YASHWI PASSARY

+1 (812) 778-5805 | ypassary.masters@gmail.com | LinkedIn

PROFESSIONAL SUMMARY

A results-driven Data Engineer with 4 years of expertise in designing and managing scalable ETL pipelines, delivering actionable insights, and driving data-driven strategies in the Healthcare domain. Seeking to pursue an MS in Data Science with a focus on machine learning and statistical analysis to transition into a leadership role as a Data Scientist, leveraging advanced analytics to develop innovative solutions and transform decision-making in healthcare.

EDUCATION

Indiana University - Bloomington   **August 2024 - May 2026**
Master's, Data Science   *GPA: 3.9*
Relevant Coursework – Applied ML, ML in Bioinformatics, Cloud Computing, Advanced Database Technology, Statistics

PROFESSIONAL EXPERIENCE

AbbVie, Janssen, Gilead | ZS Associates   **Pune, MH, India**
Senior Data Engineer   *May 2022 - August 2024*
- Spearheaded development of advanced PySpark components, including pivoting functions and history data capture, enhancing MDM capabilities and enabling efficient ETL pipelines on Palantir Foundry and AWS (Airflow, Athena, RDS, S3, EMR).
- Delivered real-time Tableau dashboards that reduced report generation time by 40%, accelerating insights into patient feedback and improving decision-making.
- Led and successfully delivered three concurrent projects with a 100% on-time completion rate, ensuring high-quality outcomes and regulatory compliance.
- Directed data forecasting initiatives by implementing business transformation logic for Power BI dashboards, driving impactful analytics.
- Recognized with the "Best Performance" award in the Digital Products and Solutions category for the impactful Janssen APAC PXP project at ZS Associates.

S&P Global | LTI Mindtree   **Mumbai, MH, India**
Data Engineer   *July 2020 - April 2022*
- Utilized Spark in Scala to distribute data processing on large streaming datasets to improve the ingestion and processing speed of that data by 87%. Oversaw product development estimation and sprint planning.
- Automated ETL processes across billions of rows of data and contributed to the process of migration of data to the cloud using Azure DevOps and Databricks which will help in downtime efficiency and faster queries.
- Led project execution of healthcare database for Covid-19 analysis using these concepts: ETL, Big Data, SQL/PL-SQL, Hadoop, PySpark, Kafka Streaming, Snowflake, AWS (S3, Lambda, DynamoDB, Quick Sight, Redshift, Glue, RDS).

PROJECTS & OUTSIDE EXPERIENCE

Web Application Development: Image Steganography | *Python, Flask, HTML, CSS* | - **Link**   **Kolkata, WB, India**
Group Leader   *May 2020 - July 2020*
- Created a web application that focused on image steganography, leveraging advanced bit manipulation techniques. Devised a secure encryption algorithm for encoding and decoding secret images within cover images. Employed Python Imaging Library (PIL) for image processing.
- Designed a robust web application using Flask for back-end architecture and seamlessly integrated HTML, CSS, and JavaScript on the frontend.

Fantasy Premier League Optimization Using Machine Learning | *Random Forest, XGBoost* |- **Link** **Bloomington, IN, USA**
Developer   *October 2024 - December 2024*
- Developed a machine learning-driven system to optimize Fantasy Premier League (FPL) decision-making by leveraging models such as Random Forest, Gradient Boosting, and Reinforcement Learning
- Designed and implemented predictive analytics for captaincy selection, dynamic fixture difficulty ratings, and team optimization strategies, achieving a mean accuracy of 83.87% in captaincy predictions
- Utilized advanced feature engineering techniques and real-time data analysis to provide actionable insights, enabling improved decision-making for FPL managers
- Demonstrated expertise in data preprocessing, model evaluation, and result visualization to deliver a comprehensive and impactful solution in sports analytics.

SKILLS

Programming & Libraries: Python (NumPy, Pandas, PySpark, Scikit-Learn, Tensorflow), SQL, Scala, JavaScript

Cloud & Data Platforms: AWS (Redshift, S3, EMR, Lambda, Glue, Athena, RDS), Azure (Data Factory, ADLS), Google Cloud Platform (BigQuery), Snowflake, Hadoop, Apache Spark, Apache Hive, Databricks, Palantir Foundry, MySQL

Machine Learning & MLOps: Regression, Clustering, Deep Learning, Neural Netwroks, Time Series Forecasting, Data Forecasting, Feature Engineering, Model Deployment, Business Transformation Logic, CI/CD Pipelines

ETL & Workflow Automation: Airflow, Jenkins, Git, JIRA

Data Visualization Tools: Tableau, Power BI, Amazon QuickSight, MS Excel

Other Skills: Data Analysis, Data Visualization, ETL Development, Workflow Automation, Business Analytics, Linux/Unix