Analysis of Happy Moments

Akshat Choube*, Harsh Yadav†, Chirag‡

April 8, 2018

Abstract

We present a model that when given a paragraph, categorizes it into one of the chosen seven categories of happiness. Furthermore, every paragraph/sentence is given a happiness index based on how much relatively positive it is. Since the traditional techniques of scoring a sentence as a sum of score of every word in the sentence is inefficient, we'd need NLP techniques such as Parts of Speech and Contextual Valence Shifting to do the same. In addition to it, we analyze the dataset further to find main reasons behind happiness in people of different age, sex, marital status, nationality, etc.

1 Motivation

To psychological researchers, happiness is life experience marked by a preponderance of positive emotion. Feelings of happiness and thoughts of satisfaction with life are two prime components of subjective well-being (SWB). Thus, understanding true source or reasons of happiness is an interesting and useful task. Understanding happiness is not straightforward, rather it highly depends on past experience of an individual, but there are some commonalities between happiness similar group can be extracted out. Understanding happiness will help us to build applications that makes people happy. Another important task is to score positivity or negativity of a sentence using some scoring criteria. Assigning scores to a sentence can help us figure out from public data like tweets and posts, the writer's subjective opinion and/or state of mind while writing it. This information is pretty useful in applications such as movie reviews, etc. Categorizing happy moments helps us to better understand their source. Thus happiness is something we all seek and is important to know its basic cause and reasons.

*111501031, CSE IIT PKD
†111501009, CSE IIT PKD
‡101501011, CE IIT PKD

2 Problem Formulation

The science of happiness is an area of positive psychology concerned with understanding what behaviors make people happy in a sustainable fashion. We want to build a model which would give us a clear and better understanding of how people express their happy moments. To achieve this, we are using a dataset named HappyDB, a corpus of 100,000 happy moments. The dataset contains answers to the queries that correlate happy moments with the past history i.e what made people happy in the past 24 hours or alternatively, past 3 months. HappyDB also contains information about demographic characteristics of the person such as age, marital status, nationality etc.

Three main tasks we plan to accomplish in this project are:

1. Analyzing happy Moments among people of different age groups, sex, nationality etc. and understanding their source of happiness.
2. Categorize each happy moment into these listed categories:
   (a) Achievement
   (b) Affection
   (c) Bonding
   (d) Enjoying the moment
   (e) Exercise
   (f) Leisure
   (g) Nature
3. Building a heuristics to score positivity or negativity of a sentence.

We show steps taken for accomplishing aforementioned tasks in following sections.

3 Algorithms/Models

3.1 Analyzing Source of Happiness among people of different age, sex, nationality etc.

Demographic.csv file contains demographic information about workers like their age, sex, marital status etc. which is used to get desired subset of population. We then look into their happy moments and collect most frequent words occurring in them.Then we visualize and analyze the results by making a word cloud. We have build a query builder using which we can query our data based on demographic or happiness information and get desired word cloud. for example What things make an Indian happy whose age is between 18 to 25 and happiness category is affection.

3.2 Categorizing Happy Moments

We trained a Long Short Term Memory (LSTM) for categorizing happy moments to seven categories mentioned below.

Category	Definition	Examples
Achievement	With extra effort to achieve a better than expected result	Finish work. Complete marathon.
Affection	Meaningful interaction with family, loved ones and pets	Hug. Cuddle. Kiss.
Bonding	Meaningful interaction with friends and colleagues	Have meals w coworker. Meet with friends.
Enjoy the moment	Being aware or reflecting on present environment	Have a good time. Mesmerize.
Exercise	With intent to exercise or workout	Run. Bike. Do yoga. Lift weights.
Leisure	An activity done regularly in one's free time for pleasure	Play games. Watch movie. Bake cookies.
Nature	In the open air, in nature	Garden. Beach. Sunset. Weather

Table 9: The categories of happy moments

We chose LSTM because unlike algorithm involving Bag of words which ignores semantics of words and loose ordering of words whereas LSTM considers it. To convert happy moment h to input layer x we need to get a sequence of words into vector representation x to feed it as the input layer to the network. This is done in the following way:

