VAIBHAV YALLA

[Portfolio](#)   +1(669)238-1423   yallavaibhav@gmail.com   [linkedin.com/in/vaibhav-yalla/](#)   [github.com/yallavaibhav](#)

PROFESSIONAL WORK EXPERIENCE

Data Engineer Coop - *John Deere Intelligent Solutions Group, Urbandale, USA*   *Sep 2022 – Dec 2022*

- **Built** an automated ETL pipeline for vehicle's rear-view video data, using Lambda, S3, Python, Moviepy, Docker, and SNS to provide video visualizations for testing model performance based on the **False Positives/Negatives, Alert precision,** environment conditions. This solution resulted in saving ~ **300** monthly labour hours and **8%** performance boost in the model when utilized by ML engineers for a better customer safety.
- **Collaborated** with a team of 5 to analyse customer data in Azure Databricks using PySpark and SparkSQL, created PowerBI dashboards that benefited **35%** of stakeholders through actionable insights & optimizing business strategies.

Analytical Engineer Intern - *UP5 Gaming, Indiana, USA*   *May 2022 – Aug 2022*

- **Developed** real-time gaming statistics web app, which helped users view the player metrics with visualizations, increasing user base & improving page views by **21%**. Tools used, HTML, CSS, Bootstrap, Python, Flask, Dash & Third-party APIs.
- **Built** an ELT pipeline for merchandize store sales data analysis using SQL, DBT and Snowflake. Delivered Tableau dashboards to stakeholders, enabling data-driven decision-making.
- **Deployed** and integrated CICD pipeline in Jenkins with docker as agent, Github for source code management, performed code review, implemented test cases (generic & singular tests cases), concurrent documentations of code using DBT.

Data Analyst - *Cognizant (Société Générale), Bangalore, India*   *Jul 2019 – Jul 2021*

- **Built** and developed dashboards using Elasticsearch & Kibana to monitor health KPI's of various applications, saved over 30 hours monthly and reduced the human error rate, which benefited in **70%** lesser tickets.
- **Collaborated** with teams to automate complex workflows by scheduling tasks in Airflow, that reduced **33%** error rate.
- **Supported** and maintained critical Linux applications (L3 support), meeting tight SLAs using ServiceNow platform.

Programmer Analyst Intern - *Cognizant (Société Générale), Bangalore, India*   *Jan 2019 – Jun 2019*

- **Analyzed** user data using Excel, SAS, Alteryx, MySQL, and Oracle to extract insights for audit & traffic data enabling efficient redistribution of server traffic. Cost savings of ~ **0.7 M** $ by enhancing architecture while ensuring security measures.
- **Created** a shell script to monitor server performance and notify employees, resulting in improved server efficiency.

Python Developer Intern - *Tata Consultancy Services, Hyderabad, India*   *Apr 2018 – Jun 2018*

- **Designed** and implemented real-time order data pipeline by utilizing AWS Kinesis Data Streams & Kinesis Agent.
- **Collaborated** with developers to deliver real-time customer request information by storing the processed into DynamoDB.

Instructional Student Assistant (Deep Learning) - *San Jose State University, San Jose, USA*   *Jan 2023 – May 2023*

- Created and graded assignments, quizzes, and exams. Assisted students regarding the doubts in the Deep learning course.

PROJECTS

Customer Behavior & Product Segmentation in H&M ([link](#)): Developed model, Customer lifetime value & churn prediction analysis to design targeted market for improving retention rate. Utilized RFM analysis and machine learning techniques to predict purchase likelihood, achieving an impressive 83% accuracy. Tech: Python, Pandas, RFM, S3, Numpy, XGBoost, Random Forest.

Formula 1 Sport Analysis: Created dashboards and BI reporting for Formula1 racing statistics using Power BI. Tech: Azure Data Factory, Python, Data Bricks, ETL, PySpark, Spark SQL, Azure, Power BI.

EDUCATION

San Jose State University: *Master of Science* - Data Analytics - GPA 3.7/4, *San Jose, California*   *Aug 2021 - May 2023*
Coursework: Mathematical Methods, Database Systems, Machine Learning, Big data Technologies, Deep Learning, Data Mining.

GITAM University: *Bachelor of Technology* - Computer Science & Engineering - GPA 3.7/4, *Vizag, India*   *Aug 2015 - May 2019*
Coursework: Data Mining & Data Warehousing, Artificial Intelligence, Data Structures & Algorithms, Cloud Computing, Cryptography.

TECHNICAL SKILLS

Skills:	Python, SQL, Unix/Linux, PySpark, AWS (Kinesis, Redshift), GCP (BigQuery), Azure (Databricks)
Databases/ Data warehouse:	MySQL, Spark SQL, HBase, MongoDB, Snowflake, Cassandra, Redshift, Big Query
Visualization:	Tableau, Looker, Power BI, Dash, Kibana, Splunk