

Disha Yadav

📍 Philadelphia, PA | 📞 +1 (267) 530 9595 | ✉ dishayadav9728@gmail.com | 💻 [GITHUB](#) | 🔗 [LinkedIn](#)

EDUCATION

Drexel University | Philadelphia, USA

Sep 2022 - Jun 2024

Master of Science in Computer Science (GPA: 3.9/4.0)

Courses: Data Structures & Algorithms, Fundamentals of Databases, Applied Database Technologies, Artificial Intelligence, Computer Vision, Cloud Computing, Data Analysis & Interpretation, Data Analysis & Interpretation

University of Mumbai | Mumbai, India

Jun 2015 - Jun 2019

Bachelor of Engineering - Computer Engineering

PROFESSIONAL EXPERIENCE

Research Assistant | Drexel University, Philadelphia, PA

Sep 2023 - Dec 2023

- Boosted data exploration by 40% through developing a user-friendly web application with Fast API, significantly reducing analysis time.
- Increased insights generation by 35% by embedding an algorithm for **automated chart recommendations** based on user data.
- Empowered users with interactive visualizations using Apache ECharts, facilitating deeper data exploration and informed decision-making.

Research Assistant | Drexel University, Philadelphia, PA

May 2023 - Sep 2023

- Improved research efficiency by 30% through detailed time series analysis, optimizing processes with trend identification
- Enhanced neural dataset analysis accuracy by 30% through the utilization of **Principal Component Analysis (PCA)** and **Independent Component Analysis (ICA)** techniques for feature extraction
- Enabled 30% better medical research decisions with advanced **statistical analysis** and visualizations, offering deeper insights.

Data Engineering Analyst | Accenture Services Pvt Ltd, India

Sep 2019 - Aug 2022

- Accelerated CRM data migration from **Azure Cloud** to **GCP** by 35% through the development of optimized **PySpark** scripts enhancing efficiency and reliability
- Spearheaded automation initiatives, slashing manual effort by 40% and ensuring seamless data migration and processing workflows
- Implemented advanced ETL techniques using Talend, **Databricks**, and **Alteryx**, achieving a 30% reduction in processing time and significant data quality enhancement.
- Mastered complex SQL queries in **BigQuery**, executing seamless data transformations and integrating disparate data sources with precision
- Engineered HiveQL ETL scripts for Azure, minimizing errors by 40% and boosting loading speed by 50%, thereby optimizing database performance
- Streamlined ETL processing through **Apache Oozie**, saving 20 hours per week and facilitating a 25% increase in data-driven decision-making
- Leveraged **Power BI**, **Tableau**, and other visualization tools to craft compelling reports and data visualizations, empowering stakeholders with actionable insights for strategic decision-making.

PROJECTS

Lyft Real-time Data Analysis — (Apache Kafka, Zookeeper, AWS EC2, AWS S3, AWS Glue, AWS Athena)

Mar 2024

- Built fault-tolerant, real-time streaming infrastructure on AWS using Apache Kafka and Zookeeper
- Engineered comprehensive data pipeline integrating **Apache Kafka**, **AWS S3**, **AWS EC2**, **Glue**, and **Athena**, enabling seamless data ingestion, processing, and analysis for actionable insights.

Stock-Market Data Engineering Pipeline — (PySpark, Azure Cloud, Azure Synapse, Databricks, PowerBI)

Feb 2024

- Created an **ELT Pyspark** script in **Azure Data Factory** to seamlessly migrate on-premises SQL data to **Azure Storage**
- Employed **Databricks** to automate data preprocessing pipelines and load data into Azure Synapse for efficient data modeling, resulting in a 30% efficiency gain and empowering actionable Power BI reports.

Loan Approval Prediction Model— (Python, Pandas, Matplotlib, Scikit-learn)

Oct 2023

- Conducted an in-depth **Exploratory Data Analysis (EDA)** on a Kaggle loan dataset, addressing outliers and missing values to ensure data quality for a high-accuracy model
- Developed a high-accuracy (80%) **Support Vector Machine (SVM)** model using strategic feature engineering.

COVID-19 Data Analysis and Prediction — (PySpark, Airflow, GCP)

Aug 2023

- Deployed **Apache Airflow**, achieving a 40% efficiency gain in ETL processing and loading, facilitating accurate analysis.
- Trained **Machine learning models** on processed data, achieving 85% accuracy in predicting trends and patterns.

SKILLS

Data Analysis: Data Visualization, Statistical Analysis, Machine Learning, Data Mining, Data Storytelling

Programming Languages: Python, Java, R, SQL, NoSQL, HiveQL

Cloud Platforms: Google Cloud Platform (GCP), Amazon Web Services (AWS), Microsoft Azure

Data Storage: AWS EC2, Snowflake, MongoDB, PostgreSQL, MySQL, Amazon DynamoDB , AWS Redshift

Data Engineering: Apache Spark, Airflow, Apache Kafka, FastAPI, Flask, Databricks, Azure Data Factory, Azure Synapse

Business Intelligence Tools: Alteryx, Talend, Power BI, Tableau, Advanced Excel

Operating Systems: Windows, Linux, Ubuntu

HONORS & CERTIFICATIONS

- AWS Certified Cloud Practitioner, demonstrating proficiency in foundational AWS Cloud concepts (2024)
- Awarded Accenture's "Employee of the Year" for exceptional performance in software development and analytics (2021)
- Certified in Tableau, Advanced Microsoft Power BI, & Alteryx, showcasing expertise in Data Visualization, Analytics and Reporting (2020)
- Published a research paper | IRJET '19: Effective Hospital Management System| TCET