Section: K18JC

Submitted To

Lovely Professional University
Jalandhar, Punjab, India.

ARTIFICIAL INTELLIGENCE

(INT 404)

EMOJI PREDICTION(🙂)

By

Name: Sourav Biswas
Registration No:11802672
Roll No:RK18JCA06

Name: Sanjay Das
Registration No:11802671
Roll No:RK18JCA05

EMOJI PREDICTION

ABSTRACT:

These days emoji play vital role in our social media chats or conversation. This emoji helps us to predict the emotion or state of mind of another person .

Emoji prediction is a fun variant of sentiment analysis. 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.

In this project, we build what's called a classifier that learns to associate emojis with sentences. Although there are many technical details, the principle behind the classifier is very simple: we start with a large amount of sentences that contain emojis collected from Twitter messages. Then we look at features from those sentences (words, word pairs, etc.) and train our classifier to associate certain features with their (known) smileys. For example, if the classifier sees the word "happy" in many sentences that also has the smiley  , it will learn to classify such messages as  .

On the other hand, the word "happy" could be preceded by "not" in which case we shouldn't rely on just single words to be associated with certain smileys. For this reason, we also look at word sequences, and in this case, would learn that "not happy" is more strongly associated with sadness, outweighing the "happy" part. The classifier learns to look at the totality of many word sequences found in a sentence and figures out what class of smiley would best characterize that sentence. Although the principle is simple, if we have millions of words of text with known smileys associated with the sentences, we can actually learn to do pretty well on this task.

ACKNOWLEDGEMENT:-

I would like to thank my mentor - Prof. Sagar Pande for his advice and inputs on this project. Many thanks to my friends and seniors as well, who spent countless hours to listen and provide feedback.

TABLE OF CONTENTS

INTRODUCTION :

1.1 Context :

This project has been done as part of my course for the INT(404) at Lovely Professional University . Supervised by Sagar Pande, I have one month to fulfill the requirements in order to succeed the module.

1.2 Motivations :

Being extremely interested in everything having a relation with the Machine Learning, the group project was a great occasion to give us the time to learn and confirm our interest for this field. The fact that we can make estimations, predictions and give the ability for machines to learn by themselves is both powerful and limitless in term of application possibilities. We can use Machine Learning in Finance, Medicine, almost everywhere. That's why I decided to conduct my project around the Machine Learning.

1.3 Idea :

As a first experience, we wanted to make my project as much didactic as possible by approaching every different steps of the machine learning process and trying to understand them deeply. The goal was to predict the emoji and know the sentiment of user. The idea was quite simple to know the sentiment through emojis . we have take some libraries that help us to find the polarity of the sentence and according to that polarity emoji was predicted.

Section: K18JC

TEAM MEMBERS :

Sanjay Das :

Contributions:-

1. Coding (joined)

2. Machine learning (joined)

3. VADER Sentiment Intensity Analysis

4. Reports(joined)

Sourav Biswas :

Contributions:-

1. Coding (joined)

2. Reports(joined)

3. Machine learning (joined)

4. emoji library

LIBRARIES :

1. NLTK(Natural language toolkit)

Natural Language Processing (NLP) is a branch of AI that helps computers to understand, interpret and manipulate human language.

NLP helps developers to organize and structure knowledge to perform tasks like translation, summarization, named entity recognition, relationship extraction, speech recognition, topic segmentation, etc.

Man is to woman as king is to _____? Meaning (king) – meaning (man) + meaning (woman)=? The answer is- queen

2. EMOJI

Emojis can also be implemented by using the emoji module provided in Python. To install it run the following in the terminal.

```
pip install emoji
```

emojize () function requires the CLDR short name to be passed in it as the parameter. It then returns the corresponding emoji. Replace the spaces with underscore in the CLDR short name.

Why NLTK is used :

The Natural Language Toolkit (NLTK) is a platform used for building python programs that work with human language data for applying in statistical natural processing (NLP). It contains text processing classification , stemming , tagging and semantic reasoning.

Built in default implementation can be used for quick prototyping , and specify implementation replaced as required . For example . here is a simple part of speech (POS) tagger based around NLTK's default word tokenizer and pos tagger.

```
>>>import nltk

>>> text=nltk.word_tokenize("and how so far")

>>>nltk.pos_tag(text)
```

