DSBDA
CASE STUDY:

PAPER PUBLICATION DATA ANALYSIS

THIRD YEAR COMPUTER ENGINEERING

Under guidance of

Prof. Ashwini Jarali

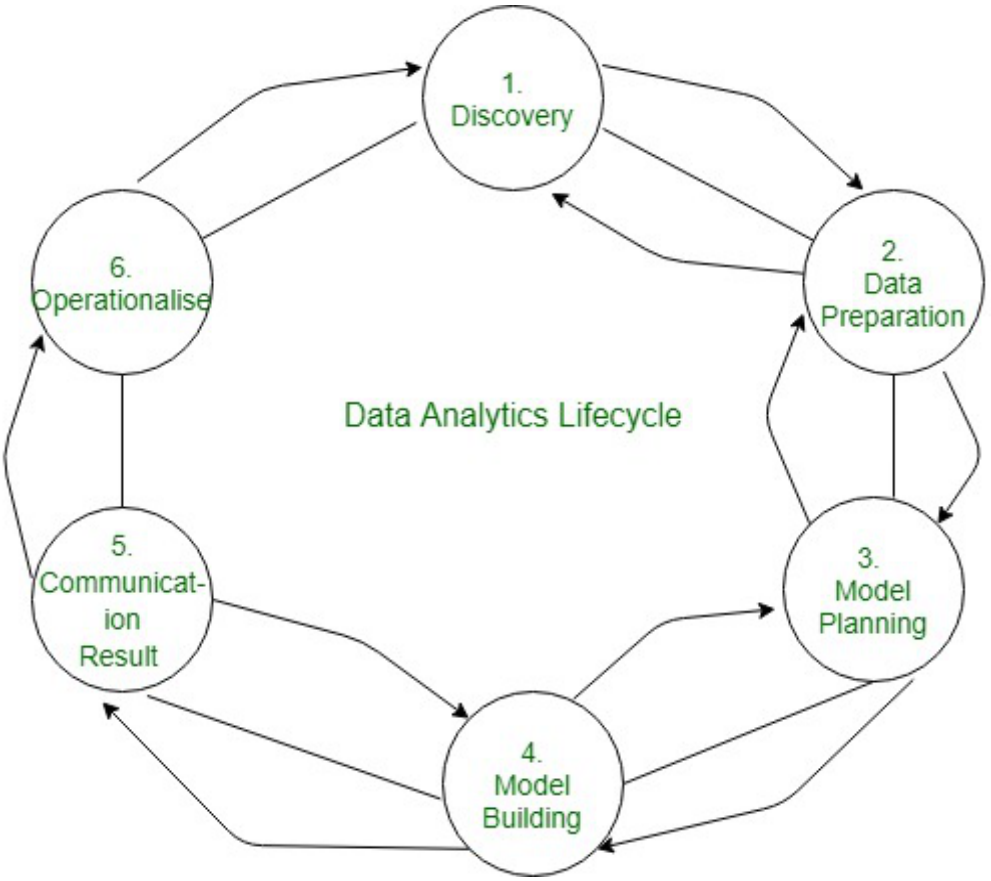
Group no.7

Mandar Deshmukh TC30
Sonal Shitole TC56
Isha Kawale TC24
Anushka Chavan TC10

Introduction:

Paper Publication data analysis is a system where user can search Publication type such as Journal,Conference, Book by Faculty Author name, Year,Domain name like Internet of things, Data Science, Artificial Intelligence, Deep Learning ,etc.

Along with users,admins can also perform search operations and through admin login journals/conferences/books details can be added as well in dataset.


1.
Discovery
2.
Data
Preparation
3.
Model
Planning
4.
Model
Building
5.
Communicat-
ion
Result
6.
Operationalise
Data Analytics Lifecycle

Phase 1: Discovery

- In this phase, we tried to learn and investigate the problem.We developed context and understanding.We came to know about data sources needed and available for the project.We formulates initial hypotheses that can be later tested with data**.**
- The topic and the dataset required for the project was given to us by our mentor.
- The dataset consists of Journal data,Conference data and Books data from different years.
- After getting the dataset we did some study and research on the topic and referred to different websites like IEEE.
- IHs(Initial Hypothesis)

  1. An idea for research can be analyzed and evaluated for the likelihood of receiving funding

  2.Emerging research topics can be classified and mapped to specific departments,faculty,author ,year,etc.

