ANKUR SHAH

work.ankurshah@gmail.com | Houston, TX | 346-542-2320
LinkedIn/ankurshah02/ tableau/ankurshah02/ GitHub/214SF website/ankurshah02.github.io/

WORK EXPERIENCE

Data Analyst **Sep 2023 – Jun 2024**
University of Houston Houston, TX

- Optimized SQL queries for the university admission website, resulting in 50% faster test score uploads to the database, enhancing efficiency for student applications.
- Developed an ETL pipeline with Pyodbc, Apache Spark, and PostgreSQL to aggregate, cleanse, transform, and load student data from 5 departments, reducing data processing time by 40% and improving data accuracy by 25%.
- Deployed Amazon S3 to store missing student data securely, then utilized Redshift to download and manually upload data to the student record database, ensuring 100% accurate reporting and analysis for 1,000+ students.
- Established a comprehensive documentation process for missing student records, enabling analysis of system performance and identification of areas for improvement, leading to a 15% increase in data accuracy.

Data Analyst Intern **Jun 2023 – Aug 2023**
Dog Hugs Cat Houston, TX

- Spearheaded data analysis using SQL and Power BI to create comprehensive reports on A/B test results for stakeholders; insights led to a 20% increase in user engagement.
- Partnered with cross-functional teams to develop time series forecasting using XGBoost, utilizing SQL queries for data extraction and analysis, resulting in 95% accurate market trend anticipation.
- Identified key trends and insights from the MapReduce analysis of Google Keywords, enabling the development of targeted SEO strategies that led to a 50% improvement in click-through rates on search engine results pages.
- Employed SQL queries and statistical analysis to identify key performance indicators (KPIs) in weekly team meetings, leading to interactive Power BI data visualizations and a 15% increase in actionable insights empowering informed decision-making.

Data Scientist **May 2021 – Jun 2022**
Trumen Technologies Pvt. Ltd. Indore, IN

- Implemented YOLO machine learning algorithm to develop a new object detection model that accurately identified 95% of sensors, improving worker efficiency in selecting the correct leveling sensor for specific locations within containers.
- Performed comprehensive business data mining using SQL, resulting in a 15% boost in customer satisfaction by identifying pivotal trends and patterns to inform product development strategies.
- Executed data cleaning and profiling processes on 10,000+ unstructured datasets using SQL and Excel, resulting in a 95% improvement in data accuracy and reliability for data analytics projects.
- Analyzed sensor data using regression techniques to pinpoint influential variables; created an interactive Tableau dashboard, enhancing real-time monitoring and leading to a 25% reduction in sensor-related errors.

ACADEMIC PROJECTS

Exoplanet Detection Using Machine Learning

- Engineered a Database for Machine Learning, merging PHLEC, Kepler, and TESS datasets, resulting in a 40% increase in exoplanet detection accuracy.
- Employed various machine learning algorithms for exoplanet detection, achieving 99% efficiency.
- Crafted Keplerformer, a transformer model optimized for time series data, achieving a validation accuracy of 93.3%.

Hospital Management System

- Applied business rules for data modeling, resulting in a highly efficient relational database system with a 98% decrease in query response time.
- Formulated complex SQL queries to efficiently create a hospital database, contributing to a 40% uptick in data precision and availability.
- Improved data retrieval processes by 40% through seamless integration between the database and frontend using XAMPP.

EDUCATION

University of Houston **Houston, TX**
Master of Science in Electrical Engineering and Data Science May 2024

Narsee Monjee Institute of Management Studies (NMIMS) **Mumbai, IN**
Bachelor of Technology in Electronics and Telecommunications May 2021

TECHNICAL SKILLS & CERTIFICATION

- **Programming Languages**: Python, R, SQL, HTML, CSS, Scala, Git
- **Databases:** NoSQL, MongoDB, Big Data, ETL, ELT, Hadoop
- **Data Processing & Visualization Tools**: Tableau, Power BI, Apache Spark, Excel, Pandas, Jupyter Notebook
- **Python Packages**: TensorFlow, NumPy, SciPy, Matplotlib, Keras, XGBoost, OpenCV, PyTorch, Scikit-Learn
- **Machine Learning:** Deep Learning, Computer Vision, Natural Language Processing, Random Forest, Artificial Intelligence (AI)
- **Cloud Platforms**: AWS, Snowflake, Databricks
- **Data Visualization Tableau and Paraview** by Hewlett Packard Enterprise Data Science **May 2023**
- **Scientific Programming with Python/R** by Hewlett Packard Enterprise Data Science **Mar 2023**