

MINI-PROJECT REPORT

*Submitted in partial fulfilment of the requirements for the award of
the degree of Master of Computer Applications under National
Institute of Technology Karnataka*

By

Piyush Gupta (214CA037)

Sumit rajput (214CA058)

Abhijeet Sabu (214CA02)

Rajkumar Sharma (214CA043)

Under the Guidance of

Dr. Pushparaj Shetty



**DEPARTMENT OF MATHEMATICAL AND
COMPUTATIONAL SCIENCE**

APRIL – 2023

Table of Contents

ABSTRACT

INTRODUCTION

METHODOLOGY

EXPERIMENTATION AND QUERY

- a) Highest cited author and his h-index (from the world)
- b) Highest publication author
- c) Highest cited authors avg. citations, and the country name.
- d) Total number of publications of the highest cited author.....
- e) Total publication in year.....
- f) Total citation per year
- g) Author(country) having highest co-authorship with Indian authors.
- h) Highest cited author from India and the university.....
- i) Comparative year wise article publication analysis of India, China and USA
- j) Total number of grants given to the field.....
- k) Country wise total number of publications

REFERANCE.....

.

Introduction

Deep learning is a branch of machine learning which is based on artificial neural network. It is capable of learning complex pattern and relationship within the data. In deep learning we don't need to explicitly program everything. It has become increasingly popular in recent years due to the advance in processing power and the availability of large database. Because it is based on Artificial neural network (ANNs) also known as deep neural networks (DNNs). These neural networks are inspired by the structure and function of the human brains biological neurons, and they are designed to learn from large amount of data.

Deep learning is subfield of machine learning that involves the use of neural network to model and solve complex problem. Neural networks consist of layers of interconnected nodes that process and transform data.

The key characteristic of deep learning is the use of deep neural networks which have multiple layers of interconnected nodes.

Abstract

Deep learning (DL), a branch of machine learning (ML) and artificial intelligence (AI) is nowadays considered as a core technology of today's fourth industrial Revolution. Due to its learning capabilities from data, DL technology originated from artificial neural networks (ANN), has become a hot topic in the context of computing, and is widely applied in various applications areas like healthcare, visual recognition, sentiment analysis, cybersecurity, and many more. However, building an appropriate DL model is a challenging task, due to the dynamic nature and variation in real-world problems and data. Moreover, the lack of core understanding turns DL methods into black-box machine that hamper development at the standard level. This article presents a structured and comprehensive view on DL techniques including a taxonomy considering various types of real-world tasks like supervised or unsupervised. Overall, this article aims to draw a big picture on DL modeling that can be used as a reference guide for both academia and industry professionals.

'Cited by' - It shows how many have referred the author's publication for their works.

'Country' - It represent the country from where the publication is done.

'Sponsor' - It shows whether or not the authors get the sponsorship for their work.

With this information in our complete database, we are in good position to analyze our database the problem statements given to us. We use 'python' and some of the libraries such as 'pandas', 'matplotlib', 'Numpy', 'pickle' etc. for better representation and analysis.

Methodology:

Now we will see the methodology that we have adopted to complete this project.

Step 1: Downloading database from scopus. We downloaded database separately for each country and again for funding details.

Step 2: Preprocessing of Main database:

In this process we removed some unnecessary attributes which are not required for our analysis and then we merge all the separate databases of countries in to one single CSV file.

Step 3 - Generating Necessary Data Required for Further Analysis

In this process we have done operations to generate necessary data for further analysis i.e. to solve questions.

This step includes sub steps:

1. Generating Authors list
2. Creating python dictionary with Author name as key and his/her corresponding database as value.
3. Creating list of Indian authors.
4. Generating Foreign Authors list.
5. Creating a dictionary with foreign author as key and number of papers published by him with Indian authors as value.

Step 4 - Finding answers to given questions.

