Twitter Activity Analysis of Users Diagnosed with Bipolar Disorder

Introduction

Bipolar disorder (BD) is associated with significant functional, cognitive, and social impairment [1]. It is a severe and sometimes underestimated disease. Up to 15% of patients commit suicide, and about half of them make at least one suicide attempt in their lifetime [2]. Nevertheless, 70% of people with this disease still are initially misdiagnosed [3].

Purpose

The main goal of this project is to find significant correlations in BD users' behavior in the social network (Twitter) with the comparison to the control group (users without BD diagnosis) that may promote more detailed research of the disease in social networks and help to monitor bipolar disorder indications in the future.

Hypothesis

I assume that bipolar users will have more significant ups and downs in arousal and valence scores and tweeting frequency. We may also expect late night posting and higher emotional volatility.

Methods

The main issue of this project is to find and analyze bipolar disorder patterns via data analysis tools. I used Python language and Pandas for my purposes. Twitter timelines by users that posted "I was diagnosed with Bipolar disorder" were scraped via the twitter api.

The /bipolar/ folder contains all the scripts used to extract and analyze the data in this project: https://github.com/di-tal/bipolar

Reference:

1. *T. Martini et al., "Bipolar Disorder Affects Behavior and Social Skills on the Internet", PLoS ONE, vol. 8, no. 11, p. e79673, 2013. Available: 10.1371/journal.pone.0079673 [Accessed 9 May 2019].*
2. *F.Ghali,"Suicide in bipolar disorder", Bipolar Disorders, vol. 18, no. 1, p. 190, 2016.*
3. *Y.-H. Huang, L.-H. Wei, and Y.-S. Chen, "Detection of the prodromal phase of bipolar disorder from psychological and phonological aspects in social media," arXiv preprint arXiv:1712.09183, 2017.*

THEORETICAL PART

Bipolar Disorder

Bipolar disorder, previously known as manic depression, is a mental disorder that causes periods of depression and periods of abnormally elevated mood. The elevated mood is significant and is known as mania, or hypomania if less severe and symptoms of psychosis are absent.

Bipolar I disorder is a bipolar spectrum disorder characterized by the occurrence of at least one manic episode, with or without mixed or psychotic features. Most patients also, at other times, have one or more depressive episodes, and all experience a hypomanic stage before progressing to full mania. The difference with bipolar II disorder is that the latter requires that the individual must never have experienced a full manic episode - only less severe hypomanic episode(s).

One of the criteria for diagnosing a manic or hypomanic episode of bipolar disorder is what we call an expansive mood. Individuals with expansive mood may behave brashly or lavishly, assume a superior or grandiose attitude, or dress and act flamboyantly. In some cases, the person may become excessively friendly to the point where the behavior seems exaggerated and extreme. Speech can often become inappropriate, such as making a crude joke at a church service or in a business meeting.

The bipolar individual may also exhibit a decreased need for sleep, spending three hours or less per night in bed. Conversations can often be frenetic and scattered. The individual may engage in more goal-oriented activities (the need to accomplish something big now) while easily being sidetracked or distracted.

Symptoms

Mood Changes

Mood changes are characterized by a sudden burst of activity, often described as being outsized or larger than life. These changes would be long-lasting rather than transient and be uncharacteristic of the natural mood state.

Symptoms may include:

- Extreme excitability
- An expansive mood
- Grandiosity and imperiousness
- Sudden shifts to extreme irritability, hostility, or even anger

Changes in Energy

It is one thing to have a sudden rush of energy; it is an another when the energy is relentless, prolonged, and overwhelming. As with mood changes, the sudden upshot of energy would not be considered normal and can switch off as quickly as it was switched on.

Symptoms may include:

- A decreased need for sleep with little apparent fatigue
- A sudden increase in goal-oriented activities (such as a project that needs to be done to the exclusion of other activities)
- Restlessness and an inability to remain still
- Persistent and often purposeless movement

Speech Disruptions

Speech disruptions are probably the easiest way to recognize a manic episode. A person may be described as having a "motor mouth" and be difficult or even impossible to interrupt.

