Assignment-CA675

October 26, 2021

1 TF/IDF on data stored on Hive Assignment

CA675

Author : Shubham Rai Student Number : 21261161

1.1 Cleaning Data

[ ]:
```
import pandas as pd
df = pd.read_csv('stackexchange_export.csv', quotechar='"')
# Remove HTML Tags
df['Body'] = df['Body'].str.replace(r'<[^<>]*>', '', regex=True)
# Keep letters
df["Body"] = df['Body'].str.replace('[^a-zA-Z]', ' ', regex=True)
# Remove new line characters
df["Body"] = df['Body'].replace(r'\n',' ', regex=True)
# Export File to be imported in Hive
df.to_csv('/home/cypherhonk/cleaned_final.csv', sep=',',␣
 ↪encoding='utf-8',index=None)
! chmod 755 cleaned_final.csv
```

1.2 Get connection to Hive

[81]:
```
from pyhive import hive
from tabulate import tabulate
import pandas as pd


host_name = "localhost"
port = 10000
user = "cypherhonk"
password = "9414161568666692861"
database="default"

def hiveconnection(host_name, port, user,password, database):
```

```
    conn = hive.Connection(host=host_name, port=port, username=user,␣
 ↪password=password,
                           database=database, auth='CUSTOM')
    return conn


conn = hiveconnection(host_name, port, user,password, database)
cur = conn.cursor()

## Usage example from https://github.com/dropbox/PyHive
```

2 Task 2 & 3

2.0.1 2.2.1) Querying top 10 posts by score

[13]:
```
cur.execute('select ID, Title, Score, DisplayName from stackexchange_view order␣
 ↪by score desc limit 10')
result = cur.fetchall()
print(tabulate(result, tablefmt='orgtbl'))
```

```
| 11227809 | Why is processing a sorted array faster than processing an unsorted
array? | 25933 | GManNickG       |
|   927358 | How do I undo the most recent local commits in Git?
| 23348 | Hamza Yerlikaya |
|  2003505 | How do I delete a Git branch locally and remotely?
| 18514 | Matthew Rankin  |
|   292357 | What is the difference between 'git pull' and 'git fetch'?
| 12834 | pupeno          |
|   231767 | What does the "yield" keyword do?
| 11551 | Alex. S.        |
|   477816 | What is the correct JSON content type?
| 10921 | Oli             |
|   348170 | How do I undo 'git add' before commit?
| 10079 | paxos1977       |
|  5767325 | How can I remove a specific item from an array?
|  9931 | Walker          |
|  6591213 | How do I rename a local Git branch?
|  9792 | Forrest         |
|  1642028 | What is the "-->" operator in C/C++?
|  9560 | GManNickG       |
```

2.0.2 2.2.2) The top 10 users by post score

[34]:
```
### Join data as usernames data was extracted and added post data pulling

cur.execute("""
    select
           OwnerUserId,
           DisplayName,
           sum(Score) as score
    from stackexchange_view
    group by OwnerUserId,DisplayName
    order by score desc
    LIMIT 10

""")
result = cur.fetchall()
print(tabulate(result, tablefmt='orgtbl'))
```

```
|  87234 | GManNickG       | 37672 |
|   4883 | readonly        | 28817 |
|   9951 | e-satis         | 26878 |
|   6068 | pupeno          | 25944 |
|  89904 | Hamza Yerlikaya | 24024 |
|  51816 | Joan Venge      | 23763 |
|  49153 | Ali             | 20203 |
| 179736 | TIMEX           | 19603 |
|  95592 | Matthew Rankin  | 19479 |
|  63051 | flybywire       | 19362 |
```

[20]:
```
top_10_user_score = []
for each_item in result:
    top_10_user_score.append(each_item[0])
top_10_user_score
```

[20]:
```
[87234, 4883, 9951, 6068, 89904, 51816, 49153, 179736, 95592, 63051]
```

2.0.3 2.2.3) The number of distinct users, who used the word “cloud” in one of their posts

[36]:
```
cur.execute("""
SELECT
     COUNT(DISTINCT owneruserid) as user_count
FROM stackexchange_view
WHERE title LIKE '% cloud %' or Body LIKE '% cloud %'
""")
result = cur.fetchall()
```

```
print(tabulate(result, tablefmt='orgtbl'))
```

| 248 |

2.1 TF/IDF

[74]:
```
### Get data for top 10 users from above list and pull data

df = pd.read_sql(f"""
SELECT

       DISTINCT owneruserid,
       displayname,
       title,
       body
from stackexchange_view
WHERE owneruserid IN {tuple(top_10_user_score)}
order by owneruserid""", conn)
```

