Ministry of Education of the Azerbaijan Republic

Khazar University

School of Science and Engineering

Major: 266147 - Software of Computer Systems and Networks

Topic: Twitter Social Network Data Mining in Healthcare

Student: Humay Huseynova

Supervisor: Dr. Mahammad Sharifov

Advisor: Behnam Kiani

Master thesis

Baku 2019

List of abbreviations:

KNN – k	k-nearest neighbors' algorithm
NB	Naïve Bayes
NHS	National Health Service
SML	Supervised Machine Learning
SVM	Support Vector Machines
HADR	High Availability Disaster Recovery
KDD	Knowledge Discovery from Data
HIPAA	Health Insurance Portability and Accountability
GPU	Graphics Processing Units
MERS	Middle East Respiratory Syndrome
IMS	Information Medical Statistics
AMA	American Medical Association

Preface

This research is done as a master dissertation to complete the requirements of a master's degree. In the beginning, I was not sure about the topic selection as I have changed my mind several times. In the end, thanks to my supervisor I have decided on this topic. The reason why I selected this topic is that social media is an inseparable part of daily life. As well as the importance of health is undeniable like water and oxygen. So, to make them together and take advantages of it is my aim. In this document, I did my best to cover all sides of the topic. Some numerical data are also gathered from the research papers, international authorities' websites.

I am grateful to all my professors and instructors who taught me during the master program, the management of Khazar University and the Dean Office. I would like to send special thanks to my supervisor Dr. Mahammad Sharifov for his support and advice. I would also like to thank Kamran Javed who supported and motivated me during this time.

Abstract

Social media offers a wide range of opportunities to share information. In recent years transparency and open sharing via social media have significantly increased, hence, social media and healthcare became of interest among researchers, physicians, and doctors.

Twitter is playing a significant role among other social media types because of its usability, so we can say that its use in healthcare is undeniable. It means that it is possible to see how much Twitter is capable to make sense in solving health problems.

The main purpose of my research is to analyze in what capacity twitter is useful for healthcare, diseases, and treatments by using needed twitter data.

I am going to use twitter mining techniques with simple Python codes to analyze tweets, to study the nature of sharing public health information on Twitter and to make a prediction for Azerbaijan environment by getting a reasonable conclusion.

Keywords: Twitter mining, social media, healthcare, diseases.

Xülasə

Sosial media informasiya paylaşmaq üçün geniş imkanlara malikdir. Son illərdə sosial media vasitəsilə edilən paylaşımlarda şəffaqlıq və açıqlıq kifayət qədər artmışdır. Bundan başqa sosial media istifadəçiləri digər mövzularda olduğu kimi sağlamlıq mövzusunda da yetərli qədər yararlı informasiya paylaşırlar. Bu da sosial media və tibb mövzusunu tədqiqatçılar, fiziklər və həkimlər üçün maraq dairəsinə çevirmişdir.

Digər sosial şəbəkələrlə müqayisədə Twitter faydalılığına görə daha yüksək əhəmiyyətə malikdir və Twitterin tibbə tətbiqinin nəticələri danılmazdır. Bu o deməkdir ki, Twitterin tibbi problemləri həll etməkdə nə qədər işə yaradığını aydınlaşdırmaq mümkündür.

Bu tədqiqatın əsas məqsədi Twitter məlumatlarını analiz edərək sistemli təhlil aparmaq, Twitterin xəstəliklərin müəyyənləşdirilməsində nə dərəcədə effektli olmasını müəyyənləşdirmək və bunlara əsaslanaraq Azərbaycan mühiti üçün təklif verməkdir.

Bunu etmək üçünTwitterin analizi üsullarından, sadə Pyhton kodlarından istifadə olunacaq, bu sahədə görülmüş işlərin tədqiqatı aparılacaq, sorğu, statistika və nəticələr veriləcək.

Açar sözlər: Twitter mining, sosial media, tibb, xəstəliklər.

Introduction

Social media is making different the means people get information, share knowledge and be in touch with each other. The main factor which leads to the mass growth of these technologies is that people easily create “user-generated content”. Twitter, YouTube, Instagram are just one of the social networks which strongly change people’s everyday life.

Twitter has some differences from other social media networks due to its practicability among the researches. As a microblog, Twitter gives a vast amount of services for example messages or tweets to the users. Twitter is a famous microblogging service that offers services to its users to send, get and read messages based on text, in other words, tweets. Though microblogging is drastically in demand, the ways to organize and search microblog data are comparatively new.

Research problem

In consequences of fast and rapid growth of technology and its high spread usage among the population, this has made easy the life of people and help them do their work in the highest effective way like before, due to receive for government the thoughts of its population related to public policy, they will need to make surveys among the citizens. These surveys will help them determine the percentage of the population which consumes a lot of resources and time as well. Apart from the area of business, when a businessman wants to introduce a new product into the market for sure, it used to have some difficulties for him to tell in advance the need for the new product in the market, although, now this has become easier via the use of social media which has made it comfortable for the government, market researchers and many other fields to get people’s mind about it.

Despite that, there are many barriers related to using this technology, because the vast majority of people are unaware of this, especially in healthcare. Twitter is not widespread in Azerbaijan like other countries. My aim with this research is to do something related to this problem in order to help people.

This work proposes results which healthcare system can take great benefits of it. The tweets, articles, and information gathered from all over the world will be helpful to determine the condition of the diseases. Apart from that, a survey which is done in order to learn people's opinion and feedback tells more things about the current problem.

Objectives

For doing this research we will collect and analyze articles related to healthcare by twitter mining and other information from well-known sources, like a database of Twitter, universities, and research websites. After gathering the needed data, the next step will be analyzing the nature of tweets according to diseases. After that, we will select the last and most needed data for our research and will see how it is useful for our aim. As well as, comparisons based on the last years and the work that was done before will be made.

Scope of the work

In this work, at the present time, the main aim is to get data related to public health via twitter, analyze it and conclude. It is doubtless that, still, there is more work to be done in the future in order to get better results and give solutions.

Table of Contents

Chapter 1

Warming up

1.1 Data mining.

Data mining is a procedure of identifying helpful or enforceable information in wide-ranging data. In other words, data mining might be defined as KDD which stands for knowledge discovery from data. it is the way of removing helpful information among a large amount of raw data. This KDD process widely maintains the latter steps: preprocess of data, mining data, and postprocess of data. Mentioned tasks can be combined, so do not need to be isolated. Data mining is an essential part of so many involved areas such as statistics, machine learning, pattern recognition, database systems, visualization, data storehouse, and data withdrawal. Data mining algorithms are widely categorized into three parts: supervised, unsupervised, and semi-supervised algorisms. For the supervised method we can give classification as an instance. In this type of data mining approaches, the data is usually separated in 2 components: to train and to test data groups with usual class tags. Next one which is supervised approach creates categorization types with using trained data and studied methods to predict. For measuring production of the classification model, this model is implemented to the trial data in order to get the correctness of classification. Usual supervised studying models cover enlisting the decision trees, k-nearest neighbors, naive Bayes classification, as well as support-vector machine. Last one-unsupervised learning approaches are planned unlabeled class of the data. The typical and usual instance of this approach is clustering. For the certain exercise, unsupervised studying approaches create method which is built on similar or dissimilar characteristics among data entities. There are some methods to measure similar or dissimilar

characteristics among data entities, such as Euclidean distance, Minkowski distance, and Mahalanobis distance. Another familiarity method for instance single euqality coefficient, Jaccard coefficient, cosine similarity, and Pearson's relation might be applied to define similar or dissimilar characteristics among data entities. Some common instances of unsupervised approach can be k-meaning, hierarchic clustering and also clustering which is built one intensity. The best way to practice semi-supervised studying approach is to use little quantity of data with labels and big quantity of data without labels. The two common kinds of semi-supervised learning are semi-supervised classification and semi-supervised clustering. The last one takes data with label to classify and data without label to purify the classification boundaries further, and the last one works with data with labels to conduct clustering. Co-Training is a common semi-supervised studying approach. Real studying approaches let the users to take important part role during the procedure with labels. Usually, users are accepted as the experts of domain. Their abilities are operated on tags for data examples. This is a machine learning approach which is reliant when it comes to classifying. The least borderline hyper plane and the utmost interest can be used as two famous studying approaches. It is not end. There are some other methods of data mining for example to detect anomaly, visualizing analytics, to select feature and so on.

1.2 Mining social media

Social media is included in a set of Internet-based applications which is made on the ideologic and technology foundation of Web 2.0 which allows the users to interact through the social network and build user-generated content. In other words, social media is a group of different kind of social media platforms for instance, classic media covers newspaper, radio, TV and non-traditional media like Instagram, YouTube Twitter, so on.

Table 1: Features of varied kinds of social media

Type	Characteristics
Online social networking	Online social networks are Web-based services that allow individuals and communities to connect with real-world friends and acquaintances online. Users interact with each other through status updates, comments, media sharing, messages, etc. (e.g., Facebook, Myspace, LinkedIn).
Blogging	A blog is a journal-like website for users, aka bloggers, to contribute textual and multimedia content, arranged in reverse chronological order. Blogs are generally maintained by an individual or by a community (e.g., Huffington Post, Business Insider, Engadget).
Microblogging	Microblogs can be considered same a blogs but with limited content (e.g., Twitter, Tumblr, Plurk).
Wikis	A wiki is a collaborative editing environment that allow multiple users to develop Web pages (e.g., Wikipedia, Wikitravel, Wikihow).
Social news	Social news refers to the sharing and selection of news stories and articles by community of users (e.g., Digg, Slashdot, Reddit).
Social bookmarking	Social bookmarking sites allow users to bookmark Web content for storage, organization, and sharing (e.g., Delicious, StumbleUpon).
Media sharing	Media sharing is an umbrella term that refers to the sharing of variety of media on the Web including video, audio, and photo (e.g., YouTube, Flickr, UstreamTV).
Opinion, reviews, and ratings	The primary function of such sites is to collect and publish user-submitted content in the form of subjective commentary on existing products, services, entertainment, businesses, places, etc. Some of these sites also provide products reviews (e.g., Epinions, Yelp, Cnet).
Answers	These sites provide a platform for users seeking advice, guidance, or knowledge to ask questions. Other users from the community can answer these questions based on previous experiences, personal opinions, or relevent research. Answers are generally judged using ratings and comments (e.g., Yahoo! answers, WikiAnswers).

A large quantity of user-generated content is made on social media networks every day. It seems this tendency is going to continue with coefficient more content for the future trend. Therefore, producers, consumers, and service providers can't easily estimate management and usefulness of heavy user-generated data. Expansion of social media is followed by challenges mentioned below: can a user be heard? (2) what resource of information user can use? (3) How can the user experience be improved? To reply to these inquiries researchers, need to see the secret information in the social media data. Mentioned challenges offer good chances for the data mining professionals to build up new algorithms and ways for social media. Data produced on these social media technologies differs from traditional data mining which is character worth. The data of social media is widely user-generated content on social media sites. Social media data are wide, fussy, allocated, not structured, and vibrant. These features present problems for data mining techniques in order to create new effective methods and ways. Social media networks are very active and repeatedly improving. The vibrant feature of the social media data is an important factor for the steady and rapid development of social media networks. Apart from this, additional curious questions attached to human behavior in social media are available and should be learned by using social media data. Also, social media is able to help to advertise people to get the right people to make maximum attain of their products within the budget of advertising. Social media helps sociologists to pick out the human behavior for instance user behaviors inside and outside of the group. Lately, social media plays an influential role in making bulk motions like Arab Spring8 and Occupy Wall Street.

Here we will give some examples related to the research problems of social media:

- ✓ Community analysis
- ✓ Sentiment analysis and opinion mining
- ✓ Social recommendation
- ✓ İnfluence modeling
- ✓ İnformation diffusion and provenance
- ✓ Privacy, security and trust

1.3 Discovering social networks in true life applications

Recently the usage of social media has been significantly increased. İt is used in a quite big sphere, such as politics issues for example election of president, bulk motion for example to organize Occupy Wall Street movements, Arab Spring, also a catastrophe and crisis reaction and help organizing. Here, we are going to give an example to mine social media for High Availability Disaster Recovery (HADR). Natural catastrophes for example earthquakes, floods, tornados) have been a big challenge for the government to face and deal with to achieve a satisfying result. During a disaster, the main aim is to save people's life by supplying effective help to the victims, then to decrease loss to the minimum level and recover the situation before the damage. Twitter as a crowdsource method has proven that use it is helpful to collect information before and after a disaster. Although, Twitter is designed to go ahead crowdsourcing. As well for rapid system, secure coordination, and trustable help organizations, some other capabilities are required. Goolsby pointed out the need for such a social media system and described the effort to build it to allow different organizations to share crowdsourced information and analyze and visualize the processed information for intelligent decision making. (Social media as the platform of crisis: The prospective of society maps or crisis maps. ACM Transactions on Intelligent

Systems and Technology). Here we are going to explain TweetTracker social media systems made to ease effective association between different organizations to effectively handle disaster crisis.

