Proteek Kumar Sanyal

Data Scientist/ML Engineer

0481709247

proteeksanyal1994@gmail.com
https://github.com/proteek-dev
https://www.linkedin.com/in/proteeksanyal/

PROFILE

Experienced Data Scientist with over 6 years of expertise in building and deploying machine learning models, developing predictive algorithms, and managing large-scale data pipelines for diverse industries. Proficient in Python, R, SQL, and AI methodologies with a strong background in agricultural technologies and bioinformatics. Passionate about using data-driven solutions to optimize agricultural outcomes and drive business growth. Experienced in collaborating with cross-functional teams and stakeholders to deliver impactful research and development solutions.

TECHNICAL SKILLS & AREA OF EXPERTISE

ML & AI Models: Regression, Recommendation Systems, Clustering, SVM, CNNs, BERT, NLP, Computer Vision, Predictive Modeling
Data Tools & Pipelines: dbt, Azure Databricks, AWS Data Services, Azure Data Factory, Bioinformatics tools
Programming Languages: Python, R, PySpark, SQL, Linux
Cloud & Deployment: AWS, Azure, Google Cloud, Docker, Kubernetes, REST APIs, CI/CD, Git
Data Science Frameworks: scikit-learn, TensorFlow, NLTK, Jupyter Notebooks, PyTorch
Data Analysis & Visualization: Pandas, NumPy, Matplotlib, Tableau, Power BI
Geospatial & Agricultural Tools: GIS, remote sensing, satellite imagery
Big Data: Hadoop, Spark, Data lakes, Data warehousing, real-time data processing
Collaboration & Leadership: Stakeholder engagement, cross-functional teamwork, communication
Agile Methodologies: Scrum, sprint planning, team collaboration

KEY PROJECT

User Liveness Detection | Truuth   *Sydney, Australia*
Integrated liveness detection models within an existing application architecture to improve user verification. Developed scalable microservices using AWS Sagemaker for model deployment, in collaboration with Site Reliability Engineers. Achieved 30% improvement in detection accuracy and 20% increase in scalability for high-concurrency environments.

Medical Image Analysis Platform | Infinite Computer Solutions   *India*
Developed CNN-based models for segmentation, classification, and anomaly detection in medical imaging. Integrated with existing systems for real-time insights, achieving 25% diagnostic accuracy improvement and 20% fewer false positives. Implemented rigorous validation across diverse populations, ensuring model robustness and reliable insights.

Predictive Modeling for Evacuation Risk Assessment | Tetra Soft India   *India*
Built predictive models (logistic regression, decision trees, random forests) to assess evacuation risk based on geographic, geopolitical, and health factors, achieving 30% accuracy improvement. Delivered clear visualizations and reports, informing strategic decisions for stakeholders like insurance underwriters and executives.

Sentiment Analysis with BERT | HCL Technologies   *India*

Developed a sentiment analysis model to evaluate user feedback and reviews. Deployed a BERT-based model as a REST API and utilized Docker and Kubernetes for seamless scalability. Improved sentiment classification accuracy by 20%, achieving 99.5% uptime through containerized deployment.

PROFESSIONAL EXPERIENCE

Data Analyst   **City of Gold Coast, Gold Coast, Australia | Jun 2025 – Present**

- Designed and maintained **Power BI dashboards** using advanced DAX and Power Query for asset, compliance, and KPI reporting.
- Automated **data refresh and report distribution** with Power Automate, cutting manual work by 40%.
- Collaborated with business partners to **translate requirements into actionable insights** and data models.
- Applied **critical analysis** to identify performance gaps and support data-driven decision making across departments.
- Enhanced **data governance and literacy**, mentoring analysts and standardising reporting practices.
- Integrated data from **Azure, SharePoint, and internal systems** to deliver unified analytical solutions.

Key Technologies: Azure Databricks, Azure Machine Learning, Microsoft Fabric, Power BI, Python, SQL, SharePoint, Power Automate.

Data Scientist/Engineer   *ICI Pty Ltd, Gold Coast, Australia | Mar 2025 - May 2025*

- Designed **Power BI dashboards** and automated reports integrating sales and marketing data.
- Analyzed campaign ROI and attribution data to guide client strategy for brands like Flight Centre and ASUS.
- Developed **ETL pipelines** using Python, SQL, and Adverity to streamline reporting workflows.
- Supported Account Service teams with actionable insights to optimize campaign performance.

Key Technologies: AWS, Python, R, ML Algorithms, Data Handling, Confluence, Adverity, MySql

ML Engineer   *Truuth, Sydney, Australia | July 2023 – Nov 2024*

- Built, tuned, and deployed machine learning models using **AWS Sagemaker**, improving product performance and scalability.
- **Led the development** of scalable data pipelines for real-time data flow and seamless integration with live applications.
- Collaborated with cross-functional teams to define success metrics and deliver data-driven solutions.
- Applied **Agile methodologie**s in sprint planning and development, maintaining iterative progress.

Key Technologies: AWS, Python, ML Algorithms, Agile, CI/CD, Confluence, Data Handling

Data Scientist   *Infinite Computer Solutions, India | Aug 2021 – Dec 2021*

- Developed deep learning models for **medical image analysis** using CNNs, integrating them with AWS Sagemaker for real-time data processing.
- Designed scalable data pipelines and ensured efficient **data processing** and storage for large datasets.
- Communicated project insights to senior stakeholders, influencing strategic decisions.

Key Technologies: AWS, CNN, SQL, Python, Flask, scikit-learn, Docker, Data management

Data Scientist   *Tetra Soft India PVT. Ltd., India | Dec 2020 – Jul 2021*

- Engineered data pipelines using **Azure Data Factory and Databricks**, supporting big data transformations.
- Developed machine learning models for **predictive risk assessment**, delivering actionable insights and reports to leadership.

- Applied Agile principles to ensure timely delivery of high-quality solutions.

Key Technologies: Azure Data Factory, Databricks, Power BI, Python, SQL, scikit-learn

Data Scientist   *HCL Technologies, India | Aug 2016 – Nov 2020*
- Built and maintained large-scale data pipelines, ensuring seamless processing and integration with cross-functional teams, utilizing **Docker and Kubernetes** for scalable deployment..
- Developed and deployed a sentiment analysis model using BERT and contributed to product enhancements.
- Communicated data-driven insights to influence business decisions and strategy.

Key Technologies: AWS, Databricks, Python, MongoDB, Data handling, Git, CI/CD, Power BI

INDEPENDENT AI PROJECTS

AI for Maternal Health & Early Outbreak Warning | Open-Source MVP   *Medium Blog*

Developed and published an open-source AI prototype inspired by UNCTAD's research to detect high-risk pregnancies and early disease outbreaks. Built an end-to-end predictive pipeline using Python and scikit-learn, with interactive dashboards for public health monitoring. The project, showcased in a Medium article, highlights practical steps for applying AI in global health.

Link: [Maternal Health](#)

End-to-End LLM Projects Using LangChain, OpenAI, Hugging Face, and FAISS   *Medium Blog*

Authored an in-depth Medium blog post that walks readers through the end-to-end process of developing large language model (LLM) projects using LangChain, OpenAI, Hugging Face, and FAISS. The blog highlights best practices, architecture, and implementation steps for building LLM applications. This post has been well-received by the data science and AI community for its practical insights and actionable information.

Link: **[Langchain](#)**

EDUCATION & QUALIFICATIONS

Macquarie University – Master of Science in Data Science   2022-2024
GLA University – Bachelor of Technology in Computer Science and Engineering   2011-2015