SAI AKHILESH ANDE

Boston, MA **|** 617-749-8460 **|** ande.s@northeastern.edu **|** LinkedIn/akhilesh1896 **|** github/akhil189 **|** Portfolio

EDUCATION

M.S. in Data Analytics Engineering, Northeastern University | Boston, MA **- GPA: 3.9/4.0**　　　　**Jan 2021 – May 2023**
　　Course work: MLOps, Machine Learning, Deep Learning, Natural Language Processing (NLP), Data Mining, Data Visualization

B.Tech. in Computer Science, Indian Institute of Information Technology | India　　　　**Aug 2015 - May 2019**

HONORS & AWARDS

Alfred J. Ferretti Excellence in Teaching Award (2023)　　　　***April 2023***
- Honored for outstanding service and contributions as a **Graduate Teaching Assistant for Machine Learning** over 5 semesters.
- Mentored **500+** students on complex ml concepts, published blogs and instructed Python labs on core data science frameworks such as **NumPy**, **Pandas, Matplotlib**, **Scipy** and **Scikit-learn** to program classical ml algorithms from scratch.

Winner (1st place) – Kaggle Datathon 2023　　　　***March 2023***
- Led a team of 3 to victory in a week-long AI Hackathon competing against 30+ teams. Built a **hybrid deep learning model** for predicting the download-worthy quality of images using image demographics, color properties, and keyword labels.

PROFESSIONAL EXPERIENCE

ML & MLOps Engineer | Tausight - *Boston, MA*　　　　**July 2023 - Present**
- Engineered an NLP model for detecting Personal Identifiable Information (PII) in electronic health records. Achieved a state-of-the-art efficacy of **94% F1-score** by fine-tuning an LLM model using **TensorFlow** and **TFX** in an **agile** workflow.
- Orchestrated MLOps best practices with **Airflow**, **ELK, DVC, MLflow,** etc. for experiment tracking, model management, and data versioning, ensuring a streamlined workflow.
- Established and managed robust MLOps pipelines, integrating **CI/CD** practices, version control, automated testing, ensuring seamless deployment of models into production environments while reducing deployment time by **25%**.

Machine Learning Co-op | Tausight - *Boston, MA*　　　　**Jan 2022 - Aug 2022**
- Conducted research and developed an **unsupervised anomaly-detection** model to identify malicious applications in healthcare systems, resulting in a reduction of false positives by **15%.** Deployed it on customer end points through **GCP** and **Airflow**.
- Built end-to-end data pipelines using **Python** and **SQL** to process multi-tera byte customer data utilizing **BigQuery**.
- Employed a wide range of advanced statistical tests, including chi-squared, z-test, t-test and ANOVA which validated the model assumptions, fine-tuned alert thresholds, and reinforced the statistical significance of anomaly alerts.
- Generated comprehensive analytic reports on unprotected health information (PHI) on customer endpoints using a diverse array of data analysis tools while ensuring compliance with data protection regulations.

Machine Learning Engineer | Youngsoft - *Hyderabad, India*　　　　**Jan 2020 - Dec 2020**
- Engineered NLP-powered **AI chatbot assistants** for healthcare products, reducing customer support costs by **30%** and enhancing customer engagement by **46%**.
- Developed them in **Python** with **Rasa** (an open-source ML framework), **REST APIs**, and deployed on servers through **SQL**, **GitHub/git**, **Docker**, and **AWS EC2** instances while integrating with social media (WhatsApp, Facebook, Telegram, etc.)
- Collaborated with cross-functional teams, including data analysts, and stakeholders, to define project goals, requirements, and deliverables, ensuring alignment with business objectives.

TECHNICAL SKILLS AND CERTIFICATIONS

Programming:	Python (Pandas, NumPy, SciPy, Dask, Matplotlib, Seaborn, Plotly), C++, R, SQL
ML Frameworks:	Scikit-learn, Keras, TensorFlow, PyTorch, HuggingFace, Spacy, NLTK, Gensim, OpenCV
ML Techniques:	Hypothesis Testing, A/B Testing, Regression, Classification, Clustering, Decision Trees, Dimensionality Reduction Neural Networks, CNN, RNN, LSTM, tf-idf, word2vec, embeddings, Transformer, BERT, GPT, NER
MLOps Tools:	git/GitHub, CI/CD, Airflow, MLflow, Data Version Control (DVC), TFX, Kubernetes, Linux, Docker, GCP, AWS
Certifications:	TensorFlow Developer Certificate

ACADEMIC PROJECTS

News Topic Recognition　　　　**Sep 2022 - Dec 2022**
- Enhanced training data quality with text processing techniques including tokenization, lemmatization, stop word removal and generated **tf-idf**, **word-2-vec, doc-2-vec, GloVe** and **FastText** word-embeddings.
- Implemented a custom **Transformer** model in **TensorFlow** for classifying news articles, achieving **91.44%** accuracy. Further fine-tuned using **BERT LLM**, resulting in an accuracy of **94.6%** close to state-of-the-art and deployed using **Docker** and **FastAPI**.

Anomaly Detection in Financial Transactions　　　　**Sep 2021 - Dec 2021**
- Addressed class imbalance (1000:17) via **stratified random sampling** and **SMOTE**. Employed various models including Logistic Regression, K-NN, Decision Tree, Random Forest, and Isolation Random Forest, achieving highest **AUROC of 0.876**.
- Designed a custom **Deep Auto-Encoder** model in TensorFlow which achieved a **Recall score of 0.85** with only a **6.5% False Positive Rate** on the malicious class and best **AUROC score of 0.956.**