- Each happy moment (set of sentences) is represented by a integer vector which is frequencies of corresponding words of sentences in a corpus consisting of all happy moments.
- An embedding layer projects each of the these vectors into a d-dimensional continuous vector space. This is done by multiplying the these vector from left with weight matrix $W \in R^{dlvl}$. Where: $|V|$ is the number of unique words in vocabulary $e_t = WX_t$.
- After the embedding layer, the input sequence of vectors becomes a sequence of dense, real valued vectors $(e_1, e_2, ..., e_T)$.
- Output y_i for each happy moment sis represented as one hot vector eg. $[1, 0, 0, 0, 0, 0, 0]$ denotes achievement.

After embedding layer we can add LSTM layer with many memory units and then the softmax output layer. Recurrent Neural networks like LSTM generally have the problem of overfitting. Dropout can be applied between layers using the Dropout layer to tackle vanishing gradient problem.

3.3 Assigning a score to each happy moment

3.3.1 Naive Model :

The initial understanding of assigning a happiness index to a sentence was only the difference between the positive and negative score of each word.

The Algorithm is as under :

1. Given a sentence, preprocess it to find the tokens in the sentence.
2. For each word in tokens
   (a) Positive score = average positive score of all words in its synset.
   (b) Negative score = average negative score of all words in its synset.
3. Net score = positive score - negative score / length of sentence

3.3.2 Final Model

The initial algorithm worked for elementary and straightforward sentences. However, the fact that a positive sentence may not only contain positive words and vice versa made us think otherwise. For example, consider following sentences:

1. It was a happy day for me because I excelled in a test. (This sentence is straightforward and would be scored correctly by naive model)
2. John is not good. (This would be scored as positive sentence by naive model as good has a high positive value)

The Final model includes concept of valence shifters and thus does not suffer from above mentioned anomaly. The Algorithm is as follows

1. Given a sentence, preprocess it to find the tokens in the sentence.
2. Using POS tagging mechanism, tag every token as a part of speech.
3. Filter tokens that have a non-zero sentiwordnet score, have POS tag in [noun, verb, adverb, coordinating conjunction] or belongs to contextual valence shifter family.
4. For each word in tokens
   (a) Positive score = average positive score of all words in its synset.
   (b) Negative score = average negative score of all words in its synset.
5. For each word in tokens
   (a) If word is presuppositional
      i. Flip the score of the next word with non-zero score.
   (b) If word is a connector

i. Edit the score of all the tokens before this word to zero

(c) If word is negative

i. Flip the score of the next word with non-zero score.

(d) If word is intensifier

i. Multiplies the score of the next word with 1.5

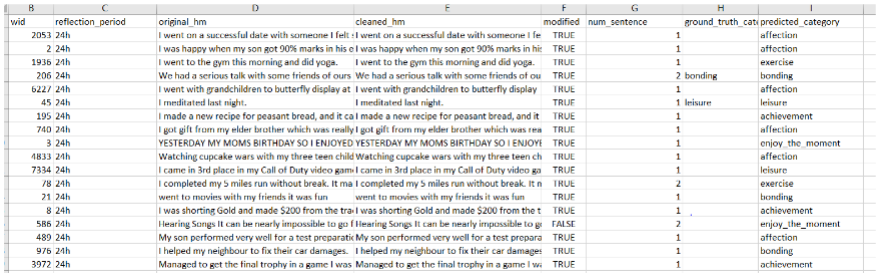
6. Score of the sentence = sum (score of tokens) / number of tokens

