Yogesh Khandelwal

Skilled professional with 6+ years of experience in dealing with life cycle of data science/engineering project

SUMMARY

- Skilled professional with proven ability in dealing with life cycle of data science project.
- Masters with specialization in Data Science and Bachelors in Computer Engineering with more than 6 years of experience in dealing data science/engineering projects in Banking, EdTech, Auto, Pharma and Insurance Industry

PROJECTS

Project 1: Hospital Readmission Prediction for a leading Health Insurance Payer

Domain: Health Insurance

Role: Data Scientist

Overview:

Led the development of a hospital readmission prediction model, achieving a **10% performance** improvement over the existing model. This initiative resulted in **$5 million** in cost savings and a **17% reduction** in the caregiver workforce for the organization.

Key Achievements:

- Implemented predictive modeling using PySpark, leveraging Databricks and Azure for scalable and efficient processing.
- Utilized mlflow for model tracking and management, ensuring reproducibility and ease of deployment.
- Enhanced organizational efficiency with a significant reduction in hospital readmissions, leading to substantial cost savings and optimized caregiver resources.

Tech Stack: Python, PySpark, Databricks, R, MLFlow, Azure

Project 2: Capital Market Response Analysis for Supply and Demand-Side Innovation

Domain: Marketing

Overview:

Conducted an in-depth analysis of the capital market response to firm activities focusing on both supply and demand-side innovation. Engineered accounting and finance variables from diverse sources such as WRDS, Capital IQ, and Bloomberg to inform the analysis.

SKILLS

Tools:
Python (pandas, seaborn, sci-kit-learn, NLTK, spaCy, Gensim), R, PySpark, Flask, Streamlit, Docker, TensorFlow, AWS, SQL, Databricks, MLFlow, PowerBI

ML Skills:
Supervised & Unsupervised Learning, Text Analysis, Statistical Analysis, Data Visualization, Topic Modeling, Deep Learning, Data Engineering

CERTIFICATES

Deep Learning Specialization - Coursera.

Social Network Theory & Analytics -IIT Hyderabad

AWS Certified Cloud Practitioner- Udemy

Credit Risk Modelling in R- DataCamp

Role: Data Scientist

Key Achievements:

- Data engineered accounting and finance variables from WRDS, Capital IQ, and Bloomberg, ensuring a comprehensive dataset for analysis.
- Employed Python, R, and SQL for data processing and analysis, facilitating a multi-faceted approach to extracting insights.
- Developed a robust text classifier achieving a **93%** accuracy rate, enabling the identification of relevant sentences in textual data.
- Utilized the text classifier to quantify the proportion of relevant sentences, providing a nuanced understanding of market sentiment.
- Contributed to strategic decision-making by uncovering insights into how supply and demand-side innovation activities influence capital market response.

Tech Stack: Python, R and SQL

Project 3: Design and develop utility for data integration and automation for business insights.

Domain: Pharmaceutical

Role: Data Engineer

Overview:

Led a comprehensive data integration and automation utility development initiative for a leading Pharma company, optimizing the extraction, transformation, and loading (ETL) process. The project resulted in streamlined data access, improved efficiency, and enhanced business insights.

Key Achievements:

- Identified and gathered data from diverse sources, including Excel, local servers, and external APIs.
- Established automated data connections, minimizing manual intervention, and ensuring consistent data retrieval.
- Developed Python scripts for data collection, reducing processing time and enhancing accuracy.
- Transformed extracted data to the required format using R and seamlessly loaded it into an Azure data pipeline.
- Collaborated with cross-functional teams to align data integration with business goals, maintaining comprehensive documentation for transparency.

Tech Stack: Python, R, Redis, Azure

Project 4: Propensity to Buy AMC (Annual Maintenance Contract) for a leading passenger vehicle manufacturer.

Role: Data Scientist

Overview:

Led a data-driven initiative to boost sales revenue at Tata Motors by analyzing the propensity of customers to purchase Annual Maintenance Contracts (AMC) for passenger vehicles. This involved comprehensive data engineering and the development of a machine learning model for predictive insights.

Key Achievements:

- Engineered a diverse dataset incorporating customer service, sales, demographic, and complaint data to create a holistic view of customer interactions.
- Built a machine learning model to predict the likelihood of customers purchasing Annual Maintenance Contracts (AMC) for their Tata Motors passenger vehicles.
- Utilized the ML model to identify potential customers with a high propensity to buy AMC, enabling targeted marketing and sales strategies.
- Implemented a ML model-based approach resulting in a remarkable **40% increase** in sales revenue attributed to AMC subscriptions.
- Contributed to strategic decision-making by providing actionable insights on customer preferences and optimizing sales efforts.
- This project demonstrated the impact of data-driven decision-making in boosting sales revenue, showcasing the effectiveness of machine learning models in predicting customer behavior and informing targeted marketing strategies.

Tech Stack: R, SAS, and SQL

EDUCATION

Department of Data Science & Forecasting, UTD, DAVV— *M.Tech with specialization in Data Science (FSP)*

Department of Engineering, UTD, DAVV— *Bachelors with specialization in Computer Engineering*