ALFIN ABRAHAM

alfinabraham01@gmail.com |+1 860 995 4511 |linkedin.com/in/alfin-abraham/ |github.com/Alfin-Abraham

EDUCATION

Seattle University **Seattle, WA, USA**
Master of Science in Computer Science, Specialization in Data Science *September 2022 – June 2024*
- **Awards:** Dean's Graduate Student Honor Roll
- **Coursework:** Machine Learning, Artificial Intelligence, Visual Analytics, Cloud Computing, Software Testing & Debugging

Medi-Caps University **Indore, MP, India**
Bachelor of Technology in Computer Science and Engineering (AI) *September 2018 – May 2022*
- **Coursework:** Operating Systems, Computer Networks, Compiler Design, Software Engineering, Python Essentials, ML, AI

WORK EXPERIENCE

Community Dreams Foundation **Wethersfield, CT, USA**
Data Analyst *September 2024 – Present*
- Led the design and implementation of a scalable synthetic data pipeline to train CDF's in-house ML-powered resume analyzer, addressing gaps in data quality, industry relevance, and privacy compliance across 35 evolving job roles.
- Overcame limitations of open-source and internal datasets by conducting in-depth stakeholder interviews (Data Analysts, Cloud Engineers, HR Admins, etc.) and leveraging LLMs (ChatGPT-4o, Claude 3.5 Sonnet) with custom prompt engineering and retrieval-based generation to produce realistic, structurally consistent, and trend-aligned resume samples tailored to each role.
- Orchestrated iterative data development sprints grounded in agile principles, delivering regular stakeholder-aligned outputs with embedded feedback loops that continuously improved fidelity, domain coverage, and downstream ML performance.
- Achieved 96.2% model testing accuracy by delivering high-quality, domain-specific training data through structural audits, bias mitigation, and metadata tagging—resulting in a performant resume analyzer that significantly accelerated candidate screening and decision-making for CDF's internal HRMS workflows.
- Collaborated with legal, HR, and engineering teams to establish and institutionalize a privacy-compliant policy for integrating anonymized real-world resumes, enabling sustainable data enrichment and long-term model adaptability.

Indian Institute of Technology Indore **Indore, MP, India**
Machine Learning Researcher *January 2022 – March 2022*
- Developed advanced Gaussian Process Regression models with Rational Quadratic and Exponential kernels to forecast peak solar irradiance (solar power potential) from complex historical time-series data, achieving 83% prediction accuracy.
- Led end-to-end data preparation, including cleaning, outlier handling, and feature optimization using Python (Pandas, Scikit-learn, NumPy), and validated model reliability through RMSE evaluation and 95% confidence intervals.
- This work contributed to improving solar panel efficiency and maximizing energy yield.

ACADEMIC PROJECTS

Real-Time ASL to Speech Translation System: Enhanced ASL Gesture Interpretation ***January 2024 – March 2024***
- Developed a real-time ASL interpreter with 46% accuracy in translating sign gestures to spoken words, integrating MediaPipe for advanced hand tracking, which enhanced gesture detection accuracy by 80%.
- Incorporated OpenAI's GPT-3.5 for context-based refinement of missed or low-confidence (threshold 0.5, scale 0-1) predictions from the core I3D model, improving overall interpretation reliability. Additionally, integrated Google's Text-to-Speech for rapid real-time, natural-sounding audio synthesis.

Stroke-Risk-Classification: Optimized Stroke Forecasting ***December 2022 – January 2023***
- Conducted comprehensive Exploratory Data Analysis (EDA) to assess data quality, distribution patterns, and class imbalance, using visualizations and statistical techniques to guide preprocessing and model selection.
- Designed and implemented an end-to-end stroke-risk prediction pipeline in Python, applying missing value treatment, outlier detection, class balancing, and feature scaling to enhance model reliability for healthcare risk assessment.
- Increased stroke prediction accuracy to 73% using K-Nearest Neighbors with 10-fold cross-validation, reducing false negatives and improving model transparency through Decision Tree visualizations, supporting early medical intervention.

SKILLS

Programming Languages: Python, HTML, CSS, JavaScript, C
Tools: Excel, Tableau, Power BI, Docker, Kubernetes
Cloud: AWS: EC2, S3, RDS, DynamoDB, Lambda, VPC, SageMaker, QuickSight, Budgets, Cost Explorer
Frameworks: Django, Flask
Libraries: spaCy, boto3, PyTorch, TensorFlow, pandas, NumPy, Matplotlib, Plotly, D3.js, scikit-learn, Seaborn
Databases: MySQL, PostgreSQL
Soft Skills: Analytical Thinking, Communication, Detail-Oriented, Problem-Solving, Teamwork, Accountability