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import warnings
import seaborn as sns
warnings.filterwarnings('ignore')
```

Downloading Database

As discussed earlier we used Scopus repository to download data base. Scopus repository allows user to download database with maximum 2000 entities directly from their website and for database with more than 2000 entities they send a link provide email address and then one can download the database from that link also in Scopus repository 20000 in Cap on entities that can be downloaded database with more than 20000 entities. also as the number of entities in the selected database increases, the availability of corresponding data attributes decreases. this means that if we select a long database for downloading then we might not receive some of the selected data attributes in the downloaded file. in our case this attribute was funding details. When we select 'funding attribute' for long database then the repository did not give corresponding data in the file.

To overcome this difficulty we downloaded database corresponding to only funding details repeatedly and then merged it to our main database.

Now after including the question that where supposed to answer we calculated that the data should be downloaded Country wise there we downloaded data Country wise there we will be able to add a Counting Column in respective Countries's database and merge those together (row wise) to get main data base.

In Country-wise database we found that if two authors publish a paper together and they are from different Countries then the same paper will be present in database of both Countries. This makes our work a lot simpler because these we can analyze the database 'as per' Country also.

So with all this in mind we used following steps to download the database.

1. Registration on Scopus with Institute email id.
2. Verifying email id and logging on to Scopus.
3. Selecting Search parameters as keyword.
4. Selecting document type as 'Article'.
5. Selecting Subject area as 'Computer Science'.
6. Then we got the query as mentioned.

	Authors	Author(s) ID	Title	Year	Source title	Volume	Issue	Art. No.	Page start	Page end	...	ISBN	CODEN	PubMed ID	Language of Original Document	Abbreviated Source Title	Document Type	Publication Stage	Open Access	Sol
0	Chen L.-C., Papandreou G., Kokkinos I., Murphy...	56118862400;6603242401;9250105400;34875408300;...	DeepLab: Semantic Image Segmentation with Deep...	2018	IEEE Transactions on Pattern Analysis and Mach...	40	4	NaN	834	848	...	NaN	ITPID	28463186.0	English	IEEE Trans Pattern Anal Mach Intell	Article	Final	All Open Access, Green	Sco
1	Shelhamer E., Long J., Darrell T.	56433480600;55457086300;7003377605;	Fully Convolutional Networks for Semantic Segm...	2017	IEEE Transactions on Pattern Analysis and Mach...	39	4	7478072	640	651	...	NaN	ITPID	27244717.0	English	IEEE Trans Pattern Anal Mach Intell	Article	Final	All Open Access, Green	Sco
2	Zhang K., Zuo W., Chen Y., Meng D., Zhang L.	57169173100;56888903800;55902679100;2339305840...	Beyond a Gaussian denoiser: Residual learning ...	2017	IEEE Transactions on Image Processing	26	7	7839189	3142	3155	...	NaN	IIPRE	28166495.0	English	IEEE Trans Image Process	Article	Final	All Open Access, Green	Sco
3	Ganin Y., Ustinova E., Ajakan H., Germain P., ...	56938634700;57190427124;57190423376;3487469480...	Domain-adversarial training of neural networks	2016	Journal of Machine Learning Research	17	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	English	J. Mach. Learn. Res.	Article	Final	NaN	Sco
4	Shorten C., Khoshgoftaar T.M.	57209776315;7006211475;	A survey on Image Data Augmentation for Deep L...	2019	Journal of Big Data	6	1	60	NaN	NaN	...	NaN	NaN	NaN	English	J. Big Data	Article	Final	All Open Access, Gold	Sco

Data Pre-processing

A. Preprocessing of main Database.

As discussed in earlier Chapter we had downloaded database separately for each Country and again for funding details. Also the repeated database had some unnecessary attribute. So these database cannot be used directly for data analysis that's why we have to do preprocessing on those.

For data preprocessing and also data analysis we used Python programming language. Python language has many libraries which means data analysis much easier. We improved the database of each Countries using 'read_csv()' function of pandas libraries and did preprocessing on it. Following is the block of each code that we used to do so.

1. Reading the csv files

(This base_dir Variable contains a string which path to database directly.

Using the block of code merged filename extension in database directly with base_dir variable and stored the resultant string in a python list.

We are doing this for every file in database directory and appending it to python list. Thus the python list will contain string which are paths of database of each Country.

After getting python to databases we imported those using pandas read-csv() function and removed unnecessary columns from them using drop() function. Then we also added a Country column in respective data base containing corresponding Country name.

The shape of data was like:-

2. Removing the uncounted attributes:-

We removed the following columns/attributes 'Author(s)', 'ID', 'Source title', 'Volume', 'Issue', 'Art No.', 'Page start', 'Page end', 'Page Count', 'DOI', 'Link', 'Document type', 'Publication stage', 'Open Access', 'Sources', 'ESD',

Also we filled all empty cells in database with 0. For doing all this we have used following block of code.

```
df.columns
```

```
Index(['Authors', 'Author(s) ID', 'Title', 'Year', 'Source title', 'Volume',  
      'Issue', 'Art. No.', 'Page start', 'Page end', 'Page count', 'Cited by',  
      'DOI', 'Link', 'Affiliations', 'Authors with affiliations',  
      'Correspondence Address', 'Editors', 'Publisher', 'ISSN', 'ISBN',  
      'CODEN', 'PubMed ID', 'Language of Original Document',  
      'Abbreviated Source Title', 'Document Type', 'Publication Stage',  
      'Open Access', 'Source', 'EID'],  
      dtype='object')
```

```
pre_process=df[['Authors','Authors with affiliations','Cited by','Year']]
```

```
pre_process.head()
```

	Authors	Authors with affiliations	Cited by	Year
0	Chen L.-C., Papandreou G., Kokkinos I., Murphy...	Chen, L.-C., Google Inc., Mountain View, CA 9...	6528	2018
1	Shelhamer E., Long J., Darrell T.	Shelhamer, E., Department of Electrical Engine...	4086	2017
2	Zhang K., Zuo W., Chen Y., Meng D., Zhang L.	Zhang, K., School of Computer Science and Tech...	3176	2017
3	Ganin Y., Ustinova E., Ajakan H., Germain P., ...	Ganin, Y., Skolkovo Institute of Science and T...	2549	2016
4	Shorten C., Khoshgoftaar T.M.	Shorten, C., Department of Computer and Electr...	1882	2019

Experimentation And Query:

