"hey" + 0.069*"b" + 0.069*"love"')
(12, '0.099*"b" + 0.099*"electr" + 0.099*"love"')
(13, '0.053*"b" + 0.053*"messag" + 0.053*"love"')
```

Output screenshot 3:

Topics and scores

```
In : lda_model_tfidf = gensim.models.LdaMulticore(corpus, num_topics=10, id2word=dictionary, passes=2, workers=4)
     for idx, topic in lda_model_tfidf.print_topics(-1):
         print('Topic: {} Word: {}'.format(idx, topic))

Topic: 0 Word: 0.062*"messag" + 0.061*"electr" + 0.058*"hey1" + 0.055*"got" + 0.055*"love" + 0.053*"spread" + 0.052*"glad" + 0.052*"thank" + 0.05
Topic: 1 Word: 0.093*"electr" + 0.067*"b" + 0.058*"nlove" + 0.055*"wait" + 0.050*"take" + 0.036*"hey" + 0.033*"love" + 0.032*"readi" + 0.032*"ass
Topic: 2 Word: 0.098*"electr" + 0.098*"love" + 0.090*"b" + 0.042*"hey" + 0.026*"electricitytalk" + 0.022*"messag" + 0.022*"one" + 0.021*"weak" + (
Topic: 3 Word: 0.085*"b" + 0.077*"londoney" + 0.065*"rt" + 0.050*"london" + 0.029*"figur" + 0.025*"help" + 0.023*"eye" + 0.021*"electricitytalk" +
ey"
Topic: 4 Word: 0.068*"electr" + 0.066*"b" + 0.053*"londoney" + 0.048*"knock" + 0.042*"hey" + 0.038*"love" + 0.030*"nlove" + 0.028*"wait" + 0.027*"
Topic: 5 Word: 0.086*"b" + 0.052*"nearli" + 0.034*"love" + 0.029*"heythank" + 0.029*"london" + 0.028*"heyb" + 0.027*"nail" + 0.026*"quit" + 0.026
Topic: 6 Word: 0.117*"b" + 0.102*"electr" + 0.059*"love" + 0.030*"heyso" + 0.027*"time" + 0.026*"amlov" + 0.025*"decodingtri" + 0.025*"might" + 0.
tcocoacai"
Topic: 7 Word: 0.113*"b" + 0.070*"help" + 0.063*"heythank" + 0.061*"messag" + 0.060*"full" + 0.054*"thi" + 0.049*"tri" + 0.028*"world" + 0.024*"h
Topic: 8 Word: 0.082*"b" + 0.073*"electr" + 0.059*"love" + 0.055*"thank" + 0.046*"hey" + 0.033*"wait" + 0.032*"spread" + 0.024*"messag" + 0.023*"
Topic: 9 Word: 0.113*"b" + 0.097*"electr" + 0.094*"hey" + 0.087*"love" + 0.021*"electricitytalk" + 0.021*"london" + 0.020*"xfxfxxd" + 0.018*"first
2*"londoney"
```

Output screenshot 4:
Topic Modeling Visualization




```
In [46]: st.grangercausalitytests(df[['tweets', 'Clean_tweet']], maxlag=1)

         Granger Causality
         number of lags (no zero) 1
         ssr based F test:         F=13.2466 , p=0.0003  , df_denom=1373, df_num=1
         ssr based chi2 test:   chi2=13.2756 , p=0.0003  , df=1
         likelihood ratio test: chi2=13.2119 , p=0.0003  , df=1
         parameter F test:         F=13.2466 , p=0.0003  , df_denom=1373, df_num=1

Out[46]: {1: ({'ssr_ftest': (13.246620283796966, 0.000283204886201661, 1373.0, 1),
            'ssr_chi2test': (13.275564100877368, 0.0002688796744834544, 1),
            'lrtest': (13.211932135223833, 0.0002781724559228756, 1),
            'params_ftest': (13.246620283797009, 0.000283204886201661, 1373.0, 1.0)},
           [<statsmodels.regression.linear_model.RegressionResultsWrapper at 0x178a9614358>,
            <statsmodels.regression.linear_model.RegressionResultsWrapper at 0x178a9614438>,
            array([[0., 1., 0.]])])}
```

Granger causality test:

```
In [52]: df.corrwith(df1.Topic)

Out[52]: tweets        -0.482029
         Clean_tweet    0.503549
         dtype: float64

In [53]: df1['Topic']=df1['Topic'].astype('category').cat.codes
         df1.corr()

Out[53]:
```

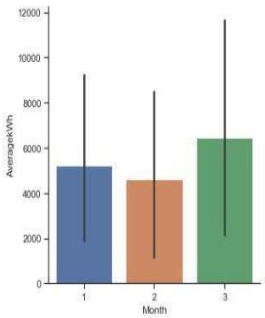
	Topic
Topic	1.0

```
In [54]: #!pip install pingouin
         import pingouin as pg
         pg.corr(x=df['Clean_tweet'], y=df['tweets'])

