**Processing Alliance Chats**

**AIM**:

We want to Process the text messages posted by our users in the alliance chat rooms.

**ABSTRACT**:

1. We extracted the text from the alliance\_chat.txt, and translated lingo text, emoticons.
2. We estimate the sentiment and the subjectivity of the text using TextBlob Python Package.
3. We identified language of the text using Compact Langauge Detector and Part Of Speech Tagger. **85%** of the text is from English. There are still challenges in translating the human text expressions into comprehensible Language (**13%** of the comments are from Unknown Language).
4. We calculate the most delighted and the most discontented users based on sentiment, subjectivity and the count of their comments from Alliance chats.
5. We also project the Most and the Least delighted Alliances.
6. We also established the Users who have highest and the Lowest on the Feel Good Ranking which takes the User’s WinCount, LostCount, User Average Might in the Army and the delight metrics.

**PURPOSE**:

1. Understanding User’s messages gives essence of their game play strategies, idea, goals.
2. It gives us their impressions on the game, on the alliances, on each other.
3. We can see the sentiments of the User, Alliances and understand who likes and is happy about the games and who would is not happy.
4. We can classify the Users geographically based on the language of communication.
5. We can correlate the actions taken by the User to the communication they had.
6. We can understand the Organization of the Alliances and their activities.
7. We can recommend happy teams/alliances for an unhappy Users.
8. We can understand the subjectivity of their language. This is useful to understand their personality. People who are more objective vs people who are more subjective.

**PROCEDURE:**

1. **PARSING THE LINGO TEXT :**

In order to process the user’s comments on alliance chat rooms, we need to understand the user’s perspective.

Users tend to use short form English words called as Lingo. The nature of the spoken language tends to cryptic as the following example suggests.

“***peopl clse 2 my hart alw get D 1st msg 4 D dy.. gud mornin*** “ which means “People close to my heart always get the first message for the day. Good Morning”

We have captured a list of Lingo words from <http://transl8it.com/largest-best-top-text-message-list/>

1. **CAPTURING EMOTICONS:**

The shorter the symbols to express more meaning and context. The better people like it. They use emoticons which conveys the feelings and impressions better than verbal language at times.

Consider the following symbols “ **~:\** “ . This means an **Elvis** and “ **~=**  “means **Lit candle, indicating a flame (inflammatory message)** in the Emoticon Language.

We used the list of emoticons from the web and compiled a list of them.

The Challenging part of the Lingo and emoticons is that they evolve over time. We need refresh our translation as time progresses with newer set of lingo lists and emoticons to translate them to proper English.

1. **CAPTURING SENTIMENT AND SUBJECTIVITY:**

**SENTIMENT ANALYSIS:**

A basic task in ***sentiment analysis*** is classifying the ***polarity*** of a given text at the document, sentence, or feature/aspect level — whether the expressed opinion in a document, a sentence or an entity feature/aspect is positive, negative, or neutral.

We use ***TextBlob library of Python Language*** to capture the Sentiment of the text. It uses a Naïve Bayesian Classifier. To put it simply, It is a mathematical tool which differentiates between positive,negative and neutral statements based on the Linguistic properties of the words in a sentence. The tool has to be fed with a known set of words which are positive in sentiment such as “***like, awesome, great, fun”***, negative words like “***hate, angry, sad***” and neutral words such as “***daily, weekly, ordinary***”. The it comes up with a mapping of the sentence based in the words which are present in the sentence also including aspects such as negation like “not happy”, increments like ”more happy” and hundred others. Then it calculates the polarity of the sentence displayed as real number between -1 to 1. -1 indicating highly negative, 1 for highly positive and 0 for neutral.

**SUBJECTIVITY ANALYSIS:**

The **linguistic** expression of somebody’s **opinions**, **sentiments**, emotions, evaluations, beliefs, speculations *(private states which are personal to themselves).* The following *"****I love my country****" has a positive subjectivity while. "****India is a country****" has* zero subjectivity.The idea is to form a mathematical model which extracts subjective information such as “ love my country” ,”I love” and generates a real number between 0 to 1 indicating 1 highly subjective to 0 for highly objective.

**PURPOSE:**

1. We can find the most objective bunch and they could be the organizer/leader of the Alliance.
2. We can find the most subjective people and obtain their happiness index.

