A
Project Report
On
"Emoji Prediction"

Prepared by

Krupali Dobariya(18DIT014)

Priyal Ramani(D19DIT082)

Under the guidance of

Prof. Dipak Ramoliya

A Report Submitted to

Charotar University of Science and Technology

For Partial Fulfillment of the Requirements for the

5th Semester Software Group Project-III((IT345)

Submitted at

Department of Information Technology

Devang Patel Institute of Advance Technology and Research

At: Changa, Dist: Anand – 388421

November 2020

CERTIFICATE

This is to certify that the report entitled "**Emoji Prediction**" is a bonafide work carried out by **Ms. Krupali Dobariya(18DIT014), Ms. Priyal Ramani(D19DIT082)** under the guidance and supervision of **Prof. Dipak Ramoliya** the subject **IT345Software Group Project-III** of 5thSemester of Bachelor of Technology in **Department of Information Technology, DEPSTAR** at Faculty of Technology & Engineering – CHARUSAT, Gujarat.

To the best of my knowledge and belief, this work embodies the work of candidate himself, has duly been completed, and fulfills the requirement of the ordinance relating to the B.Tech. Degree of the University and is up to the standard in respect of content, presentation and language for being referred to the examiner.

Prof. Dipak Ramoliya
Assistant Professor
Department of Information Technology
DEPSTAR, CHARUSAT, Changa, Gujarat.

Dr. Amit Ganatra
Principal, DEPSTAR
Dean, FTE
CHARUSAT, Changa, Gujarat.

Acknowledgement

The presentation of this report gives me the feeling of fulfillment. With immense pleasure we would like to present this report on this Dissertation report of **"Emoji Prediction".**

we would like to take this opportunity to bestow our acknowledgement to the entire person who have directly or indirectly availed us in making my project feasible and to turn it up in to successful piece of work.

we take this opportunity to express our gratitude to the following personalities who have helped us throughout our project. Without their support this project would never have been completed. we are extremely thankful to them. We are thankful to our Department HOD **Dr. Amit Nayak**, who has provided me resources from the college and we are also very thankful Internal Guide **Prof. Dipak Ramoliya** who has provided their time and never-ending support. For their support and help, we heartily thank them. This project would be incomplete without them.

Abstract

- Emojis are the incredible way of expressing yourself. Therefore, our machines should also be aware of the appropriate emoji to be used at the right time.
- In this project, the model is predicting emoji's according to a given sentences using models.
- The goal of this project is to predict an emoji that is associated with a text message.
- Emojis can be expressed to show the emotional and sentimental feelings in terms of graphical symbols which can add weightage to the text statement.
- Emojis have been widely used in social media and chat conversation on both web as well as mobile applications.
- With usage of emojis, the textual communicating statements will be more expressive & effective. Therefore, it is important to mine the relationship between the text & emoji symbols.

Table of Contents

List of Figures

CHAPTER 1: INTRODUCTION

1.1 PROJECT DEFINITION

As machine learning is currently the most famous technology, it is used in various applications.

When texting your friends, can you tell their emotional state? Are they happy? Could you put an appropriate smiley on each text message you receive? If so, you probably understand their sentiment. The goal of this project is to predict an emoji that is associated with a text message. To accomplish this task, we train and test several supervised machine learning models on a data to predict a sentiment associated with a text message. Then, we represent the predicted sentiment as an emoji.

Emojis can express diverse types of contents in a visual way, adapting to the informal style of communication in social networks. The meaning expressed by emoticons has been explored to allow or improve various tasks related to the sentiment analysis.

1.2 DESCRIPTION

- We represent a neural network methodology to predict appropriate emoji symbol for a given textual statement and the machine learning model is built by using word vector representations & deep learning frameworks.
- Model is created by using Global Vector (GloVe) embedding representation, combination of long short-term memory (LSTM) network and Recurrent neural network (RNN).
- Algorithm is able to generalize and associate words in the test set finds the proper emoji symbol even. The model becomes an accurate classifier mapping from sentences to emoji symbols.

