

SUDIPTA KUMAR HAZRA

Data Analysis || PG.D. Statistics || B.E. Food Tech and Biochem Engg. || Machine Learning || Multivariate Statistics

Phone: +353 0892034307

Email: sudiptakh037@gmail.com



[Website](#)



[LinkedIn](#)



[Github](#)

SUMMARY

Data Analytics Professional with a passion for statistical analysis and digital transformation. Proficient in Python, R, and JavaScript, with expertise in multivariate analysis, machine learning, and predictive modeling. Experienced in developing web-based solutions that integrate data science with modern frameworks, enhancing accessibility and automation.

EXPERIENCE

Walsh Scholar

[Teagasc, Ireland](#)

10/2023 - Present

Irish Agri-Food Research Institute

- Analyzed high-dimensional meat proteomics data.
- R packages used: tidyverse, tidymodels, ggplot2, limma, car.

Statistical Analyst, AI/ML Team

[Kovaion Consulting, India](#)

07/2022 - 06/2023

AI/ML Business Solution

- Built ML-driven analytics tools and web apps.
- Tools used: Python, Flask, scikit-learn, LangChain, Power BI.

Research Intern

[CSIR-CFTRI, India](#)

01/2022 - 03/2022

Food Technology Research Institute

- Analyzed multivariate NMR data to classify adulterated honey.
- Tools used: SIMCA, MATLAB.

EDUCATION

M.Res. in Food Science & Technology

[University College Cork, Ireland](#)

10/2023 - 09/2025

PGDip in Statistical Methods & Analytics

70.4%

[Indian Statistical Institute, India](#)

07/2021 - 06/2022

B.E. in Food Technology & Biochemical Engineering

76.61%

[Jadavpur University, India](#)

07/2017-06/2021



SKILLS

Data Science & Analytics

- Machine & statistical learning
- Multivariate data analysis
- Statistical modeling
- LLM-based solutions
- Chemometrics & omics data analysis

Programming Languages

- Python
- R
- JavaScript
- React.js

Softwares/Interfaces

- Azure
- Git & Github
- Power BI
- SQL & MongoDB

CERTIFICATION

[Winter School on Deep Learning,](#)

Indian Statistical Institute

[Diploma in Chemometrics,](#)
University of Copenhagen

[Web Development Bootcamp,](#)
Udemy

ACHIEVEMENT

Awarded [JXTX Foundation Scholarship](#) to present [limmaGenie](#), an AI-based assistant, at GBCC 2025, Cold Spring Harbor Laboratory, New York, USA

Received [International Grant](#) to participate in the International School of Chemometrics 2021 (Remote), University of Copenhagen

COURSES

Data Engineering and Cloud:

- [ETL and ELT in Python](#), Datacamp
- [Introduction to Microsoft Azure Cloud Services](#), Coursera

Machine Learning & Deep Learning:

- [Modeling with tidymodels](#), Datacamp
- [Winter School on Deep Learning](#), Indian Statistical Institute

Statistics and Probability for Data Science:

- [Advanced Probability: Uncertainty in Data](#), Datacamp
- [Multivariate Probability Distributions](#), Datacamp

Programming and Software Development:

- [Developing R Packages](#), Datacamp
- [Complete Full-Stack Web Development Bootcamp](#), Udemy

SCIENCE COMMUNICATION

[Presenter](#), The National Ploughing Championships Ireland 2024

Topic: AI and Animal Behaviour (with help of live facial emotion recognition demo)

<https://somu796.github.io/sudiptahazra.github.io/event/national-ploughing-championship/>

[Organizer](#), Pint of Science Dublin, 2024

Anchored and co-organized sessions related to society-focused science communication.

PROJECTS

limmaGenie: a RAG chatbot for limma statistical analysis

<https://limma-bot-cravd3dyaqhkcgf8.eastus-01.azurewebsites.net/>

[Python](#), [Azure](#)

- Developed a cloud-hosted LLM-based assistant for omics researchers performing differential expression analysis with the limma R package.
- Tech Stack: Programming language (Python), hosting (Azure Web App), user interface (ReactJS), LLM (Azure Open AI API, LangChain), data scraping (Beautiful Soup, Unstructured), Vector Database (MongoDB) and version control (Git, GitHub).

R & Bioconductor Workflow for Meat Proteomics

<https://somu796.github.io/proteomics-workflow-food-and-meat-science/>

[R](#)

- Created a R-based pipeline for Food and Meat Proteomics Data Analysis.
- Tech Stack: Programming language (R), data manipulation (tidyverse), UniProt database access (UniprotR), data container management (QFeature), protein expression analysis (limma), machine learning (tidymodel), ensemble learning (EFS), and visualization (ggplot2, pheatmap, ggvenn, UpSetR).

NLP analysis of Judicial Documents

<https://github.com/Somu796/NLP-to-analyze-Judicial-Documents>

[Python](#)

- Compared various classification algorithms to explore the advantages and limitations for text data analysis.
- Tech Stack: package management (.venv), PDF parsing (Tika), data management (Pandas), NLP & text mining (NLTK, re), machine learning (Scikit-learn, SciPy, TensorFlow), and visualization (Matplotlib).