[83]:
```
# Attach body and title fields to create one field to be
df["text"] = df["title"] + df["body"]
top_10_username = list(df["displayname"].unique())
top_10_username
```

[83]:
```
['readonly',
 'pupeno',
 'e-satis',
 'Ali',
 'Joan Venge',
 'flybywire',
 'GManNickG',
 'Hamza Yerlikaya',
 'Matthew Rankin',
 'TIMEX']
```

[76]:
```
## References taken and custom function created.
## Reference : https://medium.com/@cmukesh8688/
 ↪tf-idf-vectorizer-scikit-learn-dbc0244a911a

from sklearn.feature_extraction.text import TfidfVectorizer


# Calculate sum() of TF-IDF and get top 10 words with highest TF-IDF and select␣
 ↪only those columns
def calculate_tf_idf(df):
```

```
    vectorizer = TfidfVectorizer(stop_words='english', lowercase=True) # Remove␣
↪Stop Words
    response = vectorizer.fit_transform(df["text"]) # Use title field for TF/IDF
    df_tfidf_sklearn = pd.DataFrame(response.toarray(),columns=vectorizer.
↪get_feature_names())
    total_tf_idf = df_tfidf_sklearn.sum(axis = 0) # Remove sum of TF/IDF for␣
↪getting top 10 most used words
    top_10_list = total_tf_idf.nlargest(10) # Get top 10 words per user
    top_10_words = list(top_10_list.index) # Get list of top 10 words
    df_tfidf_sklearn[top_10_words] # Select only top 10 words as column
    return df_tfidf_sklearn[top_10_words]
```

2.2 For each of top 10 users and their top 10 words plot TF/IDF table for their correspnding comments

[80]:
```
## For all 10 users create TF-IDF table for each user for their respective␣
↪comments.

for each_user in top_10_username: ## Iterate over all top 10 users
    filtered_data = df[(df['displayname']==each_user)] # Filter data only for␣
↪selected user
    tf_idf_df = calculate_tf_idf(filtered_data) # pass on to above function
    print("For Username ID TF/IDF table : "+each_user)
    tf_idf_df.insert(0, 'usernameid', each_user)# attach username ID field to␣
↪dataframe
    display(tf_idf_df)
  ␣
↪print("**********************************************************************************\n\n\n
```

```
For Username ID TF/IDF table : readonly
```

	usernameid	python	use	list	table	branch	way	\
0	readonly	0.000000	0.000000	0.000000	0.000000	0.636117	0.000000	
1	readonly	0.108905	0.239666	0.428552	0.000000	0.000000	0.000000	
2	readonly	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
3	readonly	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
4	readonly	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
5	readonly	0.156136	0.000000	0.000000	0.000000	0.000000	0.000000	
6	readonly	0.000000	0.000000	0.000000	0.000000	0.000000	0.179925	
7	readonly	0.000000	0.000000	0.000000	0.521156	0.000000	0.000000	
8	readonly	0.127064	0.000000	0.000000	0.000000	0.000000	0.000000	
9	readonly	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
10	readonly	0.000000	0.120988	0.000000	0.310520	0.000000	0.120988	

11	readonly	0.000000	0.213146	0.000000	0.000000	0.000000	0.000000
12	readonly	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
13	readonly	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
14	readonly	0.000000	0.000000	0.000000	0.000000	0.402695	0.114986
15	readonly	0.000000	0.202521	0.000000	0.259890	0.000000	0.101261
16	readonly	0.073346	0.000000	0.000000	0.000000	0.000000	0.000000
17	readonly	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
18	readonly	0.167346	0.000000	0.439014	0.000000	0.000000	0.184138
19	readonly	0.000000	0.000000	0.137121	0.000000	0.000000	0.000000
20	readonly	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
21	readonly	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
22	readonly	0.276282	0.000000	0.000000	0.000000	0.000000	0.152003
23	readonly	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
24	readonly	0.000000	0.177000	0.000000	0.000000	0.000000	0.000000
25	readonly	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
26	readonly	0.000000	0.191672	0.000000	0.000000	0.000000	0.000000
27	readonly	0.203719	0.112081	0.000000	0.000000	0.000000	0.000000
28	readonly	0.092270	0.000000	0.000000	0.000000	0.000000	0.000000
29	readonly	0.000000	0.000000	0.202044	0.000000	0.000000	0.169489
30	readonly	0.202829	0.000000	0.000000	0.000000	0.000000	0.000000
31	readonly	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
32	readonly	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000

