

YASH RAYTHATHA

yashbraythatha@gmail.com | (973)380-6309 | San Jose, CA | [LinkedIn](#)

PROFESSIONAL SUMMARY

Enthusiastic Data Engineer with a successful history in developing data pipelines, ETL processes, and visualization tools for effective data analysis. Actively pursuing a data engineering internship to apply skills, contribute to innovative projects, and drive efficient data-driven solutions in support of company objectives.

SKILLS

Technologies:

Languages: Python, C/C++, Unix

Data Pipelines and big data engineering: Pyspark

Data Visualization: Tableau, Matplotlib, Seaborn

Database: SQL, Snowflake, MongoDB

Certifications: Azure Administrator (AZ-104) Certified

EXPERIENCE

Data Engineer

Aug 2019 - Mar 2023

HCL Technologies, Nagpur, India

- Collaborated with cross-functional teams to collect, clean, and preprocess diverse datasets, ensuring data quality and usability.
- Led the development of efficient ETL pipelines, contributing to a 20% improvement in data processing speed.
- Automated routine tasks using Python scripts, increasing team productivity, and reducing manual errors.
- Optimized spark jobs to reduce the query run time from 2 hours to 10 Minutes.
- Contributed to the documentation of data sources, methodologies, and code, ensuring comprehensive knowledge sharing within the team.

Application Support Engineer

Nov 2018 - Aug 2019

Savantis Solutions, Hyderabad, India

- Improved application performance by identifying and resolving bottlenecks.
- Streamlined incident management processes, reducing downtime, and improving customer satisfaction.
- Collaborated cross-functionally to develop, test, and deploy new features and improvements.
- Conducted root cause analysis for application failures, implementing preventive measures.

PROJECTS

Laptop Price Predictor

- Leveraged a Kaggle dataset, utilized Pandas for data cleaning, and employed Matplotlib and Seaborn for insightful visualizations to drive feature selection for predictive modeling.
- Developed and refined multiple machine learning algorithms to accurately predict laptop prices, focusing on optimizing model performance for accuracy.
- Deployed the predictive model using Streamlit, creating an interactive web application that enables real-time, user-driven laptop price estimations.

Twitter Data Analysis Using Apache Airflow

- Developed and managed a robust Apache Airflow ETL pipeline on AWS EC2 to automate the extraction, transformation, and loading of Twitter data from S3.
- Enhanced data integrity and streamlined processing by optimizing transformation and storage operations, resulting in improved efficiency.
- Leveraged Amazon QuickSight to create dynamic visualizations of the transformed data, facilitating advanced analytics and informed decision-making.

AI Powered Dots and Boxes Game

- Developed an interactive Dots and Boxes game featuring a challenging AI opponent utilizing the Monte Carlo Tree Search (MCTS) algorithm.
- Enhanced gameplay and strategy formulation by implementing MCTS, enabling the AI to simulate decisions and adapt strategies dynamically.
- Used a win-loss ratio for evaluating and optimizing AI performance, ensuring a competitive edge and engaging gameplay experience.
- Technologies Used: Pygame library, AI algorithm.

EDUCATION

Bachelor of Engineering: Computer Technology

May 2017

Rajiv Gandhi College of Engineering Research & Technology - Chandrapur, Maharashtra, India

Master of Science: Data Science

May 2025

University of New Haven, West Haven, CT