1.4 Tweet Tracker

Tweet Tracker is a means based on Twitter social network for visualizing and analyzing. The main aim of this means is to help HADR aid community to get aware of situations while calamity and catastrophe are going on. (See: S. Kumar, R. Zafarani, and H. Liu. Understanding user migration patterns across social media. Twenty-Fifth International Conference on Artificial Intelligence. Association for the Advancement of Artificial Intelligence, Palo Alto, CA, 2011.) Social media networks for example Twitter have the ability to supply information which is not possible to get from traditional media. We can give Mumbai blasts as an example: after the calamity, immediate information from the damaged area was available on Twitter. It is done via real-time Twitter tracking using special keywords or hashtags. This tool provides checking and making an analysis of the gathered tweets with sophisticated mining techniques in real time.

Twitter Tracker has three basic parts:

1. Twitter stream reader.
2. Data storage module
3. Data mining and visualization module.

The first one- Twitter stream reader is module based on collecting data which repeatedly tracks tweets via the Twitter streaming API (Application Programming

Interface). To filter tweets according to location, keywords and hashtags are possible. The main function of data storage module is to store and index gathered tweets into database so that visualization module can use it. This module is Web-based user interface. It maintains geospatial visualization of tweets which is connected to map of a special event. It sums up the tweets, conceptualize keywords which are on demand, also detect the URL address of popular tweets and define the users which mentioned in the tweets. As well as, this tool has integrated language to analyze monolingual tweets.

1.5 Fundamentals of Twitter.

We can define Twitter as a real-time, extremely social microblogging service which offers consumers to share short post updates or tweets, that seems on the timelines of users. Tweets can cover one or more objects in their characters of theme and representatives one or more places which chart to the locations in the true life. Nowadays this is equal to 280. According to the perception of users, tweet, and timeline are especially important to efficient use of Twitter's API, hence a short introduction to fundamental notions is before we are in contact with API to get few data. Before we discussed general information about Twitter so far, and in this part, we will shortly introduce tweets and timelines due to filling out an understanding of the Twitter platform. Tweets are the integrated part of Twitter, and while they are relatively accepted as brief strings of text content connected to a user's status update, there's really quite a bit more metadata there than meets the eye. Furthermore, to the textual content of a tweet, tweets have brunch with two more pieces of metadata which are of specific item: places and entities. Tweet entities are basically the user's hashtags, URL addresses, and media which might be connected to a tweet, and places are locations in the real world which can be

added to a tweet. Keep in mind that a place can be the real location in which a tweet was made, also it can be a reference to the place mentioned in a tweet.

1.6 Twitter API's overview

Twitter platform has many APIs which software developers can work with, and these APIs with high chance to create fully automated systems allow users to communicate with Twitter. Companies can take benefit from these features by understanding Twitter data, small scale projects and researches might use it in an effective way. Here we introduce some of the most popular APIs provided by Twitter:

- Tweets: searching, posting, filtering, engagement, streaming etc.
- Ads: campaign and audience management, analytics.
- Direct messages (still in Beta): sending and receiving, direct replies, welcome messages etc.
- Accounts and users (Beta): account management, user interactions.
- Media: uploading and accessing photos, videos and animated GIFs.
- Trends: trending topics in a given location.
- Geo: information about known places or places near a location.

While there are more characteristics of the Twitter APIs, we only added famous one to our list. Besides Twitter is steadily shrinking its set of services with adding APIs now and then and updating old APIs.

1.7 Python wrappers.

There are some programming languages that have enough number of wrappers and Python is one of them with the highest number of developed wrappers for Twitter API. But to see difference between them can be difficult for you, if you are not acquainted with them. Confidently by going deep into them, observing the services they offer you and checking in what extent they go with your application you can choose the best wrapper for you. In this section, we'll make comparison between the different API wrappers. Some appropriate characteristics to compare would be: number of contributors, number of received stars, number of watchers, library's maturity in timespan since first release and so on. In table 2 we describe these characteristics:

Table 2: Python libraries for Twitter API ordered by number of received stars.

Library	# contributors	# stars	# watchers	Maturity
tweepy	135	4732	249	~ 8.5 years
Python Twitter Tools	60	2057	158	~ 7 years
python-twitter	109	2009	148	~ 5 years
twython	73	1461	100	NA

TwitterAPI	15	424	49	~ 4.5 years
TwitterSearch	8	241	29	~ 4.5 years

1.8 Healthcare in the social world.

What is the role of social media in healthcare? Is this role bigger and effective? Let's look at a study related to these kinds of questions.

A study compiled by Demi & Cooper Advertising and DC Interactive Group shows that more than 90% of people ages 18-24 said they would trust health information they found on social media channels. Almost half of adults in the world use their mobile phones for searching health information. Patients are eager to share their opinions on the Internet about the services and care they got. 44% of people said that after visiting hospitals they might share their experiences about facilities despite it is positive or negative, and 42% of people said they don't fear to comment about doctors, nurses and providers on the social media. Nowadays approximately 30% of hospitals have websites and social media accounts. Moreover, 60% of doctors claim that social media plays significant role in improving healthcare quality. It is doubtless that when patients share good reviews and comments on the social media, hospitals and doctors take profit from it. While doctors must keep this information private, according to the HIPAA laws, (Health Insurance Portability and Accountability) it's important that they know about borders of using social media. Also, doctors need to be careful with the information that they tell patients on the social networks. From the security side it

is important to take it into consideration because medical information and doctor's advice should be totally secret. "The great thing with social media is it can be shared, but that's the downside (for health information)," doctors says. Because healthcare is new in this field. After couple of years to share this information for both doctors and the patients will be available in high quality when much more trustable technologies appear. Not only patients but also doctors who comment on public networks by suggesting medical advice might have liability problems, too. However, "secured patient portals are a great way to leverage mobile technology to promote healthy behavior."

Can social networks help healthcare?

According to some researchers, improving facilitations of social media and e-communication between patient and doctor will define the future of healthcare. These researchers claim that social media has a significant role to increase health and wellbeing of the patients by creating information and provide world widely available. On the other side other researchers recommend social networks be connected to provider processes and content. "Organizations can coordinate the information from social media space and connect with customers in more meaningful ways that provide value and increase trust". To catch the eye of users, the standard design for these platforms might be created with the summarized theory and fit to emotional and informational needs of users and user participation dynamics. From this point of view, officially authorized healthcare social media platforms could be evolved to support value-creating social features and functions, and rich content, typical of high-traffic sites on the open Internet, to encourage participation. All this with the added benefit of healthcare professional view and

the chance to connect with other peer patients and without doing a wide search all over the Internet websites related to cancer. “To maximize the potential of these online communities, it is thus preferable to have guidance from health professionals, who can lead, moderate, and bring into the discussion the expertise required in their off-line world”. An advanced, masterly designed social media platform can do a lot to authorized patients, and possibly developing the user satisfaction, well-being and health results. To choose website arbitrator among participants can be considered as one recommendation. In healthcare, those arbitrators can be selected from the staff of hospital.

Chapter 2

Mining Twitter

2.1 Review of the literature

The literature review was directed belonging to Twitter and healthcare research, which used a numerical way of sentiment analysis for the free-text messages (tweets). This research compared the different tools used in.

As we already know sentiment analysis makes the meaning of free-text natural language (which is, the words and attributes operated in a message) possible for checking the density of the opinions which can be positively or negatively. As well as, sentiment analysis based on social network data has been researched in a wide range. (See: Pang B, Lee L. Opinion mining and sentiment analysis. Foundations and trends in information retrieval. 2008).Sophisticated software implementations allow big numbers of messages to be operated into numerical sentiment data

quickly, like positive or negative data. These usually are based on the text classifiers and processes of machine learning.

Healthcare sentiment analysis is not a new concept. Studies shows that by using manual notation of the healthcare tweets, about 40% of messages have some positive or negative sentiment. An approach which is based on hand work was done related to examining of self-destruction data and dismissing summaries, where Cherry tried to make automation the handwork procedure with help of machine learning methods (See: Treves A J, Carnaud C, Trainin N, Feldman M, Cohen I R. Enhancing T lymphocytes from tumor-bearing mice suppress host resistance to a syngeneic tumor. Eur J Immunol. 1974; Yadav R N. Isocitrate dehydrogenase activity and its regulation by estradiol in tissues of rats of various ages. Cell Biochem Funct. 1988). They came to the result that the manual categorization of spiritual texts was rough and incoherent.

These researches show that sentiment analysis has already found a way to the recognized analysis of health care research collected from Twitter. Twitter is an easy-to-use platform because data is collected using their API. Social media such as Facebook is not convenient from this point of view as it doesn't allow developers and researchers to get the date easily like Twitter. It can be considered as a limitation of other social platforms.

Precisely proceeding the sentiment analysis of a healthcare tweets gives a chance to patients and doctors to understand the health situation (Greaves F, Ramirez-Cano D, Millett C, Darzi A, Donaldson L. Use of sentiment analysis for capturing patient experience from free-text comments posted online. J Med Internet Res. 2013).

Up to now, no study has been done which looks at all sentiment analysis methods for healthcare tweets. SentiStrength (http://sentistrength.wlv.ac.uk/) is a famous and free software source which is was based on nonspecific messages. Due to some aspects the healthcare can have a varied environment. The borders between "patient," "consumer," and "customer" is very tough to clear in the healthcare because of public NHS (National Health Service) Hence, today existing sentiment analysis methods mighty be inaccurate.

Recognition and classification

In May 2015 a study conducted. Studies needed to count in one of the below mentioned terms for title, abstract, or keywords: "Twitter" or "tweet" or "microblog" and "Sentiment" or "opinion" or "emoti" or "happi" or "Senti." There were 3 containment standards for this work. First of them is that, Twitter must be taken as main center. The reason of it was to discover study into the methods of sentiment analysis for only Twitter data. Second, healthcare should be related to as the topic. Delivery of health and healthcare, healthcare research, rules and use – all factors are added here. Last step, documents which kept numbers decided to study positive and negative sentiments, like, "−5," were added.

Suitability and Containment.

The language of the studies was English. 69 full text articles totally were collected to check eligibility. Approximately (10 papers)15 % of these texts were excluded, as they didn't take a glance at Twitter only, but also other social media. Furthermore, 36% about 25 papers again were excluded as their focus was not healthcare. Additionally, 32% overall 29 papers were rejected since their sentiment analysis was not studied completely in terms of measure and characteristics. The

comparison of the methods is based on some criteria like tool production, testing the tool for each study. To evaluate, an analogy of some quantity of interpreters who utilized manual work to comment on the tweets, if they detected anything, and the status of arrangement among them was used. What is more, the ratio of tweets that utilized for training the algorithm made comparison with the last pattern examined was as well as accessed.

Results

Overalls, in 12 papers out of 69 were shown the satisfied all the 3 inclusion factors. From 2011 until 2016 these obtained results were published with the gathered Twitter data between 2006 and 2016. Additionally, 2 papers have analyzed worldwide data. These 2 papers show that about 46% of the healthcare tweets maintain sentiments. Also, public health analysis has been done on some studies. Moreover, 3 papers conducted the sentiment analysis related to features of diseases: disease (n=1), symptoms of diseases (n=1), or treatment (n=1). At last, 2 papers examined an emergence situation of medical and a medical assembly.

Among 12 papers 5 papers total examined a manual sentiment analysis of a pattern of their data using annotators to produce the method. One of them used 13.58% (1000/7362) of their final data sample to produce method. 3 studies used about 0.7% of their whole data to produce their method (1.46%, 250/17,098; 0.55%, 2216/404,065; and 0.1%, 250/198,499). And a paper confronted the truth of the selected tools with manual annotated framework of the data. Furthermore, 2 papers among commented on reason of the sentiment analysis methods which are used.