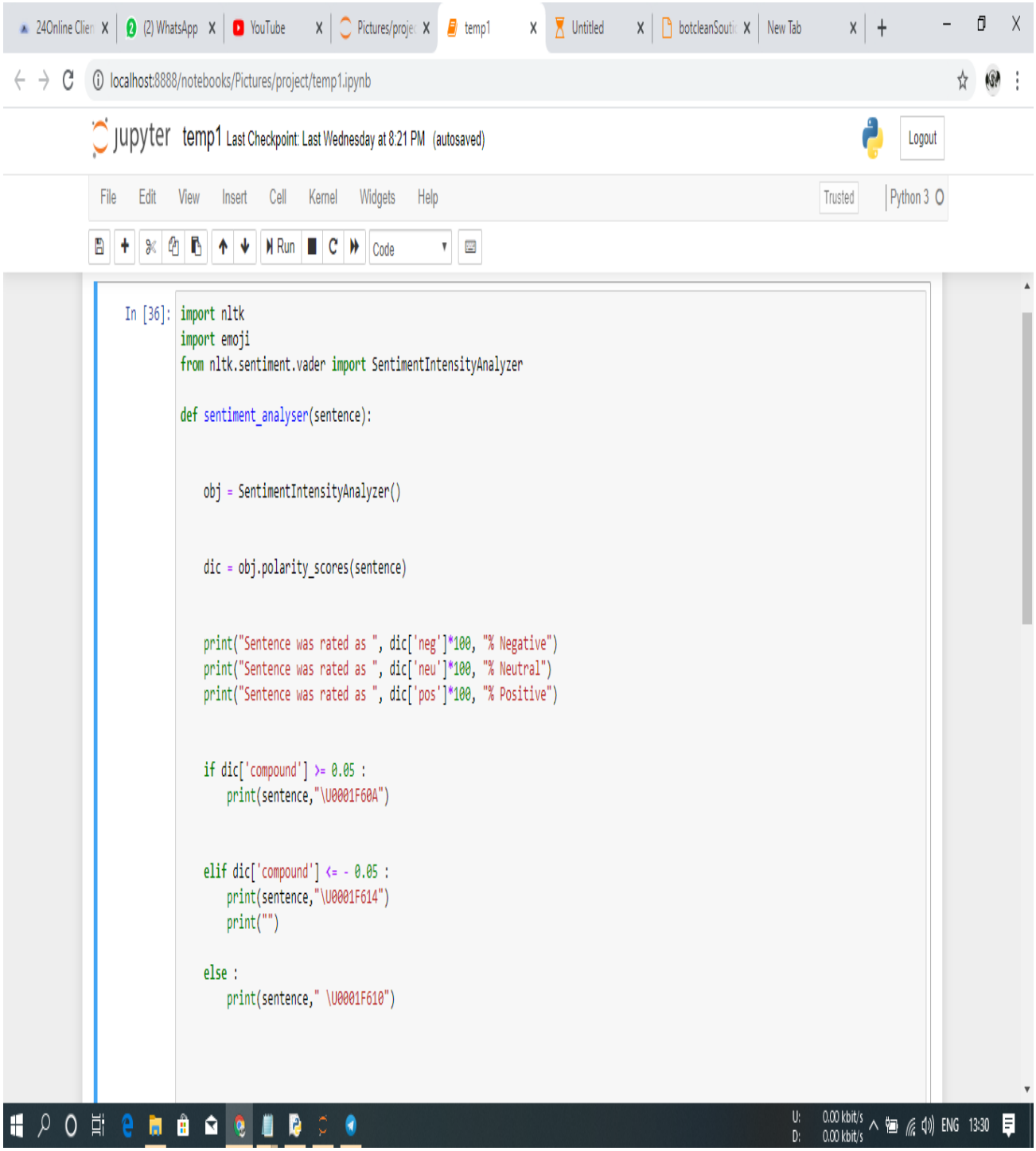
NLTK could be used for production work, although processing times might restrict it use. For faster processing . NLTK can make use of Scify and numPy snd core components could be replaced by compiled c++ implementation.