4 Implementation and Results:

4.1 Data Set

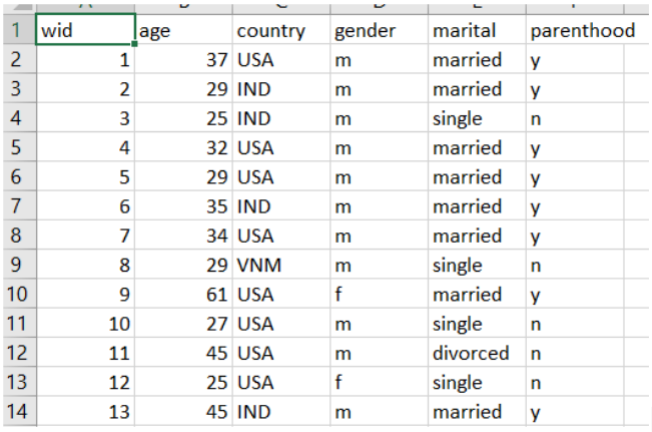
HappyDB dataset contains two files

1. Cleaned_hm :- Contains hmid (happy moment id), wid (worker's id), reflection_period, original_hm, cleaned_hm, num_sentence, ground_truth_category, predicted category.



wid	reflection_period	original_hm	cleaned_hm	modified	num_sentence	ground_truth_cat	predicted_category
2053	24h	I went on a successful date with someone I felt	I went on a successful date with someone I fe	TRUE	1		affection
2	24h	I was happy when my son got 90% marks in his e	I was happy when my son got 90% marks in hi	TRUE	1		affection
1936	24h	I went to the gym this morning and did yoga.	I went to the gym this morning and did yoga.	TRUE	1		exercise
206	24h	We had a serious talk with some friends of ours	We had a serious talk with some friends of ou	TRUE	2	bonding	bonding
6227	24h	I went with grandchildren to butterfly display at	I went with grandchildren to butterfly display	TRUE	1		affection
45	24h	I meditated last night..	I meditated last night.	TRUE	1	leisure	leisure
195	24h	I made a new recipe for peasant bread, and it ca	I made a new recipe for peasant bread, and it	TRUE	1		achievement
740	24h	I got gift from my elder brother which was really	I got gift from my elder brother which was rea	TRUE	1		affection
3	24h	YESTERDAY MY MOMS BIRTHDAY SO I ENJOYED	YESTERDAY MY MOMS BIRTHDAY SO I ENJOYI	TRUE	1		enjoy_the_moment
4833	24h	Watching cupcake wars with my three teen child	Watching cupcake wars with my three teen ch	TRUE	1		affection
7334	24h	I came in 3rd place in my Call of Duty video gam	I came in 3rd place in my Call of Duty video ga	TRUE	1		leisure
78	24h	I completed my 5 miles run without break. It ma	I completed my 5 miles run without break. It n	TRUE	2		exercise
21	24h	went to movies with my friends it was fun	went to movies with my friends it was fun	TRUE	1		bonding
8	24h	I was shorting Gold and made $200 from the tra	I was shorting Gold and made $200 from the t	TRUE	1		achievement
586	24h	Hearing Songs It can be nearly impossible to go	Hearing Songs It can be nearly impossible to gi	FALSE	2		enjoy_the_moment
489	24h	My son performed very well for a test preparati	My son performed very well for a test prepara	TRUE	1		affection
976	24h	I helped my neighbour to fix their car damages.	I helped my neighbour to fix their car damages	TRUE	1		bonding
3972	24h	Managed to get the final trophy in a game I was	Managed to get the final trophy in a game I wa	TRUE	1		achievement

2. Demographic.csv :- Contains information about worker like age, country, gender, marital, parenthood.



wid	age	country	gender	marital	parenthood
1	37	USA	m	married	y
2	29	IND	m	married	y
3	25	IND	m	single	n
4	32	USA	m	married	y
5	29	USA	m	married	y
6	35	IND	m	married	y
7	34	USA	m	married	y
8	29	VNM	m	single	n
9	61	USA	f	married	y
10	27	USA	m	single	n
11	45	USA	m	divorced	n
12	25	USA	f	single	n
13	45	IND	m	married	y

4.2 Preprocessing of data:

1. Cleansing:- HappyDB already contains cleaned version of data.
2. Text Processing:-
   (a) Tokenize
   (b) Transform to lowercase
   (c) Stemming
   (d) Removal of stop words is not necessary as they also have a sentiment value
3. Build word dictionary
4. All happy moments are tokenized and stored in a file. so we don't have to tokenize again and again.