	difference	git	process	rename
0	0.000000	0.202994	0.000000	0.000000
1	0.000000	0.000000	0.000000	0.000000
2	0.000000	0.000000	0.000000	0.000000
3	0.000000	0.000000	0.350231	0.000000
4	0.000000	0.203342	0.000000	0.000000
5	0.220471	0.000000	0.000000	0.000000
6	0.000000	0.000000	0.000000	0.000000
7	0.000000	0.000000	0.000000	0.568902
8	0.000000	0.000000	0.000000	0.000000
9	0.000000	0.000000	0.000000	0.000000
10	0.000000	0.000000	0.000000	0.338969
11	0.000000	0.238205	0.000000	0.000000
12	0.000000	0.154251	0.000000	0.000000
13	0.000000	0.000000	0.000000	0.000000
14	0.000000	0.192758	0.000000	0.000000
15	0.000000	0.000000	0.389835	0.000000
16	0.000000	0.000000	0.000000	0.000000
17	0.000000	0.000000	0.000000	0.000000
18	0.000000	0.000000	0.000000	0.000000
19	0.000000	0.000000	0.000000	0.000000
20	0.000000	0.000000	0.000000	0.000000
21	0.000000	0.000000	0.000000	0.000000
22	0.000000	0.000000	0.000000	0.000000
23	0.000000	0.000000	0.000000	0.000000

```
24    0.000000  0.000000  0.000000  0.000000
25    0.000000  0.000000  0.187558  0.000000
26    0.491935  0.000000  0.000000  0.000000
27    0.287660  0.000000  0.000000  0.000000
28    0.000000  0.000000  0.000000  0.000000
29    0.000000  0.000000  0.000000  0.000000
30    0.000000  0.000000  0.000000  0.000000
31    0.000000  0.000000  0.000000  0.000000
32    0.000000  0.000000  0.000000  0.000000
```

****

For Username ID TF/IDF table : pupeno

	usernameid	file	java	android	sql	git	way	\
0	pupeno	0.000000	0.000000	0.000000	0.000000	0.000000	0.084849	
1	pupeno	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
2	pupeno	0.000000	0.000000	0.000000	0.000000	0.000000	0.085361	
3	pupeno	0.108337	0.000000	0.000000	0.000000	0.000000	0.000000	
4	pupeno	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
5	pupeno	0.187244	0.000000	0.000000	0.277057	0.000000	0.109712	
6	pupeno	0.116987	0.000000	0.000000	0.692403	0.000000	0.000000	
7	pupeno	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
8	pupeno	0.000000	0.000000	0.130150	0.000000	0.000000	0.000000	
9	pupeno	0.000000	0.000000	0.000000	0.105631	0.000000	0.000000	
10	pupeno	0.000000	0.000000	0.275010	0.000000	0.000000	0.000000	
11	pupeno	0.312392	0.000000	0.000000	0.000000	0.000000	0.122027	
12	pupeno	0.000000	0.000000	0.352152	0.000000	0.000000	0.000000	
13	pupeno	0.082427	0.366784	0.000000	0.000000	0.000000	0.000000	
14	pupeno	0.000000	0.000000	0.000000	0.000000	0.000000	0.081616	
15	pupeno	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
16	pupeno	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
17	pupeno	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
18	pupeno	0.000000	0.000000	0.000000	0.000000	0.143193	0.088019	
19	pupeno	0.000000	0.531605	0.000000	0.000000	0.000000	0.000000	
20	pupeno	0.371197	0.000000	0.000000	0.000000	0.000000	0.000000	
21	pupeno	0.000000	0.071220	0.000000	0.000000	0.000000	0.050015	
22	pupeno	0.000000	0.000000	0.000000	0.000000	0.000000	0.079078	
23	pupeno	0.000000	0.000000	0.000000	0.000000	0.000000	0.140540	
24	pupeno	0.000000	0.362108	0.000000	0.000000	0.000000	0.000000	
25	pupeno	0.000000	0.000000	0.174881	0.000000	0.000000	0.000000	
26	pupeno	0.656317	0.000000	0.000000	0.000000	0.417075	0.085457	
27	pupeno	0.000000	0.000000	0.000000	0.000000	0.000000	0.026952	
28	pupeno	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
29	pupeno	0.000000	0.364640	0.000000	0.000000	0.000000	0.000000	
30	pupeno	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	

