Sanket Pattar

Address: London, United Kingdom
Phone: +447443814041
Email: sanketpattar57@gmail.com
LinkedIn: https://www.linkedin.com/in/sanket-pattar-851b361b2

Objective: Versatile Data Engineer and Analyst with hands-on experience in ETL/ELT pipelines and data analysis. Adept at cloud solutions and big data technologies, seeking to leverage expertise in a growth-oriented organization.

Work Experience

Data Engineer/analyst - Tech Mahindra (Project-British Telecom)

May 2021 - Aug 2022

- Led collaborations with a 3-member developer team to revolutionize the ETL process, achieving a 20% cost-cut in Azure cloud solutions and amplifying data processing speed by 30%.
- Architected and executed a robust ETL/ELT pipeline using advanced SQL, Azure Data Factory, and Azure Synapse Analytics, facilitating the migration of enterprise data (up to 20TBs) into the Azure Cloud Data Warehouse.
- Showcased proficiency in real-time data ingestion leveraging Spark (Python) on Databricks, Azure Event Hubs, and Apache Kafka, accelerating time to insights.
- Spearheaded end-to-end data engineering projects, from data extraction to loading into the data warehouse for downstream applications.
- Engaged with stakeholders to gather requirements and implement data solutions, aligning with business objectives.
- Maintained and enhanced data integration frameworks Verifying efficient data movement across platforms.
- Optimized data processing by designing pipelines that uphold best practices in data quality, scalability, and maintainability.
- Identified and addressed data discrepancies Validating data accuracy and completeness.

Junior Data Analyst - Tech Mahindra (Project-Nokia)

April 2019 - May 2021

- Executed data analysis projects using PostgreSQL and Yellowfin BI, driving network performance improvements and superior service delivery.
- Performed comprehensive analysis of network data to discern key performance indicators and trends, enabling informed decision-making and problem-solving.
- Partnered with cross-functional teams to design and maintain data dashboards using Yellowfin BI, delivering actionable insights for network service enhancements.
- Achieved a 20% reduction in PostgreSQL database query response times by optimizing queries and indexes, and streamlining network data analysis and reporting.
- Instituted rigorous data quality checks, bolstering trust in the data used for network operations and service delivery.

Skills

- Programming: Python (Advanced), C# (Basics)
- Visualization: Tableau, Excel
- Big Data Technologies: Hadoop, Spark, Kafka
- Cloud Platforms: Azure (Advanced), GCP (Basics)
- Database Management: MySQL, PostgreSQL
- Tools: SAP, JIRA
- Methodologies: Agile, Waterfall
- Specialized Skills: Data Wrangling, ETL processes, Machine Learning, Statistical Modeling with Matplotlib

Education

MSc in Big Data Science, Queen Mary University of London
September 2022-2023, Distinction

Bachelor's in Electrical and Electronics, Vishveshwarya Technology
July 2014-July 2018, First Class

Academic Projects

Machine Learning Model for Video Location Prediction

- Preprocessed a comprehensive dataset of student videos and audio recordings using Python and libraries like librosa; extracted key features like MFCCs and delved into exploratory data analysis (EDA) to comprehend feature distribution.
- Architected a machine learning model achieving 70% accuracy in predicting video locations by harnessing TensorFlow-based deep neural networks and optimizing model hyperparameters.
- Introduced a real-time monitoring system for proactive issue identification and resolution, and embraced user feedback, enhanced model accuracy and usability.
- Realized a robust and efficient solution for predicting student video locations by synergizing monitoring systems with iterative user feedback.

Neural Network Analysis of EastEnders Character Dialogue

- Engineered a Python solution to preprocess EastEnders scripts for neural network application, creating high-dimensional vector representations of character dialogues via Word2Vec, and critically analyzed Word2Vec-generated word embeddings.
- Leveraged CNNs to train a neural network model on embedded text data, capturing temporal dialogue sequence relationships and validated vector representations using metrics like cosine similarity and Euclidean distance.
- Employed t-SNE and PCA for low-dimensional vector visualization, uncovering pivotal character relationships and plot arcs.
- Trained the neural network to discern temporal dialogue sequence relationships, and visualized relationships to pinpoint character interactions and plot trajectories.

Churn detection using Bayesian machine learning models for telecommunication dataset.

- Strategic Churn Reduction: Applied Bayesian machine learning models to a telecommunication dataset, identifying key drivers of customer churn and assisting in the formulation of retention strategies that reduced customer attrition by 15%.
- Data-Driven Insights: Integrated expertise in Python, and Bayesian statistics to preprocess and analyze vast datasets, offering actionable, data-driven insights to business stakeholders, thereby informing crucial decision-making processes.
- Ethical Data Handling: Prioritized ethical considerations in data analytics, ensuring all customer data was handled with the utmost confidentiality and in compliance with industry standards, strengthening trust and rapport with clientele.
- Python, PyMC3/Stan, Bayesian Statistics, SQL/Database Management, Data Preprocessing, Bayesian Model Development, Hyperparameter Tuning, Model Interpretation, Cross-Validation, , Industry-Specific Churn Analysis, Ethical Considerations.