4.3 Task 1 Results

We generated Word clouds for some sections of population and analyzed them to get some results.(We can generate wordcloud for any section using query builder. Some of results are shown here)

4.3.1 Different Age Groups

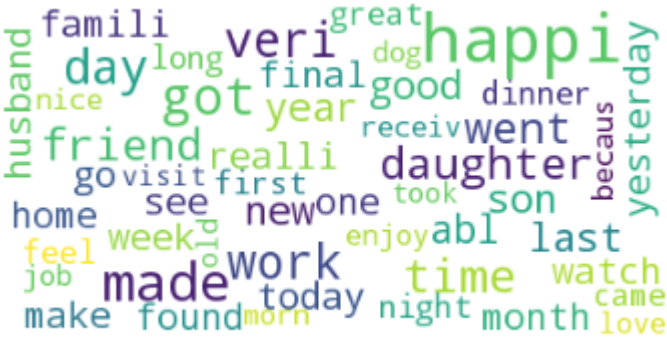
1. 18 - 25 years

   At this age major source of happiness of people are from friends, watching or playing games, success in work etc.

2. 25 - 45 years

For this age group we have words like son, daughter, wife which was absent from above age group. As seen friends still play a major role in determining happiness but it's count is less as compared to before.

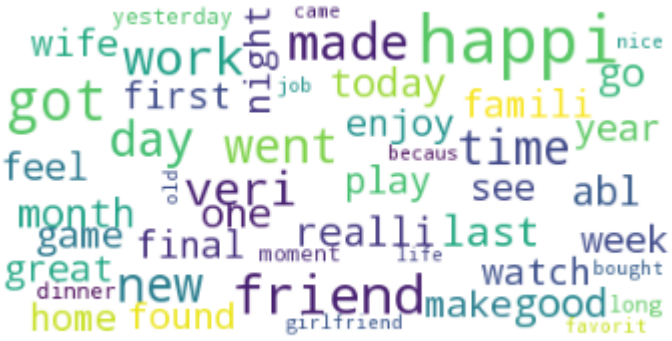
3. 45 - 60 years



Reasons like Husband, Wife, Daughter and Son (family) becomes major source of happiness. People find happiness in spending time and sharing love with family and dog.

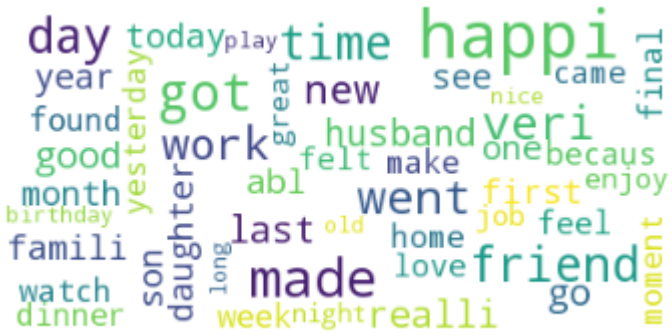
4.3.2 Based on Gender

1. Male



Things that made a man happy were friend, work, wife, playing, dinner, night, girlfriend.

2. Female

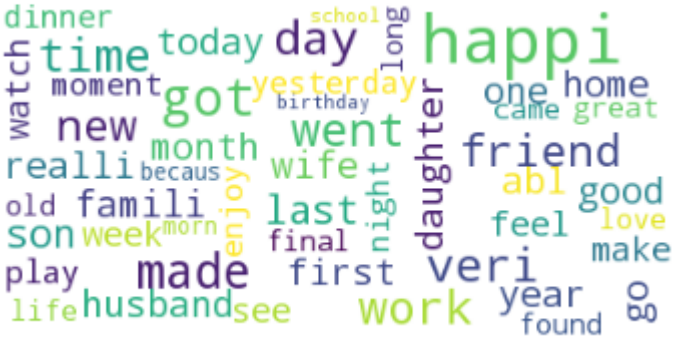


Things that make a woman happy were friend, daughter, son, husband, work, love.

