



DVPD Report

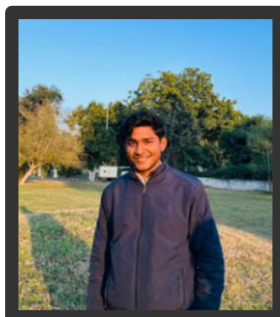
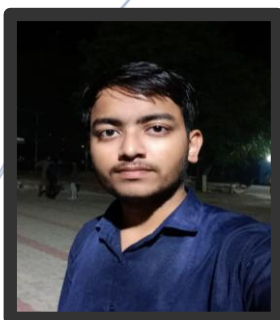
(Faculty Research Dashboard)

TEAM MEMBERS

- Kunal Kashyap(24103087)
- Amrit Noor Singh(24117003)
- Charanjot Singh(24103038)
- Ankit Jaiswal(24103013)
- Aryan Gupta(24117005)

SUBMITTED TO

Dr. Banalaxmi Brahma
Assistant Professor(Grade II)



Faculty Research Dashboard Report

1. ABSTRACT

The academic ecosystem produces extensive research outputs—publications, citations, projects, collaborations yet this information is often fragmented across multiple platforms. At NIT Jalandhar students and stakeholders lack a unified system to access faculty research profiles, publication records, and project details efficiently. This project aims to solve that gap by designing and implementing a Faculty Research Dashboard for NITJ. The dashboard provides a centralized, department-wise view of all professors along with their key research metrics such as total publications, citation counts, h-index, i10-index, co-author counts, expertise, and ongoing projects. A leaderboard section highlights top-performing faculty based on citations and publications, enabling quick identification of active researchers in each department.

To support this dashboard, an automated data collection pipeline was developed using Python to scrape and consolidate data from sources such as the IRINS portal, NITJ website, Google Scholar. The collected data is cleaned, structured and stored in a tabular format, which is then used as the backend for the dashboard. The final system allows students to discover faculty working in their areas of interest and helps the institute showcase its research output in a clear, data-driven manner.

2. Problem Statement

Research output is one of the primary indicators of an institute's academic strength. However, at NIT Jalandhar, information regarding faculty publications, citations, research interests, and projects is currently dispersed across multiple platforms and lacks standardization. Students who wish to explore research opportunities often face the following difficulties:

1. Fragmented Information Sources

Different pieces of data—such as publications, citations, and identifiers—exist across IRINS, Google Scholar, departmental directories, personal faculty pages, and academic CVs. There is no unified point of access.

2. Lack of a Central Dashboard

NITJ website not provides a consolidated dashboard that presents department-wise or institute-wide research statistics in a comparative or analytical format.

3. Limited Discoverability for Students

Students who are new to research fields struggle to identify faculty members who work in specific domains of interest due to the absence of clear categorization and filtration features.

4. Time-Consuming Manual Search

Navigating through our main site by searching each professor's profile to gather information on metrics like h-index, co-author count, or citation history is impractical.

These challenges severely restrict the ability of students to explore research pathways, apply for internships under suitable mentors, or engage meaningfully with ongoing institutional research projects. Hence, a centralized Faculty Research Dashboard is needed to bridge the gap between available research information and its accessibility.

3. OBJECTIVES

The objective of this project is to design and develop a centralized Faculty Research Dashboard for NIT Jalandhar that consolidates key research information of faculty members into a single, easy-to-access platform. The project focuses on integrating reliable data sources, automating the extraction of research metrics, organizing faculty information department-wise, and presenting it through an intuitive and visually coherent dashboard. In addition, the project aims to highlight leading researchers through metric-based leaderboards, improve accessibility for students seeking research opportunities, and ensure that the entire system is supported by a clean, structured, and validated dataset. The overall goal is to enhance research discoverability, simplify academic navigation, and provide a professional, data-driven platform aligned with institutional needs and DVPD principles.

Objectives of the Project:

- To create a unified and interactive research dashboard for all NITJ faculty members.
- To automate data extraction using Python-based scraping from reliable sources such as NITJ website, Google Scholar, Vidwan, IRNS.
- To categorize faculty profiles department-wise for clear and structured navigation.
- To collect and compute key research metrics including publications, citations, h-index, i10-index, co-author count, and expertise areas.
- To develop leaderboards that highlight top performers across departments and institute-wide.
- To preprocess, clean, validate, and convert raw scraped data into a structured CSV dataset.
- To improve student access to research information and bridge the gap between learners and faculty researchers.