CHAPTER 2: REQUIREMENTS

2.1 HARDWARE AND SOFTWARE REQUIREMENTS

For developing the model:

- Operating System: Windows 7 or higher, Linux.
- 2GB RAM for execution of testing model.
- Python IDLE(v3.7)/ Jupyter Notebook - for interpreting, debugging code and testing model.
- Google colab with good Internet connectivity (Optional/Alternative).
- Intel core i3-3200 or higher.

For executing the model:

- 2MB RAM for executionof classification.
- Intel core i3-3200 or higher.

CHAPTER 3: SYSTEM DESIGN

3.1 MAJOR FUNCTIONALITY

Pandas : With help of Pandas library we pre-process the data to filter Not Applicable data such as null values which keeps the dataset clean and maintain efficiency while working with huge data records.

Matplotlib : We used Matplotlib for data visualisation that is plotting various model output in a graph like bar, chart, scatter.

Numpy : Numpy library for easy and efficient data calculation within the dataset which also helps easy plotting of charts.

Emoji:Emoji library Predict Emoji from Sentence and Emotions.

Keras:It is used for training data using neural network algorithm LSTM and predict the emoji based on their work.

3.2 FLOW CHART

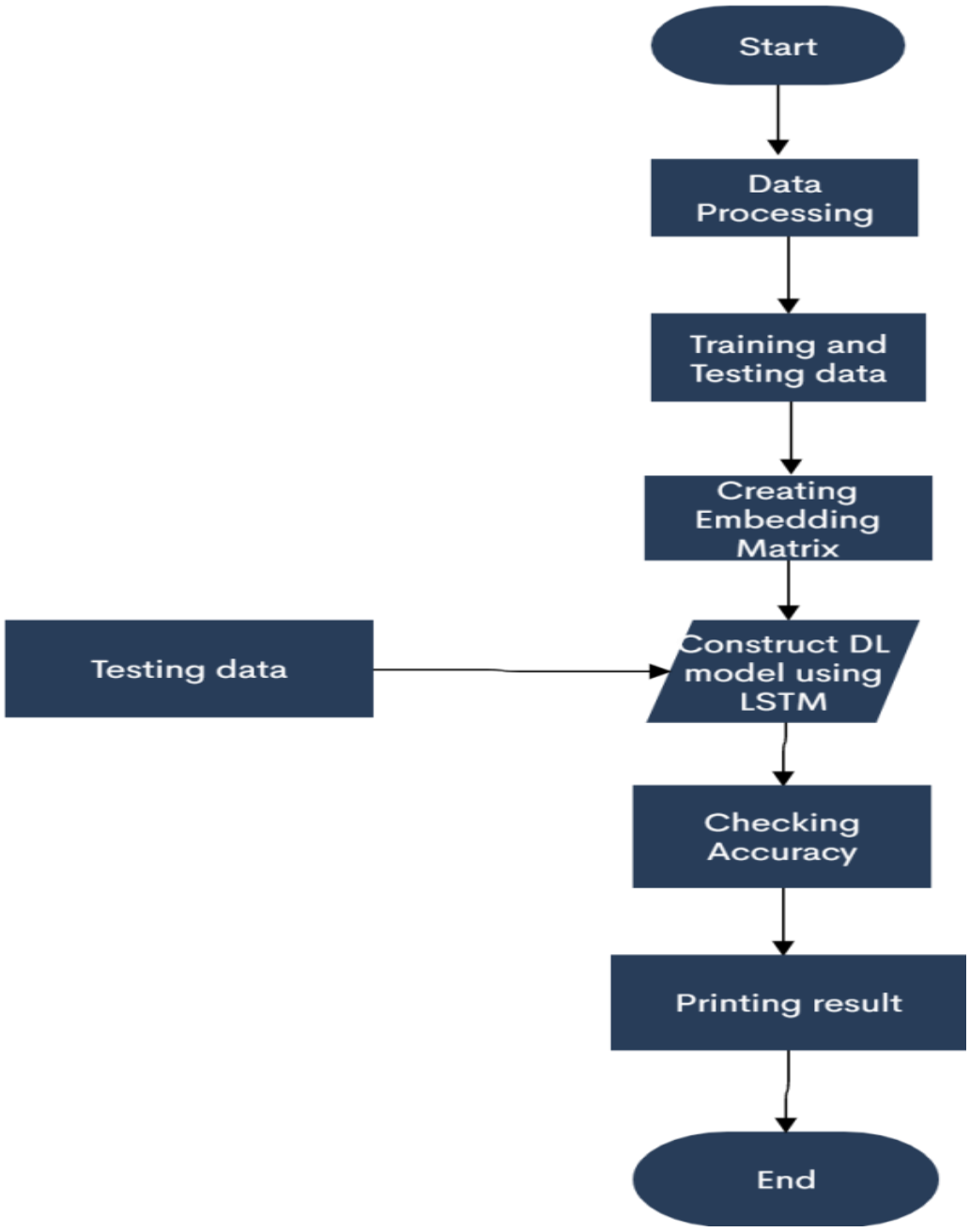

Fig 3.1 Flow Chart

3.3 SCREENSHOTS

1)Train Dataset:

	0	1	2
0	never talk to me again	3	NaN
1	I am proud of your achievements	2	NaN
2	It is the worst day in my life	3	NaN
3	Miss you so much	0	[0]
4	food is life	4	NaN

Fig 3.2Train Dataset

2)Test Dataset:

	Emotion	Unique id
52	you are failing this exercise\t	3
53	Good joke\t	2
54	You totally deserve this prize\t	2
55	I did not have breakfast	3

Fig 3.3Test Dataset

3)Emoji from unique value :

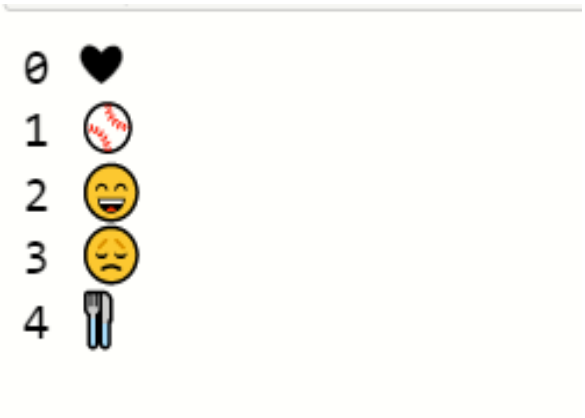

Fig 3.4Emoji from unique value

4)Data visualization(Bar Chart):

