

SUDIPTA KUMAR HAZRA

Data Analysis || PG.D. Statistics || B.E. Food Tech and Biochem Engg. || Machine Learning || Multivariate Stats Science Communication || Interdisciplinary Collaboration

Phone: +353 0892034307

Email: sudiptakh037@gmail.com



[Website](#)



[LinkedIn](#)



[Github](#)



SUMMARY

Data Science Professional with a background in statistics and food technology. Over 4 years of experience applying multivariate analysis, machine learning, and deep learning to spectroscopic and omics datasets. Skilled in Python, Scikit-learn, PyTorch, LangChain and OpenCV. Experienced in developing and deploying ML pipelines and web-based apps with modern frameworks, enhancing accessibility and automation.

PROFESSIONAL EXPERIENCE

Walsh Scholar

[Teagasc, Ireland](#) | 10/2023 - Present

Irish Agri-Food Research Institute

- Built [limmaGenie](#), a RAG-based LLM chatbot to simplify statistical analysis for omics data
- Analysing high-dimensional stress and meat proteomics data
- Stack used: R (tidyverse, car, tidymodels, ggplot2), Python

Statistical Analyst, AI/ML Team

[Kovaion Consulting, India](#) | 07/2022 - 06/2023

AI/ML Business Solution

- Built statistical and ML solutions for agri-food businesses.
- Collaboratively built computer vision and NLP solutions using OpenCV, spaCy, PyTorch, Flask, and Azure Cloud Services.
- Stack used: Python (scikit-learn, statsmodels), Azure Cloud.

Research Intern

[CSIR-CFTRI, India](#) | 01/2022 - 03/2022

Food Technology Research Institute

- Applied chemometrics to analyze multivariate NMR data for classification of 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

[Indian Statistical Institute, India](#) | 07/2021 - 06/2022

Scored 70.4%

B.E. in Food Technology & Biochemical Engineering

[Jadavpur University, India](#) | 07/2017-06/2021

Scored 76.61%

SKILLS

Data Science & Analytics

- Machine & deep learning
- Computer vision
- MLOps & Dev
- Statistical modeling
- Multivariate data analysis
- Chemometrics & omics data analysis

Programming Languages

- Python
- R
- MATLAB
- JavaScript

Tools & Platforms

- Azure
- SQL & MongoDB
- PLS Toolbox
- Power BI

CERTIFICATION

[Winter School on Deep Learning](#),

Indian Statistical Institute

[Intermediate PyTorch](#),
Datacamp

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

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

Machine Learning & Deep Learning:

- [Winter School on Deep Learning](#), Indian Statistical Institute
- [Introduction and Intermediate PyTorch](#), Datacamp

Data Engineering and Cloud:

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

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-cravd3dyaqhkegf8.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).

Facial Emotion Detection: a facial emotion recognition tool

<https://github.com/Somu796/emotion-detection-executable-app>

[Python](#), [OpenCV](#)

- Developed a real-time facial emotion detection tool using deep learning, packaged as a standalone app for accessibility in science outreach and educational settings.
- Tech Stack: image processing (OpenCV), GUI development (Tkinter), model format (ONNX), deep learning models (FER+, RFB-320), packaging (PyInstaller), version control (Git, GitHub).

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).