Somak Sanyal

Data Scientist

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

www.linkedin.com/in/somaksanyal https://github.com/somaksanyal97

Data Scientist with over 4 years of experience in developing data-driven solutions across various industries. Expertise in handling large datasets, coding, data manipulation, and building scalable machine learning models. Adept at using statistical methods, data engineering tools, and visualization techniques to drive impactful decisions. Proficient in Python, SQL, and cloud services, with a strong background in data science methodologies and a passion for continuous learning and innovation.

Skills

Deep Learning (Tensorflow, Keras, PyTorch)

Statistical Modelling

Data Pre-processing and Feature Engineering

Cloud Services: AWS, Azure

Generative AI

Databases (SQL)

Software Development & Documentation (CI/CD, Docker, Kubernetes)

Machine Learning (Predictive Analytics, Customer Segmentation, Time Series Analysis)

Python (pandas, numpy, scikit-learn)

Data Analysis & Visualization (Matplotlib, Seaborn)

Natural Language Processing (NLTK, SpaCy)

Hypothesis Testing (A/B)

Mathematics and Statistics for AI

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**, **improving text processing capabilities by 40%.**
- Leveraged advanced NLP techniques to create interactive models for **end-to-end processing of large educational datasets**, delivering actionable insights with recommendations that **enable students to choose educational institutes 33% faster.**

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 401(k) 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*

- **Gained foundational experience in Python and MySQL,** contributing to microservices-based projects with a focus on data-centric applications, problem-solving, and clear communication in Agile environments.

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 developed using Python, **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.

Projects

Credit Card Fraud Detector
- Developed a Random Forest Classifier **achieving 83.3% ROC-AUC to accurately identify fraudulent transactions** in a highly imbalanced dataset.
- Conducted extensive EDA and feature engineering, enhancing model precision and recall in fraud detection for financial security.

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 Convolutional Neural Networks (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).