ASHRAF SYED

+1 9402994044 ⬦ Denton, Texas

syedashraf2211@gmail.com ⬦ linkedin.com/in/syed-ashraf2211

EDUCATION

Master of Science - CS, University of North Texas   May 2024
GPA : 3.82
Bachelor of Technology - IT, MLR Institute of Technology   2021
GPA : 3.15

SKILLS

Programming Languages	Python, Java, C, Scala
Frameworks	TensorFlow, Spark, Flask, PyTorch
Libraries	Pandas, NumPy, Scikit-learn, Matplotlib, NLTK, Seaborn, Altair, SpaCy, Langchain, Pyaudio
Databases and Tools	MySQL, MongoDB, Jupyter, Git, AWS, Azure, GCP,Tableau, OpenAI, Huggingface
Specializations	Natural Language Processing (NLP), Deep Learning, Machine Learning (ML) AI, Data Pipeline, Big Data, Cloud Services (AWS, EMR, S3, Step Functions)

EXPERIENCE

Ally Inc   Oct 2024 - Present
Machine Learning Engineer

- Developed real-time AI-driven chatbot for customer service using LLMs achieving a 30% reduction in response time and improving customer satisfaction scores by 25
- Implemented real-time transcription pipelines leveraging AI Vortex and PyAudio, enabling transcription accuracy of 95%+ for live customer interactions.
- Integrated multi-cloud infrastructure with AWS, Azure, and GCP optimizing resource utilization and reducing operational costs by 15%.
- Technology : Python, AI Vortex, PyAudio, GCP, Azure, AWS, PyTorch, TensorFlow, Docker, Kubernetes, REST APIs, Langchain, OpenAI APIs, huggingface APIs.

University of North Texas   August 2023 - May 2024
Graduate Teaching Assistant - Machine Learning and Scientific Data Visualisation.

- Assisted in teaching machine learning concepts, including model development, hyperparameter tuning, and algorithm evaluation to over 125 students.
- Guided students in the application of Python and PySpark for large-scale data processing, enabling them to build and deploy machine learning models efficiently.
- Facilitated the use of data visualization techniques to enhance the interpretability and communication of machine learning model outputs.
- Provided mentorship on integrating machine learning models with data pipelines for real-time analytics and decision-making.

EPAM Systems   May 2021 - August 2022
Junior Data Engineer

- Engineered and optimized scalable data pipelines using Apache Spark to handle massive volumes of big data efficiently. Enhanced data processing speed by 40% through advanced Spark optimization techniques and partitioning strategies.

- Developed and automated ETL workflows, leveraging a suite of AWS services including EMR, S3, and Step Functions. This integration streamlined data pipeline operations, reducing data preparation time for downstream machine learning tasks by 50%.
- Architected robust data pipelines to process and manage complex big data sets on AWS. Utilized Scala and SQL for data transformation, achieving a 30% cost reduction and significant time savings in data processing tasks.
- Technology : Apache Spark, SQL, AWS (EMR, S3, Step Functions), Scala, Big Data, Data Pipeline, ETL.

PROJECTS

Foliar Disease Classification

- Developed a machine learning model to accurately classify foliar diseases from high-resolution plant leaf images.
- Processed and augmented a dataset of over 100,000 images, enhancing the model's ability to generalize across diverse conditions by applying techniques like rotation, scaling, and flipping.
- Employed a TensorFlow-based Convolutional Neural Network (CNN) model, incorporating deep learning techniques such as batch normalization and dropout. Achieved a classification accuracy of 98% through meticulous hyperparameter tuning and model optimization.
- Leveraged Spark's in-memory processing to expedite model training by 30%, enabling efficient use of computational resources and reducing overall latency.

Emotion Detection System

- Created a machine learning system to detect emotions in real-time from textual data, leveraging advanced deep learning algorithms.
- Curated and cleaned a dataset of 250,000+ text samples, applying NLP preprocessing techniques like tokenization, stemming, and lemmatization using NLTK and SpaCy to ensure high-quality input for machine learning models.
- Developed a Bi-LSTM network with attention mechanisms, focusing on deep learning methods to capture contextual sentiment. The model achieved 90% accuracy across multiple emotion categories, using transfer learning with pre-trained GloVe embeddings for improved performance.
- Enhanced system throughput with batch inference, optimizing deep learning model deployment for real-time applications.

Fraud Detection System

- Designed a machine learning system to identify fraudulent online transactions, applying both traditional and advanced deep learning techniques.
- Processed a large dataset with over 500,000 records, utilizing Hive for data transformation and pandas for further cleaning and preparation, ensuring the dataset was primed for machine learning models.
- Implemented and evaluated multiple machine learning algorithms—Random Forest, XGBoost, and LightGBM—to detect fraudulent transactions. XGBoost, leveraging deep learning principles, demonstrated superior performance with a Precision of 0.92, Recall of 0.48494, F1-score of 0.63523, and AUC of 0.93837.
- Conducted a comparative analysis of machine learning models, focusing on critical metrics like Precision, Accuracy, and F1-score, confirming XGBoost as the most effective model for fraud detection.

RESEARCH ARTICLES

- [Prediction of House Price Using Machine Learning Algorithms](), IEEE, 2021
- [Melanoma classification using convolutional neural network](), AIP, 2023