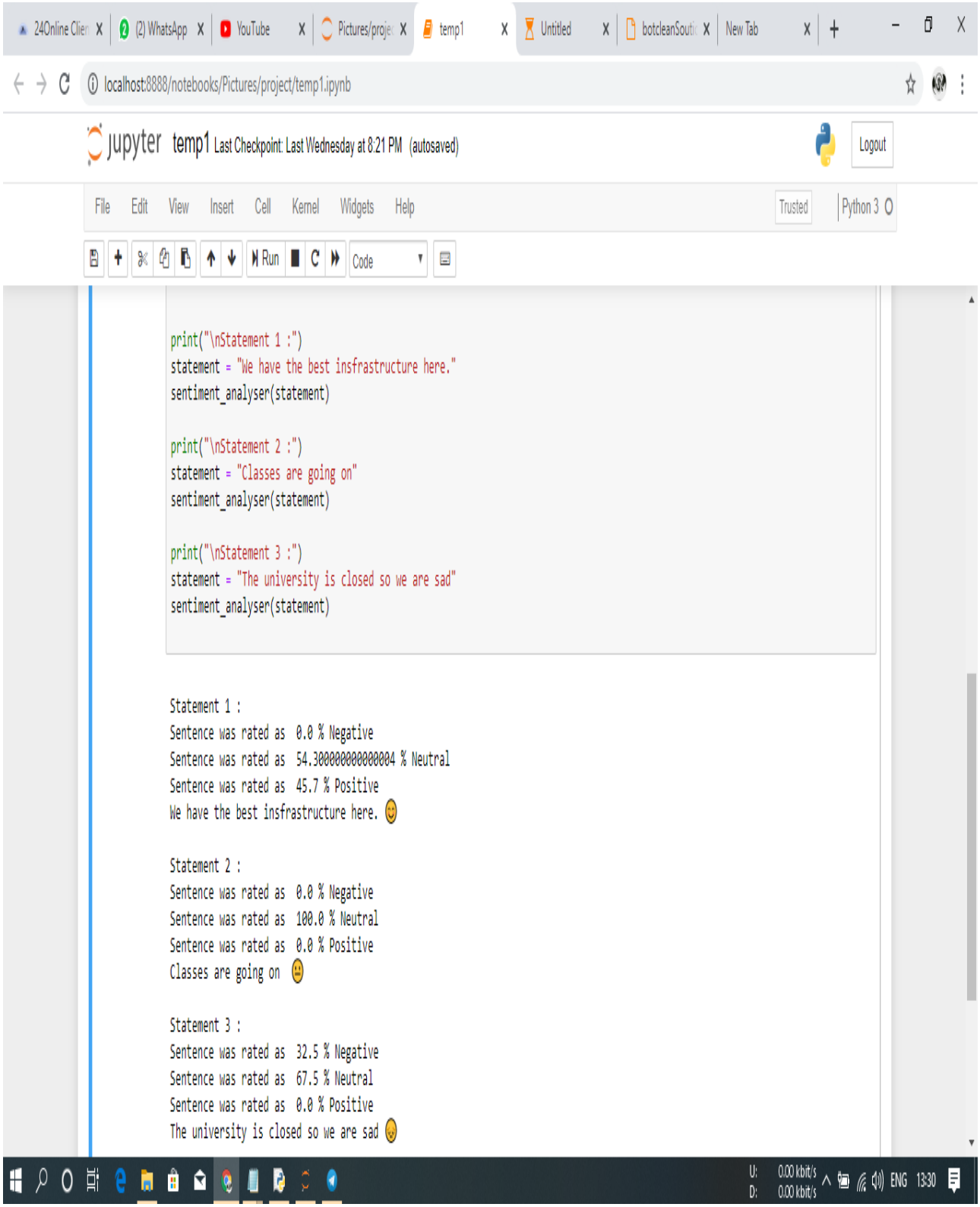
Why emoji is used :

Originally emoji were created as symbolic representation of emotions. Similar to punctuation , emoji are used to express emotion –irony and humor – and to substitute body language and tone of voice in text based communication .

In sentimental analysis the sentiment of the sentence is recognized by the emoji based upon the polarity of the sentence . emoji is the only way to know the type of sentence is it irony or a simple sentence. Emoji help us to know the expression also although it has all the features that need to know the type of sentence is this .

Screenshot :


localhost:8888/notebooks/Pictures/project/temp1.ipynb
jupyter temp1 Last Checkpoint: Last Wednesday at 8:21 PM (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3
Run
Code
In [36]:
import nltk
import emoji
from nltk.sentiment.vader import SentimentIntensityAnalyzer
def sentiment_analyser(sentence):
obj = SentimentIntensityAnalyzer()
dic = obj.polarity_scores(sentence)
print("Sentence was rated as ", dic['neg']*100, "% Negative")
print("Sentence was rated as ", dic['neu']*100, "% Neutral")
print("Sentence was rated as ", dic['pos']*100, "% Positive")
if dic['compound'] >= 0.05 :
print(sentence,"\U0001F60A")
elif dic['compound'] <= - 0.05 :
print(sentence,"\U0001F614")
print("")
else :
print(sentence," \U0001F610")


temp1 Last Checkpoint: Last Wednesday at 8:21 PM (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3
Run
Code
print("\nStatement 1 :")
statement = "We have the best insfrastructure here."
sentiment_analyser(statement)
print("\nStatement 2 :")
statement = "Classes are going on"
sentiment_analyser(statement)
print("\nStatement 3 :")
statement = "The university is closed so we are sad"
sentiment_analyser(statement)
Statement 1 :
Sentence was rated as 0.0 % Negative
Sentence was rated as 54.300000000000004 % Neutral
Sentence was rated as 45.7 % Positive
We have the best insfrastructure here.
Statement 2 :
Sentence was rated as 0.0 % Negative
Sentence was rated as 100.0 % Neutral
Sentence was rated as 0.0 % Positive
Classes are going on
Statement 3 :
Sentence was rated as 32.5 % Negative
Sentence was rated as 67.5 % Neutral
Sentence was rated as 0.0 % Positive
The university is closed so we are sad

Conclusions:-

It is our team's hope that this document will be of huge help with understanding of our little project as we have used a different approach which has proved beneficial for us and easy for us to understand the vast ocean that is Machine Leaning. We have reached the maximum accuracy of 90% after data cleaning but we will work forward to increase this accuracy little by little.

REFRENCES:-

1. **https://www.colorado.edu/linguistics/2018/05/08/emoji-prediction**
2. **https://www.kashipara.com/project/python/3538/emoji-prediction**
3. **https://www.geeksforgeeks.org/python-program-to-print-emojis/**
4. **https://github.com/vedantpuri/emoji-prediction**