```
In [ ]: pre_process.shape
```

```
Out[ ]: (9992, 4)
```

```
In [ ]: authors=[]
Citation=[]
Year=[]
country=[]
for i in range(9991):
    k=pre_process['Authors with affiliations'][i].split(';')
    for j in range(len(k)):
        m = k[j].split(',')
        if(len(m)>=3):
            auth = m[0]+m[1]
            authors.append(auth.strip())
            con = m[-1]
            country.append(con.strip())
            Year.append(pre_process['Year'][i])
            Citation.append(pre_process['Cited by'][i])
```

```
In [ ]: ACCY = pd.DataFrame(list(zip(authors,country,Citation,Year)),columns=['Authors','Country','Citation','Year'])
```

```
In [ ]: ACCY.head()
```

```
Out[ ]:
```

	Authors	Country	Citation	Year
0	Chowdhury M.E.H. Department of Electrical Engi...	Qatar	639	2020
1	Rahman T. Department of Biomedical Physics and...	Bangladesh	639	2020
2	Khandakar A. Department of Electrical Engineering	Qatar	639	2020
3	Mazhar R. Thoracic Surgery	Qatar	639	2020
4	Kadir M.A. Department of Biomedical Physics an...	Bangladesh	639	2020

a) Highest cited author and his h-index (from the world)

```
In [ ]: # Created a new dataframe 'auth_info' by first using Group By Authors then extracted total citations for i
auth_info = pd.DataFrame(ACCY.groupby('Authors')['Citation'].sum()).reset_index()
auth_info.sort_values(by = 'Citation', ascending=False)
```

```
Out[ ]:
```

	Authors	Citation
36961	Yu P.S. Department of Computer Science	2437
6669	Del Ser J. TECNALIA	2150
28291	Shazeer N. Google	2127
22984	Narang S. Google	2103
17926	Li W. Google	2103
...
10722	Gupta B.B. The Department of Computer Engineering	13
13573	Javeed D. Northeastern University	13
20427	Luo H.-S. National Engineering Research Center...	13
26358	Reghunadhan R. Department of Computer Science	13
38607	Zhang Y. College of Sciences	13

40495 rows × 2 columns

```
In [ ]: auth_info[auth_info['Citation']==auth_info['Citation'].max()]
```

40495 rows × 2 columns

```
In [ ]: auth_info[auth_info['Citation']==auth_info['Citation'].max()]
```

```
Out[ ]:
```

	Authors	Citation
36961	Yu P.S. Department of Computer Science	2437

Highest Cited Auhtor is Yu P.S.with 2437 Citations

```
In [ ]: authors_df=pd.DataFrame({
    'Authors':df['Authors'].str.split(';').explode(),
    'Titles':df['Titles'].repeat(df['Authors'].str.count(';')+1),
    'Year':df['Year'].repeat(df['Authors'].str.count(';')+1),
    'Cited by':df['Cited by'].repeat(df['Authors'].str.count(';')+1),
    'Authors with affiliation':df['Authors with affiliations'].repeat(df['Authors'].str.count(';')+1)
})

authors_df.dropna(subset=['Authors with affiliation'],inplace=True)
authors_df['Country']=authors_df['Authors with affiliation'].apply(lambda x:x.split(',')[1].strip())
```

```
In [ ]: yu_ps=authors_df[authors_df['Authors'].str.contains('Yu.P.S.')]
yu_ps=yu_ps.sort_values(by='Cited by',ascending=False)
yu_ps
```

```
Out[ ]:
```

	Authors	Titles	Year	Cited by	Authors wi
2	Yu P.S.	A Comprehensive Survey on Graph Neural Networks	2021	1906	Wu Z., Centre for Artificial Inte
73	Yu P.S.	A Survey on Knowledge Graphs: Representation, ...	2022	313	Ji S., Department of Computer Scie
329	Yu P.S.	Spatial temporal incidence dynamic graph neura...	2020	128	Peng H., School of Cyber Science an
885	Yu P.S.	Deep Learning for Spatio-Temporal Data Mining:...	2022	79	Wang S., Nanjing University of Aerc
1937	Yu P.S.	A Comprehensive Survey on Community Detection ...	2022	48	Su X., School of Computing, Macqu
3205	Yu P.S.	A Survey of Community Detection Approaches: Fr...	2023	34	Jin D., Tianjin University, Colle
5983	Yu P.S.	Higher-Order Attribute-Enhancing Heterogeneous...	2023	21	Li J., Beihang University, Beijing ,

6037	Yu P.S.	More Than Privacy: Applying Differential Priva...	2022	21	Zhu T., School of Computer Science
6137	Yu P.S.	Deep graph similarity learning: a survey	2021	21	Ma G., Intel Labs, Intel Corporat
6880	Yu P.S.	A Comprehensive Survey of the Key Technologies...	2021	19	Xia Z., Wuhan University, China, X
8182	Yu P.S.	Reinforcement-Learning-Guided Source Code Summ...	2022	16	Wang W., Department of Computer Sci