4. Methodology

The methodology adopted in this project follows a systematic, pipeline-based approach aligned with the principles taught in the DVPD course. We began by identifying and evaluating genuine, government-authorized data sources—primarily the IRINS portal—to ensure that the information collected was accurate, reliable, and of high quality. Using browser inspection tools, we analyzed the HTML structure of department pages and faculty profiles, after which a Python-based scraping pipeline was developed using Requests, BeautifulSoup, and Regular Expressions. This pipeline extracted essential research metrics such as publications, citations, h-index, i10-index, expertise, and ongoing projects. The raw extracted data underwent preprocessing using techniques learned in DVPD and Python labs, including data cleaning, normalization, validation, removal of noise, and type conversions. We verified data correctness by cross-checking values from multiple sources wherever necessary. Once cleaned, the data was transformed into a structured CSV file using pandas, enabling further processing and visualization. Basic exploratory analysis was performed to understand metric distributions and identify key features for the dashboard. Finally, this refined dataset was integrated into the dashboard, where filtering, sorting, leaderboard logic, and department-wise views were implemented. This end-to-end methodology ensured that the dashboard was data-driven, accurate, user-friendly, and methodically aligned with academic data-processing standards.

5. Let's see our Features

The Faculty Research Dashboard developed in this project incorporates a wide range of features designed to simplify the discovery, comparison, and analysis of research activities across NIT Jalandhar. Each feature has been carefully crafted to address the challenges students and stakeholders face in accessing reliable faculty research information. The dashboard brings together multiple data sources, organizes them into a structured and interactive format, and presents the information with clarity and precision. The system reflects the core DVPD principles of data accessibility, transparency, and informed decision-making.

FEATURES:

1. Centralized Research Overview

A unified interface that consolidates verified research information of NITJ faculty members, removing the need to visit multiple websites or individual profiles.

2. Department-Wise Faculty Listing

A structured categorization of faculty based on their respective departments, enabling users to browse and explore research information relevant to their academic area.

3. Comprehensive Research Metrics Display

Each faculty entry presents essential research indicators such as publication counts, citations, h-index, i10-index, and expertise fields in a clear and comparative format.

4. Research Leaderboard

A metric-driven leaderboard that highlights top-performing faculty based on a weighted score derived from citations, publications, and h-index, offering quick insight into research impact.

5. Interactive Search and Filtering

Advanced filtering and search features that allow users to locate faculty members by department, name, or research domain, improving accessibility and usability.

6. CSV Upload and Data Refresh Support

A dedicated feature allows users to upload new or updated CSV files, enabling the dashboard to refresh data instantly without modifying the backend code. This ensures scalability and long-term usability of the system.

6. Technology Architecture (Code)

The development of the Faculty Research Dashboard for NIT Jalandhar required a combination of programming tools, data processing libraries, and interactive visualization technologies. The selected tech stack ensured that the system remained efficient, scalable, and capable of handling heterogeneous academic data collected from multiple authentic sources. The approach aligns with DVPD principles, emphasizing dependable data pipelines, structured processing, and insightful visualization.

The dashboard and data pipeline were implemented primarily using Python, owing to its comprehensive ecosystem of data analytics and visualization libraries. The interactive interface was developed using Streamlit, which enabled quick deployment and a clean user experience without extensive frontend coding. For data handling and preprocessing, Pandas and NumPy were employed to structure scraped information into a reliable analytical format. Finally, Plotly was used for generating dynamic visualizations, including bar charts, pie charts, bubble charts, and layered analytics that support departmental and institute-wide insights.

Overall, this technology stack provided an end-to-end solution, covering data collection, cleaning, transformation, analysis, and dashboard presentation in a well-organized and professionally integrated environment.

LIBRARIES & TOOLS USED

The following major libraries, frameworks, and tools were utilized to build the entire research information system:

Programming & Environment

- **Python 3.x** — Core programming language for the scraper, preprocessing pipeline, and dashboard backend.

Data Collection & Parsing

- **Requests** — Fetches HTML content from IRINS web pages, nitj site, google scholar etc.
- **BeautifulSoup (bs4)** — Parses webpage structures and extracts faculty-level details such as publications, citations, and identifiers.