Speech disruptions may include:

- Rapid, pressured speech (as if you cannot get enough words in)

- Incoherent speech (often described as rambling and persistent)
- Clang associations (a serious condition in which words that sound similar are grouped together even if they don't make any sense)

Impaired Judgment

Impaired judgment can often be missed by casual observers who may dismiss the behavior as either a momentary lapse or a sudden burst of generosity, passion, daring, or goodwill. At times, the behavior may be risky, hurtful, or even dangerous.

Example include:

- Inappropriate humor and brash behavior
- Extreme impulsiveness (including gambling and risk-taking)
- An apparent lack of insight into the consequences of an action
- Reckless and extravagant spending (including the lavishing of gifts on friends, casual acquaintances, and even strangers)
- Hypersexuality and sexually provocative behaviors

Changes in Thought Patterns

Changes in thought patterns are easily spotted by those with whom the person has regular interactions. They may manifest as a sudden burst of creative insight or appear fractured and nonsensical.

Example include:

- Enhanced creativity or inventiveness (often perceived as a "breakthrough" or an epiphany)
- Flight of ideas (a rapid succession of thoughts that shoot from one idea to the next)
- Racing thoughts (a rapid stream of thought, often repetitive)
- An increased focus on religion or religious activity
- Disorientation or disjointed thinking

ANALYZED FACTORS

<u>Behavior Analysis</u>

- Tweet Timing: do BD users post more frequently at night and how consistent their activity is?
- General users' activity: how many tweets do BD users post per period compared to control group?

<u>Sentiment Analysis</u>

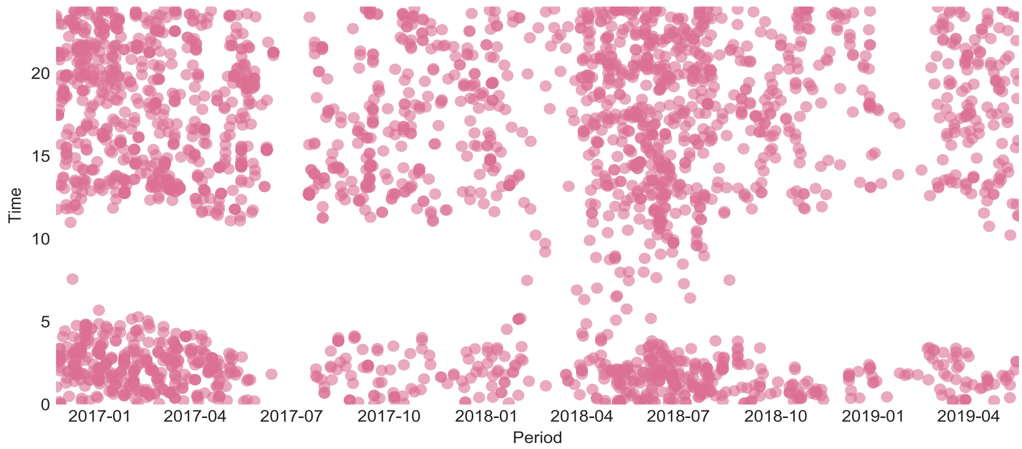
- Arousal score: do BD users have higher characteristics?
- Mood characteristics (positive/negative): would BD users post more positive or negative texts comparing with control group?

