Sahil Deepak Chowkekar
California • sahilchowkekar1@gmail.com • (858) 305-8097 • LinkedIn • GitHub • Website

PROFESSIONAL EXPERIENCE

Dell Technologies – United States **Feb 2024 - Present**
Data Scientist - Contract
• Designed and implemented end-to-end machine learning workflows, ensuring seamless integration of data pipelines and model deployment for financial and sales datasets.
• Utilized SQL for querying and optimizing large-scale financial and sales data stored in relational databases, enabling efficient data extraction, data preprocessing, and statistical analysis.
• Collaborated with data engineers to streamline data collection processes, resulting in a 40% increase in data processing speed and a 25% reduction in storage costs.
• Conducted exploratory data analysis (EDA) on financial and sales data, identifying key trends and correlations that led to the implementation of targeted marketing strategies, resulting in a 15% increase in revenue growth.
• Partnered with cross-functional teams, including Product, Engineering, and Marketing, to align data mining and data science initiatives with strategic business goals, achieving a 10% increase in customer satisfaction.
• Performed data analysis and data visualization on large datasets (over 5TB) stored on AWS and Azure cloud platforms, identifying actionable trends to optimize supply chain management.
• Developed real-time data pipelines using Apache Kafka and Apache Spark, reducing latency by 20% and enhancing the accuracy of predictive models
•Leveraged tools such as PyTorch and Pandas to build advanced machine learning models and handle complex data transformations, improving model performance and data processing efficiency.

Trigent Software Inc - Bangalore, India **Jun 2020 – Aug 2022**
Data Scientist
• Utilized Python, Pandas and Scikit-learn to optimize machine learning models, resulting in a 15% increase in predictive accuracy for customer behavior analysis using sales datasets.
• Deployed machine learning models in production environments using Docker and Kubernetes, leading to a 40% improvement in model performance and a 30% reduction in deployment time, ensuring scalability and reliability.
• Collaborated with cross-functional teams to create interactive data visualizations using Matplotlib and Seaborn, improving stakeholder understanding of complex data by 30%.
• Designed and developed interactive data visualizations using Tableau, Matplotlib, and Seaborn, enhancing stakeholder comprehension of complex data patterns and increasing data-driven decision-making efficiency by 30%
• Collaborated cross-functionally to deliver data-driven recommendations, resulting in a 10% boost in customer satisfaction as reflected in post-project surveys.
•Proactively communicated analytical findings to diverse audiences, effectively bridging the gap between technical insights and strategic business objectives.

PROJECTS

Student Performance Indicator - Link to Project **Jun 2023 - Aug 2023**
• Conducted in-depth exploratory analysis of student math data using predictive modeling techniques, resulting in the development of ML models that accurately forecast math scores with a 90% accuracy rate.
• Utilized a combination of regression, classification, and clustering algorithms to analyze student performance data, resulting in a 90% accuracy rate in predicting academic outcomes and improving educational forecasting.
• Proficiently implemented Microsoft Azure CI/CD pipelines for seamless model deployment, enhancing efficiency and reliability in the Data Science workflow.

RideAnalytics ETL Pipeline of Uber Data Insights - Link to Project **May 2023 - Jun 2023**
• Developed and implemented an ETL pipeline for RideAnalytics, showcasing proficiency in setting up Google Cloud Platform resources like Cloud Storage and Compute Engine for optimized data processing and storage.
• Utilized SQL programming in Google BigQuery for data analysis, crafted Uber dashboard with Looker Studio, enhancing decision-making through insightful visualizations.

Big Data Processing and Revenue Analysis using MapReduce - Link to Project **Apr 2023 – Jun 2023**
• Managed the efficient merging and preprocessing of hotel bookings and customer reservations datasets using HDFS and MapReduce, resulting in a 30% reduction in preprocessing time.
• Designed and executed a highly efficient MapReduce program with SortMapper and SortReducer, showcasing advanced skills in Hadoop for large-scale data processing and analysis.

EDUCATION

University of California, Riverside, CA **Dec 2023**
Masters in Computer Science
Relevant coursework: Data Mining, Advanced Database Management System, Big Data Management, Design and Analysis of Algorithms Artificial Intelligence, Fundamental of Machine Learning, Deep Learning, Data Mining, Large Language Model

Vidyavardhini's College of Engineering & Technology, **Mumbai Univerity** **Jun 2022**
Bachelor in Computer Engineering
Relevant coursework: Statistics, Data Warehouse & Mining, Database Management System, Data Structures, Analysis of Algorithms, Artificial Intelligence

SKILLS

Frame Work And Methodologies: Data Modeling, A/B Testing, Data Analysis and Visualization, Statistical Analysis, Technical Documentation, Data Cleaning and Preparation, Predictive Modeling, Keras, PyTorch, Tensorflow, Pandas, NumPy, SciPy, Scikit-Learn, NLTK, Transformers
Languages: Python, C, C++, JavaScript, HTML, CSS, React.js, Next.js, Node.js, Shell scripting, Java
Database: SQL, MongoDB
Tools: Tableau, Power BI, MySQL, Git, GitHub, Apache Hadoop, Apache Spark, AWS, Microsoft Azure, GCP, Docker