31	pupeno	0.000000	0.000000	0.000000	0.000000	0.000000	0.083149
32	pupeno	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
33	pupeno	0.000000	0.000000	0.000000	0.000000	0.728914	0.000000
34	pupeno	0.000000	0.000000	0.151689	0.000000	0.000000	0.000000
35	pupeno	0.000000	0.000000	0.337085	0.000000	0.000000	0.125140
36	pupeno	0.000000	0.000000	0.000000	0.219137	0.000000	0.000000

	like	application	data	dump
0	0.000000	0.000000	0.000000	0.000000
1	0.021477	0.029489	0.000000	0.000000
2	0.265579	0.000000	0.000000	0.000000
3	0.000000	0.000000	0.123813	0.000000
4	0.000000	0.000000	0.000000	0.000000
5	0.000000	0.000000	0.093622	0.000000
6	0.094784	0.000000	0.116987	0.323682
7	0.089201	0.000000	0.000000	0.000000
8	0.000000	0.000000	0.000000	0.000000
9	0.072300	0.000000	0.267708	0.617252
10	0.000000	0.436139	0.261367	0.000000
11	0.126552	0.000000	0.000000	0.000000
12	0.000000	0.000000	0.000000	0.000000
13	0.066783	0.000000	0.000000	0.000000
14	0.000000	0.000000	0.000000	0.000000
15	0.000000	0.026204	0.000000	0.000000
16	0.000000	0.000000	0.000000	0.000000
17	0.000000	0.000000	0.095551	0.000000
18	0.091283	0.000000	0.000000	0.000000
19	0.064529	0.177202	0.000000	0.000000
20	0.000000	0.000000	0.000000	0.000000
21	0.000000	0.000000	0.000000	0.000000
22	0.164020	0.000000	0.000000	0.000000
23	0.000000	0.000000	0.000000	0.000000
24	0.000000	0.000000	0.000000	0.000000
25	0.000000	0.000000	0.000000	0.000000
26	0.000000	0.000000	0.000000	0.000000
27	0.055903	0.000000	0.000000	0.000000
28	0.000000	0.000000	0.000000	0.000000
29	0.000000	0.000000	0.000000	0.000000
30	0.000000	0.000000	0.000000	0.000000
31	0.000000	0.000000	0.000000	0.000000
32	0.032758	0.000000	0.040432	0.000000
33	0.000000	0.000000	0.000000	0.000000
34	0.000000	0.000000	0.000000	0.000000
35	0.000000	0.356390	0.000000	0.000000
36	0.000000	0.000000	0.000000	0.000000

****

For Username ID TF/IDF table : e-satis

	usernameid	python	git	head	like	does	using	\
0	e-satis	0.077636	0.000000	0.000000	0.131321	0.155273	0.000000	
1	e-satis	0.000000	0.000000	0.000000	0.000000	0.000000	0.236308	
2	e-satis	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
3	e-satis	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
4	e-satis	0.000000	0.000000	0.000000	0.101237	0.119702	0.000000	
5	e-satis	0.000000	0.000000	0.000000	0.000000	0.227740	0.000000	
6	e-satis	0.000000	0.070489	0.000000	0.059616	0.000000	0.000000	
7	e-satis	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
8	e-satis	0.000000	0.000000	0.000000	0.034854	0.000000	0.164845	
9	e-satis	0.000000	0.075851	0.000000	0.000000	0.000000	0.151701	
10	e-satis	0.000000	0.421761	0.349226	0.000000	0.140587	0.000000	
11	e-satis	0.438262	0.000000	0.000000	0.000000	0.000000	0.000000	
12	e-satis	0.080333	0.000000	0.000000	0.203822	0.000000	0.080333	
13	e-satis	0.000000	0.000000	0.000000	0.000000	0.000000	0.308775	
14	e-satis	0.000000	0.000000	0.000000	0.169226	0.000000	0.000000	
15	e-satis	0.000000	0.000000	0.000000	0.155518	0.000000	0.000000	
16	e-satis	0.106392	0.000000	0.000000	0.224950	0.000000	0.000000	
17	e-satis	0.000000	0.000000	0.134811	0.000000	0.000000	0.000000	
18	e-satis	0.525068	0.000000	0.000000	0.000000	0.000000	0.000000	
19	e-satis	0.000000	0.000000	0.000000	0.000000	0.384412	0.000000	
20	e-satis	0.000000	0.188113	0.000000	0.000000	0.062704	0.000000	
21	e-satis	0.000000	0.000000	0.000000	0.000000	0.000000	0.118825	
22	e-satis	0.000000	0.188162	0.701109	0.000000	0.000000	0.000000	
23	e-satis	0.109937	0.000000	0.000000	0.092978	0.000000	0.000000	
24	e-satis	0.000000	0.338695	0.000000	0.000000	0.000000	0.000000	