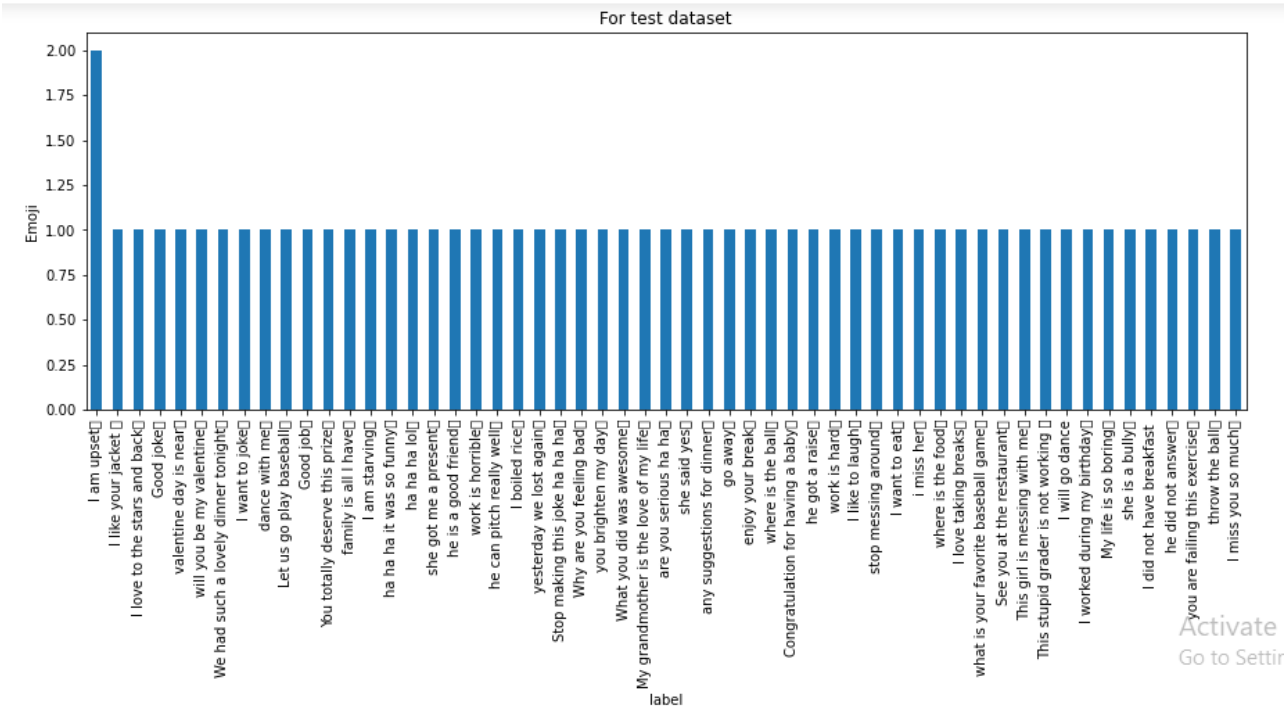

Fig 3.5Data visualization(Bar Chart)

5)Data visualization(Pie Chart):

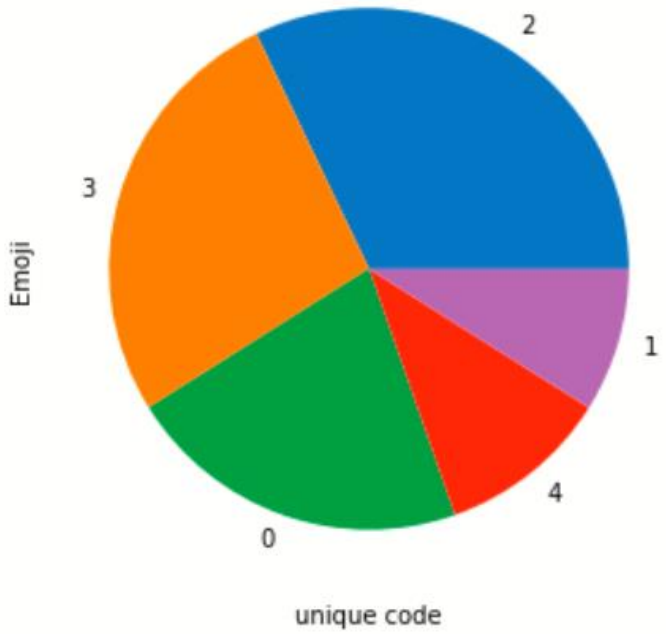

Fig 3.6 Data visualization(Pie Chart)

6) Data visualization(Line Chart):

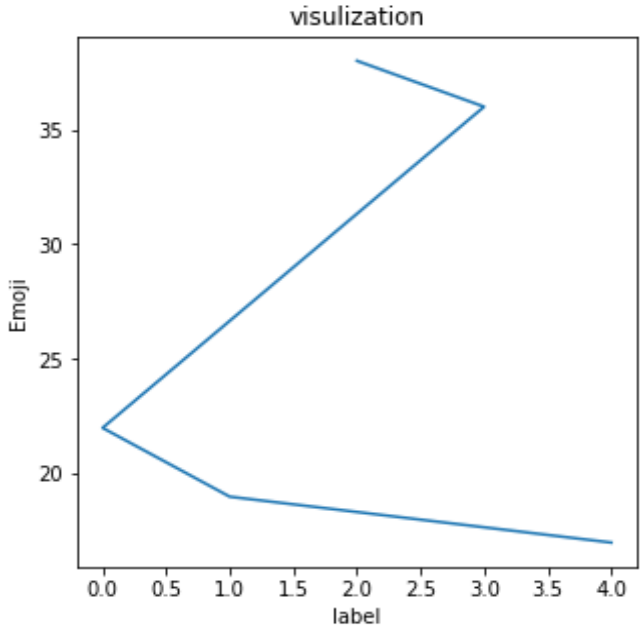

Fig 3.7 Data visualization(Line Chart)

