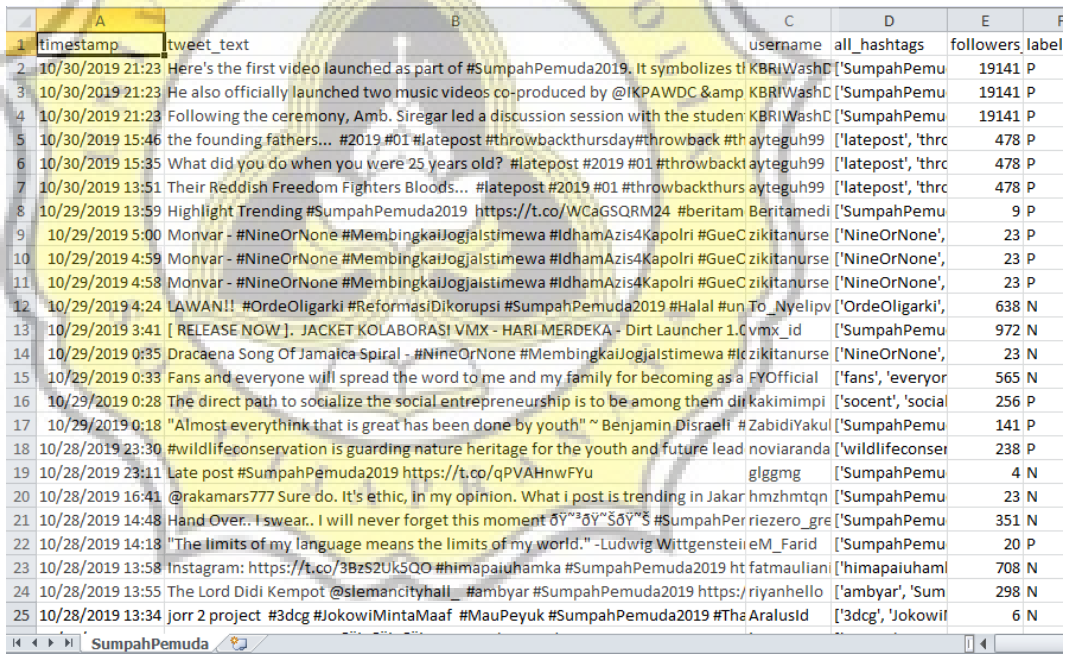
CHAPTER 5
IMPLEMENTATION AND TESTING

5.1. Implementation

```
1. import math
2. import numpy as np
3. temp_euclid = []
4. for b in uji_tfidf:
5.    i = 0
6.   e_doc = np.zeros(len(liststem))
7.    for doc in liststem:
8.       for a in list_tfidf:
9.          square = a['info'][i]['w']-b[a['term']]
10.          temp = math.pow(square,2)
11.          e_doc[a['info'][i]['doc']] += temp
12.       e_doc[i] = math.sqrt(e_doc[i])
13 .       i+=1
14.    temp_euclid.append(e_doc)
15. for d in temp_euclid:
16.    minv = min(d)
17.   result = (np.where(d == np.amin(d)))
18.    print(label[result[0]])
19.    print(minv)
20.    print(d)
```

Line 1-2 import python library math and numpy, used for calling mathematic function. Line 3 makes variable temp_euclid, Line 4 calls uji_tfidf variable as b, line 5 declares ‘i’ equal with 0. Line 6-13 calculate KNN algorithm, all array from testing document is called then the result of Tf-Idf testing is minus by Tf-Idf training and squared, after that the result is added by all texts inside document start from zero, then every words is multiple by the result erlier. Finally, data training which is score closed to data testing will shown with the rank and its result with every text in document.

