ALFIN ABRAHAM

alfinabraham01@gmail.com |+1 860 995 4511 |Portfolio |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

Medicaps 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*
- Designed and built a scalable, modular data pipeline in Python to automate synthetic resume generation, enabling high-throughput data provisioning for an ML-based resume analyzer spanning 35+ job domains.
- Engineered prompt orchestration logic using retrieval-augmented generation (RAG) with LLMs (GPT-4o, Claude 3.5), programmatically generating structurally validated, schema-driven training samples.
- Applied agile development practices with stakeholder feedback loops to iteratively improve pipeline efficiency, reduce data processing latency, and increase release velocity for training-ready datasets.
- Achieved 96.2% model testing accuracy by implementing structured audits, bias detection modules, and robust metadata tagging—enhancing downstream parsing logic and system reliability.

Indian Institute of Technology Indore **Indore, MP, India**
Machine Learning Researcher *January 2022 – March 2022*
- Developed a modular forecasting system using Gaussian Process Regression (GPR) with Rational Quadratic and Exponential kernels to predict peak solar irradiance from time-series data, achieving 83% prediction accuracy.
- Implemented an end-to-end data preprocessing pipeline in Python (Pandas, NumPy, Scikit-learn) to handle missing values, normalize features, and reduce noise—improving training stability and reproducibility.
- Validated model performance using RMSE and statistical confidence intervals, ensuring predictive robustness and production-readiness for integration into solar optimization tools.
- Contributed to improving solar panel efficiency and energy planning by enabling accurate, data-driven forecasting of solar power potential.

ACADEMIC PROJECTS

Real-Time ASL to Speech Translation System: Enhanced ASL Gesture Interpretation ***January 2024 – March 2024***
- Boosted ASL gesture detection accuracy by 80% by engineering a real-time interpreter that integrates MediaPipe for advanced hand tracking, achieving a baseline translation accuracy of 46% for complex sign gestures.
- 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 ***November 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, C, C++, C#
Tools: Excel, Tableau, Power BI, Docker, Kubernetes, Postman
Cloud: AWS: EC2, S3, RDS, DynamoDB, Lambda, VPC, SageMaker, QuickSight, Budgets, Cost Explorer
Frameworks: Django, Flask, pytest, Selenium
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