Data Cleaning & Preprocessing

- **Pandas** — Performs data loading, cleaning, transformation, and CSV structuring.
- **NumPy** — Supports numerical processing during aggregation, normalization, and trend simulations.
- **Regex (re)** — Extracts numeric values, research identifiers, and patterns embedded within raw HTML content.

Visualization & Dashboard Development

- **Streamlit** — Framework for building the interactive dashboard with navigation tabs, search features, metrics cards, and departmental analytics.
- **Plotly Express** — Generates dynamic charts such as bar graphs, pie charts, and bubble plots for exploratory and comparative visualizations.

Data Processing Techniques Used

- Missing value handling using `fillna(0)`
- Type casting using `pd.to_numeric(errors='coerce')`
- Imputation of missing categorical attributes
- Min–Max normalization for leaderboard score calculation
- Weighted scoring for research ranking:
 - 40% Citations (normalized)
 - 40% H-index (normalized)
 - 20% Publications (normalized)
- Feature engineering for “Impact Score” and department-level aggregates
- Groupby-based departmental analytics (counts, averages, distributions)

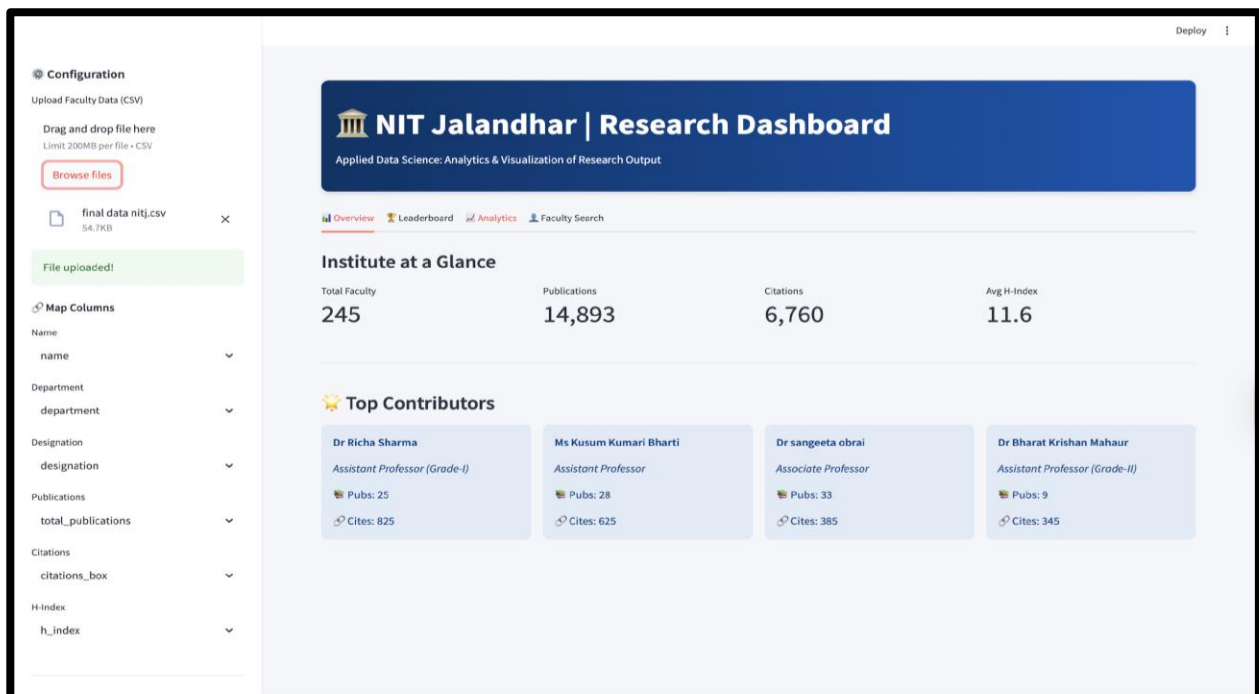
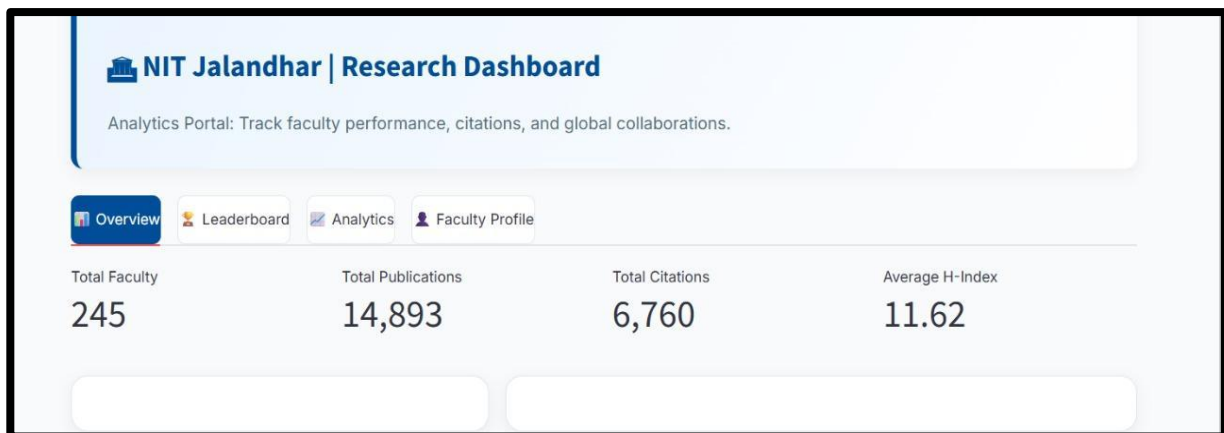
The complete implementation, including all code files, data processing scripts, and configuration modules, is provided in the GitHub repository referenced below.