PRACTICAL PART

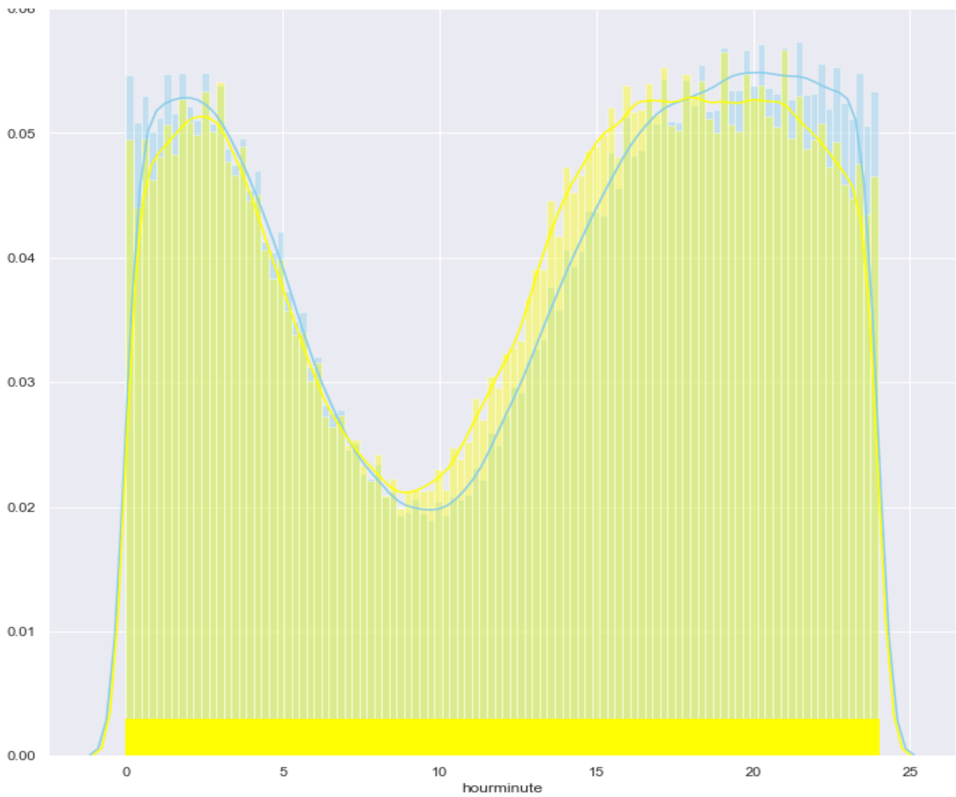
Tweet Timing

Time_of_post_ALL_DATA.ipynb

Random BD user data visualization provides some insights about the time of tweet posting and its inconsistency over months. In the plot above we see random BD user. X-scale signifies time period, Y-scale signifies 24-hours day time:



The next step is entropy and volatility calculation followed by a histogram with BD and control group comparison. Comparing bipolar and control group posting shows not a big but <u>significant difference in time of posting during the day: BD users more often post at night</u> (blue is bipolar group and yellow is control):



Kolmogorov-Smirnov test proves the results:

statistic=0.02099769021537823, p-value=0.0

Entropy Computation for bipolar and control groups

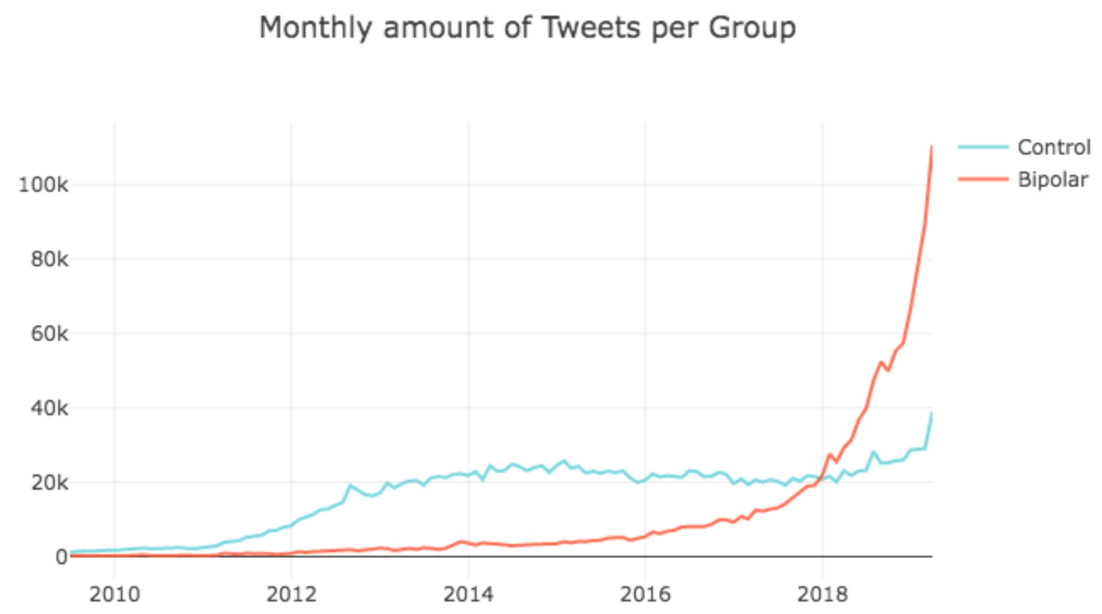
Entropy_volatility_comparison.ipynb

Shannon entropy is the average rate at which information is produced by a stochastic source of data. Shannon entropy is used as a measure for information content of probability distributions.