	use	gt	way	know
0	0.000000	0.000000	0.000000	0.069198
1	0.000000	0.000000	0.000000	0.000000
2	0.000000	0.000000	0.064273	0.000000
3	0.000000	0.000000	0.000000	0.000000
4	0.000000	0.000000	0.000000	0.000000
5	0.113870	0.000000	0.000000	0.000000
6	0.000000	0.000000	0.000000	0.062828
7	0.000000	0.000000	0.000000	0.000000
8	0.123634	0.000000	0.116494	0.146928
9	0.000000	0.000000	0.000000	0.135212
10	0.140587	0.000000	0.000000	0.000000
11	0.000000	0.000000	0.275301	0.000000
12	0.000000	0.000000	0.000000	0.143203
13	0.000000	0.000000	0.000000	0.000000
14	0.200092	0.000000	0.000000	0.000000

```
15  0.000000  0.000000  0.173264  0.081948
16  0.000000  0.654206  0.100248  0.047414
17  0.000000  0.296631  0.000000  0.000000
18  0.262534  0.000000  0.000000  0.000000
19  0.000000  0.000000  0.144885  0.205578
20  0.000000  0.000000  0.059083  0.000000
21  0.118825  0.000000  0.000000  0.000000
22  0.000000  0.000000  0.000000  0.000000
23  0.000000  0.000000  0.000000  0.000000
24  0.000000  0.000000  0.000000  0.000000
```

****

For Username ID TF/IDF table : Ali

```
   usernameid        gt     array  javascript     using       php        lt  \
0         Ali  0.225908  0.000000    0.000000  0.168649  0.000000  0.243885
1         Ali  0.089387  0.210973    0.245229  0.000000  0.000000  0.000000
2         Ali  0.000000  0.000000    0.349359  0.000000  0.000000  0.000000
3         Ali  0.000000  0.000000    0.000000  0.174570  0.223285  0.000000
4         Ali  0.000000  0.000000    0.000000  0.000000  0.000000  0.000000
..        ...       ...       ...         ...       ...       ...       ...
74        Ali  0.000000  0.000000    0.000000  0.000000  0.000000  0.000000
75        Ali  0.000000  0.000000    0.000000  0.000000  0.000000  0.000000
76        Ali  0.000000  0.000000    0.000000  0.000000  0.000000  0.000000
77        Ali  0.022841  0.000000    0.000000  0.000000  0.000000  0.024659
78        Ali  0.000000  0.000000    0.000000  0.000000  0.000000  0.000000

          id    jquery       key  file
0   0.000000  0.258347  0.000000   0.0
1   0.289498  0.102222  0.339392   0.0
2   0.000000  0.000000  0.000000   0.0
3   0.000000  0.000000  0.000000   0.0
4   0.000000  0.000000  0.000000   0.0
..       ...       ...       ...   ...
74  0.000000  0.000000  0.000000   0.0
75  0.000000  0.000000  0.000000   0.0
76  0.000000  0.000000  0.000000   0.0
77  0.000000  0.000000  0.000000   0.0
78  0.000000  0.000000  0.000000   0.0

[79 rows x 11 columns]
```