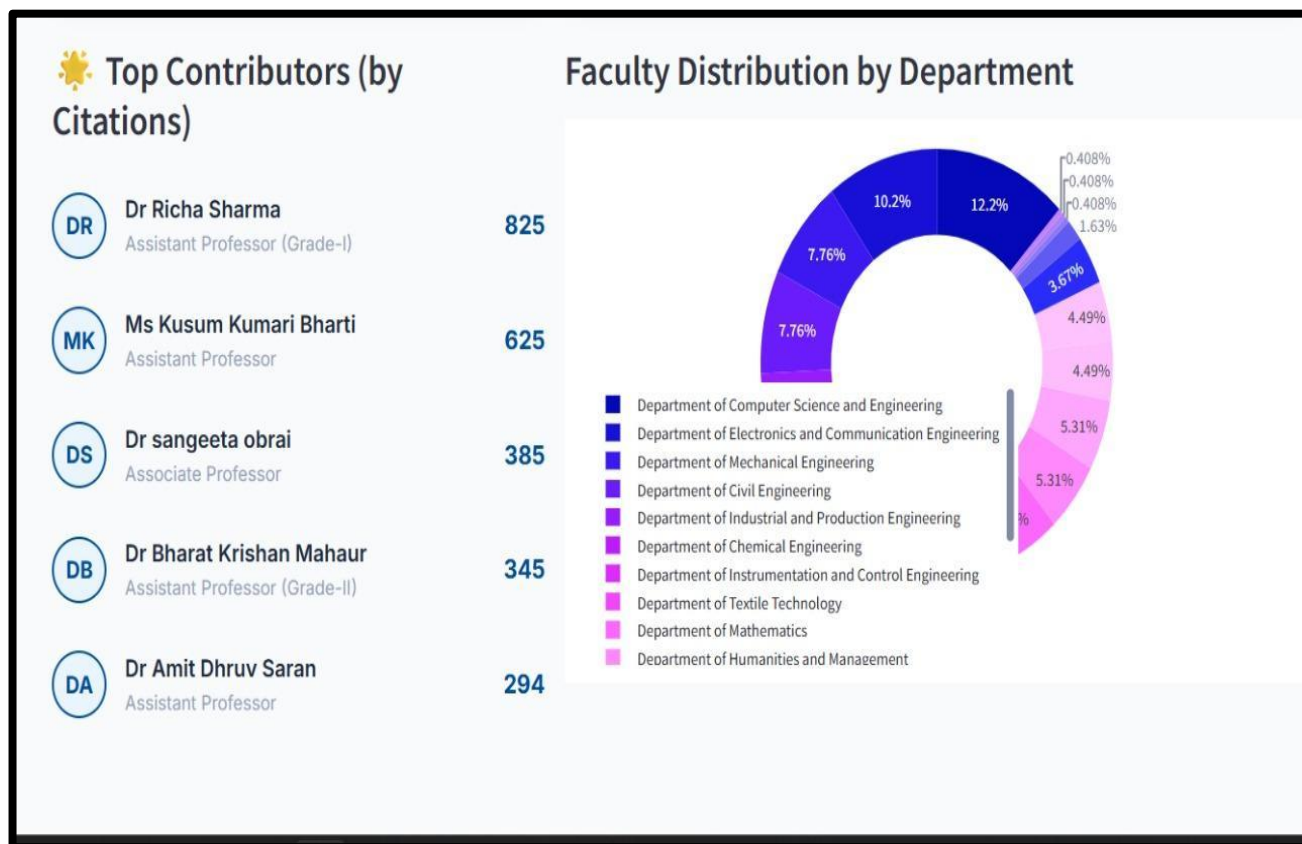