5.2. Testing



	A	B	C	D	E	F
1	timestamp	tweet_text	username	all_hashtags	followers	label
2	10/30/2019 21:23	Here's the first video launched as part of #SumpahPemuda2019. It symbolizes tl	KBRIWashD	['SumpahPemu	19141	P
3	10/30/2019 21:23	He also officially launched two music videos co-produced by @IKPAWDC &amp	KBRIWashD	['SumpahPemu	19141	P
4	10/30/2019 21:23	Following the ceremony, Amb. Siregar led a discussion session with the studen	KBRIWashD	['SumpahPemu	19141	P
5	10/30/2019 15:46	the founding fathers... #2019 #01 #latepost #throwbackthursday#throwback #th	ayteguh99	['latepost', 'thro	478	P
6	10/30/2019 15:35	What did you do when you were 25 years old? #latepost #2019 #01 #throwback	ayteguh99	['latepost', 'thro	478	P
7	10/30/2019 13:51	Their Reddish Freedom Fighters Bloods... #latepost #2019 #01 #throwbackthurs	ayteguh99	['latepost', 'thro	478	P
8	10/29/2019 13:59	Highlight Trending #SumpahPemuda2019 https://t.co/WCaGSQRM24 #beritam	Beritamedi	['SumpahPemu	9	P
9	10/29/2019 5:00	Monvar - #NineOrNone #MembingkaiJogjalstimewa #IdhamAzis4Kapolri #GueC	zikitanurse	['NineOrNone',	23	P
10	10/29/2019 4:59	Monvar - #NineOrNone #MembingkaiJogjalstimewa #IdhamAzis4Kapolri #GueC	zikitanurse	['NineOrNone',	23	P
11	10/29/2019 4:58	Monvar - #NineOrNone #MembingkaiJogjalstimewa #IdhamAzis4Kapolri #GueC	zikitanurse	['NineOrNone',	23	P
12	10/29/2019 4:24	LAWAN!! #OrdeOligarki #ReformasiDikorupsi #SumpahPemuda2019 #Halal #un	To_Nyelipv	['OrdeOligarki',	638	N
13	10/29/2019 3:41	[RELEASE NOW]. JACKET KOLABORASI VMX - HARI MERDEKA - Dirt Launcher 1.0	vmx_id	['SumpahPemu	972	N
14	10/29/2019 0:35	Dracaena Song Of Jamaica Spiral - #NineOrNone #MembingkaiJogjalstimewa #Id	zikitanurse	['NineOrNone',	23	N
15	10/29/2019 0:33	Fans and everyone will spread the word to me and my family for becoming as a	FYOfficial	['fans', 'everyor	565	N
16	10/29/2019 0:28	The direct path to socialize the social entrepreneurship is to be among them dir	kakimimpi	['socent', 'social	256	P
17	10/29/2019 0:18	"Almost everythink that is great has been done by youth" ~ Benjamin Disraeli #	ZabidiYaku	['SumpahPemu	141	P
18	10/28/2019 23:30	#wildlifeconservation is guarding nature heritage for the youth and future lead	noviaranda	['wildlifeconser	238	P
19	10/28/2019 23:11	Late post #SumpahPemuda2019 https://t.co/qPVAHnwFYu	glggmg	['SumpahPemu	4	N
20	10/28/2019 16:41	@rakamars777 Sure do. It's ethic, in my opinion. What i post is trending in Jakar	hmzhmtqn	['SumpahPemu	23	N
21	10/28/2019 14:48	Hand Over.. I swear.. I will never forget this moment ðŸ˜³ðŸ˜Šð #SumpahPer	riezero_gre	['SumpahPemu	351	N
22	10/28/2019 14:18	"The limits of my language means the limits of my world." -Ludwig Wittgenstei	eM_Farid	['SumpahPemu	20	P
23	10/28/2019 13:58	Instagram: https://t.co/3BzS2Uk5QO #himapaiuhamka #SumpahPemuda2019 ht	fatmauliani	['himapaiuhaml	708	N
24	10/28/2019 13:55	The Lord Didi Kempot @slemancityhall_ #ambyar #SumpahPemuda2019 https:/	riyanhello	['ambyar', 'Sum	298	N
25	10/28/2019 13:34	jorr 2 project #3dcg #JokowiMintaMaaf #MauPeyuk #SumpahPemuda2019 #Tha	AralusId	['3dcg', 'JokowiI	6	N

SumpahPemuda

Illustration 5.2.1: Scraped Document

Above is an example document that has been scraped from Twitter using the API key and Access Token. The data taken is timestamp, tweet, user account, number of hashtags, and total followers. Label on the right side is supervised by manual way.

Illustration 5.2.2: Document After Pre-processing

The picture above is a form of tweets in the document that has gone through the pre-processing stage. Each tweet will be divided into words where it has become just a basic word without symbols, numbers and common words.