  3.Finding faculty author wise or year wise or domain wise data of published journal/book/conference.

  4.Website will successfully display details of journals and conferences of our college according to faculty ,year ,domain,etc.

  5.Visualisation and comparison of the data based on department or year or faculty or domain will be possible .

6.System will help students and teachers who are willing to publish a paper by searching for previous data of faculty expertise ,their domain ,etc.

7.Users can get appropriate information and guidance for future work in research.

Phase 2: Data Preparation

- We have explored, preprocessed, and conditioned data prior to modeling and analysis.
- It requires the presence of an analytic sandbox, the team to execute, load, and transform, to get data into the sandbox.
- We received raw data in text format from the department.
- The dataset contained data of journals,books,conferences.
- We first converted the dataset from text to csv file.
- During the data exploration we began to notice that certain data needed conditioning.
- Some Null/NaN values needed to be removed for proper working of the project.
- Also we added some columns like Faculty_author ,Ptype ,domain for searching purposes.

Phase 3: Model Planning:

- In this phase we will explore the data to learn about relationships between variables and subsequently, selected key variables and the most suitable models.

- We will need to create a UI where users can search publication papers by providing certain filters like Publication type (Journal,Conference and Book) ,Author, Year using HTML and CSS

- System will be based on python framework -flask

  Like most widely used python libraries, the Flask package is installable from the Python Package Index (PPI). First we need to create a directory first then a virtual environment where all the project related dependencies will be loaded (including flask)

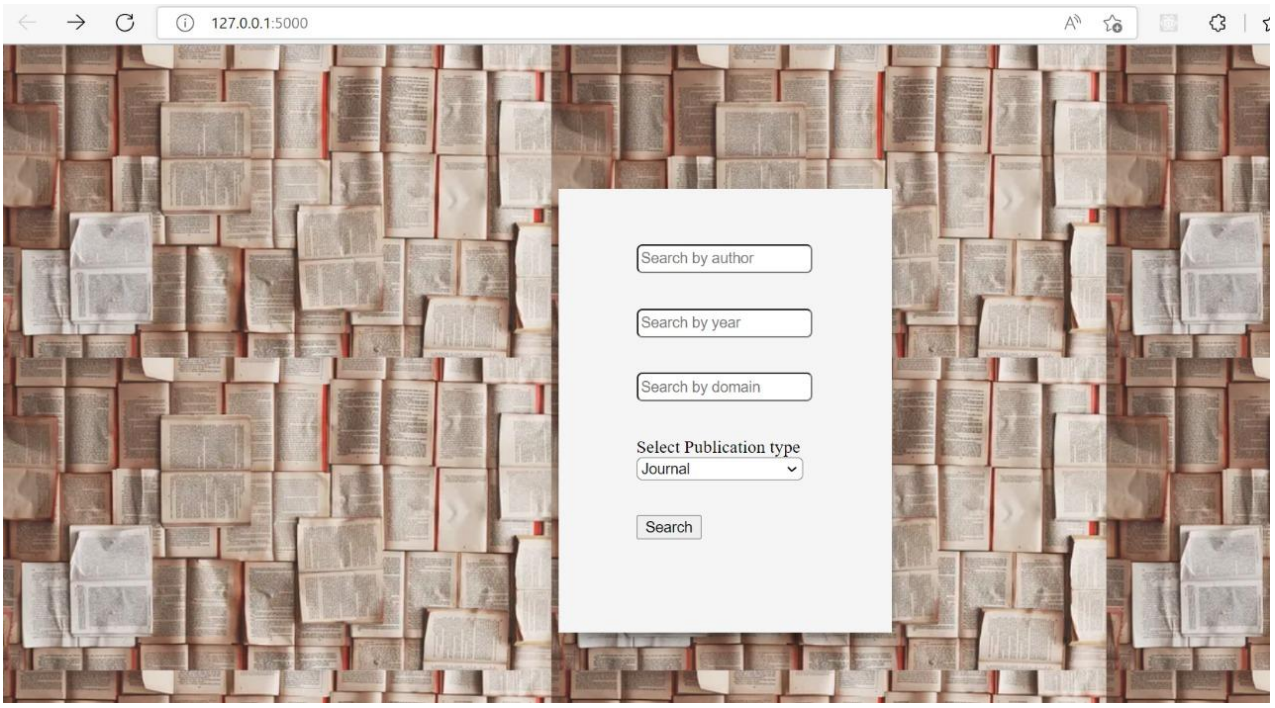
- Graphical Representation and visualization is needed on a dataset which will be done by using following libraries:
  1. numpy
  2. seaborn
  3. matplotlib
  4. pandas
  5. matplotlib figure