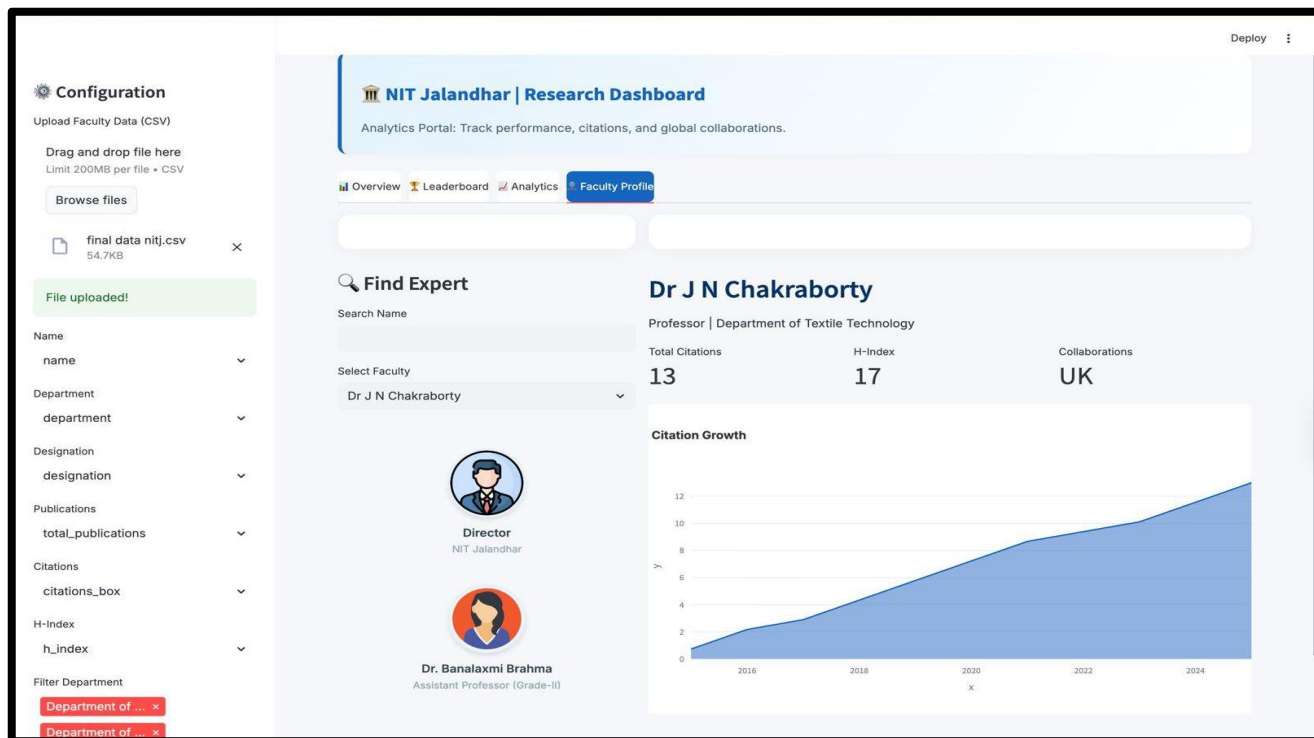
GitHub Repository:

https://github.com/DEVKING-Kunal/Faculty_Research_Dashboard_DVPD_Project

7. DASHBOARD OUTPUTS & VISUALS

This section presents the visual outputs of the Faculty Research Dashboard developed as part of the project. The dashboard contains multiple interactive components, including department-wise listings, research leaderboards, and detailed analytical charts. The images below represent different parts of the dashboard and illustrate how the system displays research information in a structured and user-friendly manner.







Research Impact Leaderboard

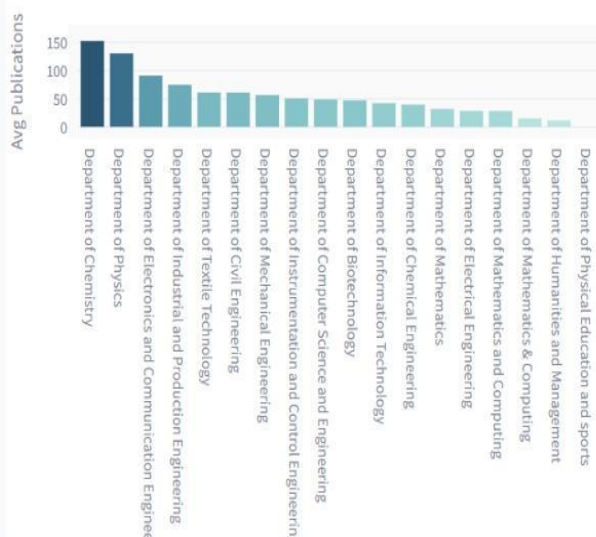
Dr Richa Sharma
Score: 46.9

Dr Suneel Dutt
Score: 62.8

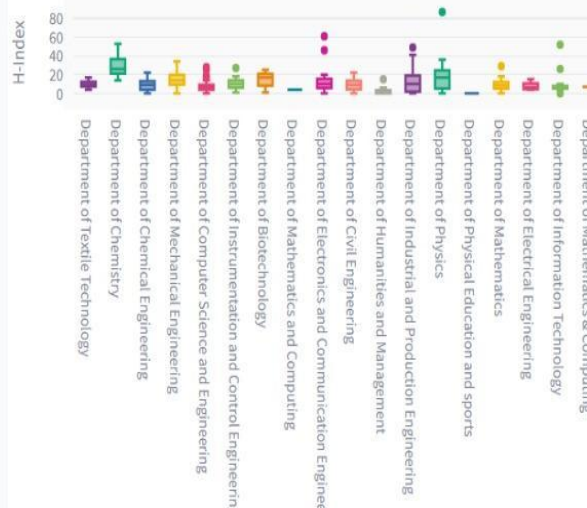
Prof Balbir Singh Kaith
Score: 39.4

Rank	Name	Department	Citations	H-Index	Impact Score
1	Dr Suneel Dutt	Department of Physics	58	87	62.8
2	Dr Richa Sharma	Department of Instrumentation and Control Engineering	825	13	46.9
3	Prof Balbir Singh Kaith	Department of Chemistry	7	53	39.4
4	Prof Binod Kumar Kanaujia	Department of Electronics and Communication Engineering	4	46	39.3
5	Dr Balwinder Raj	Department of Electronics and Communication Engineering	2	61	38.5
6	Ms Kusum Kumari Bharti	Department of Information Technology	625	10	35.9
7	Dr Vijay Kumar	Department of Information Technology	17	52	34.8
8	Dr RAVI PRATAP SINGH	Department of Industrial and Production Engineering	7	49	32.5
9	Dr Harsh Kumar Manchanda	Department of Chemistry	2	37	32.0

Average Publications per Department



H-Index Distribution (Outliers Highlighted)



8. CODE & DATASET

The code builds a Streamlit dashboard that loads or generates faculty research data, maps important columns (name, department, publications, citations, h-index), and supports image display, fuzzy search, and data scraping logs.

The dataset contains faculty research details such as department, publications, citations, h-index, i10 index, and collaboration country. If no CSV is uploaded, a dummy dataset of 150 randomly generated faculty profiles is created.

```

=== Department: Department of Computer Science and Engineering ===
Found 30 faculty
-> Scraping: Dr Harsh Kumar Verma
-> Scraping: Prof Amrit Lal Sangal
-> Scraping: Mr Deepak Kumar Gupta
-> Scraping: Dr Rajneesh Rani
-> Scraping: Dr Renu Dhir
-> Scraping: Sumit Kumar
-> Scraping: Dr Prashant Kumar
-> Scraping: Dr Armaan Garg
-> Scraping: Dr Pranjali Kumar
-> Scraping: Dr Prashant Shukla
-> Scraping: Mrs Madhurima Buragohain
-> Scraping: Dr Muktesh Gupta
-> Scraping: Dr Himanshu Verma
-> Scraping: Dr Banalaxmi Brahma
-> Scraping: Dr Shefali Arora Chouhan
-> Scraping: Dr Jagdeep Kaur
-> Scraping: Dr Aruna Malik
-> Scraping: Dr Amritpal Singh
-> Scraping: Dr Urvashi Bansal
-> Scraping: Dr Nagender Pratap Singh