7) Accuracy:

```
Accuracy: 0.6785714285714286
```

Fig 3.8 Accuracy

8) Predicted Data:

```
['she', 'got', 'me', 'a', 'present'] 😄 ♥
5
['he', 'is', 'a', 'good', 'friend'] 😄 ♥
6
['I', 'am', 'upset'] 😔 ♥
7
['We', 'had', 'such', 'a', 'lovely', 'dinner', 'tonight'] 😄 ♥
11
['work', 'is', 'hard'] 😄 😔
12
['This', 'girl', 'is', 'messing', 'with', 'me'] ♥ 😔
13
```

Fig 3.9 Predicted Data

9) Predicted Data:

```
26
['I', 'worked', 'during', 'my', 'birthday'] 😄 😔
28
['enjoy', 'your', 'break'] ⚾ 😄
29
['valentine', 'day', 'is', 'near'] 😄 ❤
32
['My', 'life', 'is', 'so', 'boring'] ❤ 😔
34
['will', 'you', 'be', 'my', 'valentine'] 😄 ❤
37
['I', 'am', 'starving'] 😔 🍴
41
['I', 'like', 'your', 'jacket'] ❤ 😄
48
['I', 'want', 'to', 'joke'] 😔 😄
49
['go', 'away'] ⚾ 😔
54
['You', 'totally', 'deserve', 'this', 'prize'] 😔 😄
55
['I', 'did', 'not', 'have', 'breakfast'] 😄 😔
```

CHAPTER 4: CONSTRAINTS AND FUTURE ENHANCEMENT

4.1 LIMITATIONS OF PROJECT

- By far our model is able to classify on specific dataset selected by us.
- By far our model is able to produce only accuracy of tested model.
- By selection of algorithm method it is still quite less accurate.
- Limited feature.
- No future prediction.

4.2 FUTURE ENHANCEMENT

- Testing and implementing with different classification techniques to increase accuracy of our model.
- Further pre-processing of data for clean and normalized efficient data usage.
- To improve on the current results, we probably, first and foremost, need access to more data for training.

4.3 OUTCOME

- Learning information gathering phase while selecting specific dataset from different dataset assets.
- Learning different libraries of Python for developing the model.
- Learning various methods of data visualization.
- Distribution of work load among teammates.

CONCLUSION

With the help of machine learning technology, it has become easy to find out relation and patterns among various data's. The work in this project mainly revolves around predicting the type of emoji which may help to know emotions using emoji of sentences in chat. Using the concept of machine learning we have built a model using training data set that have undergone data cleaning and data transformation. The model predicts the emoji with accuracy of 0.6785714285714286. Data visualization helps in analysis of data set. The graphs include bar, pie, line and scatter graphs each having its own characteristics. We generated many graphs and know the relation of emoji and emotions.

REFERENCES

- https://deepmoji.mit.edu/
- https://towardsdatascience.com/data-visualization-using-matplotlib-16f1aae5ce70
- https://www.learndatasci.com/tutorials/python-pandas-tutorial-complete-introduction-for-beginners/
- https://www.udemy.com/course/machinelearning/
- https://www.youtube.com/watch?v=iMIWee_PXl8&list=RDCMUCwB7HrnRlOfasrbCJoiZ9Lg&index=23&ab_channel=TheSemicolon
- https://towardsdatascience.com/machine-learning-recurrent-neural-networks-and-long-short-term-memory-lstm-python-keras-example-86001ceaaebc