Filtering:

Before entropy computation, I filtered out all tweets from both datasets that were posted before 01.01.2013 to make our datasets more uniform (the number of posts from control group varies less from year to year while users from bipolar disorder group mostly concentrated in 2017-2019). Furthermore, I excluded all users in both groups who had less than 1000 tweets in our dataset.

For illustration I attach a plot by CSH intern Lukas Malik who worked with the same datasets:



To measure time entropy in both groups I compared the diversity of time posting (column 'hour') per every user.

The result proves our hypothesis that people struggling with bipolar disorder have problems with sleep. Higher entropy in BD group signifies that they post more often at night (literally, their posts have more unpredictable distribution) than users from the control group:

Bipolar Group Entropy: 2.9110992417776766
Control Group Entropy: 2.857216728908001

Nonparametric test of difference between bipolar and control users regarding their entropy:
statistic=849662.0, p-value=0.00040111024592024607

T-test:
statistic=6.482584216746171, p-value=1.0671231147953156e-10

Volatility Computation for bipolar and control groups

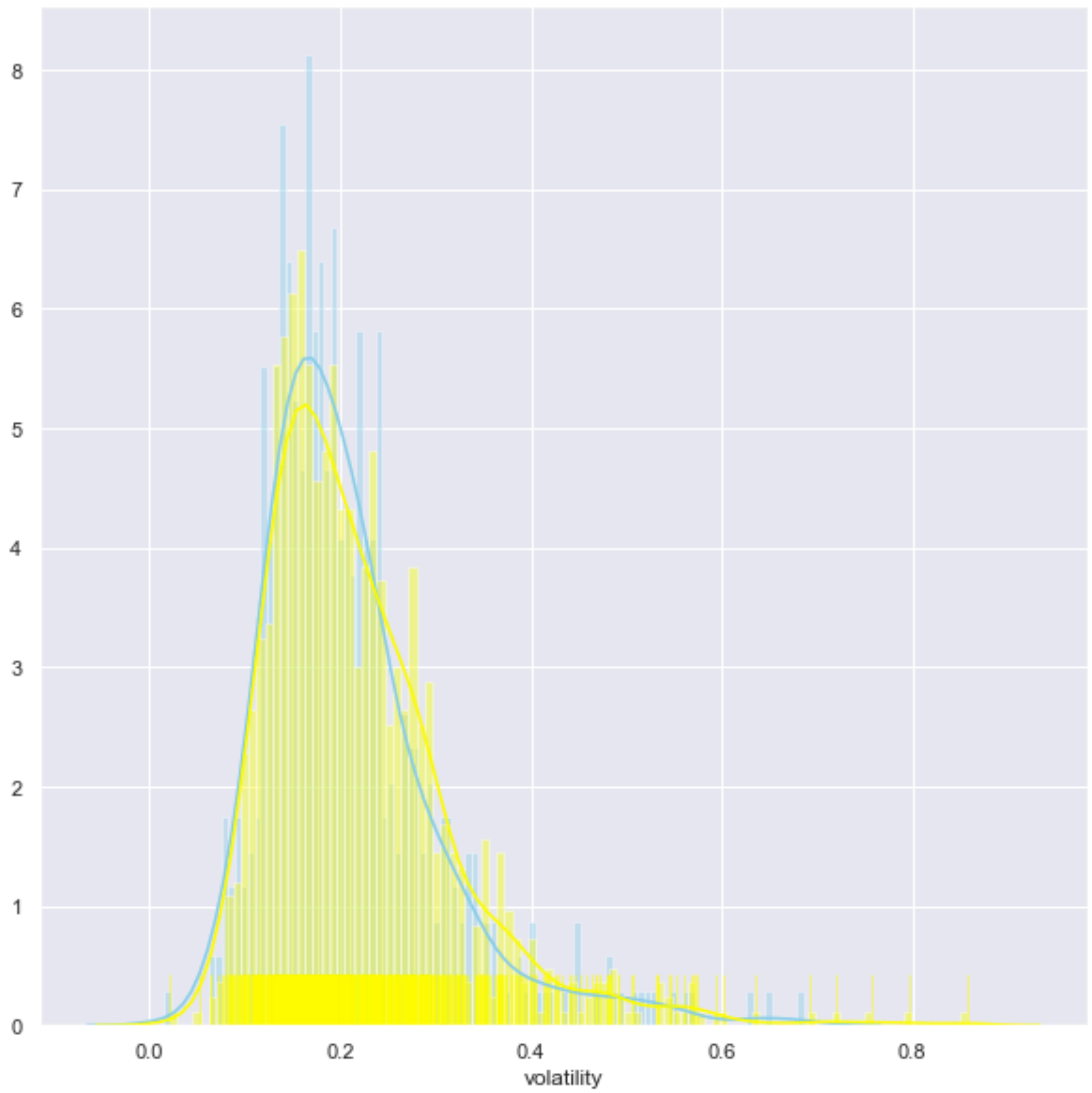
Volatility is a measurement of the consistency of the data. The lower the volatility, the less likely it is to expect dramatic changes.

