

# AFRAH FATHIMA S

Boston | MA | +1 617-397-8937 | shahabuddin.af@northeastern.edu | [LinkedIn](#) | [GitHub](#)

**SUMMARY:** Graduate Computer Science student at Northeastern University with a specialization in Data Science and AI. Experienced in building AI-enabled solutions using GenAI, LangChain, RAG, and cloud technologies like AWS and GCP. Proficient in Python, NLP, vector databases, CI/CD pipelines, and prompt engineering. Demonstrated success in automating complex workflows using LLMs and agentic frameworks. Eager to contribute to AI enablement initiatives at scale.

## EDUCATION

---

<b>Northeastern University</b> , Boston, MA Master of Science in Computer Science	Sep 2024 – Present
<b>Relevant Coursework:</b> Algorithms, Machine Learning, Essentials of Data Science, DBMS, Cloud Computing	

---

<b>Anna University</b> , Chennai, India Bachelor of Engineering in Computer Science and Engineering	Aug 2020- May 2024
--	--------------------

## TECHNICAL SKILLS

---

<b>Programming Languages:</b>	Python, SQL, R, C, Java
<b>ML and NLP:</b>	Scikit-learn, TensorFlow, PyTorch, Keras, SpaCy, CNNs, A/B Testing
<b>Statistical Methods:</b>	Hypothesis Testing, Statistical Modeling, Experiment Design, Probability Distributions
<b>AI &amp; LLM Tools:</b>	OpenAI APIs, Hugging Face Transformers, LangChain, Prompt Engineering
<b>Data Engineering &amp; APIs:</b>	Apache Spark, Airflow, FastAPI, REST APIs, PostgreSQL, MongoDB, MySQL
<b>Analysis/Visualization:</b>	Pandas, NumPy, Matplotlib, Seaborn, Plotly, Tableau, Power BI, Excel
<b>Cloud &amp; DevOps:</b>	Docker, Git, AWS(S3, SageMaker,Lambda,EC2), GCP, Terraform, GitHub Actions, Postman

## WORK EXPERIENCE

---

<b>Brainvault Technologies</b> , India   <b>Software Developer Intern</b>	Jan 2024 – Jun 2024
<ul style="list-style-type: none"><li>Built a smart assistant using <b>OpenAI APIs</b>, Hugging Face Transformers, and FastAPI to generate automated travel itineraries and perform <b>location-based RAG-style retrieval</b>, reducing planning time by 80–90%.</li><li>Integrated <b>LangChain</b> to implement agentic workflows with custom prompting and embedding-based search.</li><li>Delivered features such as automated itineraries, surprise trip planning, and nearby attraction suggestions, <b>reducing planning time by 80–90%</b> and improving client engagement.</li><li><b>Fine-tuned and evaluated LLMs (Gemini, LLaMA)</b> for domain-specific text parsing with 90–95% accuracy, automating PDF extraction workflows using <b>PDFPlumber</b> and reducing manual review time significantly.</li><li>Collaborated cross-functionally with product and engineering teams to prototype and operationalize ML/Generative AI solutions, demonstrating experience in fast-paced, startup-style environments.</li></ul>	
<b>Waycool Food and Products Pvt. Ltd</b> , India   <b>Machine Learning Intern</b>	Sep 2022 – Nov 2022

<b>Waycool Food and Products Pvt. Ltd</b> , India   <b>Machine Learning Intern</b>	Sep 2022 – Nov 2022
<ul style="list-style-type: none"><li>Traditional methods for analyzing grain size were slow and error-prone; designed a <b>CNN</b>-based deep learning model to identify, count, and measure grain sizes <b>with 85–90% accuracy</b>, significantly improving precision and consistency.</li><li>Enhanced model performance by applying advanced feature extraction, adaptive thresholding, and custom data preprocessing, <b>reducing computational time by 85%</b> without compromising accuracy.</li><li>Applied custom preprocessing, statistical validation, and performance benchmarking to deliver actionable insights from vision data.</li><li>Gained deep expertise in computer vision and <b>model optimization</b>, earning top performance scores within the intern cohort and commendation from senior data scientists for quality and impact.</li></ul>	

## PROJECTS

---

<b>NLP Agent for Legal Document Understanding</b>   <a href="#">Github-Repo</a>	Jan 2024 – Jun 2024
<ul style="list-style-type: none"><li>Built an automated <b>legal document extraction system</b> for a legal client, using Gemini, LLaMA Parse, spaCy, and regex to reduce manual review and structure legal data, supporting <b>ML workflow automation and model deployment using FastAPI</b>.</li><li><b>Fine-tuned and evaluated LLMs</b> to boost accuracy, achieving <b>85–90% extraction precision</b> and improving processing efficiency for the client.</li><li>Deployed system with <b>FastAPI</b> and optimized memory handling for long multi-page documents using a <b>RAG architecture</b>.</li></ul>	
<b>Multilingual Real-Time Profanity Detection with Audio &amp; NLP</b>   <a href="#">Github-Repo</a>	Jan 2024 – May 2024
<ul style="list-style-type: none"><li>Developed a real-time multilingual profanity filter for <b>Rajalakshmi Engineering College</b> using STT, NLP, and deep learning, achieving <b>95% accuracy in live audio censorship</b>.</li><li><b>Enhanced processing speed by 30%</b> through parallelized text analysis and TTS, enabling real-time delivery of clean transcripts or filtered audio output with <b>API deployment and real-time system monitoring</b>.</li></ul>	
<b>Analysis To Find Relationship Between Mental Trauma, Rape, and Suicide</b>   <a href="#">Github-repo</a>	Sep 2023 – Dec 2023
<ul style="list-style-type: none"><li><b>Conducted EDA for a research project</b> on population-level trauma and suicide datasets under faculty guidance, analyzing trauma, rape, and suicide data using Python to uncover key correlations.</li><li>Applied statistical analysis and hypothesis testing to validate trends and ensure actionable recommendations.</li><li>Preprocessed and visualized datasets into <b>interactive business reports</b> improving interpretability by <b>50%</b> for use by policymakers and mental health researchers.</li></ul>	
<b>Predictive Modeling for Loan and Insurance Cost Estimation</b>   <a href="#">Github-Repo</a>	Sep 2023 – Dec 2023
<ul style="list-style-type: none"><li><b>Developed ML models</b> for <b>Rajalakshmi Engineering College</b> to predict loan eligibility and insurance premiums using financial and health data.</li><li><b>Boosted accuracy by 10% (loans) and 8% (insurance)</b> by optimizing <b>SVM, Decision Trees, Random Forests, and Gradient Boosting</b>.</li><li>Conducted <b>feature importance analysis and hypothesis testing</b> to simulate <b>A/B scenarios</b>, optimizing risk and pricing decisions. Applied statistical validation to guide model improvements and business strategies.</li></ul>	