2.2 The twitter API

Twitter proposes number of APIs to satisfy programmatic permission to Twitter data. To read tweets, access user profiles, and post updates on behalf of a user are included here. On account of set up the scheme to reach Twitter data, two initial steps should be followed as below:

To register the application

To choose Twitter API client

To register it, we will use few minutes. Let's say we already did this step and signed into the account. Then we have to rectify the browser at the Application Management page at *http://apps.twitter.com* and build new application. After registering our application, the needed data for authentication can be found under the Keys and Access Tokens tab. The Consumer Key and Consumer Secret (also called API Key and API Secret, respectively) are a setting of your application. The Access Token and Access Token Secret are instead a setting for your user account. Your application can potentially ask for access to several users through their access token. The Access Level of these settings defines what the application can do while interacting with Twitter on behalf of a user: read-only is the more conservative option, as the application will not be allowed to publish anything or interact with other users via direct messaging.

Rate limits

The Twitter API limits access to applications. These limits are set on a per-user basis, or to be more precise, on a per-access-token basis. This means that when an application uses the application-only authentication, the rate limits are considered globally for the entire application; while with the per-user authentication approach, the application can enhance the global number of requests to the API. It's important

to familiarize yourself with the concept of rate limit, described in the official documentation. Available here:
https://developer.twitter.com/en/docs/basics/rate-limiting.html
It's also important to consider that different APIs have different rate limits.
The implications of hitting the API limits is that Twitter will return an error message rather than the data we're asking for. Moreover, if we continue performing more requests to the API, the time required to obtain regular access again will increase as Twitter could flag us as potential abusers. When many API requests are needed by our application, we need a way to avoid this. In Python, the time module, part of the standard library, allows us to include arbitrary suspensions of the code execution, using the time.sleep() function. For example, a pseudo-code is as follows:

```
 # Assume first_request() and second_request() are defined.
# They are meant to perform an API request.
import time
first_request()
time.sleep(10)
second_request()
```

In this case, the second request will be executed 10 seconds (as specified by the sleep() argument) after the first one.
Twitter provides more than a single API. In fact, it will be explained in the following section which there are available means to get Twitter data. To keep things simple, we can categorize our options into two classes: REST APIs and Streaming API:

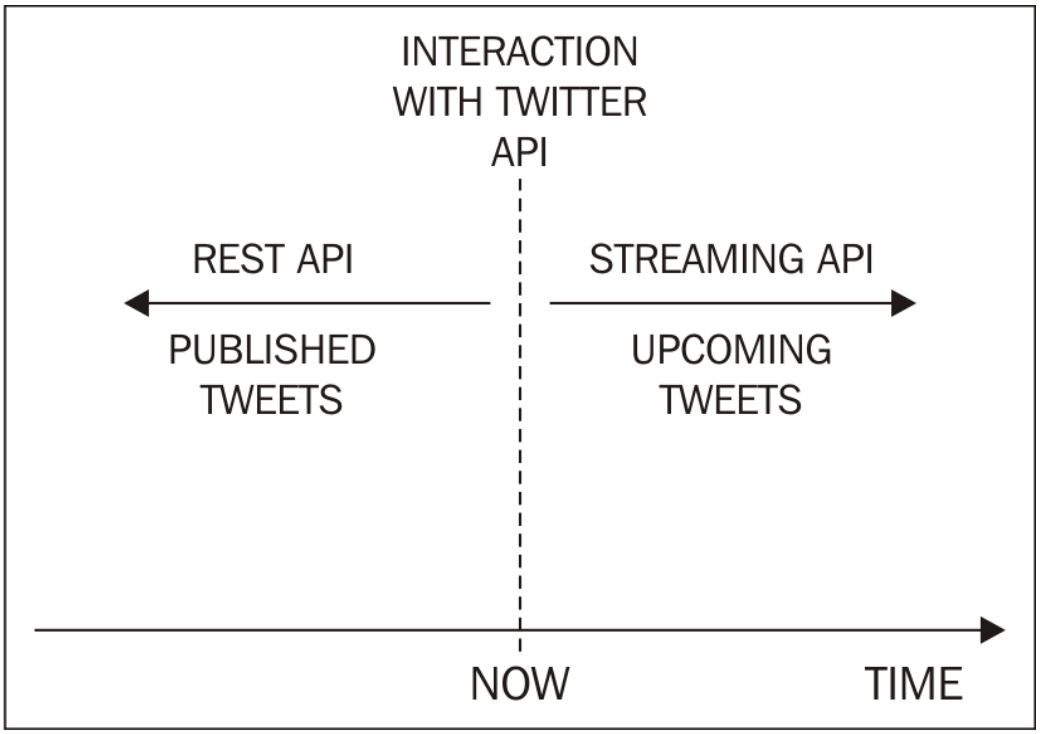


Fig 1: The time dimension of searching versus streaming

The difference, summarized in picture 1, is simple: all the REST APIs only allow you to go back in time. When interacting with Twitter via a REST API, we can search for existing tweets in fact, that is, tweets that have already been published and made available for search. Often these APIs limit the number of tweets you can retrieve, not just in terms of rate limits as discussed in the previous section, but also in terms of time span. In fact, it's usually possible to go back in time up to approximately one week, meaning that older tweets are not retrievable. A second aspect to consider about the REST API is that these are usually a best effort, but they are not guaranteed to provide all the tweets published on Twitter, as some tweets could be unavailable to search or simply indexed with a small delay.

2.3 Collecting data from Twitter

In order to interact with the Twitter APIs, we need a Python client that implements the different calls to the API itself. There are several options as we can see from the official documentation. Available here:

https://developer.twitter.com/en/docs/developer-utilities/twitter-libraries.html

None of them are officially maintained by Twitter and they are backed by the open source community. While there are several options to choose from, some of them almost equivalent, so we will choose to use Tweepy here as it offers a wider support for different features and is actively maintained.

The first part of the interaction with the Twitter API involves setting up the authentication as discussed in the previous section. After registering your application, at this point, you should have a consumer key, consumer secret, access token, and access token secret. In order to promote the separation of concerns between application logic and configuration, we're storing the credentials in environment variables. Why not just hard code these values as Python variables? There are a few reasons for this, well summarized in The Twelve- Factor App manifesto. Available here:

https://12factor.net/config.

Using the environment to store configuration details are language- and OS-agnostic. Changing configuration between different deployments (for example, using personal credentials for local tests and a company account for production) doesn't require any change in the code base. Moreover, environment variables don't get accidentally checked into your source control system for everyone to see. In Unix environments, such as Linux or macOS, if your shell is Bash, you can set the environment variables as follows:

```
$ export TWITTER_CONSUMER_KEY="your-consumer-key"
```

In a Windows environment, you can set the variables from the command line as follows:

$ set TWITTER_CONSUMER_KEY="your-consumer-key"

The command should be repeated for all the four variables we need.
Once the environment is set up, we will wrap the Tweepy calls to create the Twitter client into two functions: one that reads the environment variables and performs the
authentication, and the other that creates the API object needed to interface with Twitter.

```
# Chap02-03/twitter_client.py
import os
import sys
from tweepy import API
from tweepy import OAuthHandler
def get_twitter_auth():
"""Setup Twitter authentication.
Return: tweepy.OAuthHandler object
"""
try:
consumer_key = os.environ['TWITTER_CONSUMER_KEY']
consumer_secret = os.environ['TWITTER_CONSUMER_SECRET']
access_token = os.environ['TWITTER_ACCESS_TOKEN']
access_secret = os.environ['TWITTER_ACCESS_SECRET']
```

```
except KeyError:
sys.stderr.write("TWITTER_* environment variables not set\n")
sys.exit(1)
auth = OAuthHandler(consumer_key, consumer_secret)
auth.set_access_token(access_token, access_secret)
return auth
def get_twitter_client():
"""Setup Twitter API client.
Return: tweepy.API object
"""
auth = get_twitter_auth()
client = API(auth)
return client
```

2.4 The structure of a tweet.

A tweet is a complex object. The following table provides a list of all its attributes and a brief description of their meaning. In particular, the content of the API response for the particular tweet is fully contained in the _json attribute, which is loaded into a Python dictionary:

Table 3: Descriptions of Python library attributes

Attribute Name	Description
`_json`	This is a dictionary with the JSON response of the status
`author`	This is the `tweepy.models.User` instance of the tweet author
`contributors`	This is a list of contributors, if the feature is enabled
`coordinates`	This is the dictionary of coordinates in the GeoJSON format
`created_at`	This is the `datetime.datetime` instance of the tweet-creation time

entities	This is a dictionary of URLs, hashtags, and mentions in the tweets
favorite_count	This is the number of times the tweet has been favorited
favorited	This flags whether the authenticated user has favorited the tweet
geo	These are the coordinates (deprecated, use coordinates instead)
id	This is the unique ID of the tweets as big integer
id_str	This is the unique ID of the tweet as string
in_reply_to_screen_name	This is the username of the status the tweet is replying to
in_reply_to_status_id	This is the status ID of the status the tweet is replying to, as big integer
in_reply_to_status_id_str	This is the status ID of the status the tweet is replying to, as string
in_reply_to_user_id	This is the user ID of the status the tweet is replying to, as big integer
in_reply_to_user_id_str	This is the user ID of the status the tweet is replying to, as string
is_quote_status	This flags whether the tweet is a quote (that is, contains another tweet)
lang	This is the string with the language code of the tweet
place	This is the tweepy.models.Place instance of the place attached to the tweet
possibly_sensitive	This flags whether the tweet contains URL with possibly sensitive material
retweet_count	This is the number of times the status has been retweeted
retweeted	This flags whether the status is a retweet
source	This is the string describing the tool used to post the status
text	This is the string with the content of the status
truncated	This flags whether the status was truncated (for example, retweet exceeding 140 chars)

The attributes that hold an ID (for example, user ID or tweet ID) have a counterpart where the same value is repeated as a string. This is necessary because some

programming languages (most notably JavaScript) cannot support numbers with more than 53 bits, while Twitter uses 64-bit numbers. To avoid numerical problems, Twitter recommends the use of the *_str attributes.

We can see that not all the attributes are translated into a Python built-in type such as strings or Booleans. In fact, some complex objects such as the user profile are included completely within the API response, and Tweepy takes care of translating these objects into the appropriate model. The following example showcases a sample tweet by Packt Publishing in the original JSON format as given by the API and as accessible via the _json attribute (several fields have been omitted for brevity).

Firstly, the creation date in the expected format, is as follows:

{

"created_at": "Fri Oct 30 05:26:05 +0000 2015",

The entities attribute is a dictionary that contains different lists of labeled entities. For example, the hashtags element shows that the #Python hashtag was present in the given tweet.

2.5 Present utilization of Twitter in Healthcare

Nowadays Twitter has a leading role over medicine due to some advantages which it offers. Healthcare suppliers, who have about 300-350 followers and tweet once or twice in a day, are doing good job. The majority of tweets of medical providers usually gives general information about public health, as well as news related to medical technologies and researches about treatment methods. For example, Picture 2 shows tweet of Dr. Sheila Sahni, who is a cardiologist and has roughly 7000 followers and 8500 tweets. He shares a last work about heart attack indications. Doctor compared men and women according to various reasons and

signs. As we mentioned above that, this kind of tweet make the profits of Twitter maximum with sharing new study and giving information to people, without breaking the patient privacy rules. Although, physicians who are new in their work, or have less experience, they shared their interaction process with patients, patients' X-ray results and some private information. Doctors should remove patients' personal information and data from tweet to avoid of privacy breaking when they want to share it. A study showed that a group consisted of 500 doctors who had about 200 followers, approximately in 5% of their tweets privacy issues of patients were broken. Doubtlessly, these violations of patient privacy considered as unprofessional, inappropriate and illegal. Furthermore, doctors who had less than 300 followers seldom had contradiction "Tweet is just my opinion, not advice". This is essential factor in order to notify patients about the nature of tweet to avoid confusion and wrong information. Hence, studying tweets of health masters reported that more experienced and more proficient doctors usually have enough followers. Experienced doctors as well as are eagerly to use Twitter to increase the efficiencies mentioned before, for example, to make health information available to people, to decrease privacy of bum steer, and unprofessionalism.

