SHUBHANKAR JOSHI

Portland, OR +1 (541)-250-1125 joshishu16@gmail.com linkedin.com/s-joshi16 github.com**/**shubhu163

EDUCATION

Master of Science in Computer Science - GPA – 4.0   **Sept 2023 – Sept 2025**
Oregon State University, Corvallis, OR.
Relevant Coursework: Machine Learning, Deep Learning, DBMS, Optimization for AI, Algorithms, Cyber Security, & NLP.

Bachelor of Technology in Instrumentation & Control - CGPA - 9.26 / 10   **Aug 2019 – May 2023**
Savitribai Phule Pune University, Pune, India.
Relevant Coursework: Data Structures, Web Technology, Project Management, Data Science, Networks & Statistics.

EXPERIENCE

Research Assistant, Oregon State University, Corvallis, OR.   **Sept 2023 – Present**
- Analyzed **50M+ ribosomal sequences** from 9 diverse human tissues, performing transcriptome alignment and extracting key insights on protein translation using **Bowtie and Pysam**, enhancing **Data Warehousing** capabilities.
- Developed end-to-end LLM pipelines using Transformer and CNN which improved model reliability by **20% within 3 months**.
- Streamlined model deployment with PyTorch Lightning and Wandb, reducing manual QA **by 15%** in the **first 2 months.**
- Achieved a Pearson correlation coefficient of **0.837** between predicted and true ribosomal coverage in the **test set,** demonstrating strong **Predictive Analytics.**
- Optimized model parameters using **grid search** and **Optuna**, leading to a **10%** increase in model performance metrics.
- **Designed NLP-based QA** models to process unstructured biomedical data and created real-time dashboards in Power BI and Looker, delivering actionable insights to **stakeholders in weekly review cycles**.
- Collaborated with health system teams to integrate QA pipelines into EHR search workflows, building scalable ML solutions.

Data Analyst Intern, Landmark Techedge Pvt Ltd, Pune, India.   **Aug 2022 – May 2023**
- Extracted & managed **5GB** daily production data from **MySQL, HDFS, and Excel** to create **Dashboards on Tableau.**
- Developed ETL processes **leveraging SQL Server, SSIS, and ADF**, increasing data processing efficiency by 25% in 3 months.
- Built Tableau dashboards that led to a 19% revenue increase in Q4 2022 through sales strategy optimization.
- Conducted data validation, integrated **Alteryx** for financial forecasting, and supported product strategy with KPI metrics.
- Partnered cross-functionally with marketing, and business teams to **deliver measurable improvements in 6 months.**
- Led technical solutions for finance automation tooling and content management system enhancements.

Software Engineer & Operations Lead, Samarth Enterprises, Pune, India.   **Nov 2020 – July 2022**
- **Modernized** manufacturing operations by developing internal software tools using **Python and Excel VBA** to automate inventory tracking, reducing stock management time by 40%.
- **Built real-time QC dashboards** in Power BI for wire harnesses and regulators, boosting defect detection by 20%.
- Digitized supplier records & invoice **workflows led to cutting paperwork by 60**% and improving traceability.
- **Led 10+ technicians**/interns during ramp-ups, managing tasks, escalations, **and Piaggio delivery timelines.**

PROJECTS

LLM Fine-Tuning for Clinical QA | Transformers, HuggingFace, NLP, PyTorch   **April 2024 – July 2024**
- Fine-tuned BERT and GPT-2 on biomedical QA datasets; implemented extractive highlighting and served models via REST API and validated generation quality using QA metrics.

Bird Classification Robustness | Python, PyTorch, Torchattacks.   **Jan 2024 – March 2024**
- Developed an adversarially robust image classifier using ResNet with PGD/FGSM defenses. Achieved 98.5% precision.
- Applied A/B testing to compare against baseline models and measured the success of adversarial generalization

Real-Time ML Inference Pipeline for Fraud Detection | Python, PyTorch, Torchattacks.   **Sept 2023 – Dec 2023**
- Built scalable fraud detection system using FastAPI and Docker, integrated with AWS Lambda for low-latency predictions.
- Performed data mining on behavioural features, and improved detection accuracy by 11%.

SKILLS

LLMs, XGBoost, AutoML, Reinforcement Learning, Predictive Modeling, Sentiment Analysis, Hypothesis Testing, A/B Testing, Statistical Modeling, Tokenization, Summarization, Text Classification, Question Answering, Named Entity Recognition, Information Retrieval, Python, SQL, R, Java, C++, Bash, MATLAB, PyTorch, TensorFlow, Scikit-learn, HuggingFace, FastAPI, Ray, Spark, Jenkins, CI/CD, Airflow, Docker, AWS (S3, SageMaker, Lambda), GCP (Vertex AI, BigQuery), Azure, Tableau, Power BI, Looker, Excel (Macros, KPIs, Graphing), Google Cloud, MySQL, PostgreSQL, MongoDB, Data Warehousing, Data Architecture, Data Governance, Communications, Stakeholder Management, Leadership, Presenting, Problem-Solving, Strategic Thinking, Time Management, Coding, Data Mining, Health Systems, Actionable Insights

EXTRACURRICULAR & PUBLICATIONS

Led a 20-member team as Aaroh Music Club Head, organized events, and performed as a singer. (sing.shubhankar.)
Ocular Disease Recognition using Ensemble Techniques: Published on IEEE Xplore in ICCUBEA - 2023**.**
User Survey of Music Streaming Apps: Selected in HWWE 2023 and currently in the process of publication.