

0481709247

Proteek Kumar Sanyal

Data Scientist/Analyst

proteeksanyal1994@gmail.com<https://www.linkedin.com/in/proteeksanyal/>

PROFILE

With over five years of experience in managing, implementing data, specializing in developing and implementing Data Science, Data Engineering and AI models.

Skilled in generating tangible business value with proficiency in Azure Databricks, Generative AI, Computer Vision and creating proof-of-concept prototypes, with a robust background in the entire Software Development Life Cycle (SDLC) focusing on requirements gathering, analysis, design, and development.

Holds a Master of Data Science from Macquarie University, with a practical focus on business solutions.

TECHNICAL SKILLS & AREA OF EXPERTISE

Machine Learning Algorithms: Regression, Recommendation systems, Linear & Logistic Regression, Random Forests, SVM, Naive Bayes, k-NN, Clustering (K-means, hierarchical clustering), Naive-Bayes.

Deployment & Production: Model deployment (AWS, Azure, GCP), Docker, REST APIs, Git, CI/CD pipelines, Kubernetes

Deep Learning: CNNs, Transformers, BERT, NLP, CV

Databases & Big Data: SQL, MongoDB, Neo4j, Databricks

Programming Languages: Python, R, PySpark, Pytorch, Linux

Data Handling: Data preparation, Feature engineering, Data visualization (Power Bi)

Python Data Science Libraries: NumPy, Pandas, Matplotlib, Scikit-Learn, Seaborn, Tensorflow, Pytorch

Soft Skills: Confluence, Communication, Stakeholder management, Problem-solving, Team leadership, Mentorship.

EDUCATION & QUALIFICATIONS

Macquarie University – Master of Science in Data Science	2022-2023
GLA University – Bachelor of Technology in Computer Science and Engineering	2011-2015

PROFFESIONAL EXPERIENCE

ML Engineer – Intern | Truth July 2023 to Nov 2023

- Enhanced product robustness, scalability, and long-term usefulness.
- Completed high-level **Python** coding for **Machine Learning models**.
- Boosted design and deployment success by integrating advanced **ML techniques**.
- Extracted trends from **large datasets**, supporting business decisions.
- Integrated fine-tuned models with TestApp stream for **live app deployment**.
- Committed to team success by **collaborating** across departments to deliver practical solutions.

Project: User Liveness

- Conducted SOTA **literature** review on face anti-spoofing.
- Developed exploratory **data analysis** insights for large datasets.
- Fine-tuned data processing **pipelines** and integrated them with the current live app.
- Utilized **AWS** services for scalable storage and data processing, including S3 for data storage.

Majorly Used: AWS, Python, ML Algorithms, Confluence, Deep learning, Data Handling, Statistics, Soft Skills

Data Scientist | Infinite Computer Solutions (India) Limited

Aug 2021 to Dec 2021

- Analyzed structured and unstructured data for actionable **business insights**.
- Improved data science processes through enhanced storage, integration, and reporting strategies.
- Designed instruments and applications for optimized **data collection**.
- Utilized **Databricks** to implement scalable ML models for efficient data processing.

Project: Edison Health Link (EHL) Platform

- Developed **deep learning models** for medical image analysis tasks, including segmentation and detection.
- Utilized **CNN & computer vision** architectures for feature extraction from imaging data.
- Integrated deep learning algorithms with imaging systems for real-time insights.
- Employed **Databricks** to orchestrate data workflows and implement **ML models**, ensuring efficient data processing and real-time analysis.
- Collaborated with cross-functional teams to deliver business-focused solutions.

Majorly Used: Databricks, Azure, Python, Deep Learning, GIT, Soft Skills

Data Scientist | Tetra Soft India PVT. Ltd.

Dec 2020 to Jul 2021

- Delivered solutions using data-driven approaches and software needs insights.
- Created **algorithms** to handle special user and system requirements.
- Built scalable products using the latest technologies and approaches.
- Engineered data solutions using **Azure Data Factory** and **Databricks**, focusing on data processing and workflow automation, enabling accurate predictions.
- Led the data migration and warehousing projects, enhancing data integration and storage solutions.
- Collaborated closely with project stakeholders to gather and analyze requirements, ensuring alignment with business objectives.
- Conducted functional testing to validate the effectiveness of data solutions, **ensuring robustness and accuracy**.

Project: Passport 360

- Developed predictive models for assessing evacuation risk using **machine learning techniques** and managed the database schema, optimized **SQL** queries for performance enhancements.
- Presented project outcomes to key clients, facilitating workshops to discuss data-driven strategies and improvements.
- Conducted **feature engineering** to transform raw data into informative variables.
- Presented model results through insightful **visualizations**, reports, and presentations using **Power Bi**.
- Utilized Azure services, including Azure Data Factory for data pipeline automation and **Azure Databricks** for big data processing and analytics.
- Implemented holistic data science solutions using innovative approaches.

Majorly Used: CosmosDB, Databricks, SQL, Python, Linux, Data Handling, CI/CD, Power Bi, Soft skills

Data Scientist | HCL Technologies

Aug 2016 to Nov 2020

- Drafted software requirements and managed design and testing processes.
- Executed functional and integration testing and **software delivery support**.
- Built and tested **automation tools** for infrastructure provisioning.
- Explored data profiles and limitations to identify data relationships.
- Applied **analytical skills** to solve challenges without prior patterns.
- Managed database operations and maintenance, ensuring data integrity and security.
- Presented visualizations, reports, and presentations using **Power Bi**.
- Participated in the **development** and testing phases of software projects, employing a structured approach to software delivery.
- Coordinated with cross-functional teams to achieve project milestones and deliver on customer expectations.

Projects

Grasp Tool:

- Parsed documents using deep learning & **OCR** models with **Python**.

- Utilized **Azure** services, for data training & Processing.
- Created Web APIs with ASP.NET and C#.
- Integrated **RabbitMQ** for message queuing and **ELK stack** for logging.
- Utilized **AWS** for scalable infrastructure, employing services like S3 for data storage and **EC2** for compute resources. And employed **Databricks** to orchestrate data workflows and implement **ML models**, ensuring efficient data processing and **real-time analysis**.
- Developed holistic data science solutions by integrating technical expertise with innovative approaches.

Voice Platform:

- Developed **Microservices** for web request handling and **S3** bucket management.
- Implemented **BERT model** for sentiment analysis.
- Highlighted sentiments through accurate text-based sentiment analysis.
- Leveraging Databricks service, using of **Dockers & Kubernetes** for deployment.

Majorly Used: AWS services, Databricks services, Python, MongoDB, Data handling, Git, CI/CD, Power Bi, Soft skills

INDUSTRY KNOWLEDGE: Financial Services, Insurance, Healthcare, Storage