> <u>Filtering:</u>
>
> For computing volatility of both groups, I applied additional filters to datasets. I excluded users who tweet rarer than 40 posts per month and who have less than 12 months of posting in their Twitter history.

In the end, I have got almost 2 times more control group users than bipolar users (993 vs. 513).

<u>Volatility characteristic doesn't align with our hypothesis. The results of volatility for tweet timing show higher volatility for the control group than for the bipolar group:</u>

> Bipolar Group Volatility: 0.209863
> Control Group Volatility: 0.22005



General Users' Volatility Activity

UsersActivity.ipynb

Filtering:

For general users' activity I also filtered out all tweets that were posted before 01.01.2013 to make our datasets more uniform.

Since our datasets contained only data with months when users posted anything, 'empty' months (when used didn't post anything) were missing. I had to fill in the gaps in months with zeros (add '0-value' months):

```
# ADD MISSING MONTHS WITH 0 POSTS
fillna_bipolar = bipolar_per_month.set_index('date').groupby('id').res
ample('MS').asfreq()
fillna_control = control_per_month.set_index('date').groupby('id').res
ample('MS').asfreq()
```

```
# FILL NAN WITH 0 VALUES TO POSTS COLUMN AND PREVIOUS ROW VALUE TO MON
THS COLUMN
fillna_bipolar['amount_of_months'].fillna(method='ffill', inplace=True
)
fillna_control['amount_of_months'].fillna(method='ffill', inplace=True
)
fillna_bipolar['posts_per_month'].fillna('0', inplace=True)
fillna_control['posts_per_month'].fillna('0', inplace=True)
fillna_bipolar.head()
```

Filtering:

Since users who had less then one-year history would not bring us significant results, I excluded them from the datasets (however, users who had at least two months of posting even with 10 months of break in the middle were included in the datasets).