```

```

import streamlit as st
import pandas as pd
import plotly.express as px
import plotly.graph_objects as go
import numpy as np
import base64
import os
from thefuzz import process # Import for Fuzzy Search

def img_to_bytes(img_path):
    img_bytes = None
    if os.path.exists(img_path):
        with open(img_path, "rb") as image_file:
            encoded_string = base64.b64encode(image_file.read()).decode()
            ext = img_path.split('.')[-1].lower()
            mime_type = "jpeg" if ext in ['jpg', 'jpeg'] else ext
            img_bytes = f"data:image/{mime_type};base64,{encoded_string}"
    return img_bytes

st.set_page_config(
    page_title="NITJ Research Admin",
    page_icon="🔍",
    layout="wide",
    initial_sidebar_state="expanded"
)

```

```

# -----
# [DATA LOGIC]
# -----
def generate_dummy_data():
    departments = ['CSE', 'ECE', 'ME', 'Civil', 'Physics', 'Chem', 'Textile', 'IPE', 'ICE']
    names = [f"Dr. Faculty {i}" for i in range(1, 151)]
    # Adding Banalaxmi MAM manually to dummy data for testing if no CSV
    names.append("Dr. Banalaxmi Brahma")

    countries = ['India', 'USA', 'UK', 'Canada', 'Germany', 'Australia', 'Japan', 'France']
    weights = [0.6, 0.1, 0.05, 0.05, 0.05, 0.05, 0.05, 0.05]
    data = []
    for name in names:
        dept = np.random.choice(departments)
        pubs = np.random.randint(5, 200)
        cites = int(pubs * np.random.uniform(5, 60))
        h_index = int(np.sqrt(cites) * np.random.uniform(0.8, 1.2))
        i10 = int(h_index * 1.5)
        country = np.random.choice(countries, p=weights)

```

This project is built on the strong foundation we gained in class, especially in **KNN, data preprocessing, feature selection, and data classification**. The concepts we learned—cleaning data, normalizing values, and applying similarity-based methods—directly shaped how we designed the workflow and analysis in this system, turning classroom theory into a practical application

9. CONCLUSION

The Faculty Research Dashboard developed in this project successfully resolves the issue of scattered and difficult-to-access research information at NIT Jalandhar. By combining automated web scraping, structured data preprocessing, and an interactive dashboard interface, the system brings together essential research metrics—such as publications, citations, h-index, i10-index, co-author counts, and expertise—into a single unified platform. This significantly improves accessibility for students, faculty, and administrators, while also supporting better academic decision-making.

The analysis of the compiled dataset reveals important patterns within the institute's academic structure. Most faculty members belong to junior and mid-level designations, especially Assistant and Associate Professors, and the largest departments—such as Computer Science, Electronics, Mechanical, and Electrical Engineering—reflect NITJ's strong technical orientation. Research productivity increases with seniority, with Professors showing higher output, and the Director standing out as a major outlier with exceptionally high publication and citation counts. Visualizations used in the dashboard clearly highlight these trends, especially the departmental distribution and the publication-by-designation comparison.

Overall, the project demonstrates how DVPD concepts—data sourcing, preprocessing, EDA, visualization, and dashboard building—can be applied to create a meaningful real-world system. The dashboard not only centralizes faculty research information but also provides valuable insights into NITJ's academic landscape and can be extended further for trend analysis, collaboration networks, and other research-oriented features.

10. REFERENCES

1. **National Institute of Technology Jalandhar (Official Website)**
<https://www.nitj.ac.in/>
2. **IRINS – Indian Research Information Network System (Faculty Research Profiles)**
<https://irins.org/>
 NITJ IRINS Portal: <https://nitj.irins.org/>
3. **Google Scholar – Research Citation and Publication Metrics**
<https://scholar.google.com/>
4. **Python Official Documentation (Libraries and Modules Used)**
<https://docs.python.org/>
5. **Streamlit Documentation (Dashboard Development)**
<https://docs.streamlit.io/>
6. **Plotly Documentation (Visualizations)**
<https://plotly.com/python/>
7. **Dataset Link**
<https://docs.google.com/spreadsheets/d/1HVL75v3VavB8q8LURnKg4G9oHtmw62An2mtxJwaMCYM/edit?usp=sharing>