```
In [ ]: h_index = max([min(i+1, row['Cited by']) for i, row in yu_ps.iterrows()])
print("h-index of Yu P.S who is the highest cited author is : ",h_index)
```

h-index of Yu P.S who is the highest cited author is : 128

h-index of Yu P.S who is the highest cited author is : 128

```
In [ ]:
```

b) Highest publication author

```
In [ ]: auth_pub, count = np.unique(ACCY['Authors'].str.split('D'), return_counts=True)
```

```
In [ ]: auth_pub = list(auth_pub)
count = list(count)
```

```
In [ ]: high_pub = pd.DataFrame({'Author':auth_pub, 'Total Publications':count}).sort_values(by=['Total Publicatio
```

```
In [ ]: high_pub.head()
```

```
Out[ ]:
```

	Author	Total Publications
15141	[Khan M.A. , eapartment of Computer Science]	40
1725	[Acharya U.R. , eapartment of Electronics and C...	22
27514	[Sarkar R. , eapartment of Computer Science and...	19
28132	[Shankar K. , eapartment of Computer Applications]	18
28207	[Sharif M. , eapartment of Computer Science]	15

```
In [ ]: print("Highest Publication Author = ",auth_pub[count.index(max(count))][0])
```

Highest Publication Author = Khan M.A.

Higest Publication Author = Khan M.A.

c) Highest cited authors avg. citations, and the country name.

```
In [ ]: tot_citation = auth_info['Citation'].max()
print('Total Citations Highest cited Author = ',tot_citation)
tot_publications = 112
```

Total Citations Highest cited Author = 2437

Total Citations Highest cited Author = 2437

```
In [ ]: avg_citation = tot_citation/tot_publications
print('Average Citations of Highest Cited Author = ', avg_citation)
```

Average Citations of Highest Cited Author = 21.758928571428573

Average Citations of Highest Cited Author = 21.758928571428573

```
In [ ]: country = ACCY[(ACCY['Authors']=='Yu P.S. Department of Computer Science')]['Country'].iloc[0]
print('Country of Higest Cited Author =',country)
```

Country of Higest Cited Author = United States

Country of Higest Cited Author = United States

d) Total number of publications of the highest cited author

```
In [ ]: tot_pub_hca = ACCY[(ACCY['Authors']=='Yu P.S. Department of Computer Science').shape[0]
print('Total number of publications of the highest cited author = ',tot_pub_hca)
```

Total number of publications of the highest cited author = 6

Total number of publications of the highest cited author = 6

e) Total publication in year

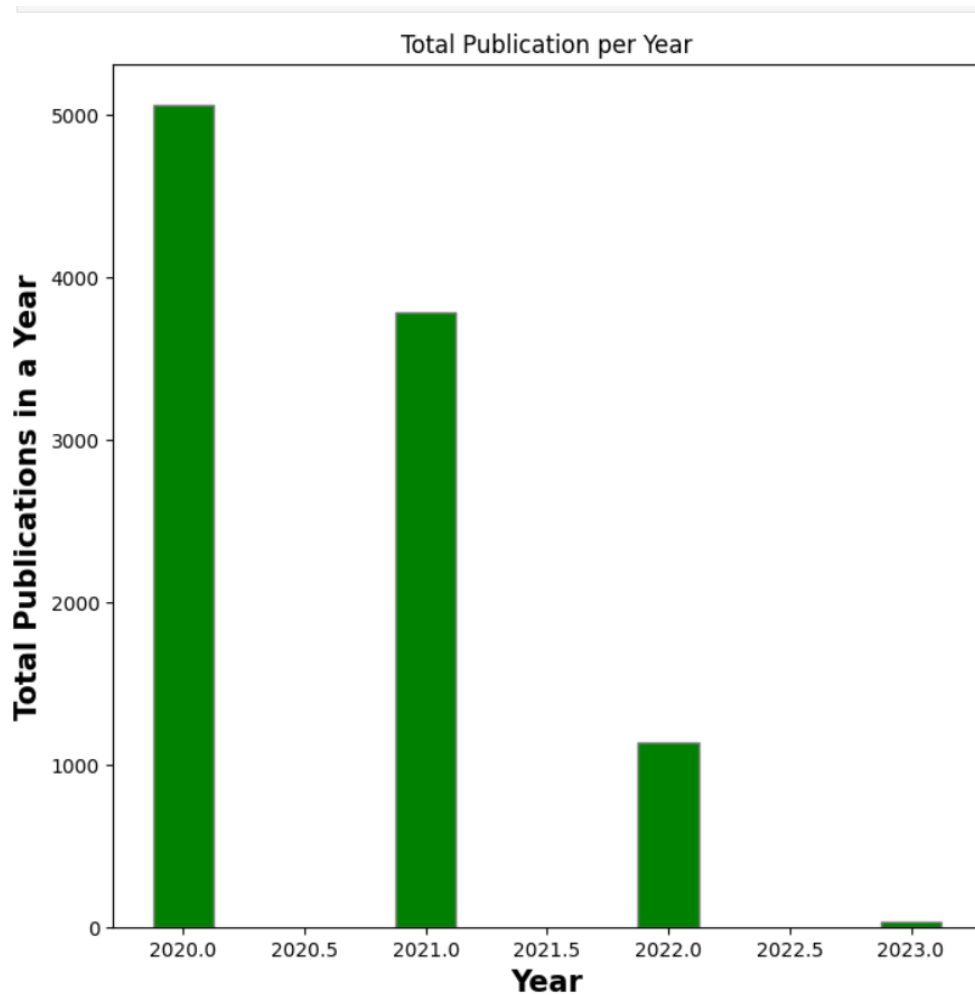
```
In [ ]: df_tot_pub_per_year = pd.DataFrame(df['Year'].value_counts()).sort_values(by = 'Year').sort_index().reset_index()
df_tot_pub_per_year.rename({'index': 'Year', 'Year' : 'Total Publications in a Year'}, axis=1, inplace=True)
df_tot_pub_per_year
```