Fig 2: Sample tweet

On other side, Twitter is like a bridge to share and connect physicians to professionals in different areas. For instance, conferences on researches boomed Twitter's capacity to compose absorbing and develop influence. Physicians may practice learning of clinical tests via Twitter, accept patients. Moreover, Twitter has influence on quality of the care and availability for the health information. Multi-tiered sentiment analysis model, data-mining, and artificial intelligence applications are also helpful to medicine to stud tweets of population. Thanks to resource allocation of this real-time information, doctors can make scheme for cures. Former works showed the forecasting strength of Twitter in public health from tracking infectious disease outbreaks, natural disasters, drug use, and so on. It is obvious that before mining big dataset was real trouble because of low computational power. Recent progress in graphics processing units (GPUs) created effective and momentary mining and analysis. Studying the present usage

of the Twitter in healthcare has showed that numerous doctors use Twitter efficiently to give knowledge about their works amidst their collaborators and patients. Although, some are not aware of the adequate role of Twitter. So, there is knowledge deficit is which is complicated. It requires enough experience of using social media and learning its side effects among doctors and healthcare providers.

2.6A disease emergence announcement framework for smart home panel.

Some diseases like MERS (Middle East Respiratory Syndrome), Ebola are spread in many countries. For these countries, outbreaks of diseases became emergent worry. Government can take action when emergence of disease spread out in a big amount. When it comes to this issue, social media has great role. Mining Twitter for these kinds of problems is not so easy as there are numerous problems:

1. To filter noisy information,
2. To create dictionary
3. Sentiment analysis
4. To build authenticity.

The system was built to make a framework for mining and removing information from twitter concerned to outbreak of disease. This system uses many programming interfaces to define user and give information to user about concerned problems related to health. The location of user also is taken into account. That panel is included in a smart home dashboard system, which let the user know about disease outbreak information. This panel gives an efficient human-home connection dialogue to the user.

Related Work

Social media mining for useful data can't be considered as a newfangled idea. Proper amount of people had some research on this topic.

Additionally, researches show that social media works quicker that traditional media being the source of large information. Twitter catches trend among these social media. Laaksonen (Laaksonen C, Jalonen H, Paavola J. Utilising Social Media for Intervening and Predicting Future Health in Societies, Proc. 5th International Conference on Well-Being in the Information Society 2014) researched the relation between social media and health and utilizing social media for predicting health on a population level. They suggested the use of sentiment analysis to detect symptoms of collective negative emotions in social media feeds. Here we will look at mined tweets using sentiment analysis. Paul and Dredze (Paul, M. J., & Dredze, M. A Model for Mining Public Health Topics from Twitter. Technical Report. Johns Hopkins University 2012) successfully attempted to discover mentions of over a dozen ailments from over one and a half million health related tweets and then incorporate prior knowledge to apply several tasks like localizing ailments according to locations and examine attributes and drug utilization.

Although, the main difference between these researches is that here we will use two approaches: online and offline. Firstly, the hashtags, which are on trend, will be defined. Later according to hashtags, keywords will be selected. This will let us to see if these keywords contain useful information like tags, which never used before related to symptoms or illnesses.

Twitter Mining Framework

In this system, Twitter is used to mine and extract disease outbreak information. During last years, some works have been done in order to take out useful data

related to public health as we talked about upper. Now the idea is like this: using Twitter Search and Streaming API and compound of an online and offline ways to supply the disease outbreak information to the users. Let's look at offline method first. Offline is good at mining for Twitter Search API with well-known ailments. As we know, Twitter Search API gets access to old tweets. Online mode identifies and uses commonly used terms to define and categorize newer and unfamiliar ailments.

Offline mode:

In the offline mode, the disease names are used to define general key words, which happens in the tweets related to ailments. The system takes the Twitter Search API to access a group of tweets, which suits the ailment names, that administrator supplies. A Parts-of-Speech (POS) tagger was used to section the tweet text and categorize or tag its parts of speech (Singular/Plural Nouns, adjectives, etc.). Among the regained messages, density of the words was enumerated. Confirmed catalog of commonly used keywords have included to the database in order to use mining streaming tweets while online method.

Online mode:

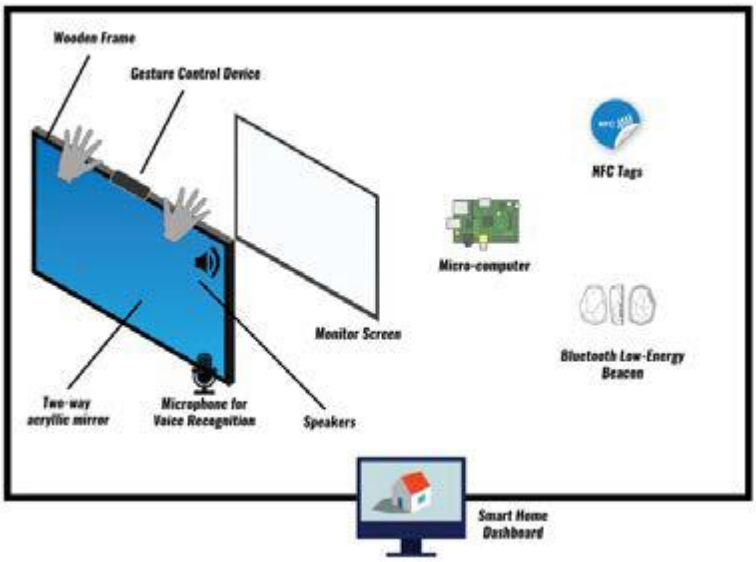
It should be taken into consideration that at what times it is effective to mine the Twitter Streaming API. Due to define these times, we allow the program mine tweets from the Streaming API during a whole day. Once the density of happening data was made, this information was graphed in durations of two hours to define the best time. Firstly, the system will get the common keywords from the database, then use them to get tweets live from Twitter Streaming API for two hours. Then it will take the retrieved tweets and discard the tweets that don't have an embedded location. After that, it performs sentiment analysis on filtered tweets to get a float value of subjectivity in a range of [0.0, 1.0] and discard tweets above 0.5 range. After that, the mined tweets are segmented into words and compared with the list

of diseases in the database. If there is a match, the tweet is stored as a new case in the database containing the matched disease, two-digit country code (extracted from embedded location in the tweet), a count of tweets from that country mentioning that disease, and the date. If the tweet did not match any disease from the database, the tweet's text is inserted into a table in the database that is later analyzed by the administrator and decide if its another name for a stored disease, or a new disease was mentioned.

Smart Home Dashboard Architecture

In order to implement a smart home dashboard, we used the Magic Mirror project recently developed by Xonay Labs5 as a baseline for our proposed work. The architecture of the smart home dashboard with its main components is shown in Figure 3.

Fig 3: Main components of smart home panel

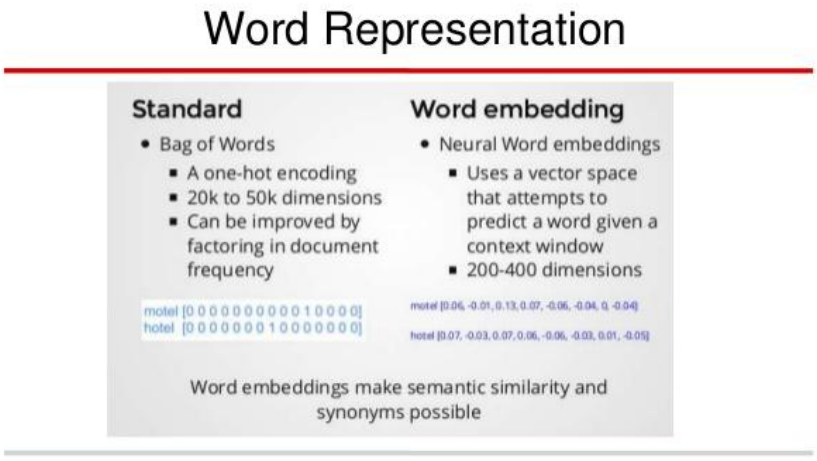


2.7Mining drug-related Twitter data: Language patterns and benefits.

Twitter data contains drug names in terms of keywords as well as their widespread spelling forms were gathered. This data has plenty of drug-related information, useful or useless. According to this data, language models were developed to help the data mining tools and methods to make progress in the special area. Two models were generated:

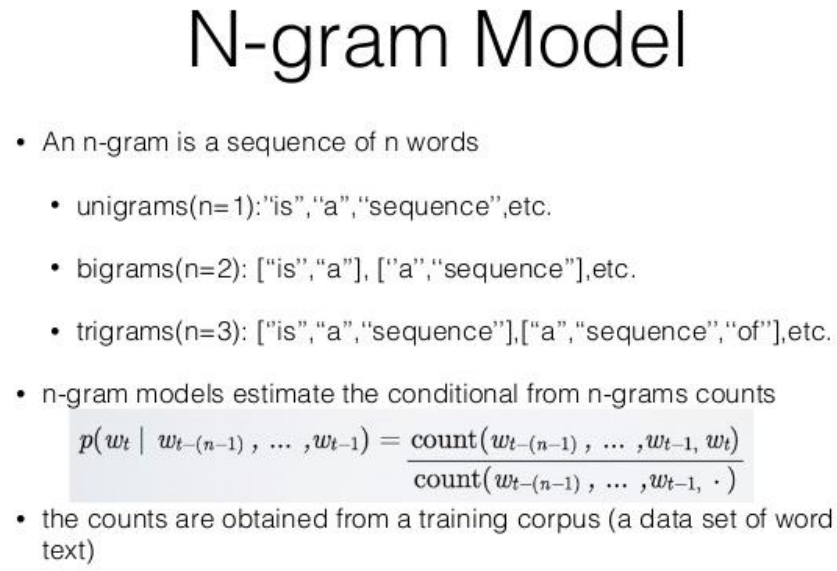
1. Distributed word representations

Fig 4 shows word representation:



2. Probabilities of n-gram sequences.

Fig 5 shows general overview of n-gram model:



Importance of data

Unrefined data carrying drug-related information is useful to develop sample system. In the field of pharmacy and toxicology number of tasks used to access the user sentiments related to the drugs, to calculate the efficiency of different drugs and for other kind of tasks significant to the wider public health society.

The distributed word representations were composed according to modifying plenty of parameter compositions, hence making sure that the varied patterns possess dissimilar kinds of the semantic information. Thus, to build systems based

on mining information connected to receipt of remedies, we may employ these models.

The other model which is n-gram language model takes consecutive possibilities of word circumstances, for text classification and text normalization, this model is good choice.

To download tweets and load two varied group of models Python scripts are used. The two groups of language models come along unrefined dataset. the first model is a group of models built on the distributed semantics, that cover semantic features with representation of word symbols like tight vector, on the other hand the second group is built on n-gram sequences, holding consecutive samples.

Features of data.

This data set consists of 267,215 Twitter posts, which are tweeted between four-month duration, from November 2014 until February 2015. These tweets talk about more than 250 drug-related keywords. Semantic and consecutive features of these tweets are included to language models.

We can see the monthly deploy of data via diagram 1. This graph shows that the amount of drug-related tweets gathered are enough stable for four months period. December had maximum number of drug-related posts. Why? Reason can be holiday. The table shows as well as the periodicity of the upper 10 drug-related key words in the data pattern which is delivered. Keyword which had been found in high frequency is "*Adderall*". Some drug-related key words were talked about in the set of data, showing knowledge related to classification of drugs like antipsychotics, narcotic pills, stimulants, anti-depressants, pain killers, sedatives and some others.

