Proteek Kumar Sanyal 0481709247

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

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

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

PROFILE

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*

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 (Edison Health Link)*

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, actionable insights during image acquisition and interpretation.

**Predictive Modeling for Evacuation Risk Assessment** | *Tetra Soft 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**. Conducted advanced feature engineering to enhance model insights. Delivered clear visualizations and reports, informing strategic decisions for stakeholders like insurance underwriters and executives.

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

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.

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 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.