

BIRZEIT UNIVERSITY

AI PROJECT.

TWEET EMOTION DETECTION

PREPARED BY:

SALEH KHATIB-1200991

DIAA TAHBOUB-1200136

TAHER HASAN-1191740

DR: DR. AZIZ QAROUSH

DATE:2023/2/17

Formalization:

In this project, we are talking about determining the type of Tweet, whether the Tweet is positive or negative, so we are talking about dealing with texts, so we resorted to using NLP (Natural Language Processing), where we did this sequence of steps:

- 1- Divide tweets into sentences and words by tokenize functions.
- 2- Finding the lemma of a word depending on its meaning and context.
- 3- Find the most frequent words and expressions by frequency distribution.

```
D:\python\python.exe C:/Users/Owner/Desktop/projects/AI/pro2/FeturuNeg.py
('.', 668)
('!', 209)
('♥', 81)
(77, 'اهم ساليه انه عمل 🔥')
(77, 'صدر طبيعى 🔴')
(77, 'مفدوش...', '▲')
('4/12/- الاتحاد محمد بالهبوطه', 50)
('انتكل ولكن من آتصعب أن تصفع يدك على قلبك كـ', 47)
('رأيه المهر مختلف بيور .. فلا جرنا يدوسه', 46)
('أصل مع طيفك الذين إنتعشتك عشان ما احزم بعيادك على طول ياجعلها في ذمتك ما وحشتكم ❤#تصفيبي...', 44)
("ياما : عطا .. ولا أحد : حاجة !", 33)
("ياما : وفا .. ولا عليه "ضفور" ياما اتجز .. ولا لقـ", 22)
('☀️', 22)
('+ أكثر من رتبة للتغريبة الواحدة ✨ وخلال ساعـه...', 19)
('كان البشر أكثر حرية ☀️ #هايد', 18)
('...', 18)
('سجد الله وبروبيه أتحزن حذر أكثر من بذر لعام 🌾 ابتداء من ربـلـس', 17)
('☀️', 16)
('☀️', 16)
('الـ', 15)
('☀️', 15)
('🔴🔴🔴', 12)
('الـ', 12)
```

Fig 1: Negative File Sentence Tokenize

```
D:\python\python.exe C:/Users/Owner/Desktop/projects/AI/pro2/FeturuNeg.py
('!', 4066)
('ء', 3692)
('♥', 3561)
('...', 3393)
('س', 2458)
('.', 2085)
('neg\\t', 1909)
('و', 1806)
('الله', 1586)
('ا', 1554)
(' Dio ', 1441)
('!', 1362)
('ما', 1306)
('على', 1283)
('يا', 1011)
('،', 951)
('☺', 912)
('☀', 877)
('،', 877)
(871, 'كل')
(838, 'بس')
(824, 'بـ')
(778, 'ـ')
an opened file in the Project View
File Control Run Python Packages TODO Python Console Problems Terminal Services
10:39 CRLF UTF-8 4 spaces Python 3.10
Type here to search
O 9°C ENG 10:58 PM 2/17/2023
```

Fig 2: Negative File Word Tokenize

```
D:\python\python.exe C:/Users/Owner/Desktop/projects/AI/pro2/FetruPos.py
('!', 655)
('تـىـ كـانـواـ يـشـعـونـ رـيـالـ مـدـرـيدـ هـذـهـ التـنـصـرـ', 167)
('!', 115)
('إـهـدـارـ غـرـبـ .. لـفـرـسـ سـهـلـةـ جـدـاـ', 48)
('خطـ دـفـاعـ يـعـاـسـ كـثـيرـاـ', 48)
('كـنـوـ .. إـمـكـنـ', 48)
('أـ نـرـاـكـمـ ♥ـ', 29)
('سبـاحـ الـجـيـرـ ☺ـ #ـجـيـنـ عـنـدـ الـغـ', 28)
('ـ رـفعـ مـوـتـ الـموـسـيقـيـ فـيـ الـآـيـانـ', 22)
('▲ـ', 20)
('ـ', 18)
('ـ', 16)
('ـ', 15)
('ـ', 15)
('ـ يـافـنـ تـمـلـكـ عـطـاءـ جـمـاـ أـكـالـيلـ إـمـتـنـانـ عـلـىـ هـيـلـةـ شـكـرـ', 15)
('ـ يـعـطـيكـ العـافـيـهـ', 15)
('ـ', 15)
('ـ', 15)
('ـ', 15)
('ـ الـنـاسـ الـبـرـ طـلـبـ وـاحـتـلـتـ دـيـ عـارـدـةـ أـنـ الشـفـلـ الـحـدـ يـادـوـ حـيـنـاـ!', 14)
('ـ اـسـفـاطـ النـظـامـ', 14)
('ـ', 13)
('ـ', 12)
('ـ الـدـاعـوهـ سـابـيهـ بـاـ حـبـبـيـ ☺ـ وـ قـيلـهـاـ اـسـاءـ لـخـدـالـلهـ لـ إـلتـزـامـ بـالـجـيـادـ وـ لـ اـحـترـامـ', 12)
('ـ', 12)
```

Fig 3: Positive File Sentence Tokenize

```
D:\python\python.exe C:/Users/Owner/Desktop/projects/AI/pro2/FtruePos.py
('!', 4883)
('..', 3783)
(3513 , 'و')
(2828 , 'و')
(2762 , 'ال')
('.', 2392)
(2248 , 'ع')
('pos\\lt', 2031)
(1991 , 'ا')
(':', 1908)
('😊', 1783)
('❤', 1770)
('.', 1647)
(1362 , 'ي')
('!', 1211)
(1015 , 'ل')
('ـ', 999)
(998 , 'ك')
('❗', 916)
(810 , 'لـ')
(783 , 'الله')
(773 , 'الله')
('`', 757)
```

Fig 4: Positive File Word Tokenize

From the previous steps, we deduce the following features:

1- Emoji:

This feature takes the emoji inside the sentences and determines the weight of the positive and the negative in them. If the positive weight is dominant over the negative, we take the answer as 1, and vice versa we take -1, but if they are equal, we take 0.