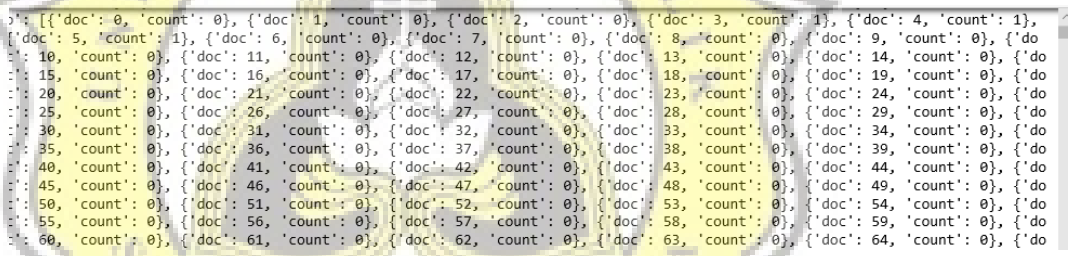


Illustration 5.2.2: Term Frequency Result

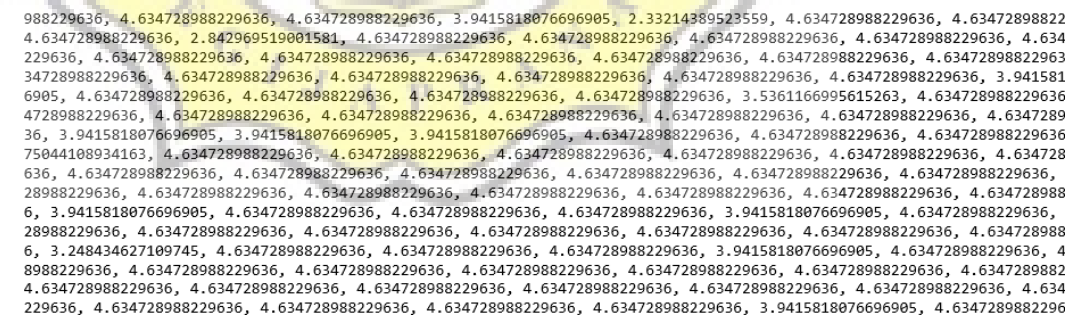


Illustration 5.2.3: Inverse Document Frequency Result

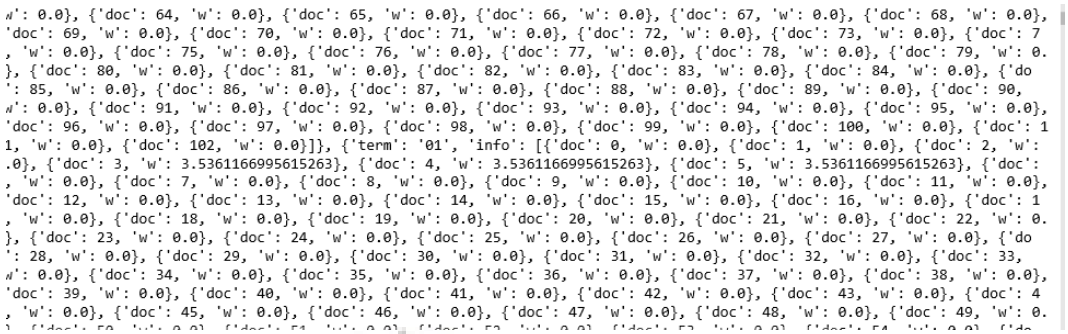


Illustration 5.2.4: Tf-Idf Result

The above results are Tf-Idf from the data to be tested. It may took a few minutes to calculate the document using python because of how massive the document is. Save the result for k-Nearest Neighbour calculating.