(While, this is not the state of the art sentiment-subjectivity analyzer, the TextBlob package is reasonably good for basic purpose.)

**DELIGHTED USERS:**  The sentiment sign gives the overall satisfaction of the Users. The more

Subjective the comments are the more the magnification of the sentiment. Our criteria for Delighted Users is the

Sum of User sentiment with subjectivity and the total count of comments.

**MOST DELIGHTED USERS**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| UserID | Sentiment | subjectivity | count | Delight |
| 1886619 | 1534.679 | 5858.805 | 18536 | **1.66664E+11** |
| 104283 | 1051.138 | 7489.556 | 28901 | **2.27525E+11** |
| 3765281 | 2309.957 | 6930.6 | 20577 | **3.29425E+11** |
| 578124 | 2148.754 | 7827.686 | 21083 | **3.54611E+11** |
| 1493831 | 2508.977 | 10743.93 | 29136 | **7.85398E+11** |

**LEAST DELIGHTED USERS**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| UserID | sentiment) | subjectivity) | count | Delight |
| 54297 | -528.05 | 3313.5 | 9518 | **-16653596347** |
| 938001 | -11.2991 | 142.6878 | 439 | **-707774.8943** |
| 2011251 | -4.19302 | 143.4544 | 958 | **-576243.8672** |
| 1531995 | -19.2011 | 45.92803 | 108 | **-95241.91846** |
| 8680783 | -4.02475 | 36.40456 | 302 | **-44248.81436** |

**MOST DELIGHTED ALLIANCE**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| AllianceID | sum(SI.sentiment) | sum(SI.subjectivity) | count | Happy |
| 21114|2 | 9831.675 | 58812.22 | 169784 | **9.81729E+13** |
| 22587|21 | 10858.63 | 39995.48 | 136743 | **5.93869E+13** |
| 8647|2 | 9369.567 | 39571.57 | 144412 | **5.35434E+13** |
| 24946|1 | 8557.666 | 37518.25 | 135858 | **4.36197E+13** |
| 19992|2 | 9653.061 | 37044.53 | 120680 | **4.31543E+13** |

**LEAST DELIGHTED ALLIANCE**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| AllianceID | Sentiment | Subjectivity | count | Delight |
| 19753|2 | 119.3221 | 499.5728 | 2274 | **135553311.9** |
| 24231|21 | 155.6131 | 521.3021 | 2031 | **164757603.3** |
| 1807|1 | 183.8791 | 489.6527 | 2399 | **215998536.9** |
| 9792|2 | 159.7405 | 565.3114 | 2829 | **255467474.1** |
| 24109|21 | 176.0031 | 584.8421 | 3513 | **361607276.9** |

**USER FEEL GOOD:**

Happy users are giving happy comments but that necessarily doesn’t mean they feel great about the gaming. They might like the game but the best gaming experience is obtained when you **“win”** the match or gather that **Mighty Army.** So we established a Feel Good criteria to differentiate it from the regular delight metric.

Our criteria for ***Feel Good*** is that the Users have maximum positive sentiment, maximum subjectivity, maximum count of the comments, maximum games won, minimum lost count and maximum might is possessed by the User .

The might mentioned here is the average might of the user in the matches.

**MAXIMUM USER FEEL GOOD**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| userid | Subjectivity | Sentiment | count | wincount | lostcount | might | **FEEL GOOD** |
| 1838917 | 62.20588 | 21.60867 | 276 | 90 | 66 | 6066.394 | **145599.7** |
| 546844 | 681.8848 | 130.7929 | 2623 | 4694 | 268 | 20.54291 | **90928.66** |
| 73494 | 1863.931 | 468.8063 | 5334 | 4776 | 726 | 18.39706 | **74512.99** |
| 115101 | 1284.604 | 266.7061 | 4215 | 6519 | 228 | 8.21677 | **51696.78** |
| 1199889 | 949.4609 | 194.8951 | 3498 | 6606 | 343 | 7.54146 | **47238.24** |

**MINIMUM USER FEEL GOOD**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| userid | Subjectivity | Sentiment | count | wincount | lostcount | might | **FEEL GOOD** |
| 1798595 | 1.24444 | 0.33403 | 56 | 199 | 310 | 11.02179 | **-1217.16** |
| 69510 | 2497.926 | 508.9735 | 8246 | 105 | 208 | 10.46587 | **-1073.09** |
| 9882719 | 0 | 0 | 1 | 1 | 105 | 9.29937 | **-960.14** |
| 11267 | 755.123 | 131.1225 | 2565 | 219 | 306 | 10.85216 | **-939.993** |
| 9550217 | 0 | 0 | 1 | 0 | 91 | 6.63193 | **-596.526** |

These Users are clearly not delighted. But there is no direct correlation whether they might quit or not. That inference needs data about user who actually quit and modelling the information to predict who will quit.

