

Vineeth Kumar Vellala

vellalavineethkumar@gmail.com ❖ +44 7824 864338 ❖ G41, 5 Burley Road, Leeds, United Kingdom

SUMMARY

Data Science professional with nearly 4 years of hands-on experience in data management, preprocessing, feature engineering, and model development. Currently pursuing an MSc in Data Science and Analytics, specializing in the application of generative AI for synthetic data generation and customizing large language models. Proven ability to propel data-driven solutions for business growth, ensuring data integrity and compliance. Passionate about mentoring and actively seeking opportunities to leverage my expertise in data science to guide and inspire the next generation of data scientists in a dynamic and challenging environment.

WORK EXPERIENCE

NHS Digital (NHSx)

Mar 2024 – Present

Dissertation Partner – part-time

Leeds, UK

- Leveraged statistical analysis and Generative AI techniques to generate synthetic EHR data, reducing reliance on real patient data by 75%, while ensuring analytical utility.
- Validated synthetic data, achieving a 95% match rate with key health indicators from genuine EHR data.
- Ensured 100% GDPR compliance in synthetic data generation methods, safeguarding patient privacy.
- Presented monthly updates to NHS stakeholders driving iterative feedback and project advancement.

Informatica

Jan 2020 – Sept 2023

Informatica Data Engineer/Technical Consultant – full-time

India

- Analysed and processed over 10M+ records using advanced data processing techniques, improving data retrieval and management for analytical tasks.
- Structured and managed large datasets across 100+ tables, optimizing data workflows for data scientists.
- Prepared and standardized data from over 10 source systems using advanced data cleansing techniques, ensuring consistency and quality for downstream systems.
- Leveraged feature engineering techniques and fine-tuned an XGB model, achieving 93% precision in matching fuzzy records for data deduplication processes.

Samsung R&D

Jan 2018 – Jan 2019

Data Science Project Trainee – Internship

India

- Led the development of an 'Object detection and tracking for 360° videos' project in collaboration with Samsung architects, ensuring alignment with industry standards through the implementation of YOLO v3 and Deep SORT algorithms.
- Implemented a detection (YOLO v3) and tracking (Deep SORT) pipeline, reducing manual work by 90%.
- Earned the "Project of the Year" award and had research findings published at the international conference [PReMI 2019](#).

SKILLS

Data Science and Analytics: Statistical Analysis, Hypothesis Testing, A/B Testing, Time Series Analysis, Experiment Design, and Model Evaluation.

Machine Learning and AI: Generative AI, Machine Learning (Scikit-learn, TensorFlow, Keras, PyTorch), Deep Learning, Feature Engineering, Data Mining, Computer Vision, Natural Language Processing (NLP).

Programming and Scripting: Python, R Programming, SQL, Shell Scripting.

Data Management and Processing: ETL (Extract, Transform, Load), Data Quality, Data Cleaning, Data Pipelines, Data Modelling, Data Management.

Big Data Technologies: Hadoop, Spark, Apache Kafka.

Cloud Computing: AWS, Azure.

Data Visualization: Matplotlib, Seaborn, Tableau.

Database Management Systems: MySQL, PostgreSQL, MongoDB.

Version Control: Git.

Soft Skills: Quantitative Analysis, Problem-Solving, Collaboration, Adaptability, Mentoring.

EDUCATION

University of Leeds

Expected September 2024

MSc Data Science and Analytics, School of Mathematics

Leeds, UK

In-depth coursework and dissertation covering

- Statistical Theory and Methods, Programming for Data Science.
- Data Science and case studies, Big Data and Consumer Analytics.
- Deep Learning, Applied Machine Learning.
- Business Analytics and Decision Sciences.

KLE Technological University

Aug 2016 – July 2020

Bachelor of Engineering in Computer Science

India

- Secured a top 5% CGPA of 9.18 with coursework in Data Structures, DBMS, Data Mining, Big Data, etc.
- Winning Team, Smart India Software Hackathon (MHRD, Govt of India) 2019 for the project Real-Time Multi-Person Recognition and Tracking for CCTV camera footage in High Security Areas.
- Undergraduate Research Fellow at AI and Data Analytics Lab, provisioning NVIDIA DGX to 4000+ students.

CERTIFICATIONS & INTERESTS

Certifications

- Deep Learning Specialisation by Coursera
- Machine Learning for Engineering and Science Applications by IIT, Madras.
- Master Data Management SaaS Implementation practitioner by Informatica University.

Interests

- Regular reader of advancements in Gen AI and Data Science.
- Actively mentor schoolgirls promoting education in STEM as part of the Leeds Institute for Data Analytics (LIDA) Open Data Science for Schools initiative.
- Trek Organiser for the "Get Out, Get Active" program at the University of Leeds.

Note: No Visa Sponsorship needed, I am on the PSW scheme and can self-sponsor.

References available upon request.