```
Out[ ]:   Total Publications in a Year  count
0                               2020   5056
1                               2021   3784
2                               2022   1131
3                               2023     29
```

```
In [ ]: plt.figure(figsize=(8,8))
plt.bar(df_tot_pub_per_year['Total Publications in a Year'], df_tot_pub_per_year['count'],
        color = 'green', width = 0.25, edgecolor = 'grey')

plt.xlabel('Year', fontweight = 'bold', fontsize = 15)
plt.ylabel('Total Publications in a Year', fontweight = 'bold', fontsize = 15)
plt.title('Total Publication per Year')

plt.show()
```

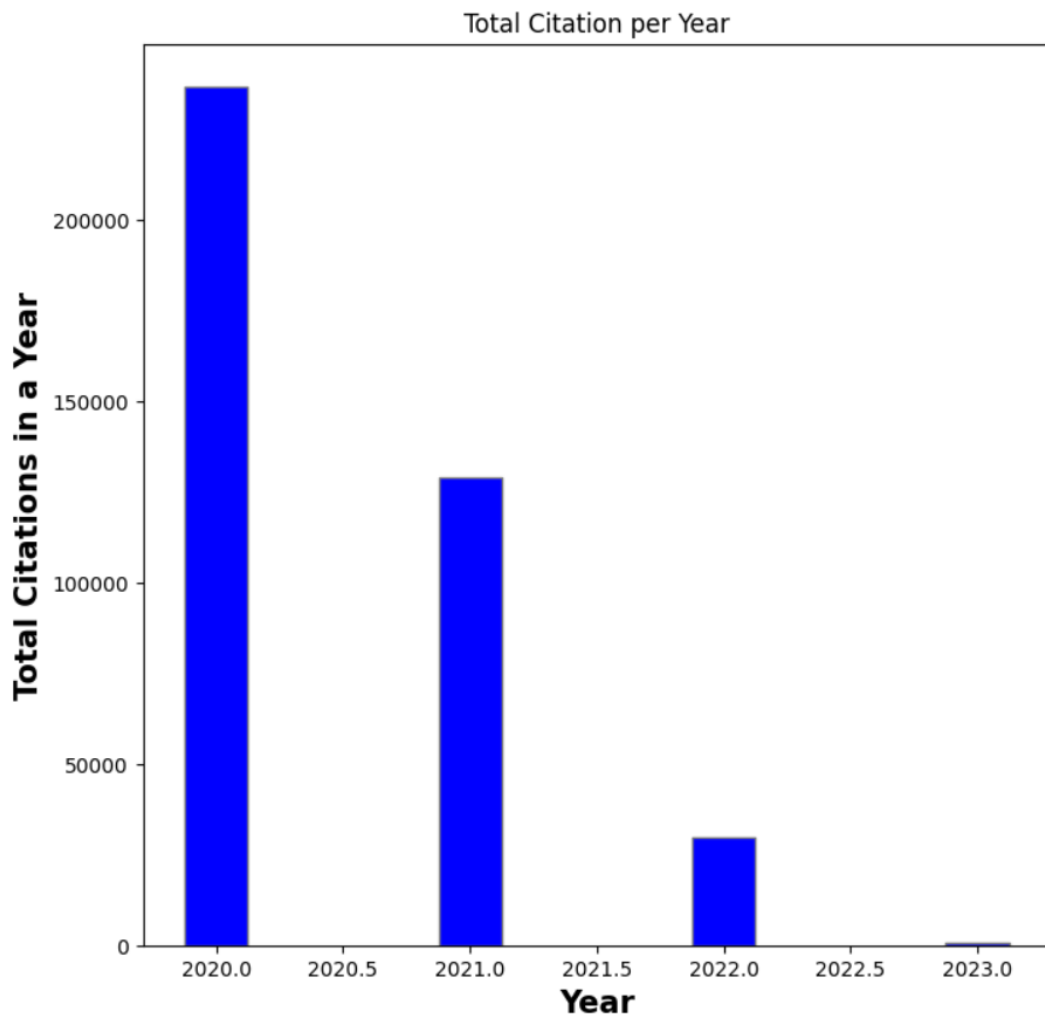



f) Total citation per year