1. **CAPTURING LANGUAGE:**

Identifying the Language of the user helps us with geological structure of the User base. We can perform language based user analytics. We use the ***Chromium compact language detector*** from Google package for python.

The following are the most messaging language profiles.

**X AXIS**: SPOKEN LANGUAGES

**Y AXIS**: NUMBER OF MESSAGES IN EACH LANGUAGE

**DATA**: Alliance\_chat.txt Total number of Messages : 3404022.

**OBSERVATION:**

1. **60%** of the chats are in **English**.
2. **38% of the chats are Unknown.**
3. Less than 1% for each of the remaining languages.

**CHALLENGES:**

1. **Unknowns are predominantly English words without a proper grammar.**

This is a real issue. People don’t follow grammar most of the times when they are communicating. This impedes our Language detection tool. Yet the meaning is conveyed through those words.

The following English sentences are tagged as Unknowns by the Chromium Language Detector tool.

1. “where too?”
2. “stupid keyboard”
3. “how you been”

Apparently these phrases do convey the meaning even though they fail the grammar.

1. **Ignore Language??**

There is a significant portion of the comments (1072) which are ignored by the language detector.

There are words phrases which are not even English. Yet they are means of communication.

1. “pffft” means giggle.
2. “yourrrrrrr greaaaaaaatt” – you are great

The challenge with the Language identification or in general the text mining is that User comments are highly random. Gibberish could fill up the conversation and our tools might find it difficult to identify them precisely. English language in general is subjective as well makes the Text mining even more difficult.

1. **PARTS OF SPEECH TAGGING ON THE MESSAGES:**

Using the TextBlob package, we tag the words of the sentences in the message to their parts of speech {verb, noun, pronoun, adverd, adjective, conjunction etc}. They are useful to understand the user’s impression on the quality of the game/other users/alliances etc. For example the adjectives such as “slow, fast” and adverbs such as “beautiful, unique” gives us quality of the object being described.

***Language Detection revisited:***

In addition to this it also solves the language detection problem one more step. In the Last section we tagged the incomplete sentences in English as “Unknown” via the CLD classifier. Now we classify them again into English as we are able to detect the parts of speech they belong to. But there is a catch here.

The words ***‘ahhh Zebenyaaaaa’ is tagged as a Noun****. {Well they are nouns no doubt.}*

But anything other than a Noun is an Identifiable English word. Thus we tagged all the messages with parts of speech other than nouns as ‘English’. This reduces the number of messages with unknown languages.

Thus the new graph for Language detection is the Following.

**X AXIS**: SPOKEN LANGUAGES

**Y AXIS**: NUMBER OF MESSAGES IN EACH LANGUAGE

**DATA**: Alliance\_chat.txt Total: 3404022.

**OBSERVATION:**

1. The Percentage of comments in English is **85%**
2. While the percentage of comments with unknown Language is **13%.**
3. The remaining comments are distributed in an almost linear manner in decreasing percentage <1% across various languages shown.

We now have a better say on the Language of the Users. We can have advanced Language based user diagnostics.

**CONCLUSION:**

1. We extracted the text from the alliance\_chat.txt, and translated lingo text, emoticons.
2. We estimate the sentiment and the subjectivity of the text using TextBlob Python Package.
3. We identified language of the text using Compact Langauge Detector and Part Of Speech Tagger. **85%** of the text is from English. There are still challenges in translating the human text expressions into comprehensible Language(**13%** of the comments are from Unknown Language).
4. We calculate the most delighted and the least delighted users based on sentiment, subjectivity and the count of their comments from Alliance chats.
5. We also project the Most and the Least delighted Alliances.
6. We also established the Users who have highest and the Lowest on the Feel Good Ranking which takes the User’s WinCount, LostCount, User Average Might in the Army and the delight metrics.