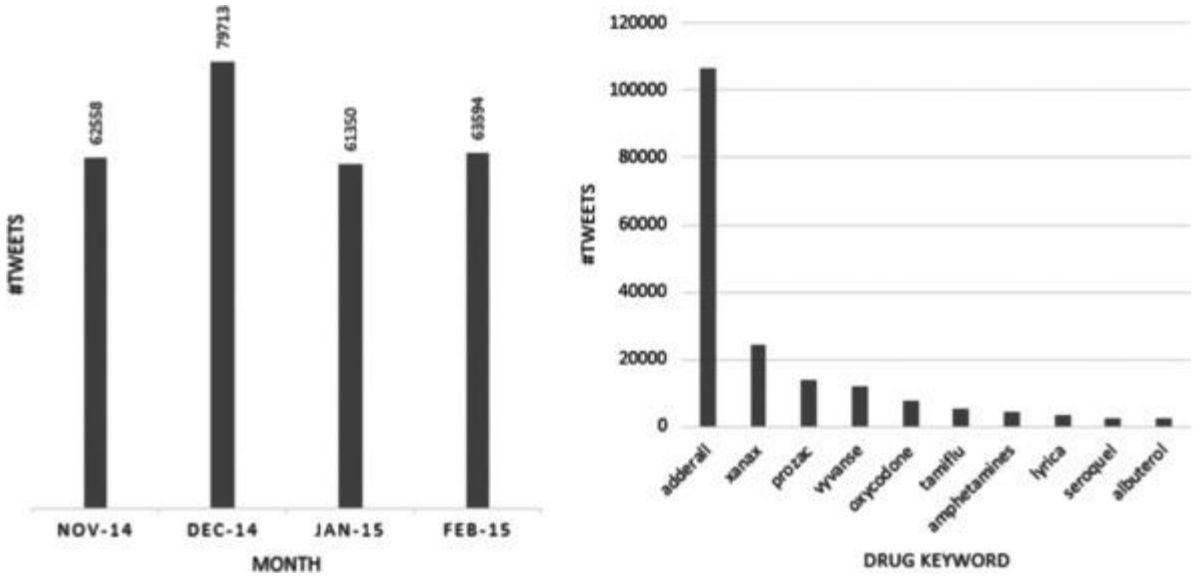


Fig 6: Diagram on the left side shows distribution of tweets for four months Diagram on the right shows the upper ten drug-related keywords on Twitter for the same time duration.

Data collection

To start the gathering of Twitter data, beginning step was to define the drugs which are in the area of interest. Drugs which have different group of conditions and large percentage of usability were associated. Hence, the two factors were taken into consideration after getting advice professionals of pharmacology:

- ✓ Drugs have constant conditions
- ✓ Drugs have upper diffusiveness in terms of usage

In the first measure, receipted drugs for some cases were selected. Here involved but not limited to nicotine addiction, coronary vascular disease, depression, Alzheimer's disease, chronic obstructive pulmonary disease, hypertension, asthma, osteoporosis and type 2 diabetes. Second step was completed with the first list with drugs inclined to abuse, antivirals and biologics. Statistics which used for the second specification related to commonness of usage, were taken from the IMS

(Information Medical Statistics) Health's highest 100 drugs according to their sale amount between 2013-2015.

There was a problem regarding to collecting drug-related Twitter data: People frequently misspell names of drugs. Table 4 represents the tweets which mentions two drugs Seroquel and Adderall with different misspelling forms as well us the right forms. This table says that, to collect data by using right spelling form of the drug can't be enough, because only one keyword has more probability of failure which can lead to undesired results. For this reason, misspelling generator was created. This generator builds all common wrong-spelled drug forms according to the right form. given. The misspelling generator works with 3 steps.

- All lexical forms of the keyword with a Levenshtein distance are built. Mathematically, the Levenshtein distance between two strings, a and b (of length |a| and |b| respectively), is given bylev a,b(|a|,|b|) where:

$$\operatorname{lev}_{a,b}(i,j) = \begin{cases} \max(i,j) & \text{if } \min(i,j) = 0, \\ \min \begin{cases} \operatorname{lev}_{a,b}(i-1,j) + 1 \\ \operatorname{lev}_{a,b}(i,j-1) + 1 \\ \operatorname{lev}_{a,b}(i-1,j-1) + 1_{(a_i \neq b_j)} \end{cases} & \text{otherwise.} \end{cases}$$

Fig 7. Equation for Levenshtein distance.

Here, 1(ai≠bi) is the indicator function equal to 0 when ai≠bi and equal to 1 otherwise, and leva, b(i,j) is the distance between the first i characters of a and the first j characters of b.

- Since these variants can be phonetically very dissimilar from the original keyword, a filter is applied which only keeps misspellings that have the same phonetic representations as the original keyword.

- Finally, to obtain a smaller set of misspellings, the Google custom search API is used to identify those that are frequently used as input to the search engine. Table 4 shows an arbitrary example of the tweets which connected to drugs. The tweets seem to show several kinds of knowledge related to the drugs. According to the scheme, different kinds of drug-related information might be mined from this database, allowing this become worth source.

Seroquel

same here. i take lamictal, effexor xr, klonazepam, seroquil. i know i could not function without them.

stomach hurts from missing a day of my pill, guess that also explains why i was a bit moody today. seroquil... it's not as bad as paxil thoug

when i was taking seroquel it sedated me so much i was incapable of being fully conscious before 3pm.my psychiatrist didn't care

i got expired zyprexa seroquill n lithium

this child is on two medications that counteract each other. serroquil and trazadone makes for a dangerous combination

the seroqul lamictal effoxer xr. seems not to be working.. anyone know of a good therapist.. lmao :D

Adderall

i take aderall for my adhd.i have 2 explain 2 them in blood tests since it has amphetamines in it.legal speed

zero sleep thanks to adderall

i am with you though, if nfl has amphetamines associated with adderal and vyvanse, fighters should not be allowed to do coke haha

not only antidepressants but other medications as well. i was misdiagnosed with adhd & was put on aderol and seroquel for 10 yrs

and if thats not bad enough she accomplished this with excessive amounts of adderol and vyvanse

which do you think currently has a higher demand, amphetamines (aderoll, vyvance) or cocaine? especially considering time of yr

Fig 8: Pattern tweets which contain right and wrong forms for Seroquel and Adderall drugs.

```
prozac is good for that too. pretty much any anti anxiety pill
will put you out if you want to sleep. clonazepam is good too.

railing adderall at 4 in the morning is my aesthetic

also adderall prevents me from having any feelings other than
tired rage.

taking a vyvanse and doing something with my life tonight.
i didn't accomplish anything sobbing all day.

i feel like adderall is cocaine for broke college students

i've read novels with lesser plots than this xarelto commercial.

honestly god bless whoever created adderall u dont know me but
i love u very much

been recommended paroxetine for severe tourettes, any chance
of retweet to see others experiences? had akathesia in past.

i am ok my neurologist took me of tecfidera because i had
really bad side effects!

my prescription of tamiflu cost $150 without insurance, was
$50 with coverage. apparently treating the flu is only for
the wealthy

basically body wants/need to purge, you take imodium you end
up much sicker for longer. :( i almost took imodium before i
puked.

snorted 2 15 mg oxycodone ($24)

sos i need adderall delivered to yorktown mall plz and thank
u
```

Fig 9: Pattern tweets holding drug-related information. Gathered during November of the year 2014 and February of the year 2015.

Language design built on deploying semantics

Several expression-level models were made with help of the word2vec tool. This tool presently used in creating distributed word representations which gained from the data of the natural language. The tool offers an algorithm in order to build a

dictionary from an unlabeled dataset. Vector presentations of the words are learnt with preparing a not-deep neural networks consisted of 2 layers. Officially, t-series of terms and their text windows w were given. The impartial function is to set the parameters H that maximize P(w|t;H). The context, which is a symmetric window of terms around an input word, can be varied to modify properties of the distributed models (*e.g.*, varying window sizes may impact the capture of syntactic and semantic properties. Similarly, the vector sizes and various other parameters influence the qualities of the models.

For the generation of the models, several parameters were varied including vector sizes and context window sizes. For the vector sizes, models between the sizes 200 and 400 were generated. For the different vector sizes, we generated models using context windows within the range. These models can be used to explore a variety of research tasks that can benefit from drug related knowledge generated directly from the users, such as exploring associations between drugs and adverse reactions, identifying drug abuse related signals, and, from a more NLP/data science perspective, the effects of different parameters such as context window and vector sizes in capturing semantic and syntactic properties. Such distributed word representation models are already being applied for research utilizing other sources of noisy health-related data, such as clinical reports and such models have been generated from texts from other domains such as published literature and generic social media. However, perhaps due to the absence of available drug-related chatter data from social media, there are currently no such models available for this domain. We, therefore, believe that these models, along with the set of unlabeled data, will aid drug-related data mining tasks and will complement the resources generated from other domains (*e.g.*, clinical reports). For example, while clinical reports hold information discovered in clinical settings, social media

chatter may hold information that are expressed by users in non-clinical settings. We present sample utilities of the models in the next section.

Language model based on n-gram sequences

Sequential, n-gram language models were generated. These language models capture the probabilities of n-gram sequences and such models have been applied in the past for tasks such as lexical normalization. However, to the best of our knowledge, no such sequential probabilistic models are currently available for drug-related social media posts. Given a sequence of terms w1n=w1…wn the model learns n-gram sequence probabilities, such that approximations can be made for a sequence w1…wm *via* the expression P(w1m)=∏k=1mP(wk|wk−1). To generate the n-gram language models, we used the KenLM n-gram, language modeling tool. We have made available a set of n-gram language models (*n*=2–4) from this data.

Accessibility of language models and the data

Here the utilities of the models which we talked about will be described. The results which have good quality described here with low settings and with no optimization.

The probability of using divided semantic language were checked in order to investigate connection among the drugs and harmful reactions. Last study has showed that methods based on co-occurrence based to define drug-adverse reaction combinations. Those methods basically trusted on sample initiation and recognition or handmade regulations for finding drug-adverse reaction combinations and are restricted in many ways, especially for unmanageable social media data, like skills to manage wide amount of text and exploring combinations which do not track regulations. One of the belongings of the distributional semantics model is that, the skills to catch semantic combinations among the

conditions based on co-occurrence inside the huge-sized corpus. The words and phrases which learned by models are described like vectors in spaces that are high in dimensions. Vectors which are alike in terms of context seem close in the semantic field. For this reason, new measurement was developed: a single cosine likeness measurement to compare the similarities of drug keywords with group of conditions showing adverse reactions.

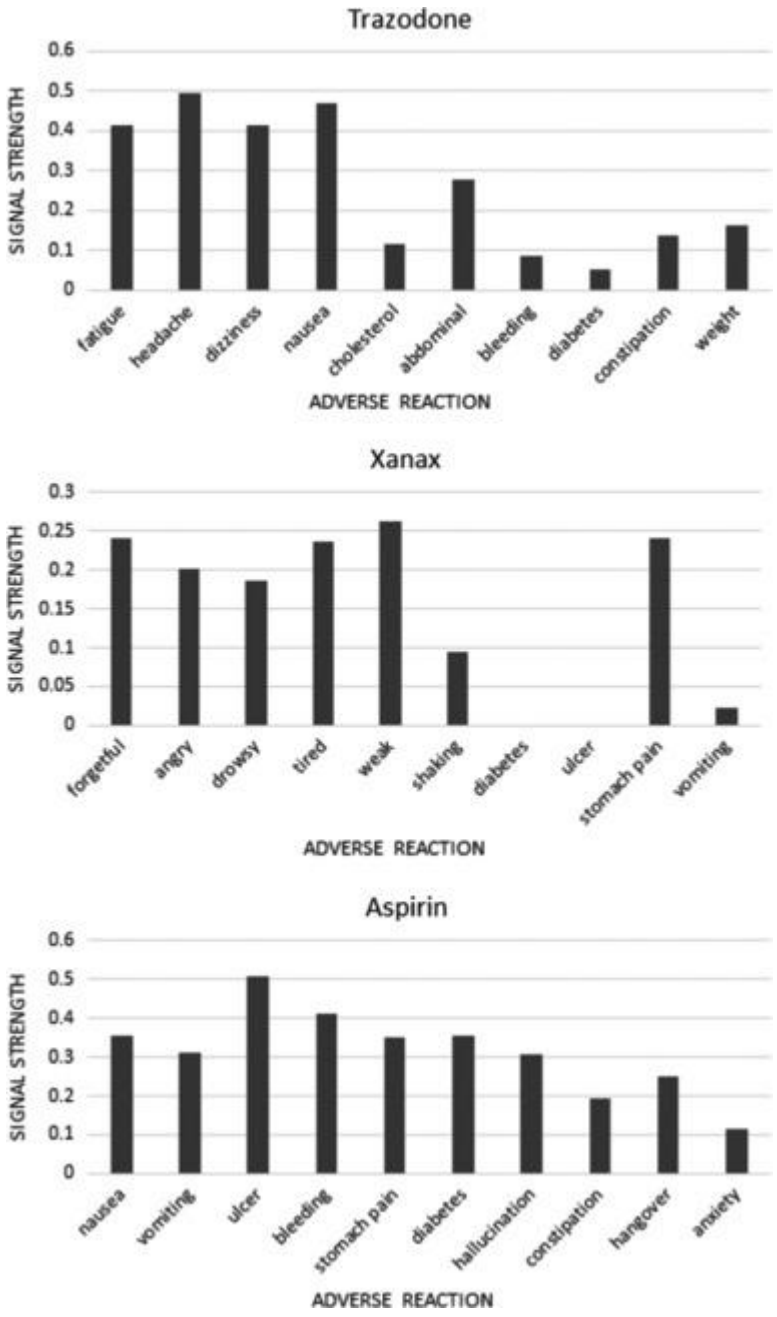
In order to fulfil this experiment 3 drugs were selected: trazodone, aspirin, Xanax. These drugs are included in the most used drug list. You can see in the picture above (Figure 9). For this distributed representation model was used in these terms: vector size=400, context window size=9. After that, cosine likeness marks between the drug keyword and the group of adverse drug reaction conditions were contrasted. To be clear, for the multi-word adverse reactions, middle likeness for the whole conditions were calculated. Specifically, to make comparison recognized adverse reaction showing keywords for all drugs which mentioned with the adverse reactions which are unknown to be combined to the drugs. The main concentration was the adverse reactions which are not critical.