Phase 4: Model Building:

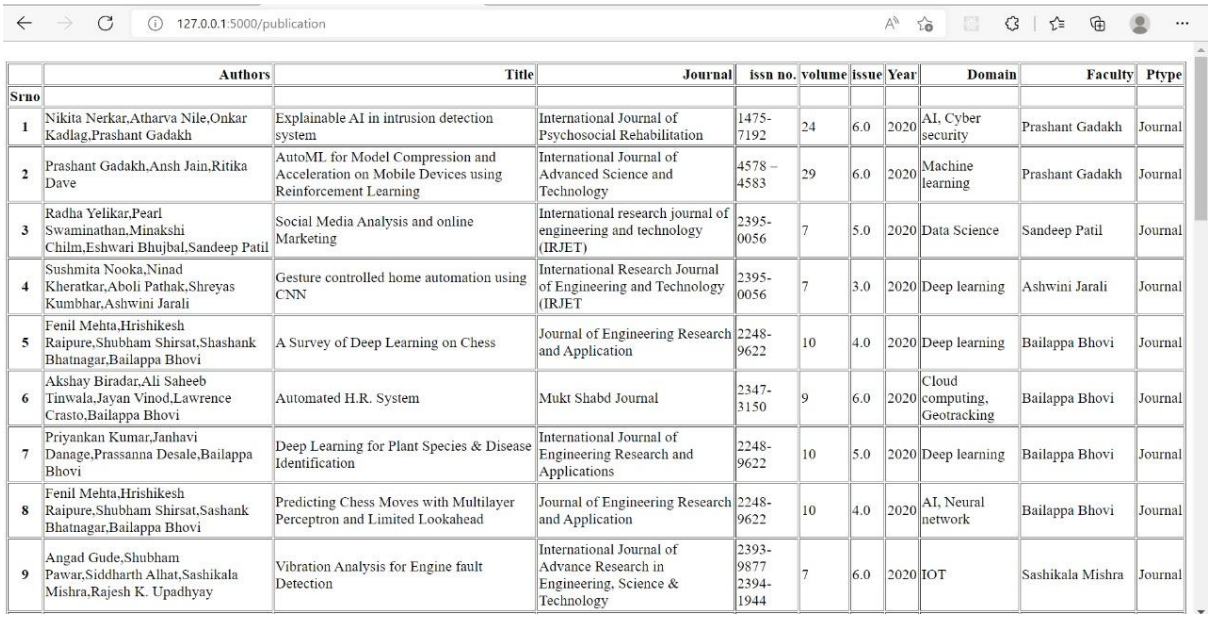
- First we have prepared a form in which user can fill certain filters like Publication type (Journal,Conference and Book) ,Author, Year.The form will redirect to the particular Journal,Conference , Book ,Author, Year
- Users can search publication papers by providing certain filters like Publication type (Journal,Conference and Book) ,Author, Year.
- Graphical Representation is provided in the form of pie charts ,bar graphs ,etc. for better understanding.