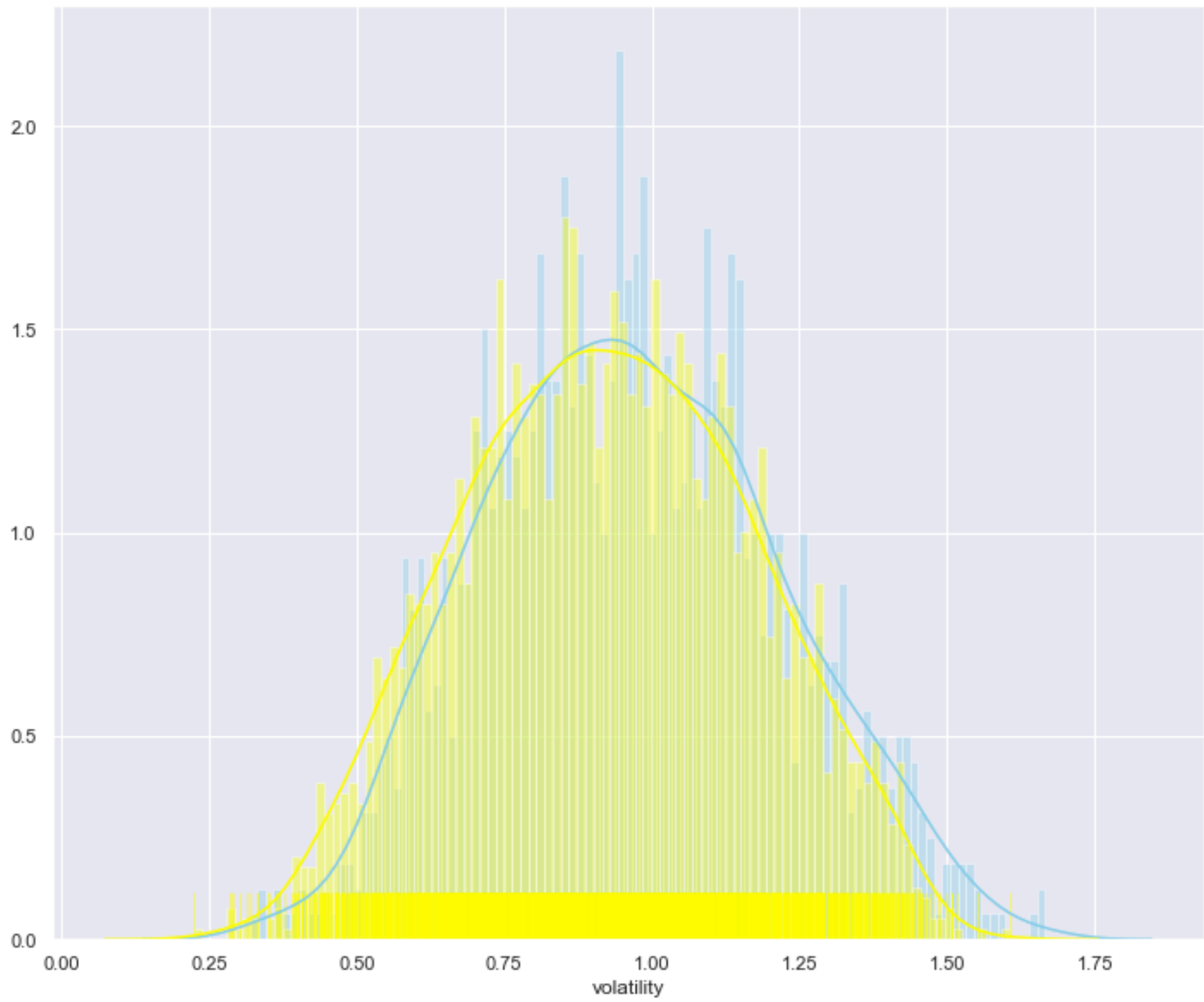
To measure volatility in users' activity I created additional column with amount of posts per each month for every user. I also normalized data (the number of posts per month) by Z-score. This let me to compute the consistency of posting activity and compare both groups: bipolar and control.

Although histogram doesn't show to us impressive insights, calculation shows us higher volatility in activity of bipolar users than control group users that were proved by t-test as well:

Bipolar Volatility Activity: 0.966729
Control Volatility Activity: 0.923236

T-test:
statistic=array([5.13668407]), pvalue=array([2.92986847e-07])

This result may signify that users diagnosed with bipolar disorder have higher variability in amount of posting during a period of time. Which aligns with our hypothesis.