Table 6 represents the cosine likeness deploy for those three drugs and 10 negative response expressions for the each one. For Trazodone, the first four negative response terms (fatigue, headache, dizziness, and nausea) are considered as negative responses, on the other hand the rest six reactions have been selected randomly which are not are randomly chosen adverse drug reactions not familiar with combinations of the drug. Besides, having no first-class adjustment, a basic cosine likeness computes between drug word vector and negative responses obviously show probable combinations.

For Aspirin, 5 negative reactions in the left side of table, show to the left in the figure represent common negative responses and last five reactions shows negative reactions that combinations are not given. Despite of the combinations are not

powerful like Trazodone last 5 combinations have obvious tendency of there is a clear trend of bottom marks values.

Figure 10: Negative responses for medicines



For Xanax, a group of negative responses and less familiar negative responses were selected from website of Drugs.com (https://www.drugs.com/). Figures 7 represents familiar results related to the drug, which have very powerful remarks for the words of left side.

Results which acquired of distributed semantic model show encouraging performance, in particular, it is better to take into consideration that, the results are carried out with no preprocessing and postprocessing of data, also signal adjustment, and any other conditions. For the future work, there are some probabilities as they are not studied here: in what extent the size of vectors and windows influences signal correctness.

Serial language examples to sort text.

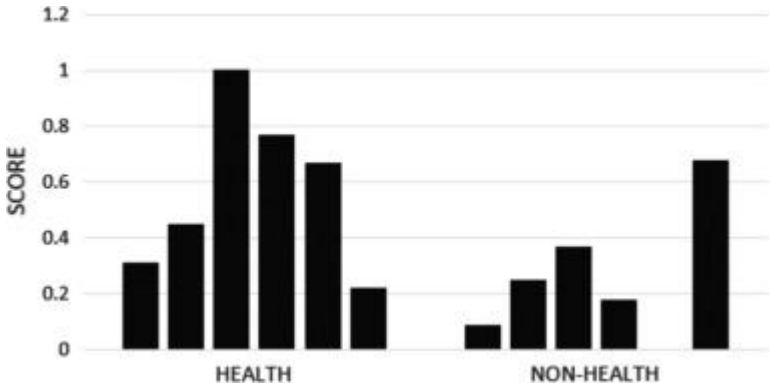
Implementation area of n-gram language patterns can be enough big, as well as range of task. For reasonable time, n-gram models are accessible, hence, these models are used to make lexical normalization. For example, translating, categorizing texts. Text classification related to health information have been tested in order to employ probability n-gram models. The date which gathered from Twitter are attached to receipts of drugs. Then we can say that, there are subcategories. These tweets are subcategory of whole health related tweets and likelihood, this can assist to define wider group of heath related tweets.

Firstly, manually commented data which is related to health was carried from research. (M.J. Paul, M. Dredze A Model for Mining Public Health Topics from Twitter (Technical Report) Johns Hopkins University,2011). The data set consists of 5128 tweets and approximately 35 % out of them are health tweets. The main aim was to analyze the effectiveness of the consecutive models for this experiment rather than to analyze effectiveness in text classification. Consecutive of parts in the series is presented, therefore, according to the equation which described above,

the possibility of the sequence will be evaluated. The character of the data predicts somethings related to the problem: tweets which related to health, will have strong possibilities. For proving this, a little number of tweets with annotation were selected and results generated to use n-gram model. To evaluate the score of the patterns very basic tool was used. In the first step, for every pattern tweet, complete value figures were created with the help of tetra gram model. Second step was to measure values to match this sequence {0, 1}

Fig 11 shows a pattern of possibilities carried for a tweet group, which gathered from health-related dataset. However, the influence of this model was evaluated for the little quantity of conditions, the diagram gives this option: according to tetra-gram model health-related tweets are likely to get larger score. This kind of characteristic can effortlessly be combined with the issues related to text classification. For the future use, this presentation would develop.

Fig 11: Possibilities of health and non-health related tweets



Chapter 3

Opportunities and challenges of Twitter for Healthcare.

3.1 The advantages of Twitter related to the Medicine

However, the imperfections are obvious, most analysists tend to praise social media in terms of encouraging method pay higher attention to patients as well as exclude impediments in order to understand cure. Mass media tools like TV let the patients be in contact with doctors and adjust via television shows, still the audience interest is no active, also the information resources are under need. The practicality of social media networks and especially Twitter is completely varied from TV. Because Twitter users are able to help to conversations as their number are getting higher, People who have Twitter account can easily share messages consisted of maximum 280 letters. To create their own account is enough to post their experiences for professionals, in comparison with information which has restricted resources and have to keep audience. Many accounts can provide medical information, while there can only be a certain number of medical shows on television.

A last work with more than two thousand doctors which we have described above, confirmed this. The physicians basically share tweets at least once in a day with 300 followers per every of them.

Twitter's reputation in healthcare has boomed during last times, the main cause of it is that, this social media is professional platform for healthcare professionals to achieve a wider viewership including doctors, interns, medicine students and clients. People with Twitter account might get the needed knowledge. Furthermore, the feedback from purchasers are significantly increased, now they are more active since they straight response to the tweets and by reacting or sharing

they support their doctors. To mark the most helpful and correct tweet clients reply, retweet, and favorite. Feedback and response from the users give opinion to professionals so that they can see which kind of information they are looking for and they consider as reliable. Furthermore, the hashtag function demonstrate that what clients are interested in what topics, and this promotes groups and set of tweets to focus on the same ailment and discussion topic. Some researchers confess that, they utilize Twitter in order to let the people about their researches, as Twitter has high prospective for it.

To let the research people, doctors, professionals to come together and divide up about news on medicine, cures, medical issues which demand new solutions, as well as curious works. In order to decrease problems in healthcare cooperation is highly precious since barriers to collaboration in medicine is extremely valuable, as regular structural and may extend between controls in the fields like computer science bioengineering. From the research point of view Twitter works like evaluation field for edition article. Also, Twitter works like evaluation field for edition article because tweet usually keeps crucial thoughts and comments related to the news. According to the clients, to post tweet related to the articles seems more easier rather than leaving opinion on the official site. Twitter attaches biomedical fellows, apart from this, Twitter works for making the researches progress further available to the doctors. Associating research fellows and physicians is highly significant and helpful, since physicians explore recent knowledge by being in attach to the research fellows and after that they take profit of this information in order to make decision about client cures. Furthermore, the knowledge change process between professionals encourage the research in effective and individual areas. Hence, the basic modifications which Twitter supplies are the two-path transmission channel between resource and customer of information. In addition to this, the two-path transmission channel via Twitter

supports stakeholders who usually don't have one in the conventional healthcare interplays.

It is obvious advantage of Twitter that it attaches large number of people with help of tweets, also Twitter has capability to shake up public health struggle, as well as spreading health news, splitting information related to ailments, or cooperating assistance tries. Let's look at one example: in 2013 when Typhoon Haiyan happened in the Philippines, the help of tweets over Twitter was enough big. This assisted to workers to define damaged locations which need urgent help.

A strong instance of Twitter usage is that most of the advantages mentioned before may likely take place below of one post on the Twitter. For instance, in the cardiology assembly of one hospital, a person can reply to the tweet related to higher blood pressure issue which that hospital posted. Then discussion begins on the last research works which shows genetic correlation among higher blood pressure and special genes. Other user can ask other question: how this kind of research helps to increase clinical results as well as other tests by encouraging to share the research? Then doctors who is professional in this field, can reply to client within his experience. As well as other doctor or researcher can put the link which is related to medical assembly of blood pressure overcoming methods. Furthermore, a client can ask question that how to fight with hypertension issue in home environment. There are different range of professionals and circumstances which is inimitable, so the process of changing knowledge is highly profitable.

When all is said and done, Twitter as a social media is rich with the capability to make the sharing of information on Twitter has the potential to create a chatty and cooperative aura for the related people such as, clients, doctors, and fellow researchers. In such a way, doctors implementing every day are joined to the last researches, fellow researcher might define clinical issues to be solved, and customers might perceive their situations much better. The standard of service

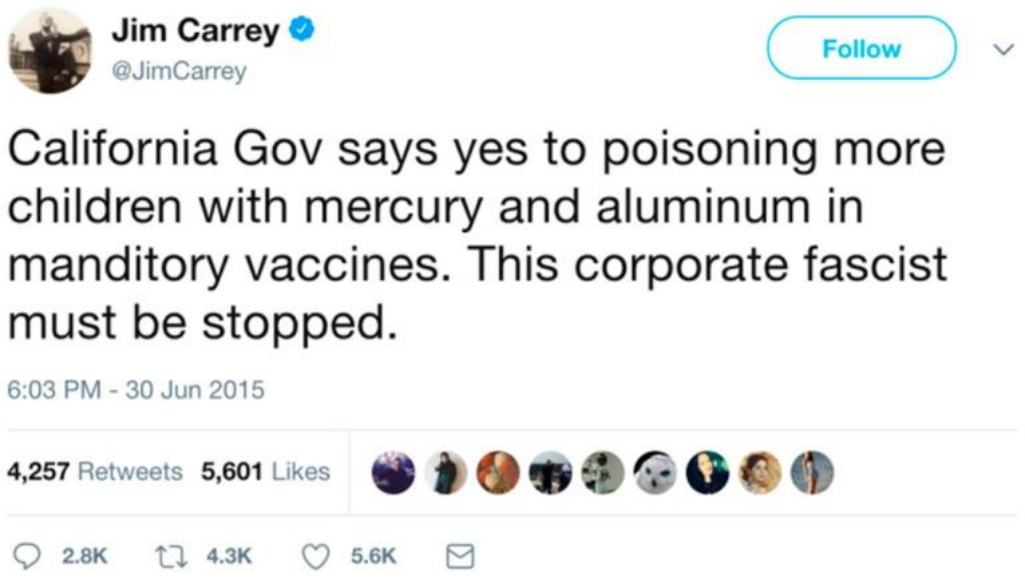
which sent to clients via higher openness and correspondence with social media might be increased with the help of access to information.

3.2 The adverse effects of Twitter in Healthcare.

Meantime it is doubles that, Twitter's profitable side to client service in terms of information is available, there are negative sides as well as. Researches show that approximately 20 % of the total information on tweets is incorrect. Higher portion of incorrect information in the tweets creates problem for the fruitful and useful tweets related to medicine, and they found that about 20% contained inaccurate information. Another negative side is, there is not available any tool or system to examine the wrong information. But it trusts on willing donations by medical researchers and physicians. They attempt to rectify misinformation. It means, people who are not professional share their opinion without any problem since they don't have any concern about the result of it. These kinds of scripts create real issue for clients, undergraduates, because unlike professionals they face with who may face with obstacles in order to differentiate resources.

As an example of hazardous wrong information brought about because of non-professionals sharing thoughts with no scientific confirmation is the famous people's action against of vaccination. Some famous people's tweets most remarkably Jenny McCarthy and Jim Carrey, are related to side effects of vaccines. Their tweet says that, vaccination can lead to autism. It is misinformation. Because it doesn't have any scientific confirmation. We can see that tweet in the Figure 12.

Fig 12: Sample tweet



Since he is famous people, Jim Carrey's account is verified. But he is not professional in this field. From this point of view Twitter might be an "idea boomer" symbolizing shared thoughts more than stable realities, because to share tweet and quote is very easy. The action against of vaccines unfortunately became popular, partly due to tweets which had wrong information and the replication impact. To sum up, many places are facing problems with revival of ailments which before extinguished like rubeola.