Notice that happiness among man and woman are substantially different.

4.3.3 Based on Marital status

1. Married

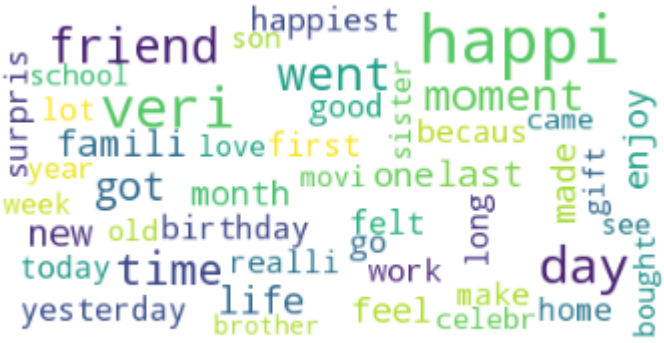


Things that makes married people happy are friend, family, work, time, daughter, son, etc.

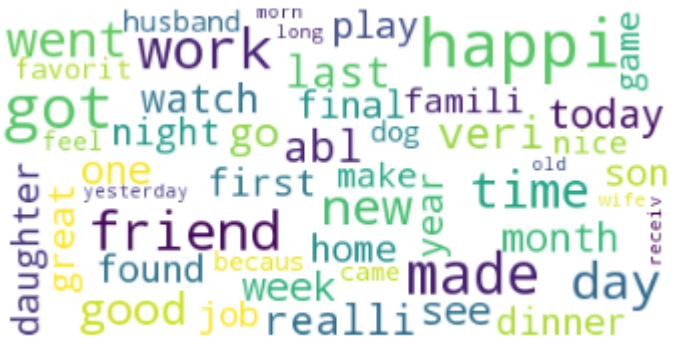
2. Single

4.3.4 Based on nationality

1. India



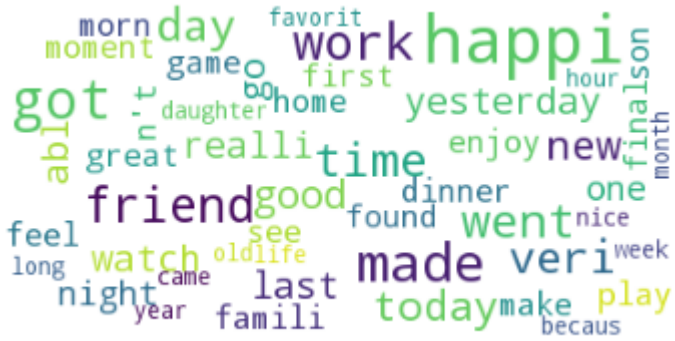
2. USA



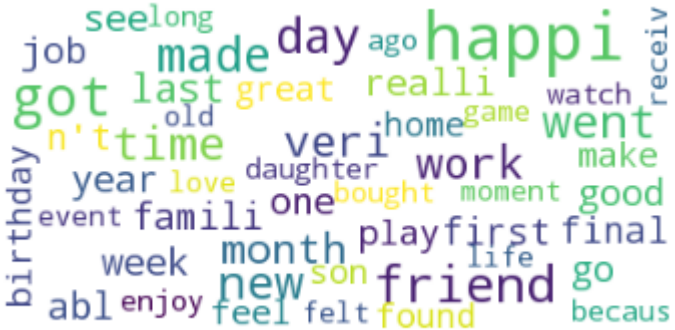
Friends and family are common source of happiness. words like dog, dinner, game etc. don't appear for Indians. Words like brother, sister, movie don't appear for Americans.

