

SOHIL ANANTH R

Leicester, United Kingdom — +44 7747 75648 — sohilananth109@gmail.com

[LinkedIn- Sohil](#) — [GitHub- SxR24](#)

PROFILE

MSc Bioinformatics candidate (University of Leicester, graduating Sep 2026) specialising in data analysis, statistical validation, and large-scale computational workflows. Proficient in building end-to-end data pipelines, classification models, and analytical dashboards on real-world biomedical datasets (single-cell RNA-seq, clinical prescriptions, customer risk stratification). Skilled in Python, R, SQL, and reproducible computational practices (Git, Conda, HPC environments). Proven ability to translate complex bioinformatics analysis into actionable insights for non-technical stakeholders.

TECHNICAL SKILLS

Languages & Machine Learning: Python, R, SQL, XGBoost, scikit-learn, Pandas, NumPy, feature engineering, model validation, predictive modelling

Data Processing & Pipelines: ETL workflows, data cleaning, reproducible workflows, Jupyter, Streamlit, FastAPI, Git/GitHub, Conda, statistical hypothesis testing

Bioinformatics & Genomics: Single-cell RNA-seq (scRNA-seq), STAR, Salmon, TopHat, Seurat, scran, Biopython, PCA, Leiden/SNN clustering, UMAP/t-SNE, GRCh38, SRA Toolkit, sequence analysis

Statistics & Validation: Hypothesis testing (t-test, chi-square, Spearman/Pearson correlation), model evaluation, cross-validation, statistical significance testing, power analysis

Data Visualisation & Business Intelligence: Power BI, Matplotlib, Seaborn, interactive dashboards, R Shiny, decision-support reporting

Reproducibility & Infrastructure: HPC environments (SLURM job scheduling), version control workflows, containerisation, automated analysis pipelines

RELEVANT PROJECTS

Single-Cell RNA-seq Classification & Pipeline Modernisation

Feb 2026 – Apr 2026

MSc Steered Research Project

- Reproduced and extended end-to-end scRNA-seq classification across 734 single-cell libraries (587 GB dataset), achieving 99.86% concordance with published reference matrix; validated pipeline congruence with paired t-tests and chi-square statistical tests.
- Re-engineered legacy alignment pipeline: STAR and Salmon alignment, scran size-factor normalisation, multi-metric quality control (QC), Seurat/Leiden clustering; achieved 30x speedup in alignment time while improving cell-quality filtering sensitivity.
- Classified cells into progenitor, neuronal, and zonal subpopulations using PCA dimensionality reduction, SNN/Leiden graph clustering, marker-gene overlays, and Spearman correlation to bulk references; built interactive R Shiny portal with lazy-loading and in-memory caching for stakeholder-friendly exploration.
- Demonstrated reproducibility best practices: Conda environments, Git version control, modular code architecture, automated QC reporting. Code: github.com/SxR24/BS7120-GRP_ASSIGNMENT.

Telco Customer Churn Intelligence

Feb 2025 – Apr 2025

End-to-End ML Pipeline & Dashboard

- Built and validated XGBoost classification model for customer risk stratification; implemented train/test/validation split strategy with cross-validation and feature-importance analysis to identify high-value retention targets.
- Engineered 40+ derived features via domain-driven feature engineering; performed exploratory data analysis (EDA) and statistical validation (correlation analysis, chi-square tests for categorical relationships).
- Translated model predictions into Power BI decision-support dashboard for non-technical stakeholders; version-controlled complete workflow on GitHub (Python, XGBoost, Pandas, Power BI).

NHS Real-Time Drug Spend Analytics

Mar 2025 – Apr 2025

Data Pipeline & BI Solution

- Engineered ETL pipeline over 109M+ prescriptions and £878M spend dataset; integrated multi-source clinical records, resolved data quality issues, and surfaced prescribing signals.
- Built interactive Power BI dashboard for non-technical stakeholders; Python (Pandas, NumPy), Streamlit prototype, SQL queries for data extraction and aggregation.

Advanced Sequence Motif Finder

Nov 2024 – Feb 2025

Bioinformatics Web Application

- Developed and deployed motif-search tool for DNA/RNA/protein sequence analysis; implemented efficient pattern-matching algorithms using Biopython and Pandas.
- Live deployment: sequencemotiffinder.streamlit.app. Tech stack: Python, Biopython, Streamlit, interactive visualisations.

PROFESSIONAL EXPERIENCE

Bioinformatics Intern — BioBerry

2024

Drug Development & Computational Modelling

Applied computational analysis and bioinformatics modelling to drug-development workflows in biomedical R&D environment.

Business Development Executive — Etiqu*Sep 2024 – Jan 2025**Client Engagement & Problem-Solving*

Engaged with clients on technical and business requirements; developed communication and solution-design skills; balanced professional delivery alongside independent research projects.

Python Tutor — Bishop Heber College*Apr 2021 – Nov 2021**Technical Education & Communication*

Delivered structured tutoring on quantitative and programming concepts to mixed-ability learners; created clear, accessible learning resources and notes.

Vice President, Bioinformatics Department — Bishop Heber College*2022 – 2024**Leadership & Coordination*

Coordinated academic and departmental initiatives across multiple stakeholders; demonstrated ownership, reliability, and collaborative project delivery; fostered inclusive teamwork within department.

EDUCATION**University of Leicester — Leicester, UK***Sep 2025 – Sep 2026**MSc Bioinformatics, School of Biological Sciences*

Coursework and examinations complete; dissertation in progress. Specialisation: single-cell genomics, statistical analysis, reproducible computational workflows.

Bishop Heber College — Tamil Nadu, India*2021 – 2024**BSc Bioinformatics, CGPA: 7.8/10*

First-class equivalent (UK 2:1); foundation in molecular biology, bioinformatics algorithms, statistics, and programming.

CERTIFICATIONS & ADDITIONAL INFORMATION

Certifications: Google Data Analytics Professional Certificate (Coursera) — Crash Course on Python (Google & Coursera)

Languages: English (Fluent), Tamil (Native), Hindi

Availability: Full-time, immediate