

# MD TANVIR SARDARR - Aspiring Data Scientist

Address: Chester, United Kingdom

Email: [tanvir.xf@gmail.com](mailto:tanvir.xf@gmail.com) | Phone: +447438888381 | Portfolio: <https://tanvirs.uk> | LinkedIn: <https://www.linkedin.com/in/tanvir153/> | GitHub: <https://github.com/mdtanvirsardarr/>

## Professional Summary

Aspiring Data Scientist focused on building trustworthy and interpretable AI systems, with expertise in model calibration, interpretability, and end-to-end deployment with rigorous MLOps practices.

## Education

- **MSc Data Science, University of Chester, United Kingdom** (Oct 2024 - Oct 2025)  
**Modules:** Principles of Data Science, Machine Learning, Statistical Programming, SQL and NoSQL Databases, Enterprise Development, and Research Methods.  
**Dissertation focus:** A web-based Student dropout risk prediction system with calibrated Random Forest probabilities, SHAP explanations, and Monte Carlo simulation.
- **BEng Software Engineering, Zhengzhou University, China** (Sept 2018 – Aug 2022)  
87% marks, CGPA 3.45 / 4  
**Key subjects:** Programming in C and Python; Data Structures and Algorithms, DBMS, Software Requirements Engineering, Computer Networks, Operating Systems, AI fundamentals, Calculus, Probability and Statistics.

## Certifications

IBM Data Analyst Certificate – in progress

Microsoft Power BI Certificate – in progress

Bangladesh University of Engineering & Technology (BUET) - Python Bootcamp 2023

## Core Skills

**Programming:** Python

**Data & EDA:** Pandas, NumPy, SciPy

**ML:** scikit-learn pipelines; RF, Linear/Logistic Reg, KNN, SVM; XGBoost, cross-validation; hyperparameter tuning; class-imbalance handling

**Time Series:** ARIMA/SARIMAX, ETS, VAR

**Evaluation & Calibration:** ROC-AUC, PR-AUC, F1, MAE/RMSE; Brier, Reliability diagrams, ECE

**Data & APIs:** SQL/SQLite/SQLAlchemy; MongoDB; JSON/CSV; Flask/REST

**Viz & Workflow:** Matplotlib; Plotly, APScheduler

## Projects

### Web-Based Student Dropout Risk Prediction System-

- Calibrated Random Forest with reliability diagrams and expected calibration error; SHAP for global and local explanations; request-level audit logs and CSV or JSON export.
- Role-aware UI for students, staff, and admins. Result on held-out test set: ROC-AUC 0.86, F1 0.80, precision 0.78, recall 0.82.
- Stack: Python, scikit-learn, SHAP, Flask, SQLAlchemy, SQLite, Pandas, NumPy, Matplotlib.

### Web-Based Lifestyle Disease Risk and Recommendations -

- Calibrated decision tree for risk scoring and a KNN recommender and suggests practical lifestyle actions grounded in BMI rules.
- Secure Flask service with CORS, pagination, CSV export, and SHAP feature attributions.
- Stack: Python, scikit-learn, Flask, SQLite, SQLAlchemy, pandas, NumPy.

### Web-based Burnout Guard system-

- Fusion of structured HR features with VADER or TextBlob sentiment extracted from short texts.
- XGBoost champion model compared to logistic regression baseline using PR-AUC and Brier; calibrated probabilities logged to SQLite for audit.
- Representative run recorded XGBoost ROC-AUC near 0.76 and logistic regression ROC-AUC near 0.758.
- Role-aware UI for users and admins.
- **Stack:** Python, scikit-learn, XGBoost, VADER, Text Blob, Flask, SQLite.

### Dermatology Data Analysis - supervised and clustering study

- Pipeline includes EDA, scaling with StandardScaler, and cross-validation; comparison of Random Forest, KNN, k-means, DBSCAN, and GMM.
- Results: Random Forest reached near-perfect accuracy on full and histopathology subsets; KNN recorded about 98.65%.
- Stack: Python, scikit-learn, pandas, NumPy, Matplotlib; JSON figure logs.

## Achievements

Outstanding Student Award- 2021 (Zhengzhou University)

## Leadership & Campus Involvement

Students' Union Election, Former President Candidate - May 2025 (University of Chester)

## Additional Information

Languages - English, Chinese, Bangla

## References

Available on request