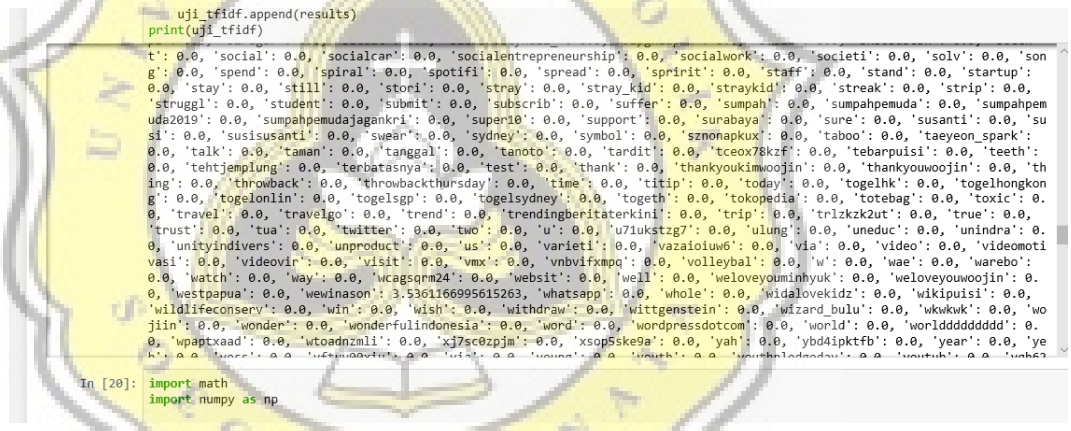


Illustration 5.2.5: Tf-Idf Data Training Result

Data training have to be tested, therefore the data will be calculated using k-Nearest Neighbour algorithm.

```
for b in uji_tfidf:
    i = 0
    e_doc = np.zeros(len(liststem))
    for doc in liststem:
        for a in list_tfidf:
            square = a['info'][i]['w']-b[a['term']]
            temp = math.pow(square,2)
            e_doc[a['info'][i]['doc']] += temp
        e_doc[i] = math.sqrt(e_doc[i])
        i+=1
    temp_euclid.append(e_doc)
```

Illustration 5.2.6: k-Nearest Neighbour Algorithm Formula

```
35    N
Name: label, dtype: object
3.6677901234544
[28.20944475 30.2869862  26.91860624 26.22601631 26.03299687 24.63159175
 21.9825822  24.00201844 24.00201844 24.00201844 22.7077744  30.99783465
 25.23908671 30.53867578 26.16648811 22.61612582 25.34638632 21.10109557
 27.17809803 22.44445871 26.95813852 21.611339   22.7077744  23.70257759
 22.9179938  20.44013428 22.05001051 22.39411754 20.78968418 21.46131714
 28.8126595  29.71979126 27.15248676 22.09564798 20.54466055  3.66779012
 21.34967357 21.53381065 21.74124174 28.74564928 23.27512434 23.0951984
 23.96122933 26.84321319 27.48867804 21.23903335 24.28008426 26.40227532
 20.72969288 23.83362707 28.77532088 21.24148958 21.74124174 30.56625021
 26.62647993 24.58433714 20.5094913  22.2297617  21.34967357 20.54466055
 26.49065834 23.25671849 20.727176   20.72969288 28.19195583 21.40645839
 20.46185328 23.20765229 22.44445871 27.02530277 20.54466055 20.54466055
 20.58327426 34.07144471 21.74124174 21.64053333 20.54466055 32.31068273
 23.17592999 23.89569465 20.0269901  20.20235412 26.52651009 20.068471
 20.63204969 21.88850713 29.1571092  21.3953349  21.73884197 20.54466055
 20.8069197  21.82715601 21.02912952 21.68592586 21.3953349  20.86046649
 21.02912952 21.70771866 28.71425097 22.54979452 20.73306515 20.81073468
 21.74124174]
50    P
Name: label, dtype: object
4.696394914598546
[26.39535438 28.60495574 25.01106399 24.26408646 24.05532968 22.53125101
 19.6004051  21.84121904 21.84121904 21.84121904 20.41041347 29.35656166
 23.1938237  29.89607223 24.19973265 20.30839991 23.31053966 18.60640196
 24.36480241 20.11704965 25.05360637 19.18311584 20.41041347 21.51172109
 20.64404029 17.8533204  19.67599866 20.06086888 18.25247732 19.01394471
 27.03907002 28.00371319 25.26261065 19.72712903 17.97289729 23.91753601
 18.88784057 19.09573149 19.32934434 26.96765307 21.03980448 20.84058898
```

Illustration 5.2.7: k-Nearest Neighbour Algorithm results

The number on top is where the number of row data testing is inside data training and the side of it is its label. Below it is the value of data testing and the like-table number is every value of data training. If the value of training is near testing value, it declared as its name. .