SANKET SHAH

sanket89@gmail.com

https://sanket891.github.io

Lansdale, PA (267-283-8624)

Introduction

Azure Data Science Certified professional with over 10 years of experience implementing data-driven solutions focused on retail and healthcare. Experienced in Data Science, Data Engineering and ML Ops roles of translating business questions into actionable insights. Hands on experience with supervised and unsupervised learning algorithms, deep learning, text analytics utilizing NLP and time series forecasting models. Advanced PySpark skills in building data pipelines, executing spark SQL queries and deploying Spark MLlib models. Proven track record in the full lifecycle of the model building process from business requirements to model deployment and tracking of business results. Seeking a pivotal leadership position in Data Science to contribute significant value to the organization.

Technical Skills

- Expert in manipulating data from acquisition to reporting.
- 8+ years of experience in SQL, Tableau, MicroStrategy, Microsoft Office
- 6+ years of experience in Python, Spark, Impala, Hive
- Hands on experience in building data science solutions in Azure Cloud, Databricks and Streamlit
- **Tools**: Jupyter Notebooks, Cloudera Data Science Workbench, SAS Enterprise Miner/ Guide, Palantir Foundry, Databricks, Google Colab, SQL server management studio, Visual Studio Code, Docker, FastAPI
- **Libraries:** Pandas, Scikit-learn, NumPy, Spark MLlib, Keras, FB Prophet, XGBoost, CatBoost, LightGBM, Haystack (Q&A NLP), Streamlit, NetworkX, Seaborn

Professional Experience (last four roles)

Qurate Retail Group - (QVC Inc) - West Chester, PA

Sr. Manager Data Science, July 2023 – Present
- Assisted a lead director in the launch of US Data Science organization as a separate function in recruiting 7 data scientists and evaluating 30+ candidates.
- Managed and led a high-performing team of 3 data scientists, overseeing their work and ensuring the successful execution of projects and objectives.
- Oversaw a major data science project from a consulting firm, taking ownership of the model and process after the handoff; resulting in a 50% improvement in efficiency.
- Spearheaded cross-functional projects, employing a consultative approach to understand departmental needs, align data science solutions with business goals, and secure stakeholder buy-in.
- Established and maintained robust relationships with internal IT partners, external vendors and consultants, ensuring seamless translation of business requirements into effective data solutions.
- Pioneered the adoption of using Agile methodology for deploying machine learning models in Databricks
- Collaborated with cross-functional teams to develop proof of concepts using large language models

Sr. Data Scientist, Machine Learning, Apr 2022 – July 2023
- Built a product-based recommendation engine in Azure/Databricks for customer service team for upselling and cross selling products. Utilized tools like FastAPI, Databricks serverless model API, Azure Event Hub, Azure Web App, AD Authentication, Key Vault, Docker and Streamlit.
- Developed hourly sales forecasting model and deployed it as a Streamlit app on Azure Web App for planning and programming team to assess sales performance.
- Developed a process to find similar products based on text description (use case: out of stock products) utilizing pre-trained models from HuggingFace Model Hub and sentence transformer python package.
- Converted and enhanced existing sales attribution code from SAS to Python by streamlining and optimizing existing process.
- Mentor other Data Scientists throughout the organization and serve as a primary point of contact for data science related questions.

WCG Clinical - Plymouth Meeting, PA
Sr. Data Scientist, Jan 2021 – Apr 2022
- Apply Machine learning techniques on clinical trials data sourced from public, proprietary and vendor supplied sources. i.e. clinicaltrials.gov, EHR data, behavioral survey ratings data, investigators/ sites
- Developed a prototype for Questions and Answering system applying NLP, Python and Haystack (deepset.ai) by scraping articles from public websites and internal PDF files to help internal consultants utilize vast amount of text data.
- Developed and streamlined a process to resolve name and address entities using Spark MLlib, resulting in savings of 10+ hours of manual fuzzy lookups per month.
- Maintain existing Machine learning models/ reports and collaborate with other team members to bring conceptual ideas to tangible product using Machine Learning
- Mentor Jr. Data Scientist and other analysts on data science tools and concepts

Wawa Inc - Media, PA
Data Scientist, Dec 2017 – Jan 2021
- Applied XGBoost algorithm to HR predictive modeling application to design viable action steps on employee retention that the department can implement.
- Developed and deployed customer lifetime value model to predict customer's 12-month value based on transaction history and proposed a new segmentation methodology that digital marketing team can use for retention and promotion strategy.
- Partnered with a cross-functional team (pricing manager, pricing analyst and data engineer) to automate the process of ingesting approximately 12 M competitor price observations per week. Replicated vendors' web UI into a Tableau dashboard which reduced query time to seconds and improved efficiency by 10 folds.
- Utilized facebooks 'prophet' library to forecast transactions and sales dollars for 850+ stores saving ~100 hours of finance department time.
- Applied machine learning techniques such as regression, decision trees, XGBoost, neural networks, clustering, survival curves etc. on various projects using Jupyter notebooks and Cloudera Data Science Workbench

Senior Customer Analyst, Aug 2015 – Dec 2017
- Automated recurring reports for digital marketing, resulting in savings of 7 hours per month.
- Served as part of a team that released MicroStrategy BI tool which enabled self-service reporting for other departments- Savings of 90+ hours of analyst's time a week.
- Conduct customer loyalty A/B tests for Wawa rewards and share actionable findings with stakeholders.

CSL Behring – King of Prussia, PA
Data Analyst, Dec 2012 – Aug 2015
- Developed a tool in Excel using VBA for field sales force to print a report for a specific prescriber which was used in the sales pitch – improving sales force efficiency by 10%
- Utilized RFM analysis to rank prescribers which improved sales force targeting efficiency by 8% annually Reduced report development time by 4 folds after introducing Tableau across the Finance department.
- Process and aggregate HER/prescriber data from multiple channels including Specialty Pharmacies, Distributors, Medical Labs and hub to examine trends and generate customer/ patient level insights.

Education

- Master of Science in BI & Analytics – 2012 - 2014   **Saint Joseph's University,** Philadelphia, PA
- Bachelor of Arts in Economics – 2007 - 2011   **Saint Joseph's University,** Philadelphia, PA

Certifications

- Microsoft Azure Data Scientist Associate (I610-7590) - (Feb 2023)
- Microsoft Azure AI Fundamentals (I042-7659) - (Nov 2021)
- Microsoft Azure Fundamentals (I027-7543) - (Nov 2021)
- AWS Certified Cloud Practitioner (#1315336) - (09/2020)