Proteek Kumar Sanyal 0481709247

Data Scientist [proteeksanyal1994@gmail.com](mailto:proteeksanyal1994@gmail.com)

<https://github.com/proteek-dev>

<https://www.linkedin.com/in/proteeksanyal/>

PROFILE

Data Scientist with over 5 years of experience in building and deploying machine learning models, managing large-scale data pipelines, and collaborating across teams to deliver business solutions. Proficient in Python, SQL, and various data science frameworks including AWS Sagemaker, pandas, TensorFlow, and scikit-learn. Demonstrated leadership and stakeholder management skills, with a strong understanding of Agile development methodologies and big data principles.

TECHNICAL SKILLS & AREA OF EXPERTISE

**ML & AI Models**: Regression, Recommendation Systems, Clustering, SVM, CNNs, BERT, NLP, Computer Vision

**Data Tools & Pipelines**: dbt, Snowflake, Azure Databricks, AWS Data Services, Azure Data Factory

**Programming Languages**: Python, R, PySpark, SQL, Linux

**Data Science Frameworks**: pandas, scikit-learn, TensorFlow, NLTK, Jupyter Notebooks, PyTorch

**Cloud & Deployment**: AWS, Azure, Docker, Kubernetes, REST APIs, CI/CD pipelines, Git

**Data Analysis & Visualization**: Exploratory Data Analysis (EDA), Feature Engineering, A/B Testing, Regression Analysis, Power BI, Tableau

**Big Data**: Data lakes, data warehousing, distributed computing, real-time data processing

**Collaboration & Leadership**: Stakeholder engagement, cross-functional teamwork, communication skills

**Agile Methodologies**: Scrum, iterative development, sprint planning, team collaboration

KEY PROJECTS & ACHIEVEMENTS

**GovHack2023**

* Runner up for Cost-of-Living Pressure NSW sponsored by Department of Customer Service (New South Wales) and honourable mention for using open data to support Tasmanians sponsored by GovHack (Tasmania).

**User Liveness Detection | Truuth**

* Developed microservices for integrating liveness detection models into existing app architecture using AWS Sagemaker.
* Collaborated with site reliability engineers to ensure the scalability and reliability of deployed solutions.

**Edison Health Link (EHL) Platform | Infinite Computer Solutions**

* Developed REST APIs for medical image analysis models, integrating them with existing imaging systems for real-time insights.
* Used Docker and Kubernetes for containerization and orchestration of the models.

**Predictive Modelling for Evacuation Risk Assessment | Tetra Soft India**

* Designed and developed microservice-based predictive models to assess evacuation risks, utilizing SQL for data extraction and analysis.
* Automated data pipelines to enhance the scalability and reliability of the system.

**Sentiment Analysis with BERT Model | HCL Technologies**

* Developed a BERT model for sentiment analysis, implemented as a REST API for user reviews and feedback.
* Utilized Docker and Kubernetes for containerization and deployment, ensuring scalability and high availability.

PROFFESIONAL EXPERIENCE

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

* Built, tuned, and deployed machine learning models using AWS Sagemaker, improving product performance and scalability.
* Maintained data pipelines and integrated them with live applications, ensuring real-time data flow and reliability.
* Led cross-functional meetings to align teams on success metrics and business goals, facilitating smooth project delivery.
* Applied Agile methodologies in sprint planning and development to maintain fast-paced, iterative progress.

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

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

* Developed and deployed deep learning models for medical image analysis using CNNs, integrating them with AWS Sagemaker for real-time processing.
* Led the design of scalable data pipelines, ensuring efficient data processing and storage for large datasets.
* Communicated project outcomes and insights to senior stakeholders, influencing decision-making and strategy refinement.

*Key Technologies*: AWS Sagemaker, CNN, SQL, Python, Flask, scikit-learn

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

* Engineered and maintained data pipelines using Azure Data Factory and Databricks, supporting big data transformations.
* Developed machine learning models for risk assessment, presenting insights and reports to key clients and leadership teams.
* Utilized Agile principles to manage project timelines, ensuring timely and high-quality deliverables.

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

**Data Scientist  
HCL Technologies,** India **| Aug 2016 – Nov 2020**

* Developed and maintained data pipelines for large-scale data processing, working closely with cross-functional teams to meet project goals.
* Built and deployed sentiment analysis models using BERT, contributing to product development and enhancement.
* Managed stakeholder expectations and presented data-driven insights, influencing business decisions and strategic direction.

*Key Projects*:

* **Grasp Tool**: Built OCR models for document processing, integrated with Azure and AWS for scalable deployment.
* **Voice Platform**: Implemented sentiment analysis using BERT and managed cloud services with Databricks and Kubernetes.

*Key Technologies*: AWS services, Databricks services, Python, MongoDB, Data handling, Git, CI/CD, Power Bi, Soft skills

EDUCATION & QUALIFICATIONS

Macquarie University – Master of Science in Data Science 2022-2024

GLA University – Bachelor of Technology in Computer Science and Engineering 2011-2015

ADDITIONAL INFORMATION

* Experienced in Agile methodologies, proficient in sprint planning, backlog management, and iterative development.
* Strong stakeholder management skills, capable of leading and influencing teams to deliver complex data-driven solutions.
* Proficient in developing data pipelines and integrating them with client-side systems to ensure seamless data flow.