Phase 5: Communicate Results

- **GUI - search page**



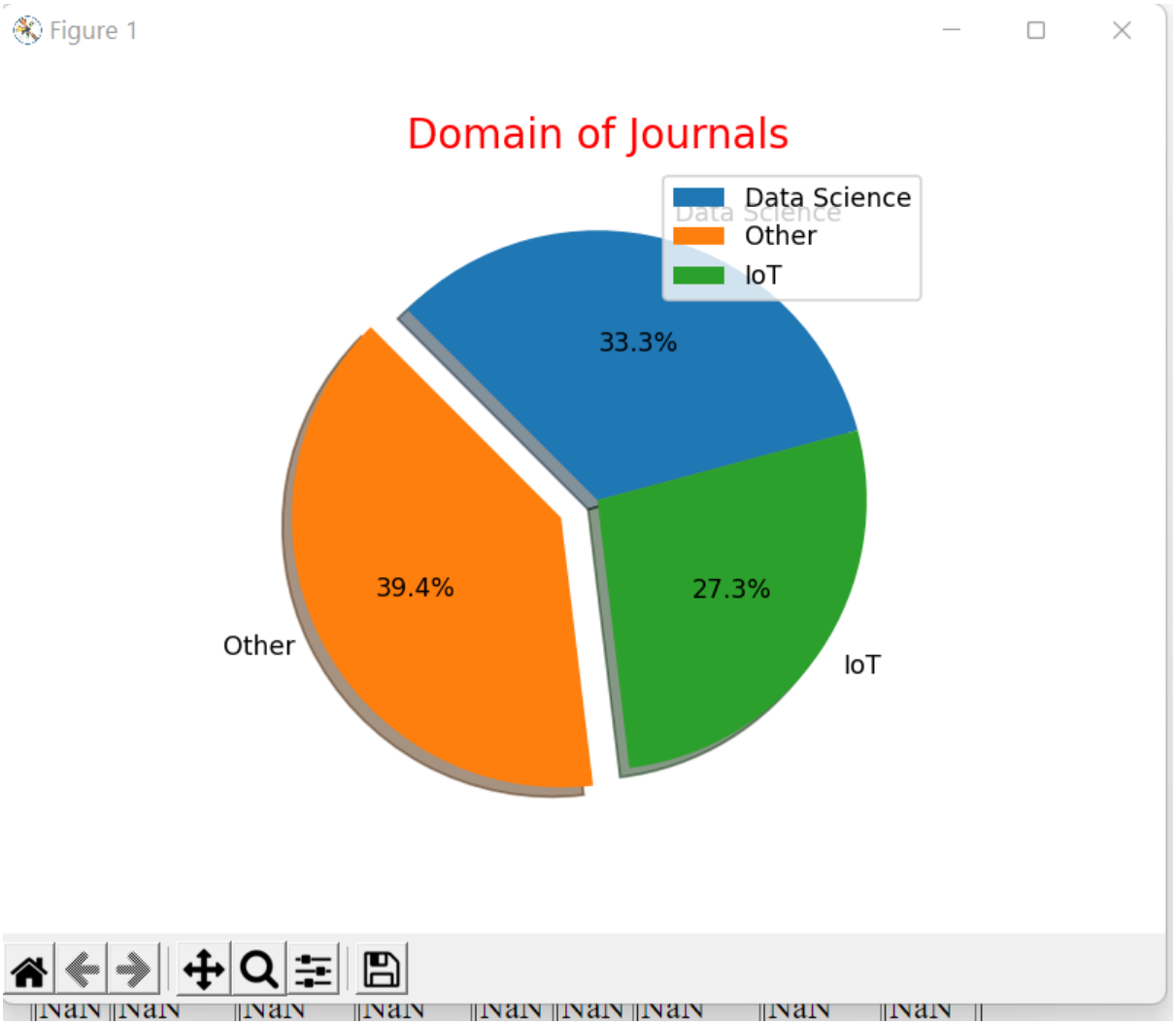
- **Tabular representation of filtered/searched data.**