Famous peoples' disapproval of vaccinations are awful instances of non-professionals sharing opinion on health-related topics. However common problem to differentiate delegates of commercial concerns from neutral doctors is very tough and complicated. On the other hand, this issue got worsened with unknown and unverified accounts, their intentions and professionalism are doubtful. So, a possible danger is the toughness in defining physicians on the Twitter exactly and relying verified accounts. As well as, because of the little quantity of letters

wanted, tweets are usually short and have to skip main information, which might often lead to misunderstandings. Especially for complicated studies, it is not easy to talk about moments, warnings, and restrictions of the tweets. For researchers, the mentioned issue might create misapprehension of recent developments for patients. For physicians, same issues might be highly hazardous, since opinions of confirmed doctor can be accepted as health advice.

Apart from this, other problem related to share information on Twitter for healthcare is Twitter may transfer as a bulky information. Since information is heavy in quantity, to gather and analyze all data is almost impossible. Almost for each hashtag experts and non-professionals can supply their opinion, with unlawful and wrong demands decreasing the effect of the right information. Furthermore, when it comes to challenges of the correctness of information, responds and feedbacks can fail to notice, thus the last results and solutions can become invisible. For these conditions like this, less information can be helpful.

As well as, deflections of experts can be considered as crucial important problems related to Twitter, like other social media platforms. Former works showed elevated ratios of unqualified tweets by obviously verified accounts like physicians. Unqualified aptitudes added blasphemy, objections related to clients, annulment of client security, and disagreement of interest. These deflections are caused because of characteristics of Twitter as well as users who have deficit of experience.

The last possible damage concerned to usage of Twitter in medicine is that, doctors spend some part of the day on Twitter. It is obvious that, to share knowledge and opinions is good and beneficial. On the other hand, doctors should be careful in terms of time. Furthermore, more than 80% of tweets which tweeted by confirmed accounts, have been posted during work time, between 9 a.m. to 5 p.m.; it means

while spending time on Twitter doctors are losing their clients if they don't balance it carefully. Doctors' durability was showed clearly, resistance has been demonstrated explicitly, owing to comments related to shortage of collaboration on the social media because of shortage of time and concern of interruption by their main academic job. If small rate of physicians chose to be part of Twitter discussion, for obvious recommendation and correct information on the Twitter would be uncommon.

If we look at problems again related to share information, the possible issues are as follows:

- ✓ Peak level of misinformation;
- ✓ Complexities related to confirming reliability of resources;
- ✓ Enormously large data availability
- ✓ Worries about being professional or not;
- ✓ Spending time of doctor.

3.3 Recommended solutions.

Medical experts should frequently talk to clients, colleagues, and research fellows. Since social media is turning into the main component of communication, medical experts of our time should develop more sophisticated capabilities related to social media. Worries with inappropriate information, security, and non-professional activities require obvious clear principles and rules for doctors.

In spite that there is demand for rules and principles, the present rules by the American Medical Association (AMA) are uncertain, they say to doctors that rely on your own judgement. The large portion of non-professionals and wrong

information, together with the shortage of stable detectors for medicine experts in the “bios” of their Twitter accounts, demonstrate a shortage of official social media knowledge between doctors. The shortage of information is not astonishing, since conventional instruction for medicine experts exclude efficiently using maximum advantage of social media for medical and research aims.

Therefore, healthcare experts should be officially educated for suitable and efficient social media usage. This education is an important since social media is turned into integral part of the everyday lifetime. Analogue attempts notified people related to college and experts, sportsmen, because presently, they got to main factors on the social media sites. Official commands connected to medical institutions and all the way through practice can be prosperous, since overall medical experts pass the same course of study.

Furthermore, it is better if AMA propose new rules in order to use social media by medical experts. These rules must underline that, medical experts should keep distinct individual and expert accounts, keep away from blasphemous, racists and sexist opinions, comments, they should esteem the security and anonymity of clients, keep away from encouraging special outcomes or drug, and add obvious comments of connections and disagreements in Twitter accounts. Medical experts must test to fight strongly with bum steer since they spread the information. In addition to this, the rules must inform experts to revise the tweets in advance sharing, since revising tweets in advance sharing is very important to maintain client security and make sure master attitude. Furthermore, to utilize Twitter “bio” efficiently and expertly is very important to let the clients to discover reliable resources of information.

Since the offered remedy to train medical experts on social media manners as long as the education period will have possibility to decrease dangers connected to

doctors on social media, training experts might be a recurrent and constant operation. Applying any training program, as it is good in comparison with the present whole shortage of education, is an act forward to the desired solution, it is good to consider that it is beginning. Furthermore, the differences among the kinds of medical experts, as well as physician, medicine students, nurses, researchers and other people, is not discussed here, since they need to be worked deeply to get reasonable result.

As a result, Twitter shares the joint aims:

- To modify the comprehension of healthcare.
- Not to see it in the black box form, but to see something more attainable.
- To let the clients to percept the situation and give advanced choices.

Because Twitter in the healthcare is able to give consent to clients to get information related to their health and care. Along with this, the stability between availability and wrong information should be maintained.

Chapter 4

The situation of Twitter and healthcare in Azerbaijan.

4.1 Some statistics of diseases for Azerbaijan

The data in the table 5 below was taken from the State Statistics Committee of Azerbaijan. This data gives general information about diseases in Azerbaijan between 2013-2017. According to table, the state of diseases during these five years has changed in a positive or negative way.

Diseases of skin and sub skin tissues were decreased from 51713 to 43029. As well as, disease of injuries, poisoning and some other results of impact of external reasons were decreased from 121199 to 118472.

Unfortunately, other in other diseases reduction wasn't noticed.

The big growth has been detected in the blood diseases and other hematogenic disturbances. The rise is almost 50% which is alarming. According to medicine, the main reason of blood diseases are genes. Many people are unaware of their genes, like which diseases their parents have, or information related to blood.

Table 4:

	2013	**2014**	**2015**	**2016**	**2017**
Person					
All diseases	**1739584**	**1852918**	**1824086**	**1867071**	**1875652**
of which:					
certain infectious and parasitic diseases	112429	123445	114955	116827	124457
neoplasms	10894	11011	11118	11761	12082
blood diseases and other hematogenic disturbances	48238	59473	64078	67831	67779
endocrine, nutritional and metabolic diseases	48549	51099	53371	53246	47476

mental and behavioural disorders	10804	12749	7764	10058	10281
diseases of the nervous system [3)]	63307	74355	72961	73045	72306
diseases of the eyes and adnexa	51153	66112	62274	67651	69261
diseases of the ears and mastoid	39428	45008	42305	46034	46188
diseases of the circulatory system	129970	140433	134225	142277	143182
diseases of the respiratory system	731015	741198	752669	749438	735922
diseases of the digestive system	138522	152362	156197	163009	160801
diseases of skin and subskin tissues	51713	52305	43683	43936	43020
diseases of the osteomuscular systems and connecting tissues	21360	27262	28282	29393	29486
diseases of the urogenital system	89327	95643	98413	104694	107333
pregnancy, childbirth and the puerperium	51954	55232	51501	57581	64670
some cases arising during perinatal period	7106	7897	7495	7722	8387
congential anomalies (defect of growth), deformations and metabolism of chromosome	2796	3786	3334	3439	3665
other symptoms and out-of norm cases not classified under other heading and detected during clinical research and laboratory examination	9820	9752	9214	9958	10884
injuries, poisoning and some other results of impact of external reasons	121199	123796	110247	109171	118472
Number of diseases per 10 000 population					
All diseases	**1871.2**	**1968.0**	**1914.2**	**1937.2**	**1926.9**
of which:					
certain infectious and parasitic diseases	120.9	131.1	120.6	121.2	127.9
neoplasms	11.7	11.7	11.7	12.2	12.4
blood diseases and other hematogenic disturbances	51.9	63.2	67.2	70.4	69.6
endocrine, nutritional and metabolic diseases	52.2	54.3	56.0	55.2	48.8
mental and behavioural disorders	11.6	13.5	8.1	10.4	10.6
diseases of the nervous system [3)]	68.1	79.0	76.6	75.8	74.3
diseases of the eyes and adnexa	55.0	70.2	65.3	70.2	71.2
diseases of the ears and mastoid	42.4	47.8	44.4	47.8	47.5
diseases of the circulatory system	139.8	149.2	140.9	147.6	147.1
diseases of the respiratory system	786.3	787.2	789.8	777.6	756.0
diseases of the digestive system	149.0	161.8	163.9	169.1	165.2

diseases of skin and subskin tissues	55.6	55.6	45.8	45.6	44.2
diseases of the osteomuscular systems and connecting tissues	23.0	29.0	29.7	30.5	30.3
diseases of the urogenital system	96.1	101.6	103.3	108.6	110.3
pregnancy, childbirth and the puerperium [4)]	193.0	205.7	192.6	216.3	243.8
some cases arising during perinatal period [5)]	411.5	463.2	450.9	484.2	582.3
congential anomalies (defect of growth), deformations and metabolism of chromosome	3.0	4.0	3.5	3.6	3.8
other symptoms and out-of norm cases not classified under other heading and detected during clinical research and laboratory examination	10.6	10.4	9.7	10.3	11.2
injuries, poisoning and some other results of impact of external reasons	130.4	131.5	115.7	113.3	121.7

4.2 Survey on Twitter use in Azerbaijan and its results

As we know people are not eager to use Twitter in our country. In order to observe the Twitter usage among population in Azerbaijan, we conducted the survey. The survey covered the questions listed below:

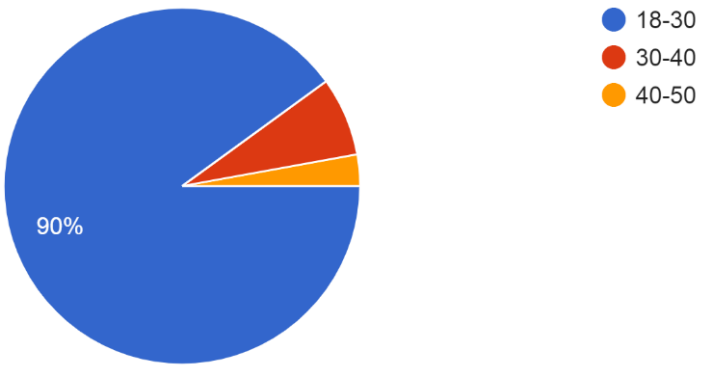
- What is your age (range was given: 1) 18-30, 2) 30-40, 3) 40-50.
- Your sex (male or female)
- Your job, or qualification (in case you are a student) 1) Engineer 2) Teacher 3) Doctor 4) Others
- Do you use Twitter? (yes or no)
- Do you trust on reliability of Twitter information? (yes or no)
- In your opinion, which areas are Twitter applied on in the world? (Give your opinion)

- What is the reason of Twitter's low usage in Azerbaijan? (Give your opinion)
- What do you think about applying Twitter in medicine for Azerbaijani case? (Give your opinion)

Results were like this:

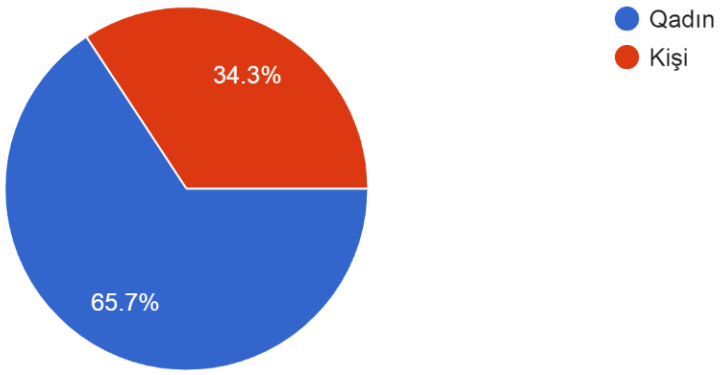
For the first question, percentages were 90%, 7% and 3%. 90 % of the respondents were included in the age group of 18-30. It means that, young people use Twitter rather than other age groups. Surprisingly, only 3% of 40-50 years old people use Twitter which in comparison with world statistics is pretty low. 7% of correspondents' age were between 30-40.

Fig 13:



For the second question, 65,7% of respondents were women, rest of them were men. It is matching with world statistics, as they say women tend to use social media rather than men. According to statistics, 62 % of women about 40 million more than men use Twitter in each month in the world.