****

```
For Username ID TF/IDF table : Joan Venge

    usernameid    python    string      like      list        gt     class  \
0   Joan Venge  0.193545  0.000000  0.000000  0.132596  0.000000  0.000000
1   Joan Venge  0.000000  0.000000  0.000000  0.000000  0.000000  0.000000
2   Joan Venge  0.000000  0.000000  0.051889  0.000000  0.000000  0.000000
3   Joan Venge  0.000000  0.000000  0.000000  0.000000  0.000000  0.000000
4   Joan Venge  0.000000  0.000000  0.000000  0.000000  0.000000  0.358144
..         ...       ...       ...       ...       ...       ...       ...
61  Joan Venge  0.000000  0.000000  0.153097  0.000000  0.000000  0.128018
62  Joan Venge  0.000000  0.000000  0.000000  0.169950  0.000000  0.000000
63  Joan Venge  0.344646  0.000000  0.000000  0.000000  0.000000  0.000000
64  Joan Venge  0.000000  0.000000  0.000000  0.100046  0.328351  0.000000
65  Joan Venge  0.000000  0.121922  0.079756  0.000000  0.000000  0.000000

        want  value  function    index
0   0.000000    0.0       0.0  0.00000
1   0.000000    0.0       0.0  0.00000
2   0.071552    0.0       0.0  0.28987
3   0.000000    0.0       0.0  0.00000
4   0.000000    0.0       0.0  0.00000
..       ...    ...       ...      ...
61  0.000000    0.0       0.0  0.00000
62  0.153303    0.0       0.0  0.20702
63  0.000000    0.0       0.0  0.00000
64  0.000000    0.0       0.0  0.00000
65  0.000000    0.0       0.0  0.00000

[66 rows x 11 columns]

**************************************************************************


For Username ID TF/IDF table : flybywire

   usernameid      file    python      want        vs       use      java  \
0   flybywire  0.000000  0.124444  0.000000  0.000000  0.000000  0.000000
1   flybywire  0.000000  0.000000  0.000000  0.000000  0.000000  0.000000
2   flybywire  0.099270  0.111844  0.000000  0.000000  0.000000  0.000000
3   flybywire  0.000000  0.000000  0.000000  0.000000  0.000000  0.000000
4   flybywire  0.000000  0.119751  0.000000  0.000000  0.000000  0.000000
5   flybywire  0.000000  0.000000  0.000000  0.228055  0.000000  0.000000
6   flybywire  0.000000  0.000000  0.116217  0.000000  0.000000  0.000000
7   flybywire  0.000000  0.000000  0.000000  0.000000  0.000000  0.000000
8   flybywire  0.000000  0.000000  0.123087  0.000000  0.000000  0.000000
```

9	flybywire	0.000000	0.000000	0.112601	0.000000	0.000000	0.000000
10	flybywire	0.175813	0.000000	0.082066	0.000000	0.000000	0.000000
11	flybywire	0.000000	0.000000	0.000000	0.000000	0.000000	0.657184
12	flybywire	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
13	flybywire	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
14	flybywire	0.401779	0.000000	0.000000	0.000000	0.158003	0.000000
15	flybywire	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
16	flybywire	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
17	flybywire	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
18	flybywire	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
19	flybywire	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
20	flybywire	0.236095	0.000000	0.110204	0.000000	0.000000	0.000000
21	flybywire	0.000000	0.146583	0.000000	0.153493	0.306986	0.000000
22	flybywire	0.000000	0.321788	0.000000	0.000000	0.000000	0.000000
23	flybywire	0.076939	0.000000	0.287305	0.000000	0.000000	0.000000
24	flybywire	0.241039	0.000000	0.000000	0.000000	0.000000	0.000000
25	flybywire	0.071814	0.000000	0.067043	0.000000	0.000000	0.000000
26	flybywire	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
27	flybywire	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
28	flybywire	0.171867	0.000000	0.160447	0.000000	0.000000	0.000000
29	flybywire	0.000000	0.000000	0.086403	0.000000	0.054596	0.000000
30	flybywire	0.000000	0.000000	0.000000	0.246946	0.493891	0.000000
31	flybywire	0.000000	0.000000	0.000000	0.255143	0.000000	0.000000
32	flybywire	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
33	flybywire	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
34	flybywire	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
35	flybywire	0.315679	0.000000	0.098234	0.000000	0.000000	0.000000
36	flybywire	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
37	flybywire	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
38	flybywire	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
39	flybywire	0.000000	0.000000	0.000000	0.000000	0.000000	0.142146
40	flybywire	0.000000	0.000000	0.000000	0.271833	0.135916	0.484679
41	flybywire	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
42	flybywire	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
43	flybywire	0.000000	0.487702	0.000000	0.000000	0.000000	0.000000
44	flybywire	0.000000	0.330355	0.000000	0.000000	0.000000	0.000000
45	flybywire	0.000000	0.000000	0.000000	0.217766	0.000000	0.000000
46	flybywire	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
47	flybywire	0.297907	0.000000	0.139056	0.000000	0.175732	0.000000
48	flybywire	0.000000	0.000000	0.069671	0.000000	0.000000	0.000000