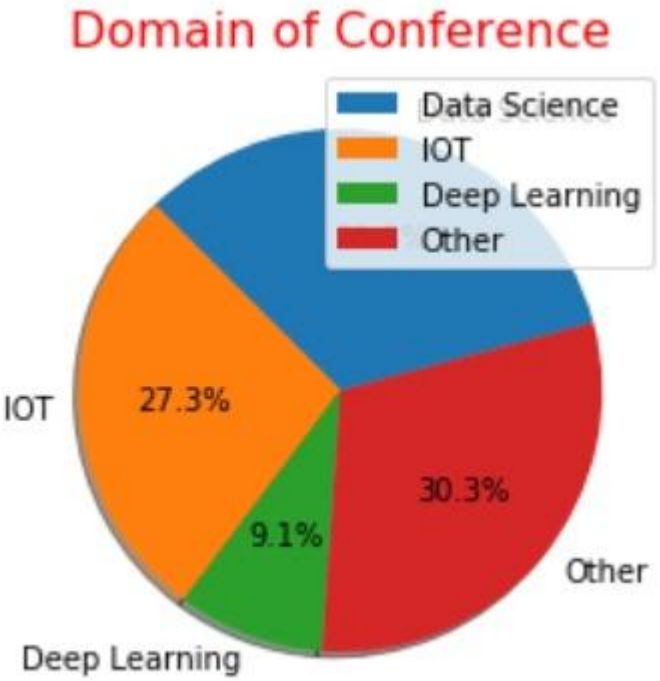
	Authors	Title	Journal	issn no.	volume	issue	Year	Domain	Faculty	Ptype
Srno										
1	Nikita Nerkar,Atharva Nile,Onkar Kadlag,Prashant Gadakh	Explainable AI in intrusion detection system	International Journal of Psychosocial Rehabilitation	1475-7192	24	6.0	2020	AI, Cyber security	Prashant Gadakh	Journal
2	Prashant Gadakh,Ansh Jain,Ritika Dave	AutoML for Model Compression and Acceleration on Mobile Devices using Reinforcement Learning	International Journal of Advanced Science and Technology	4578 – 4583	29	6.0	2020	Machine learning	Prashant Gadakh	Journal
3	Radha Yelikar,Pearl Swaminathan,Minakshi Chilm,Eshwari Bhujbal,Sandeep Patil	Social Media Analysis and online Marketing	International research journal of engineering and technology (IRJET)	2395-0056	7	5.0	2020	Data Science	Sandeep Patil	Journal
4	Sushmita Nooka,Ninad Kheratkar,Aboli Pathak,Shreyas Kumbhar,Ashwini Jarali	Gesture controlled home automation using CNN	International Research Journal of Engineering and Technology (IRJET	2395-0056	7	3.0	2020	Deep learning	Ashwini Jarali	Journal
5	Fenil Mehta,Hrishikesh Raipure,Shubham Shirsat,Shashank Bhatnagar,Bailappa Bhovi	A Survey of Deep Learning on Chess	Journal of Engineering Research and Application	2248-9622	10	4.0	2020	Deep learning	Bailappa Bhovi	Journal
6	Akshay Biradar,Ali Saheeb Tinwala,Jayan Vinod,Lawrence Crasto,Bailappa Bhovi	Automated H.R. System	Mukt Shabd Journal	2347-3150	9	6.0	2020	Cloud computing, Geotracking	Bailappa Bhovi	Journal
7	Priyankan Kumar,Janhavi Danage,Prassanna Desale,Bailappa Bhovi	Deep Learning for Plant Species & Disease Identification	International Journal of Engineering Research and Applications	2248-9622	10	5.0	2020	Deep learning	Bailappa Bhovi	Journal
8	Fenil Mehta,Hrishikesh Raipure,Shubham Shirsat,Sashank Bhatnagar,Bailappa Bhovi	Predicting Chess Moves with Multilayer Perceptron and Limited Lookahead	Journal of Engineering Research and Application	2248-9622	10	4.0	2020	AI, Neural network	Bailappa Bhovi	Journal
9	Angad Gude,Shubham Pawar,Siddharth Alhat,Sashikala Mishra,Rajesh K. Upadhyay	Vibration Analysis for Engine fault Detection	International Journal of Advance Research in Engineering, Science & Technology	2393-9877 2394-1944	7	6.0	2020	IOT	Sashikala Mishra	Journal

- **Domain of Journals Pie-chart.**

  A pie chart is a type of graph that records data in a circular manner that is further divided into sectors for representing the data of that particular part out of the whole part. Each of these sectors or slices represents the proportionate part of the whole.

