Data Scientist

NUMAAN FAROOQ

(+91) 7006431667 | numaanfarooqds@gmail.com
LinkedIn | GitHub | Portfolio

Summary

Highly skilled Data Scientist with a focus on bioinformatics, machine learning, and predictive modeling. Expert in deep learning, computer vision, and natural language processing (NLP) for solving complex problems. Experienced in applying advanced techniques for medical image segmentation, predictive modeling, and data-driven insights. Proven ability to deliver high-accuracy models and optimize performance through analytical rigor and innovation. Strong collaborator with a passion for using data to drive impactful solutions.

Work Experience

Machine Learning Intern

Unified Mentor | January 2025 – February 2025

- Designed and implemented scalable machine learning pipelines for predictive analytics tasks.
- Optimized machine learning models to enhance prediction accuracy by 15%.
- Conducted data cleaning and preprocessing for large datasets, ensuring quality input for training models.
- Leveraged TensorFlow, Scikit-learn, SQL,Python for model development and deployment.

Bioinformatics Intern

Dept. of Fisheries, SKAUST | February 2025 - March 2025

- Conducted karyotyping and DNA extraction for aquatic species, achieving 95% sample purity
- Developed Python scripts for sequence alignment and phylogenetic tree visualization
- Created interactive dashboards for genomic data analysis using Plotly and Matplotlib
- Collaborated on research papers analyzing genetic diversity in fish populations

Managed Network Expert (Data Analysis)

Chegg | December 2024 – Present

- Providing expert solutions for statistical and data analysis queries, maintaining a 95% accuracy rate.
- Analyzed complex datasets and created visual representations to assist students with advanced data concepts.
- Collaborated with the global network to implement best practices in data analysis solutions.

Machine Learning Intern

Cognifyz Technologies | November 2024 – December 2024

- Collaborated on predictive modeling projects to optimize business performance metrics.
- Conducted data cleaning and preprocessing for large datasets, ensuring quality input for training models.
- Applied machine learning techniques to generate actionable insights from unstructured data.
- Collaborated on predictive modeling projects to optimize business performance metrics.
- Enhanced model accuracy by 12% through feature engineering and hyperparameter tuning.
- Used SQL, Numpy, Pandas, Python.

Data Science Intern

Cognifyz Technologies| September 2024 – October 2024

- Collaborated on predictive modeling projects to optimize business performance metrics.
- Conducted data cleaning and preprocessing for large datasets, ensuring quality input for training models.
- Applied machine learning techniques to generate actionable insights from unstructured data.
- Collaborated on predictive modeling projects to optimize business performance metrics.

- Enhanced model accuracy by 12% through feature engineering and hyperparameter tuning.

Brain Tumor Segmentation Project
Personal Project | *Oct 2023 – May 2024*

- Developed and implemented a U-Net-based deep learning framework for segmenting brain tumors from multimodal MRI scans.
- Leveraged TensorFlow, Scikit-learn, and Python for model development and deployment.
- Achieved high segmentation accuracy by leveraging advanced preprocessing and augmentation techniques.
- Utilized TensorFlow and Keras for model development, optimizing performance through extensive hyperparameter tuning.

Retinal Vessel Segmentation Project
Personal Project *| 2023 – July 2024*

- Designed a U-Net architecture to segment retinal vessels from fundus images effectively.
- Focused on preprocessing techniques such as image resizing, normalization, and contrast enhancement to improve model performance.
- Conducted experiments with various learning rates and optimizers to achieve optimal accuracy.

Skills

- **Programming:** Python, R, SQL
- **Machine Learning Frameworks:** TensorFlow, Keras, PyTorch
- **Data Analysis:** Pandas, NumPy, Scikit-learn
- **Visualization:** Tableau, Matplotlib, Seaborn
- **NLP & CV:** OpenCV, NLTK
- **Cloud Platforms:** AWS, Google Cloud
- **Tools:** Git, Jupyter, Hadoop

Projects

Plant Disease Prediction System | *Deep Learning, Computer Vision*

- Developed CNN model achieving 94.2% accuracy on Plant Village dataset (38 classes)
- Implemented data augmentation pipeline using TensorFlow Keras (rotation, flip, zoom)
- Optimized MobileNetV2 architecture with 87% recall for early disease detection
- Deployed model as REST API using Fast API.

Brain Tumor Segmentation and Prediction

- Developed a U-Net-based segmentation model for identifying brain tumor regions with 93% accuracy.
- Applied ensemble techniques for tumor classification to enhance diagnostic precision.
- Leveraged TensorFlow, Python, and advanced augmentation strategies for robust performance.

Retinal Vessel Segmentation

- Built a segmentation model using U-Net to isolate retinal vessels in fundus images with 94% accuracy.
- Improved model performance through advanced preprocessing and optimizer experimentation.

Fake News Prediction

- Created an NLP pipeline to classify fake news articles using machine learning algorithms.
- Achieved an accuracy of 88% with a 34% lift, analyzing large textual datasets for insights.

Cervical Cancer Prediction

- Designed a predictive model assessing cervical cancer risk based on clinical data.
- Applied statistical techniques to extract insights and enhance prediction reliability.

Breast Cancer Prediction

- Developed an early-diagnosis machine learning model achieving high predictive accuracy.
- Extracted actionable insights through exploratory data analysis and feature selection.

Education

BACHELORS (HONS) IN BIOTECHNOLOGY – University of Kashmir – Srinagar, Jammu & Kashmir June 2022-June 2026
Majors: Biotechnology, Biological Data Analytics, Bioinformatics, Bacteriology
Average CGPA- 7.5

Certifications

- IBM- Introduction to Cloud Computing
- IBM- What is Data Science
- IBM- Introduction to Web Development with HTML, CSS, JavaScript
- IBM - Tools for Data Science