	gt	standard	instance	output
0	0.000000	0.000000	0.000000	0.000000
1	0.000000	0.000000	0.000000	0.000000
2	0.000000	0.000000	0.000000	0.000000
3	0.000000	0.000000	0.000000	0.000000
4	0.239501	0.000000	0.000000	0.000000
5	0.000000	0.000000	0.000000	0.000000

6	0.280514	0.000000	0.000000	0.000000
7	0.000000	0.000000	0.000000	0.000000
8	0.000000	0.000000	0.000000	0.000000
9	0.000000	0.000000	0.000000	0.000000
10	0.198083	0.000000	0.000000	0.000000
11	0.000000	0.000000	0.000000	0.000000
12	0.000000	0.000000	0.000000	0.000000
13	0.000000	0.000000	0.864911	0.000000
14	0.301781	0.500837	0.000000	0.293212
15	0.000000	0.000000	0.000000	0.000000
16	0.000000	0.217125	0.000000	0.000000
17	0.000000	0.000000	0.000000	0.000000
18	0.000000	0.000000	0.000000	0.000000
19	0.000000	0.000000	0.000000	0.000000
20	0.000000	0.000000	0.000000	0.000000
21	0.000000	0.000000	0.000000	0.000000
22	0.000000	0.000000	0.000000	0.000000
23	0.000000	0.000000	0.000000	0.000000
24	0.000000	0.000000	0.000000	0.000000
25	0.000000	0.000000	0.000000	0.283009
26	0.000000	0.000000	0.000000	0.000000
27	0.000000	0.000000	0.000000	0.000000
28	0.000000	0.482042	0.000000	0.225767
29	0.052138	0.000000	0.000000	0.364736
30	0.000000	0.000000	0.000000	0.000000
31	0.000000	0.000000	0.000000	0.000000
32	0.000000	0.000000	0.317777	0.000000
33	0.000000	0.000000	0.000000	0.000000
34	0.000000	0.000000	0.000000	0.000000
35	0.000000	0.000000	0.000000	0.000000
36	0.000000	0.000000	0.000000	0.000000
37	0.000000	0.000000	0.000000	0.000000
38	0.000000	0.000000	0.000000	0.000000
39	0.114201	0.000000	0.000000	0.000000
40	0.000000	0.000000	0.000000	0.000000
41	0.000000	0.000000	0.000000	0.000000
42	0.000000	0.000000	0.000000	0.000000
43	0.000000	0.000000	0.000000	0.000000
44	0.000000	0.000000	0.000000	0.000000
45	0.000000	0.000000	0.000000	0.000000
46	0.000000	0.000000	0.000000	0.000000
47	0.000000	0.000000	0.000000	0.000000
48	0.084083	0.000000	0.000000	0.000000

****

For Username ID TF/IDF table : GManNickG

	usernameid	lt	copy	data	int	idiom	gt	\
0	GManNickG	0.000000	0.523691	0.00000	0.000000	0.418953	0.000000	
1	GManNickG	0.126037	0.000000	0.00000	0.252073	0.000000	0.252073	
2	GManNickG	0.431818	0.000000	0.51101	0.172727	0.000000	0.129546	

	quot	swap	arraysize	array
0	0.000000	0.314215	0.000000	0.079656
1	0.252073	0.000000	0.000000	0.000000
2	0.086364	0.000000	0.283895	0.194318

