

AFRAH FATHIMA S

Boston | MA | +1 617-397-8937 | shahabuddin.af@northeastern.edu | LinkedIn | GitHub | Portfolio

SUMMARY

Computer Science graduate with a Master's degree from Northeastern University, specializing in Data Science, Machine Learning, and AI. Experienced in developing predictive models, conducting statistical analysis and deploying ML solutions using Python, SQL, TensorFlow, PyTorch, and scikit-learn. Hands-on experience with end-to-end data pipelines, API development (FastAPI, Flask), and cloud platforms (AWS, GCP) gained through internships and academic projects.

EDUCATION

Northeastern University, Boston, MA

Sep 2024 – Present

Master of Science in Computer Science

Relevant Coursework: Algorithms, Machine Learning, Essentials of Data Science, DBMS, Cloud Computing, Web Development

Anna University, Chennai, India

Aug 2020 – May 2024

Bachelor of Engineering in Computer Science and Engineering

TECHNICAL SKILLS

Programming Languages: Python, SQL, C, R, Bash, Java

ML and NLP: Scikit-learn, TensorFlow, PyTorch, Keras, SpaCy, CNNs, A/B Testing

Statistical Methods: Hypothesis Testing, Statistical Modeling, Experiment Design, Probability Distributions

Data Engineering & APIs: Apache Spark, Airflow, FastAPI, Flask, REST APIs, PostgreSQL, MongoDB, MySQL

AI & LLM Tools: OpenAI APIs, Hugging Face Transformers, LangChain, Prompt Engineering

Analysis / Visualization: Pandas, NumPy, Matplotlib, Seaborn, Plotly, Tableau, Power BI, Excel

Cloud & DevOps: Docker, Git, AWS (S3, SageMaker, Lambda, EC2), GCP, Terraform, GitHub Actions, Postman

WORK EXPERIENCE

Khoury College of Computer Sciences, Boston | Graduate Teaching Assistant

Sep 2025 – Present

- Mentored students in **object-oriented programming**, **software design patterns**, and Java development through interactive lab sessions and hands-on coding exercises.
- Conducted **code reviews** and provided **debugging assistance** to improve students' programming logic, code structure, and software design quality.
- Guided students in applying **unit testing**, version control, and **collaborative development** practices to build scalable, maintainable software systems.
- Supported students during office hours with design decisions, debugging, and conceptual understanding of advanced OOP principles and design patterns.

Brainvault Technologies, India | Software Developer

Jan 2024 – Jun 2024

- Designed and developed an AI-powered travel assistant for TravelNet Solutions using **Python**, **OpenAI APIs**, **Hugging Face Transformers**, **FastAPI**, **MySQL**, BeautifulSoup, and Nominatim API, automating itinerary creation, location mapping, and surprise trip suggestions.
- Delivered features that **reduced travel planning time by 80–90%** while enhancing client engagement, demonstrating direct business impact.
- Implemented **end-to-end LLM-based extraction workflows with Gemini, LLaMA, and PDFPlumber**, achieving **90–95%** field-level accuracy and significantly reducing manual data processing.
- Collaborated cross-functionally with product, design, and engineering teams to scope, prototype, and deploy AI-powered tools, **ensuring reliability, scalability, and operational efficiency**.
- Leveraged infrastructure automation and production deployment** practices to streamline internal workflows and accelerate solution rollout.

Waycool Food and Products Pvt. Ltd, India | Machine Learning Engineer

Sep 2022 – Dec 2022

- Designed and implemented a **CNN-based computer vision** pipeline to automatically identify, count, and measure grain sizes, improving accuracy to **85–90%** and significantly enhancing consistency over traditional methods.
- Optimized model performance through advanced feature extraction, adaptive thresholding, and custom preprocessing, **reducing computational time by 85%** while maintaining high precision.
- Conducted **data preprocessing**, **statistical validation**, and performance benchmarking, transforming raw vision data into actionable insights for quality control and process improvement.
- Collaborated with senior data scientists, iteratively refining the model and pipeline, and **earned top performance scores** and a commendation for quality, impact, and technical rigor.
- Gained practical expertise in **computer vision**, **deep learning**, and **model optimization**, demonstrating readiness to apply ML solutions in production and real-world environments.

PROJECTS

AI-Driven Financial Risk Management System | GitHub Repo

Sep 2025 – Dec 2025

- Built an end-to-end credit risk prediction pipeline processing 30K financial records, achieving **ROC-AUC of 0.78–0.82** through ensemble modeling and enabling automated risk stratification for lending decision support.
- Applied **SMOTE for class imbalance handling** and conducted feature importance analysis to identify key default predictors, improving model recall by 15% and providing actionable insights for risk assessment strategies.
- Built an **ensemble stacked model** combining multiple classifiers, achieving **81.9% accuracy** and generating interpretable AI-powered risk explanations for transparent decision-making in financial workflows.

Online Purchase Intent Prediction | GitHub Repo

Sep 2025 – Dec 2025

- Designed and deployed **machine learning system** predicting e-commerce customer purchase probability with **86.5% accuracy**, enabling **data-driven marketing interventions** with projected **15–20% reduction in cart abandonment**.
- Implemented **intelligent risk assessment engine** analyzing browsing behavior (page views, bounce rates, engagement metrics) to generate **real-time actionable recommendations** with **confidence scoring** and **intervention timing optimization**.
- Created **comprehensive analytics dashboard** with **live performance monitoring**, model comparison visualizations, and prediction history tracking; integrated **3 ML models** with **ensemble methodology** for robust predictions.

Predictive Modeling for Loan and Insurance Cost Estimation | GitHub Repo

Sep 2023 – Dec 2023

- Developed predictive analytics solutions for automated loan eligibility assessment and insurance premium estimation, improving forecast accuracy by **10% (loans)** and **8% (insurance)** to optimize underwriting and pricing strategies.
- Conducted **feature importance analysis**, **hypothesis testing**, and **A/B scenario simulations** to optimize risk assessment and pricing strategies, delivering data-driven recommendations for business decision-making.
- Applied **statistical validation and model evaluation techniques** (cross-validation, ROC curves, confusion matrices) to ensure robust, actionable insights for underwriting and premium calculation.