

Somak Sanyal

Data Scientist / Machine Learning Engineer

✉ sanyalsomakncl@gmail.com ☎ +44 7585692594 📍 Newcastle upon Tyne, U.K.

🌐 www.linkedin.com/in/somaksanyal 🐙 <https://github.com/somaksanyal97>

🔗 <https://sites.google.com/view/somaksanyal>

Data Scientist with 4 years of experience in machine learning, NLP, and generative AI. Proven track record in developing and optimizing models for business solutions. Seeking to leverage my expertise in data science and machine learning to enhance customer experience through data-driven insights and innovation.

Skills

Machine Learning • Deep Learning • Python • Natural Language Processing • Generative AI • Large Language Models • Databases (SQL) • Data Pre-processing and Feature Engineering • Data Analysis and Visualization • Distributed Computing • Databricks • Spark • Docker • Kubernetes • Mathematics and Statistics for AI • Big Data Analytics

Professional Experience

- 2024 – Present **Data Science Intern**, *Calnestor Knowledge Solutions* 🔗
- Developing and fine-tuning a **Gen AI based Tiny Language Model** for **Calnestor's Bharat Plus AI platform** 🔗, focused on enhancing text processing capabilities.
 - Creating interactive models to analyze and summarize educational data, providing actionable insights for institution recommendations.
- 2019 – 2022 **System Engineer**, *Tata Consultancy Services* 🔗
- Implemented and assessed classification algorithms for **Wells Fargo's new pension scheme, achieving 92% model accuracy.**
 - Leveraged customer features and 401K fund data for nuanced loan eligibility evaluation, improving decision-making accuracy by 30% and enhancing loan approval efficiency by 25%.
 - Served as a Database Administrator for a major UK retail chain, ensuring robust data management for over 100K employees, requiring critical thinking and attention to detail.
- 2018 – 2019 **Intern**, *Infosys* 🔗
- Completed training in Python and MySQL, contributing to a microservices-based project with an emphasis on clear communication and problem-solving within an Agile environment.

Education

- 2022 – 2023 **MSc. in Advanced Computer Science**, *Newcastle University, U.K.* 🔗
- GRADE: MERIT**
- RELEVANT MODULES: **Machine Learning, Engineering for AI (Big Data Analytics), Internet of Things, Cloud Computing**
- PROJECTS**
- 'Prediction of Disease Progression from Biomedical Data'** 🔗
- Developed a multi-class classification model achieving 66.6% accuracy with real patient data encompassing 224 cases.

- Systematically evaluated best combinations of data features, algorithms, and parameters to optimize model performance, enabling timely interventions based on early disease progression identification.

Link for Video Demonstration. [🔗](#)

Secure Ticketing System Implementation:

- Directed a team to deploy a secure ticketing system, reducing fraud risks by 30% and enhancing event attendee experience with a 50% faster check-in process.

2015 – 2019

B.Tech in Computer Science & Engineering, KIIT University, India [🔗](#)

GRADE: MERIT

RELEVANT MODULES: **Computer Graphics, Computational Intelligence, Database Management Systems, Design and Analysis of Algorithms**

PROJECT

Speed Detection of vehicles using image processing with OpenCV.

Publications

"EmoRecogSet: A Benchmark Singular Dataset for Integrated Face Recognition and Emotion Detection."

(Pre-print submitted in April '24)

- Developed a 117,000-image dataset for face recognition and emotion detection, with applications in data-driven model validation.
- Utilized CNN and transformer models to validate the dataset, demonstrating significant potential for AI research and development.

Accomplishments

- Recipient of **Vice-Chancellor's International Scholarship** [🔗](#) from **Newcastle University, U.K.** for pursuing Master's programme in Advanced Computer Science (2022 -2023).