ABHIJIT CHALLAPALLI

abhijitchallapalli99@gmail.com | +1- 945-400-9320| LinkedIn| GitHub |Portfolio| Arlington, TX

EDUCATION

University of Texas at Arlington Arlington

Master's in data science | GPA: 4.00/4.00 08/2022 – Expected 05/2024

EXPERIENCE

Graduate Research Assistant – The University of Texas at Arlington| Python, ML, Linux. **10/2021 – Present, USA**

- **Led** a critical **data migration and integration** project from **DSpace to BePress**, employing **Open Refine** and **Python** for **metadata cleaning** and **quality assurance**, **resulting** in the seamless **transition** of over **one-million** records.
- **Reported** data visualization and analysis with **Plotly, Seaborn, Tableau, and PowerBI**, generating **over 50** comprehensive **dashboards** and **reports** to drive project decisions.
- **Automated** data extraction and manipulation in **Unix/Linux** for over **15,000 files**, automating JSON data preparation and improving import **efficiency by 40%**.
- **Leveraged AWS services (S3, Lambda, RDS, Glue)** to manage and process terabytes of data, enhancing storage and data processing capabilities.
- **Developed** and **implemented** a **machine learning** model to automate the classification and tagging of academic papers, improving **metadata accuracy by 30%** and reducing manual **processing time** by **50%**.
- **Managed root cause analysis** and **quality control** for data processes, significantly enhancing **data reconciliation** and synchronization across platforms.

Data Engineer - Capgemini Technology Services India Ltd |Python, ETL, TWS, Linux. **02/2021 – 07/2022, India**

- **Delivered** intricate **ETL pipelines** for **data marts with 100% success** for **Barclays**, meeting stringent deadlines and maintaining **code excellence**.
- **Orchestrated** in-depth **data analysis and business intelligence** enhancements for **JetBlue**, **enhancing** functionality and performance by **40%.**
- **Extracted** and transformed data from warehouses for **Barclays**, translating business needs into **50+ reports.**
- **Proficiently developed and designed** automated test scripts within the ETL framework using **Ab Initio for JetBlue**, resulting in streamlined and **60% efficiency gains** in testing processes.
- **Restructured data integration** services for **Barclays** to securely transfer **payment data & invoices** between **4 applications.**
- Implemented **data models** and **flow diagrams** for efficient data processing, achieving streamlined **data integration**.

SKILLS

- **Programming Languages:** Python, C, HTML
- **Databases & Frameworks:** SQL, MySQL, OpenSearch
- **Platforms & Applications:** Azure (Databricks, Data Lake, SQL, VS Code, GitHub), AWS (S3, EC2, Athena, Redshift, Glue, IAM, DynamoDB, Lambda, Kinesis, SageMaker), Snowflake, Apache (Spark, Airflow, Kafka), Data mining, Data Marts, Tableau, PowerBI, OpenCV, Keras, TensorFlow, NetworkX, Pandas, NumPy, Matplotlib, Scikit-learn, MATLAB, NLTK, Google Cloud, Weka, Data structures, ETL Ab initio, Talend, Informatica, data mapping, Hadoop, Docker, Data Flow.

PROJECTS

Facial Recognition Enhancement System | Python, OpenCV, TensorFlow, Keras **2024**

- **Leveraged** TensorFlow and OpenCV for developing a deep learning model, **achieving 95%+ accuracy** on the LFW dataset.
- **Enhanced** profile face recognition **accuracy by 15%**, specifically tailored for **ResNet-50/VGG-16** architecture adjustments.
- **Optimized** the model using **250,000+ images**, **resulting** in a **20% reduction** in **false positives** and a **25% improvement** in detection latency, showcasing significant **performance improvements**.

Sarcasm Detection on Twitter Datasets| Python, NLTK, Google Colab, TensorFlow, Keras. **2023**

- **Detected** sarcasm and complex emotions in customer reactions using a BiLSTM-DNN model.
- **Used** (BiLSTM-DNN) Bi-directional Long Short-Term Memory Deep Neural Networks model.
- **Initiated** Hyper-parameter Tuning using Grid-Search-CV with **85% accuracy**.

Classification Project (Titanic Survival Prediction) | Python, TensorFlow, Pandas, NumPy, Matplotlib **2022**

- **Conducted exploratory data analysis** on the Dataset, consisting of **887 records**, revealing insightful patterns and trends.
- **Cleansed** the Dataset by **validating and refining data**, ensuring **95% accuracy** in alignment with column specifications.
- **Built different models** for classifying the data and calculated accuracy for each model. Observed that Decision tree and KNN classifiers have highest accuracies.

ACHIEVEMENTS

- **Research Paper** presented in JETIR **Journal Number: 63975, ISSN: 2349-516**
- **Secured** Rank 1 in Capgemini Training.
- **Data Challenges Lead** for "DATATHON" 24hour coding challenge organized by ACM and UTA Libraries.

CERTIFICATIONS.

- **Python Data Structures**
- **End-to-End Data Engineering Project**
- **Data Engineering with AWS**
- **Python for Everybody**
- **What is Data Science?**
- **Supervised Machine Learning**