4.3.5 Based on Reflection Period

1. 24 hours

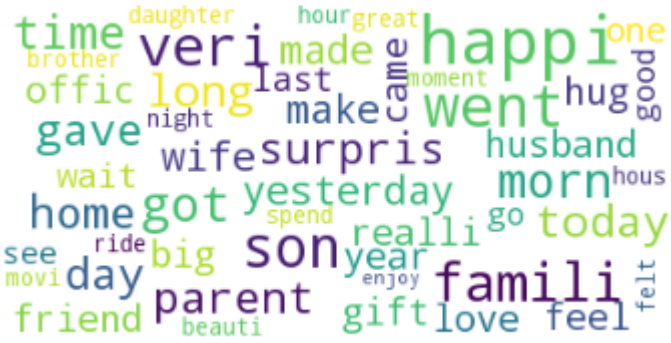


2. 3 months



Words like birthday, bought don't appear for 24 h for obvious reasons. But these two have a lot of similarity.

4.3.6 Another Word cloud



These are happy words related to affection for Indians where son occurs pretty much often than daughter.

4.4 Implementation of Task 2

LSTM model was trained and predictions were stored in a csv file name after_prediction.csv. Prediction accuracy 80.3 % was achieved which is good considering little ambiguity in various happy moments. for example : I was happy because we celebrated my mom's birthday yesterday, this can be affection as well as enjoy the moment. We had less training data, accuracy can be improved by increasing training data size.

Our Prediction results on some data is shown below

	hmid	wid	reflection_period	original_hm	cleaned_hm	modified	num_sentence	ground_truth_category	predicted_category	our_prediction
0	27673	2053	24h	I went on a successful date with someone I fel...	I went on a successful date with someone I fel...	True	1	NaN	affection	bonding
1	27674	2	24h	I was happy when my son got 90% marks in his e...	I was happy when my son got 90% marks in his e...	True	1	NaN	affection	affection
2	27675	1936	24h	I went to the gym this morning and did yoga.	I went to the gym this morning and did yoga.	True	1	NaN	exercise	nature
3	27676	206	24h	We had a serious talk with some friends of our...	We had a serious talk with some friends of our...	True	2	bonding	bonding	bonding
4	27677	6227	24h	I went with grandchildren to butterfly display...	I went with grandchildren to butterfly display...	True	1	NaN	affection	enjoy_the_moment
5	27678	45	24h	I meditated last night.	I meditated last night.	True	1	leisure	leisure	leisure
6	27679	195	24h	I made a new recipe for peasant bread, and it ...	I made a new recipe for peasant bread, and it ...	True	1	NaN	achievement	achievement
7	27680	740	24h	I got gift from my elder brother which was rea...	I got gift from my elder brother which was rea...	True	1	NaN	affection	affection
8	27681	3	24h	YESTERDAY MY MOMS BIRTHDAY SO	YESTERDAY MY MOMS BIRTHDAY SO	True	1	NaN	enjoy_the_moment	affection

4.5 Implementation of Task 3

Applying the above algorithm to rate sentences using contextual valence shifters on our happydb, we get the following result:

1. Arranging with most positive first

```
[(1675,
  "I was reminded of why I love Eli Manning so much. He's honestly a talented quarterback, but I had forgotten all ab
out him for so long, then when I saw a funny gif of him, it brought back so many happy memories.",
  0.8659755859374999),
 (257,
  'my partner buy a big cake for me its very nice and tasty .i never expect the cake from my partner he really loves
me lot .the cake is very delicious and very cold .the taste was very different compared to other cake .the day was we
nt very interesting and i spent most of the time with my partner .we went to movie and we had a fun .i really miss my
partner and after that most of the day i love to spent with him.',
  0.6123633322178023),
 (257,
  'i very eagerly waiting to see my friend engagement function .she were really very cute and beautiful .she were ver
y happy to have the engagement with his loved one .my friend character were really good and adjustable his partner we
re also very kind and humble .they were made for each other .they were loved a long years and they were in the living
together relationship then now they were understand well and now they get soon get married.',
  0.5582345642839746),
 (600,
  'My husband bought my favourite sweet that really made me very happy',
  0.3066856971153846),
 (2324,
  'The beautiful weather outside makes  me happy. ',
  0.29993872549019607),
 (693,
  'I got to eat some really delicious noodles that my mom made.',
  0.2821657509157509),
 (5046,
  'I found a poet that I really love a lot on twitter. ',
  0.27819113756613756),
 (2,
  'I was very happy when my family brings me beautiful gift for me',
  0.26410018552875697),
 (9360, 'My son told me that he loved my beautiful face', 0.24738636363636363),
```