2- Greetings and prayers:

In this feature, we check the tweets if they contain expressions of greeting and supplication. If it contains expressions, the answer is 1, and if it does not, the answer is 0.

3- Positive&Good:

In this feature, we check the tweets if they contain words denoting positivity and goodness. If it contains the words, the answer is 1, and if it does not, the answer is 0.

4- Swear&Angry:

In this feature, we check the tweets if they contain expressions of swearing and anger. If it contains the words, the answer is 1, and if it does not, the answer is 0.

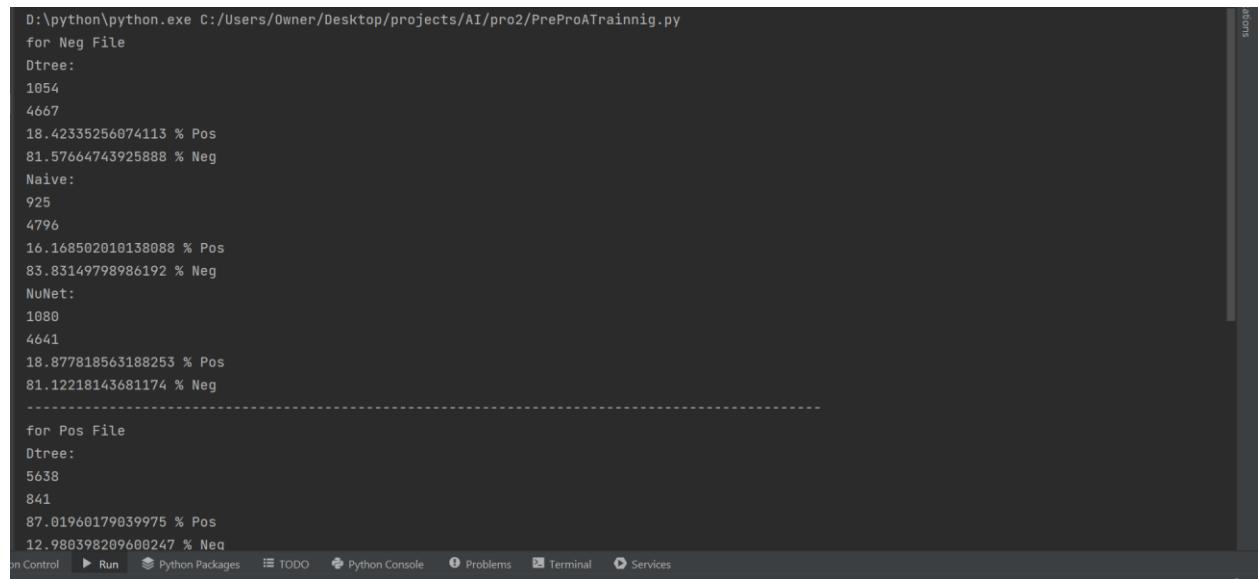
5- Negative&Bad:

In this feature, we check the tweets if they contain words denoting badness and negative. If it contains the words, the answer is 1, and if it does not, the answer is 0.

From these features we can convert tweets to samples in the Data Set.

After that we apply our 3 classifiers which are:

Decision tree, Naive bayes, Neural network.



```
D:\python\python.exe C:/Users/Owner/Desktop/projects/AI/pro2/PreProATrainnig.py
for Neg File
Dtreet:
1054
4667
18.42335256074113 % Pos
81.57664743925888 % Neg
Naive:
925
4796
16.168502010138088 % Pos
83.83149798986192 % Neg
NUNet:
1080
4641
18.877818563188253 % Pos
81.12218143681174 % Neg
-----
for Pos File
Dtreet:
5638
841
87.01960179039975 % Pos
12.980398209600247 % Neg
```

Fig 5: Output Of Training & Testing1.

```

for Pos File
Dtree:
5638
841
87.01960179039975 % Pos
12.980398209600247 % Neg
Naive:
5410
1069
83.50054020682204 % Pos
16.49945979317796 % Neg
NuNet:
5655
824
87.28198796110512 % Pos
12.71801203889489 % Neg
for the DTTree we have:
precision: 0.842498505678422
recall: 0.8701960179039975
F1-value: 0.8561233011920129

Process finished with exit code 0

```

File Control Run Python Packages TODO Python Console Problems Terminal Services

an opened file in the Project View 237:16 CRLF UTF-8 4 spaces Python 3.10

Type here to search 9°C 11:43 PM 2/17/2023

Fig 6: Output Of Training & Testing1.

I hope you enjoy Our report.

Best wishes.