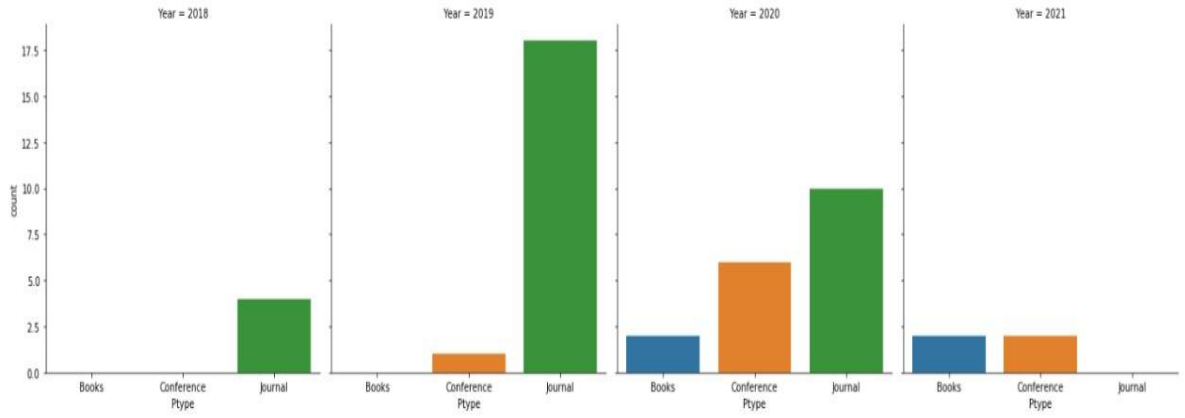


- **Domain of Conference Pie-chart.**

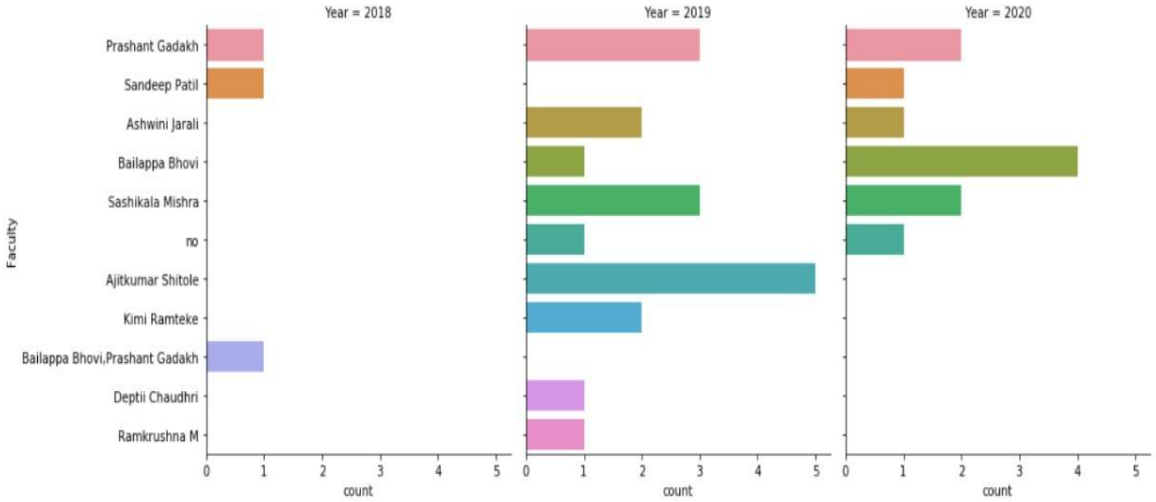


Year wise publication of Journal, Conference and Books.

If we are working with data that involves any categorical variables like in this case -faculty author or publication type, the best tools to visualize and compare different features of data would be categorical plots(catplot).



Faculty wise count of publications.



Phase 6: Operationalize

- In this last phase we have communicated the benefits of the project more broadly and set up a pilot project to deploy work in a controlled way before broadening the work to full enterprise of users.
- This approach will enable the team to learn about performance and related constraints of the model in a production environment on a small scale , and make adjustments before full deployment.
- We can use this analysis on the college website to get a better idea about the research going on in the college.

- Also data visualization can be done by further adding more columns like department ,references ,etc.