Sentiment Analysis

semantic_valence_arousal.ipynb

For sentiment analysis I wanted to prove two characteristics of bipolar users:

1) They post more emotional texts (they have higher arousal score)
2) Their negative or positive posts are more significant in emotional way (they have higher positive and negative valence)

Filtering:

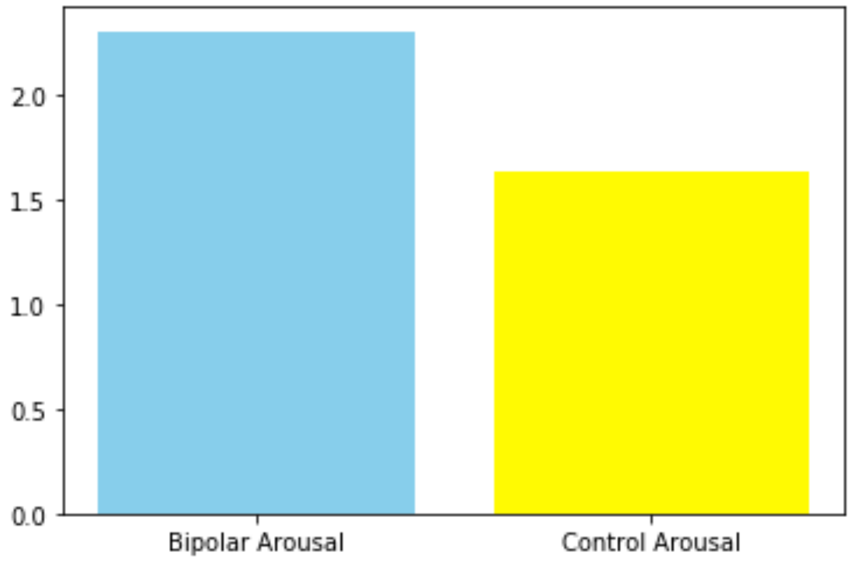
For sentiment analysis I excluded retweets.

I used standard NLTK library and public functions to compute mean arousal as well as positive and negative valence for both groups.

I received results that proved both hypotheses:

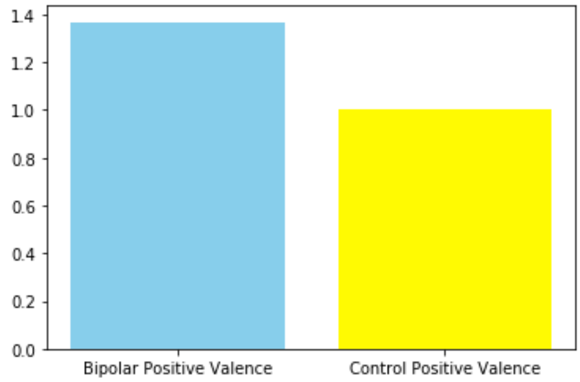
Mean Bipolar Arousal: 2.3076814464458337

Mean Control Arousal: 1.6345723829574528



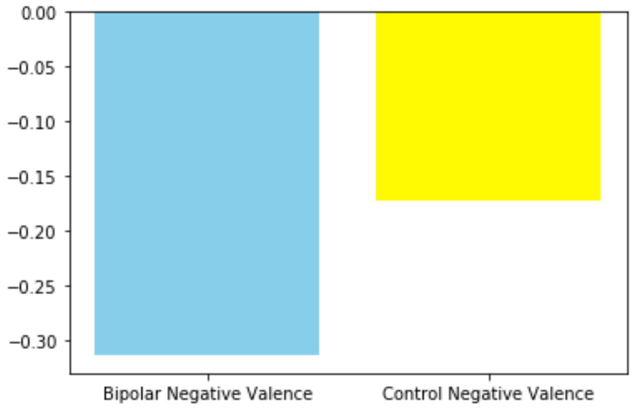
Bipolar Positive Valence: 1.3683190707782253

Control Positive Valence: 1.0041989043480273



Bipolar Negative Valence: -0.3138192668779847

Control Negative Valence: -0.17296477049013181



CONCLUSION

Results of this project show that some significant symptoms of bipolar disorder can be recognized by the modern tools of data analysis. We have proved that time of posting (higher entropy, i.e. night tweeting), uneven amount of publications (higher volatility for general activity) and more emotional texts (higher arousal score as well as higher negative and positive valence) correlate with the illness.

This is a promising sign for psychiatrists that can use social network analysis of their patients for orientation in diagnosis.

I believe that this project may promote more detailed research of the disease in social networks and help to monitor bipolar disorder indications in the future.