```
In [ ]: df_tot_cit_per_year = pd.DataFrame(df.groupby('Year')['Cited by'].sum()).sort_values(by = 'Year').sort_index().reset_index(  
df_tot_cit_per_year.rename({'Cited by': 'Total Citations in a Year'}, axis=1, inplace=True)  
df_tot_cit_per_year
```

```
Out[ ]:   Year  Total Citations in a Year  
0  2020             236579  
1  2021             129025  
2  2022              29644  
3  2023               763
```

```
In [ ]: plt.figure(figsize=(8,8))  
plt.bar(df_tot_cit_per_year['Year'], df_tot_cit_per_year['Total Citations in a Year'],  
color = 'blue', width = 0.25, edgecolor = 'grey')  
  
plt.xlabel('Year', fontweight = 'bold', fontsize = 15)  
plt.ylabel('Total Citations in a Year', fontweight = 'bold', fontsize = 15)  
plt.title('Total Citation per Year')  
  
plt.show()
```



h) Highest cited author from India and the university

```
In [ ]: ACCY[ACCY['Country']=='India'].groupby('Authors')['Citation'].sum().sort_values(axis=0, ascending=False).reset_index()
```

Out []:

	Authors	Citation
0	Gupta D. Maharaja Agrasen Institute of Technology	1377
1	Shankar K. Department of Computer Applications	902
2	Khanna A. Maharaja Agrasen Institute of Techno...	881
3	Bhat M.M. Lelafe IT Solutions	674
4	Shah J.L. Higher Education Department	674
...
2410	Bhawal S. School of Computer Engineering	13
2411	Singh V. School of Computer Science	13
2412	Vakharia V. School of Technology	13
2413	Vaichole T.S. School of Computer Science and E...	13
2414	Gupta S. Department of Computer Science Engine...	13

2415 rows × 2 columns

```
In [ ]: print('Higest Cited Author from India : ',ACCY[ACCY['Country']=='India'].groupby('Authors')['Citation'].sum().sort_values
print('Total Citations : ',ACCY[ACCY['Country']=='India'].groupby('Authors')['Citation'].sum().sort_values(axis=0, ascend
```

Higest Cited Author from India : Gupta D. Maharaja Agrasen Institute of Technology
Total Citations : 1377

i) Comparative year wise article publication analysis of India, China and United States:

```
In [ ]: df_new = ACCY[(ACCY['Country']=='India') | (ACCY['Country']=='China') | (ACCY['Country']=='United States')]  
df_new
```

```
Out[ ]:
```

	Authors	Country	Citation	Year
19	Yu P.S. Department of Computer Science	United States	1906	2021
20	Cui Z. Department of Civil and Environmental E...	United States	278	2020
21	Henrickson K. Inrix Inc.	United States	278	2020
22	Ke R. Department of Civil and Environmental En...	United States	278	2020
23	Wang Y. Department of Civil and Environmental ...	United States	278	2020
...
48116	Chen Y. National Engineering Research Center f...	China	13	2022
48117	Zhang Z. National Engineering Research Center ...	China	13	2022
48119	Sadeghi Eshkevari S. Senseable City Laboratory	United States	13	2022
48120	Cronin L. Department of Civil and Environmenta...	United States	13	2022
48121	Pakzad S.N. Department of Civil and Environmen...	United States	13	2022

25992 rows × 4 columns

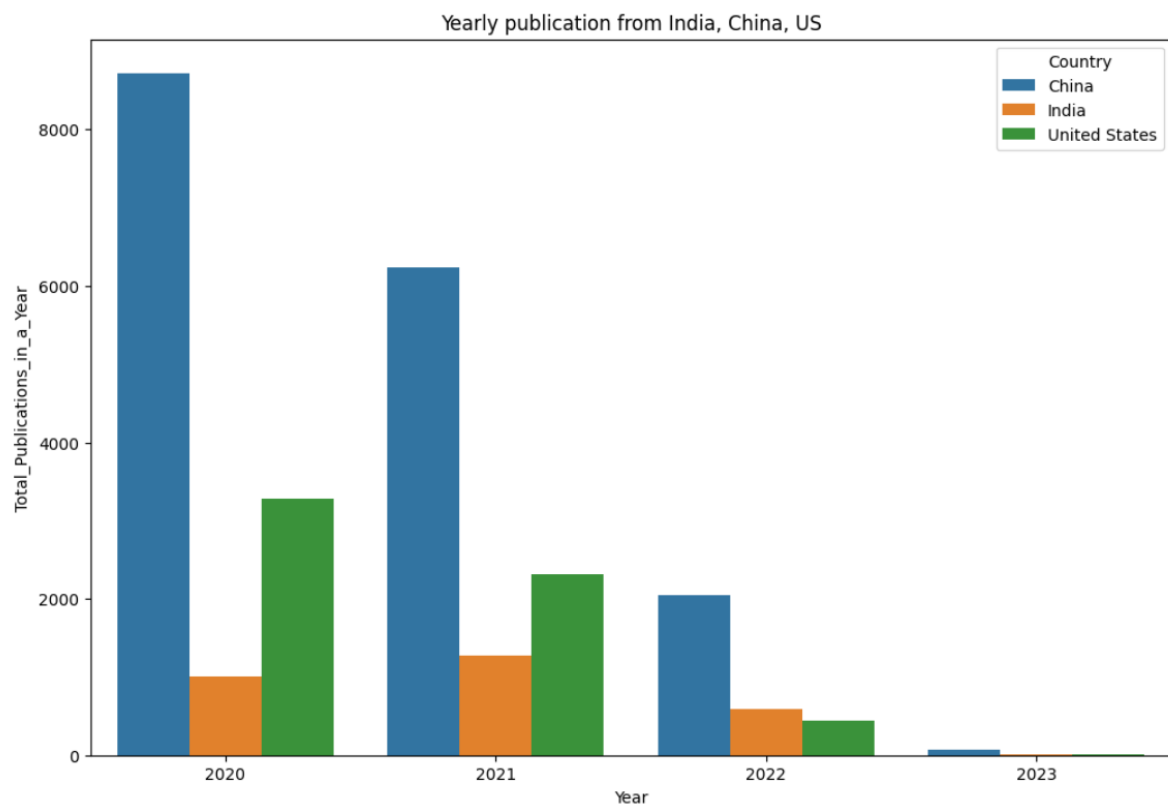
```
In [ ]: df_1 = df_new.groupby(['Country', 'Year'])['Country'].agg(Total_Publications_in_a_Year = 'count').reset_index()  
df_1
```