2. Test cases:

   (a) Presuppositional words:

```
# Presuppositional : recover
score_sentence ('It was nice to see my mother recover from a disease')
```

```
The score for the sentence is :  0.08107142857142857
```

```
# Presuppositional : barely
score_sentence ('I am barely happy these days')
```

```
The score for the sentence is :  -0.09801136363636363
```

   (b) Intensifiers:

```
# Intensifier : amazing
score_sentence ('Scored two amazing goals while playing football')
```

```
The score for the sentence is :  0.05056390977443609
```

```
# Intensifier : very
score_sentence ('I was very happy to see my son come first in class in a test')
```

```
The score for the sentence is :  0.1564073542198542
```

```
# Intensifier : deeply
score_sentence ('I was deeply saddened to learn about the untimely death of my neighbour')
```

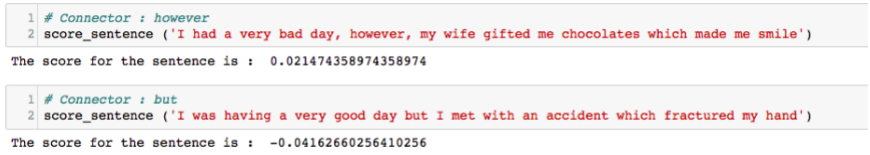
```
The score for the sentence is :  -0.06048044217687075
```

   (c) Negatives:

```
# Negative
score_sentence ('I am never sad')
```

```
The score for the sentence is :  0.20833333333333334
```

(d) Connectors:



5 Contribution:

1. **Akshat Choube :** Problem formulation, literature survey on classification of sentences, LSTMs, Context valence shifters. Designing Query Builder and Analyzer, Training LSTM and optimizing code.
2. **Harsh Yadav:** Problem formulation, literature survey on scoring sentences, syn-sets of a word, SentiWordNet, POS tagging, context valence shifters. Designing and implementing Algorithm for assigning score to sentences.
3. **Chirag:** Data collection, literature survey on similar work done in past eg. Italian tweets, preprocessing of data , analyzing Word clouds and drawing conclusions, found bugs and suggested improvements.

6 Conclusion

We find happiness in daily trivial and fundamental activities which motivates us to move further in life. Thus performing the tasks planned for this project made us understand psychology of happiness better and made us understand true sources of various kinds of happiness. On the technical side, this project introduced us to state of art NLP techniques and deep learning.

7 References :

1. HappyDB: A Corpus of 100,000 Crowdsourced Happy Moments : Link : https://arxiv.org/pdf/1801.07746.pdf
2. Predicting happiness: user interactions and sentiment analysis in an online travel forum : Link : https://link.springer.com/content/pdf/10.1007%2Fs40558-017-0079-2.pdf
3. Sequence Classification with LSTM Recurrent Neural Networks in Python with Keras . Link: https://machinelearningmastery.com/sequence-classification-lstm-recurrent-neural-networks-python-keras/

4. Learning Word Vectors for Sentiment Analysis by Andrew L. Maas, Raymond E. Daly, Peter T. Pham, Dan Huang,Andrew Y. Ng, and Christopher Potts

5. Measuring Similarity between sentences by Thanh Ngoc Dao (thanh.dao@gmx.net) , Troy Simpson (troy@ebswift.com)

6. Bag-of-words model : Wikipedia

7. Contextual Valence Shifters by Livia Polanyi and Annie Zaenen.