****

For Username ID TF/IDF table : Hamza Yerlikaya

	usernameid	file	timer	java	new	table	\
0	Hamza Yerlikaya	0.470020	0.000000	0.000000	0.000000	0.000000	
1	Hamza Yerlikaya	0.066477	0.000000	0.053078	0.379920	0.000000	
2	Hamza Yerlikaya	0.356757	0.000000	0.000000	0.000000	0.000000	
3	Hamza Yerlikaya	0.000000	0.000000	0.000000	0.000000	0.000000	
4	Hamza Yerlikaya	0.000000	0.000000	0.156744	0.000000	0.000000	
5	Hamza Yerlikaya	0.000000	0.000000	0.026395	0.000000	0.533399	
6	Hamza Yerlikaya	0.000000	0.744404	0.073674	0.210937	0.000000	
7	Hamza Yerlikaya	0.000000	0.000000	0.000000	0.000000	0.000000	
8	Hamza Yerlikaya	0.000000	0.000000	0.284629	0.000000	0.000000	
9	Hamza Yerlikaya	0.000000	0.000000	0.000000	0.000000	0.000000	

	bits	byte	iterate	application	files
0	0.000000	0.000000	0.000000	0.268619	0.104470
1	0.000000	0.000000	0.000000	0.000000	0.000000
2	0.000000	0.000000	0.000000	0.000000	0.211455
3	0.000000	0.000000	0.000000	0.000000	0.084239
4	0.527915	0.527915	0.527915	0.000000	0.000000
5	0.000000	0.000000	0.000000	0.000000	0.000000
6	0.000000	0.000000	0.000000	0.000000	0.000000
7	0.000000	0.000000	0.000000	0.254880	0.099127
8	0.000000	0.000000	0.000000	0.000000	0.000000
9	0.000000	0.000000	0.000000	0.000000	0.000000

****

For Username ID TF/IDF table : Matthew Rankin

```
       usernameid   install       pip    python        lt        gt    branch  \
0  Matthew Rankin  0.000000  0.000000  0.000000  0.000000  0.000000  0.000000
1  Matthew Rankin  0.000000  0.000000  0.358589  0.000000  0.000000  0.000000
2  Matthew Rankin  0.000000  0.000000  0.000000  0.000000  0.043561  0.571753
3  Matthew Rankin  0.427251  0.284834  0.000000  0.000000  0.000000  0.000000
4  Matthew Rankin  0.000000  0.000000  0.000000  0.650278  0.544984  0.000000
5  Matthew Rankin  0.235094  0.211585  0.094038  0.000000  0.000000  0.000000
6  Matthew Rankin  0.000000  0.000000  0.000000  0.000000  0.000000  0.000000
7  Matthew Rankin  0.349045  0.319958  0.319958  0.000000  0.000000  0.000000

      flask   version  debugger   started
0  0.555725  0.555725  0.000000  0.000000
1  0.000000  0.000000  0.495842  0.495842
2  0.000000  0.000000  0.000000  0.000000
3  0.000000  0.000000  0.000000  0.000000
4  0.000000  0.000000  0.000000  0.000000
5  0.000000  0.000000  0.000000  0.000000
6  0.000000  0.000000  0.000000  0.000000
7  0.000000  0.000000  0.000000  0.000000
```

****

```
For Username ID TF/IDF table : TIMEX

    usernameid    python      want  user    string        gt  dictionary  \
0        TIMEX  0.000000  0.000000   0.0  0.000000  0.000000         0.0
1        TIMEX  0.000000  0.000000   0.0  0.000000  0.000000         0.0
2        TIMEX  0.000000  0.130509   0.0  0.000000  0.179279         0.0
3        TIMEX  0.000000  0.000000   0.0  0.000000  0.000000         0.0
4        TIMEX  0.000000  0.000000   0.0  0.000000  0.000000         0.0
..         ...       ...       ...   ...       ...       ...         ...
110      TIMEX  0.113904  0.000000   0.0  0.000000  0.000000         0.0
111      TIMEX  0.000000  0.031423   0.0  0.090197  0.000000         0.0
112      TIMEX  0.000000  0.000000   0.0  0.000000  0.000000         0.0
113      TIMEX  0.000000  0.000000   0.0  0.000000  0.423683         0.0
114      TIMEX  0.000000  0.000000   0.0  0.000000  0.000000         0.0

           lt       use  javascript    return
0    0.000000  0.000000     0.00000  0.000000
1    0.000000  0.000000     0.00000  0.000000
2    0.187309  0.000000     0.00000  0.000000
3    0.000000  0.000000     0.00000  0.000000
4    0.000000  0.000000     0.17604  0.000000
..        ...       ...         ...       ...
110  0.000000  0.000000     0.00000  0.000000
111  0.000000  0.000000     0.00000  0.048591
```

```
112  0.000000  0.000000     0.00000  0.000000
113  0.000000  0.076049     0.00000  0.000000
114  0.000000  0.000000     0.00000  0.000000

[115 rows x 11 columns]
```

****