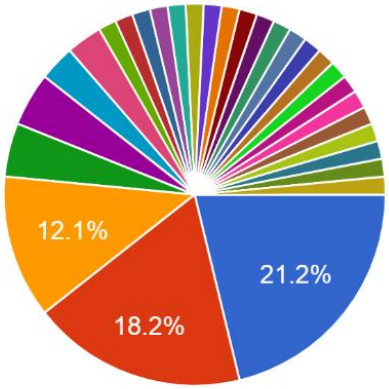
Fig 14:



For the third question, results were very different and interesting. The job of correspondents was like this:

- ✓ Engineer -21,2%
- ✓ Teacher -18,2%
- ✓ Doctor -12,1%
- ✓ Marketing- 10%
- ✓ Accountant -4,5%
- ✓ Economist- 3%
- ✓ Others -31% totally. Their jobs were different like bank worker, designer, office assistant, manager, housekeeper and so on.

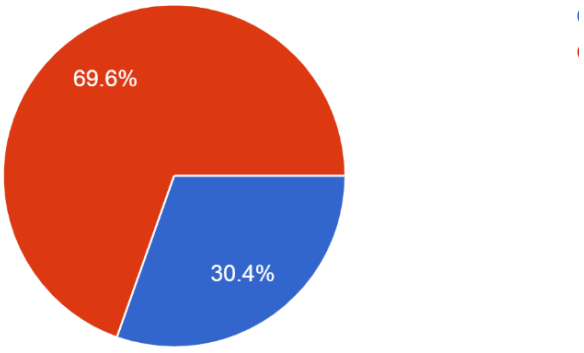
Fig 15:



These results are matching with today's situation. People who use Twitter more is engineers as they need developer account in order to use Twitter's API, or their researches are depend on somewhat Twitter in terms of data mining. In the world statistics, among the Twitter users, engineers and economists are on the top places. But in Azerbaijan only 3 %of economists have Twitter account.

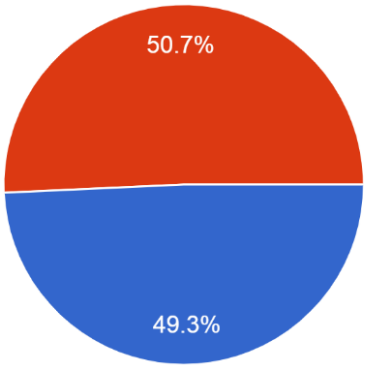
For the fourth question, people said 69,6% of them use Twitter. It can't be considered as pretty good results, but we can't say it is very low however almost all people have Twitter account in the developed countries.

Fig 16:



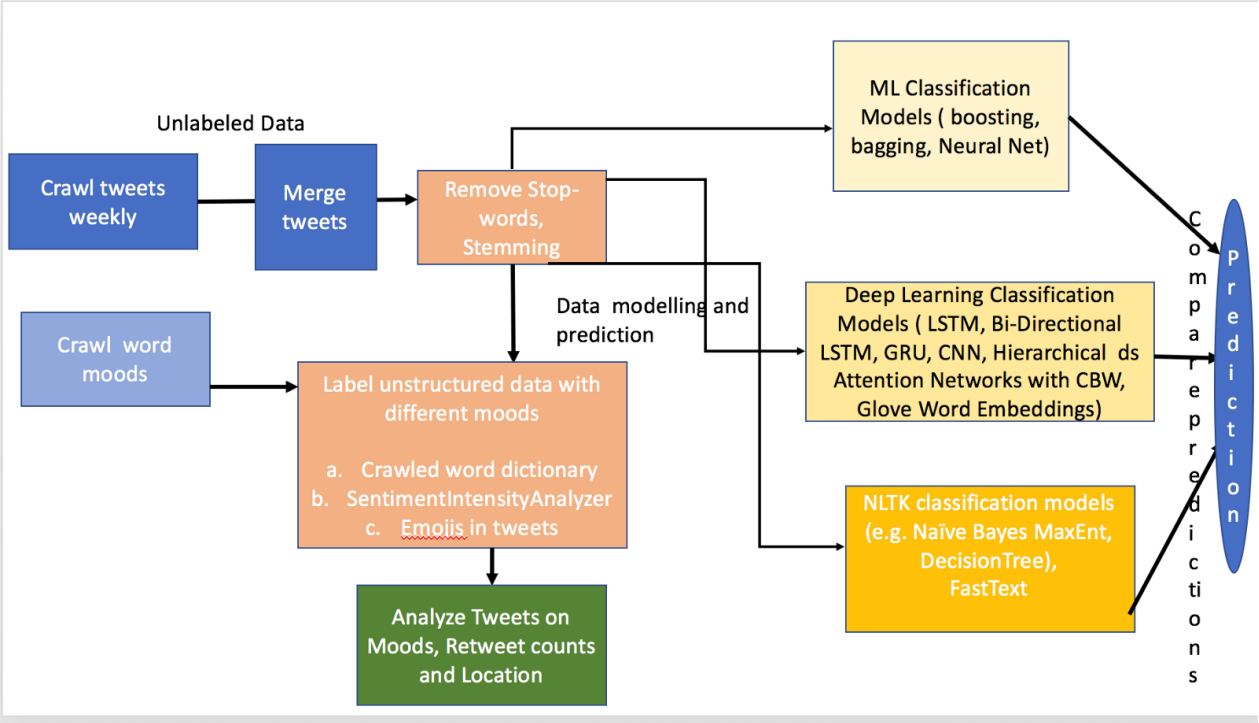
For the fifth question, 50,7 % people said they trust on Twitter information. Rest of people have doubt whether Twitter data is reliable or not. As we discussed in the 3rd chapter, only 20% of tweets contain misinformation. Why Azerbaijani people do not trust on Twitter data? Reason is, they think social media platforms can't be source of information as some portion of users are ordinary people. To some extent, it is true. But there are enough reliable sources of information which many users take benefit of it.

Fig 17:



For the sixth question, what do you think about Twitter's implementations in the world? The diverse range of answers was conducted. About 23 % said they don't have information about it. Probably. They are respondents who don't use Twitter. 5% of them think that Twitter is applying on all areas in the world. About 25% and the biggest part said, the main implementation of Twitter is politics. Why people think like this? Reason is, it is common view that social media have been made to keep people under control in case they share posts against government. People see social media as well as Twitter as a management tool. As an example, we can say USA elections. In 2019, they studied sentiment analysis of the coming congress elections of Lok Shobha. They used deep data analysis apart from positive and negative sentiments. The table below shows their common strategy:

Fig 18: Sentiment analysis structure for Twitter data



Another example can be Trump's elections. Data analysts say that, Trump's twitter post was saying many things about upcoming elections. As he is very active user, and his tweets retweeted many times. It allowed data mining experts to track his performance. They said the results of elections were same with out predictions based on Twitter.

Therefore, it claims that, if Twitter is used in right way, benefits of it are inevitable. For the seventh question, that why Twitter is not used like other social media sites? 12% of them said, they don't know. 10 % people think that Twitter is professional platform, that's why people don't use it. For them, it is boring social media type

and it is not eye catching like Facebook and Instagram. People think, our population tend to use social media websites which are enjoying, interesting. Another interesting reply was that, they said the main part of Twitter users are youth, and they are not interested in professional platforms however they are student- potential researcher.

For the eighth question, about implementation of Twitter to Azerbaijani medicine 28% of people don't have an opinion about it. 13% of them think that it is not good idea, because social medias are created in order to share information about everyday life, not social life and serious topics. Roughly 30% of people claim that, Twitter's implementation on Azerbaijani medicine can have beneficial results as nothing is done so far related to this topic. They say, people need to be informed about implementation. Advertisement of medicine in Azerbaijan is very low. So, there is no website or platform which shows general knowledge, trends, and patient information.

4.3 Conclusion

Trends show that, Twitter mining will hit all areas as soon as possible. Few years ago, social media was just a dream, now it is part of our life. Like this, Twitter mining is becoming integral portion of lifetime. For this reason, to take advantages of it to the fullest should not be postponed.

As a result of my research, I came to conclusion that, we should start mining Twitter as well as other social media platforms in order to produce useful information related to healthcare. It is true that, in the beginning it will be tough as Azerbaijan environment is not suitable for this.

On the other hand, the survey shows there are enough people who think to mine Twitter must be beneficial. It is good sign that, these people are doctors and engineers. From this point of view, if Azerbaijani doctors and computer engineers come together, they can achieve reasonable results in terms of healthcare.

I would like to recommend starting from training and educating medical experts of Ministry of Health of Azerbaijan. This is first part of work in terms of medical professionals. Second part of our work is engineers who will work together with medical professionals. For this part I recommend Innovations Agency of Azerbaijan which is under State IT Development Fund and High-Tech Park LLC. The first disease which we should take into consideration is blood diseases such as anemia. As we have discussed before, these diseases are widespread in Azerbaijan and during last years we observe growth in it. Thus, to educate people in this topic, get feedback from them about current health issues will be beneficial. As well as, doctors, young researchers, medical students can share their researches with people and learn their opinion about it.

References:

1. Hawkins J, Brownstein JS, Tuli G, Runels T, Broecker K, Nsoesie EO, McIver DJ, Rozenblum R, Wright A, Bourgeois FT, Greaves F. Measuring patient-perceived quality of care in US hospitals using Twitter. BMJ Qual Saf. 2016;
2. Mastering Social Media Mining with Python, Marco Bonzanini, 2016;
3. Mohammad S. 9 – Sentiment analysis: detecting valence, emotions, and other affectual states from text. Emotion Measurement. 2016;
4. Tighe PJ, Goldsmith RC, Gravenstein M, Bernard HR, Fillingim RB. The painful tweet: text, sentiment, and community, 2016;
5. Afyouni S, Fetit AE, Arvanitis TN. #DigitalHealth:exploring users' perspectives through social media analysis. Stud Health Technol Inform. 2015;
6. J Med Internet Res. Structure analyses of tweets pertaining to pain. 2015;
7. Kent E, Prestin A, Gaysynsky A, Galica K, Rinker R, Graff K, Chou Ws. "Obesity is the new major cause of cancer": connections between obesity and cancer on facebook and twitter. J Canc Educ. 2015;
8. Laaksonen C, Jalonen H, Paavola J. Utilising Social Media for Intervening and Predicting Future Health in Societies, Proc. 5th International Conference on Well-Being in the Information Society 2014;
9. Greaves F, Ramirez-Cano D, Millett C, Darzi A, Donaldson L. Use of sentiment analysis for capturing patient experience from free-text comments posted online. J Med Internet Res. 2013;
10. Paul, M. J., & Dredze, M. A Model for Mining Public Health Topics from Twitter. Technical Report. Johns Hopkins University 2012;
11. Liu B, Zhang L. A survey of opinion mining and sentiment analysis. Mining Text Data. 2012;
12. M.J. Paul, M. Dredze A Model for Mining Public Health Topics from Twitter (Technical Report) Johns Hopkins University,2011;

13. S. Kumar, R. Zafarani, and H. Liu. Understanding user migration patterns across social media. Twenty-Fifth International Conference on Artificial Intelligence. Association for the Advancement of Artificial Intelligence, Palo Alto, CA, 2011;
14. R. Goolsby. Social media as crisis platform: The future of community maps/crisis maps. ACM Transactions on Intelligent Systems and Technology 2010;
15. Pang B, Lee L. Opinion mining and sentiment analysis. Foundations and trends in information retrieval. 2008;
16. Yadav R N. Isocitrate dehydrogenase activity and its regulation by estradiol in tissues of rats of various ages. Cell Biochem Funct. 1988;
17. Treves A J, Carnaud C, Trainin N, Feldman M, Cohen I R. Enhancing T lymphocytes from tumor-bearing mice suppress host resistance to a syngeneic tumor. Eur J Immunol. 1974;
18. http://sentistrength.wlv.ac.uk/
19. https://stackabuse.com/accessing-the-twitter-api-with-python/
20. https://www.academia.edu/2931894/Mining_Social_Media_A_Brief_Introduction
21. https://mashable.com/2012/12/18/social-media-mobile-healthcare/
22. https://www.hhs.gov/hipaa/index.html
23. https://pdfs.semanticscholar.org/2136/54653a9862bc857aff9f753572c689d40020.pdf
24. https://www.stat.gov.az/source/healthcare/?lang=en