```
Out[ ]:
```

	Country	Year	Total_Publications_in_a_Year
0	China	2020	8717
1	China	2021	6231
2	China	2022	2041
3	China	2023	72
4	India	2020	1009
5	India	2021	1276
6	India	2022	589
7	India	2023	16
8	United States	2020	3281
9	United States	2021	2311
10	United States	2022	435
11	United States	2023	14

```
In [ ]: plt.figure(figsize=(12,8))  
sns.barplot(x="Year",  
            y="Total_Publications_in_a_Year",  
            hue="Country",  
            data=df_1)  
plt.title('Yearly publication from India, China, US')
```

```
Out[ ]: Text(0.5, 1.0, 'Yearly publication from India, China, US')
```



j) Total number of grants given to the field

```
In [ ]: df.loc[:, ['Titles', 'Funding Details', 'Authors']]
```

Out []:

	Titles	Funding Details	
0	Can AI Help in Screening Viral and COVID-19 Pn...	Qatar National Library; Qatar National Researc...	Chowdhury M.E.H.; Rahman
1	Recent advances in deep learning for object de...	NaN	Wu
2	A Comprehensive Survey on Graph Neural Networks	National Science Foundation, NSF, (1763325, CN...	Wu Z.; Pan S.; Chen F.; I
3	Traffic Graph Convolutional Recurrent Neural N...	USDOT	Cui Z.; Henric
4	A Survey on Deep Learning for Named Entity Rec...	NaN	I
...	
9995	Hybrid beamforming with relay and dual-base st...	NaN	Alsunbuli B.N.; Fakhru
9996	A Hybrid Intelligent Framework to Combat Sophi...	National Natural Science Foundation of China, ...	Javeed D.; Gao T.
9997	EHPE: Skeleton Cues-based Gaussian Coordinate ...	NaN	Liu H.; Liu T.;
9998	Input estimation of nonlinear systems using pr...	MIT Senseable City Lab Consortium; Pennsylvani...	Sadeghi Eshkevari S
9999	DECIMER 1.0: deep learning for chemical image ...	Google	Rajan K.; Z

10000 rows × 3 columns

```
In [ ]: print("Total number of grants given to the field of 'Deep learning' are",df1['Funding Details'].notnull().sum())
```

Total number of grants given to the field of 'Deep learning' are 6993

Total number of grants given to the field of 'Deep learning' are 6993

k) Country with total number of publication:

```
In [ ]: cwp = df_new['Country'].value_counts()
```

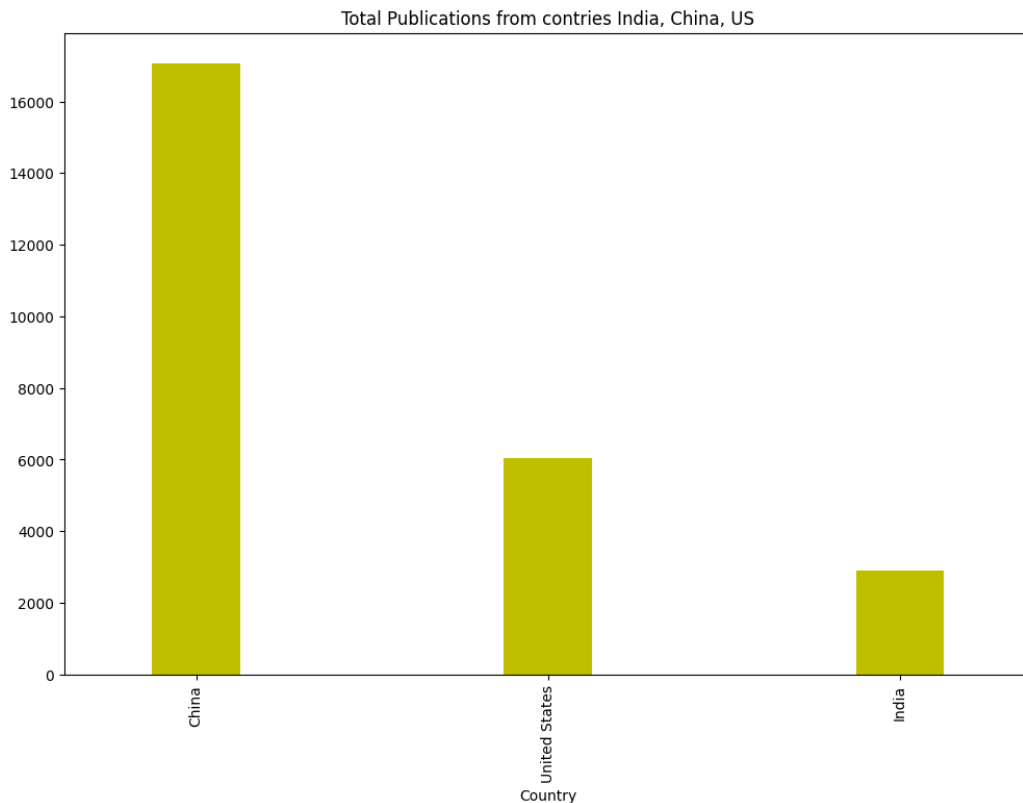
```
In [ ]: cwp
```

```
Out[ ]: Country
China      17061
United States  6041
India       2890
Name: count, dtype: int64
```

China: 17061 , United States: 6041 , India: 2890

```
In [ ]: plt.figure(figsize=(12,8))
cwp.plot(kind='bar', color = 'y',width=0.25)
plt.title('Total Publications from contries India, China, US')
```

```
Out[ ]: Text(0.5, 1.0, 'Total Publications from contries India, China, US')
```



g) Author having highest co-authorship with indian authors.

```
In [ ]: authors_df=pd.DataFrame({
    'Authors':df['Authors'].str.split(';').explode(),
    'Titles':df['Titles'].repeat(df['Authors'].str.count(';')+1),
    'Year':df['Year'].repeat(df['Authors'].str.count(';')+1),
    'Cited by':df['Cited by'].repeat(df['Authors'].str.count(';')+1),
    'Authors with affiliation':df['Authors with affiliations'].repeat(df['Authors'].str.count(';')+1)
})

authors_df.dropna(subset=['Authors with affiliation'],inplace=True)
authors_df['Country']=authors_df['Authors with affiliation'].apply(lambda x:x.split(';')[-1].strip())
```

```
In [ ]: authors_df.columns
```

```
Out[ ]: Index(['Authors', 'Titles', 'Year', 'Cited by', 'Authors with affiliation',
              'Country'],
              dtype='object')
```

```
In [ ]: indian_authors=authors_df[authors_df['Country']=="India"]
indian_authors.assign(Authors=indian_authors['Authors'].str.split(';').explode('Authors'))
```

	Authors	Titles	Year	Cited by	Authors with affiliation	Country
49	Khan A.I.	CoroNet: A deep neural network for detection a...	2020	674	Khan A.I., Department of Computer Science, Jam...	India
49	Shah J.L.	CoroNet: A deep neural network for detection a...	2020	674	Khan A.I., Department of Computer Science, Jam...	India
49	Bhat M.M.	CoroNet: A deep neural network for detection a...	2020	674	Khan A.I., Department of Computer Science, Jam...	India
55	Dargan S.	A Survey of Deep Learning and Its Applications...	2020	316	Dargan S., Department of Computational Science...	India
55	Kumar M.	A Survey of Deep Learning and Its Applications...	2020	316	Dargan S., Department of Computational Science...	India
...
9982	Venugopalan A.	Applying deep neural networks for the automati...	2021	13	Venugopalan A., Department of Computer Science...	India
9982	Reghunadhan R.	Applying deep neural networks for the automati...	2021	13	Venugopalan A., Department of Computer Science...	India
9983	Das R.	High granular and short term time series forec...	2022	13	Das R., Dept. of Computer Science and Engineer...	India
9983	Middya A.I.	High granular and short term time series forec...	2022	13	Das R., Dept. of Computer Science and Engineer...	India
9983	Roy S.	High granular and short term time series forec...	2022	13	Das R., Dept. of Computer Science and Engineer...	India

2487 rows × 6 columns

```
In [ ]: indian_author_cnt = indian_authors.groupby('Authors').size()
In [ ]: highest_count_index=indian_author_cnt.sort_values(ascending=False).index[1]
In [ ]: highest_cnt_author=highest_count_index.split(';')[0]
In [ ]: print("Author with Highest Co-Authorship with Indian author is :",highest_cnt_author)
```

Author with Highest Co-Authorship with an Indian author is : Gupta D.

Author with Highest Co-Authorship with an Indian author is : Gupta D.

REFERENCES

1. <https://www.scopus.com>
2. <https://apps.webofknowledge.com/>
3. <http://networksciencebook.com/translations/en/resources/data.html>
4. <https://pandas.pydata.org/docs/>