Sushma Suda

+1(312)-912-4276 | ssuda5@uic.edu | linkedin.com/in/sushmasuda

EDUCATION

University of Illinois, Chicago, IL
Master of Science in Management Information Systems   3.9/4 GPA

VNR VJIET, India
Bachelor of Technology in Electrical Engineering   3.9/4 GPA

SKILLS

Languages: Python (Pandas, NumPy, FlaskSciPy), Java, C/C++

Web Technologies: HTML/CSS, Javascript, Xml, Json, Rest, Ajax

Databases: (Relational) MySQL, Oracle SQL, Amazon Redshift, (NoSQL) MongoDB

Frameworks: React, Node.js, Django, Flask, Apache Airflow

Tools & Technologies: IntelliJ, Eclipse, Visual Studio, Tableau, Power BI, Agile, Linux, Git, GitLab, Apache Spark

Cloud Technologies: AWS, Azure.

Certifications: Certified System Administrator (CSA), ITILv4, AWS Cloud Security

WORK EXPERIENCE

Research Assistant

University Of Illinois at Chicago   **Oct 2022 – Present**

- Implemented robust data engineering solutions using Python, Apache Airflow, Azure Virtual Machine and Jupyter Hub to pull incremental data from API and to schedule stored procedure.
- Collaborated with researchers to provide tailored data solutions for research papers, clinical trials, and grants at the cancer center, ensuring alignment with project objectives. Managed Data Mining requests and established Azure pipelines for scheduled data transfers to database tables, ensuring seamless data access and integrity for research projects at the cancer center.
- Extracted, transformed, and loaded EMR and Diagnosis data for various cancers such as breast, lung, kidney, and Head and Neck from the EPIC System employing Alteryx ETL tool.
- Developed and maintained a Tableau dashboard report for cancer data, integrating ICD and CPT codes to provide comprehensive insights and facilitate data analysis for research purposes.
- Utilized advanced Microsoft Excel capabilities such as VLOOKUP, HLOOKUP, and PivotTable to identify and report data gaps, ensuring data completeness and accuracy.

Data Engineer

Deloitte, USI   **Jan 2021 – Aug 2022**

- Designed and built real-time data pipelines, data models, and a data visualization dashboard for executive management to support analytics.
- Performed a robust restart mechanism for failed pipelines by leveraging Metadata Schema and Audit Pipeline, ensuring data integrity and minimizing disruptions in data processing.
- Build a Power BI dashboard to track job application across employer and employee facilities.
- Migrated 300+ on-premises database tables to Azure SQL DB using DMS service, enabling scalability, improved data security, and optimized performance for 60% cost-effective operations.
- Acquired expertise in data profiling and wrangling, facilitating a 60% reduction in time taken to convert raw data into clean, well-structured datasets with team of 3 people.

PROJECTS

HTS Classification System

- Developed a web application for classifying HTS codes, hosted on Microsoft Azure.
- Utilized REST API Services, Web Environment, and an SQL Database, each in separate VPC subnets.
- Configured and deployed the application using GitHub repositories, Visual Studio, and Azure

YouTube Live sentiment analysis

- Designed and built an end to end solution for determining the sentiment of a YouTube Live video stream's chat.
- Using NLTK (Vader), PySpark we performed analytics on the live YouTube chat stream obtained using the YouTube API, using Java and a Redis Pub/Sub for Python to extract the Live Feed.
- Achieved 78% accuracy for our model built using a distributed Pipeline of Cleaning, TF-IDF and then Logistic Regression