Out[54]:
```

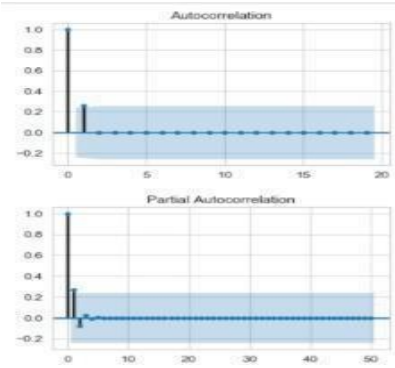
	n	r	CI95%	r2	adj_r2	p-val	BF10	power
pearson	1377	-0.108	[-0.16, -0.06]	0.012	0.01	0.000059	107.325	0.98

Output screenshot 5:
SDGE Data set monthly usage of electricity.

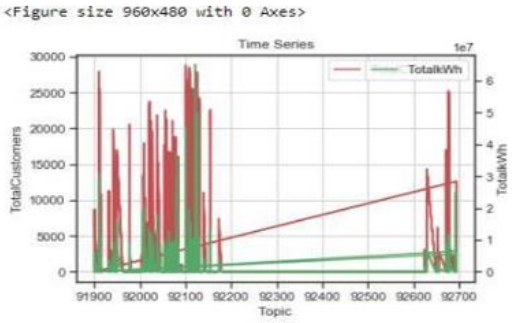
```
In [58]: import seaborn as sns
         sns.factorplot(x='Month',y='Averagekwh',data=Electricdata,kind='bar')

Out[58]: <seaborn.axisgrid.FacetGrid at 0x178ab769ef0>
```



Output screenshot 6:
Autocorrelation and Partial Autocorrelation



Output screenshot 7:
Time Series:



8. Conclusion

Theoretical support for our design was suggested, assuming a link between i) event and text ii) text and word vector iii) word vector and topic iv) topic and event and v) real world event and phenomena. Now we are showing evidence of these partnerships. Previous study has reported that injuries lead social media sites to report to people. Likewise, the conversion of text into word vector has formerly been stated. The most probable terms are coherent within each subject and have wide differences between the subjects. Thus, it is possible that topics on the Social Media Network will be generated using LDA from text. The event is then used to create an idea of the realities of the world under scrutiny. We conduct case studies using Twitter data to measure the level of use of force to support our case. The conclusions are then compared to topics produced by professional developers and keyword analysis.

9. Future Work

As several such models exist, a more thorough analysis of this model to other current models may be proposed in future work. Moreover, as described in the introduction, for more guided event detection, this model could be used. The textual research in this work may be improved by all thing considered the synonyms and coupled definitions by words embedding, which immediately groups identical words together. Additionally, it is also possible to consider other data modalities, such as social media metadata, videos, and images. Since this knowledge has a geographical component, related evidence for another area of the planet will also be analyzed by

future research to find out if the styles we have seen are true elsewhere. Finally, it may be beneficial to explore the same method of additional use, such as water.

References

[1] Todd Bodnar, Matthew L. Dering, Conrad Tucker, Kenneth M. Edu. "Using Large-Scale Social Media Networks as a Scalable Sensing System for Modeling Real-Time Energy Utilization Patterns", IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2017.

[2] Todd Bodnar, Conrad Tucker, Kenneth Hopkinson, Sven G. Bilen. "Increasing the veracity of event detection on social media networks through user trust modeling", 2014 IEEE International Conference on Big Data (Big Data), 2014.

[3] www.dataquest.io

[4] ieeexplore.ieee.org

[5] www.guru99.com

[6] Lanham, Michael J., Geoffrey P. Morgan, and Kathleen M. Carley. "Social Network Modeling and Agent-Based Simulation in Support of Crisis De-Escalation", IEEE Transactions on Systems Man and Cybernetics Systems, 2014.

[7] Submitted to Asia Pacific University College of Technology and Innovation (UCTI).

[8] www.coursehero.com

[9] dl.acm.org

[10] Sakaki, Takeshi, Makoto Okazaki, and Yutaka Matsuo. "Tweet Analysis for Real-Time Event Detection and Earthquake Reporting System Development", IEEE Transactions on Knowledge and Data Engineering, 2012.

[11] arxiv.org

[12] Submitted to Pennsylvania State System of Higher Education

[13] Submitted to University of Wales Institute, Cardiff

[14] Submitted to Savitribai Phule Pune University

[15] aminer.org

[16] cps-vo.org

[17] skeletalmusclejournal.biomedcentral.com

[18] www.grin.com

[19] Magdalini Eirinaki, Malamati D. Louta, Iraklis Varlamis. "A Trust-Aware System for Personalized User Recommendations in Social Networks", IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2014.

[20] scholar.afit.edu