Amit Kheterpal

akheter1@asu.edu [GitHub]

Tempe, Arizona (open to relocation) (623) 209-4223 [LinkedIn]

EDUCATION

W. P. Carey School of Business at Arizona State University August 2022 – May 2023
Master of Science in Business Analytics (MSBA), GPA: **4.0/4.0** *Tempe, AZ*
- *Relevant Coursework:* Introduction to Enterprise Analytics, Descriptive and Predictive Supply Chain Analytics, Data-Driven Quality Management, Data Mining-1, Data Mining -2, Analytical Decision Modeling I, Marketing Analytics.

Guru Gobind Singh Indraprastha University August 2014 – May 2018
Bachelor of Technology, Electrical and Electronics, First Division with **Distinction** *Delhi, India*
- *Relevant Coursework:* Introduction to C, Data Structures, Statistics and Probability

PROFESSIONAL EXPERIENCE

Data Science Intern - Capstone Project December 2022 – May 2023
Avnet Inc. *Phoenix, Arizona*
- Led development of Python-based supply chain forecasting solution, incorporating **Prophet**, **Auto ARIMA**, and **Exponential Smoothing** models. Achieved **85%** accuracy, optimizing supply chain efficiency and forecasting outcomes.
- Created interactive Python dashboards using **Streamlit** Library, optimizing inventory management and improving operational efficiency by displaying performance indicators and insights from time series models.

Statistical Analyst November 2020 – July 2022
ComScore Technologies Ltd. *Pune, India*
- Led migration from **SAS** to **Python** for sample selection and weights generation, incorporating a raking algorithm for weighting and leveraging **Spark**. Automated the process using **Jenkins**, improving efficiency and achieving **$40,000** in annual cost savings.
- Resolved panel health issues, including data inconsistencies, missing values and outliers utilizing **SAS**, **Python**, and **PostgreSQL** by analyzing and problem-solving, reducing data anomalies by **50%** and improving overall data accuracy and reliability.
- Monitored and analyzed long term trends in market-specific key metrics such as demographic information and market penetration, enabling prompt identification and resolution of potential issues contributed to enhanced data accuracy and reliability.
- Spearheaded the automation and implementation of **Python-based QA** processes for sample selection and weight generation, resulting in a notable **50%** increase in daily production efficiency.
- Developed a **Python-based web crawler** to automate the collection of holiday data, optimizing its use in the attrition model.
- Utilized advanced statistical analysis techniques to conduct **ad-hoc root cause analysis** for optimizing sample selection processes.

Big Data Engineer July 2018 – November 2020
Infosys Ltd. *Pune, India*
- Spearheaded migration of large datasets to **AWS** cloud servers for a **Fortune 500** Fintech client, enhancing data accessibility, scalability, and security through effective project management and technical expertise, reducing data storage cost by **25%.**
- Implemented an **AWS**-based data solution, utilizing **Airflow** for workflow orchestration. Streamlined data ingestion from **S3**, processing with **Glue** and **PySpark**, and storage of outputs in **Redshift**, achieving a significant **40%** increase in processing speed.
- Lowered AWS Glue job run-time by **10%** through refactoring Spark code, boosting data processing performance.
- Optimized over **100**+ SQL scripts for efficient data validation and extraction from **AWS Athena** and **Snowflake** database.
- Collaborated with clients to streamline production processes utilizing **AWS lambda** trigger decreasing human labor by **25%.**

PROJECT EXPERIENCE

- **Predicting Hotel Booking Cancellations:** Developed and implemented a high-performing ensemble-based classification solution, utilizing **LR**, **KNN**, **DT**, **RF**, and **ANN** models, resulting in a **97%** accuracy rate, enabling proactive customer churn prediction and targeted retention strategies to drive business growth.
- **Real-Time Parking Space Detection:** Developed and fine-tuned a custom **YOLOv5** model for parking space detection, leveraging transfer learning techniques to achieve high accuracy and robust performance.
- **Sentiment Analysis and Topic Modeling:** Employed **NLP** modules such as Vader Lexicon and Latent Dirichlet Allocation to perform sentiment analysis and topic modeling on smart watch brand reviews from Reddit API.
- **Lean Six Sigma Project:** Implemented Lean Six Sigma methodologies to reduce SLA time by **15%** and generate **$1.2M** cost savings. Leveraged **Tableau** for data visualizations, effectively identifying process bottlenecks and presenting compelling data-driven stories that enabled actionable insights and informed decision-making.

SKILLS & ACTIVITIES

- **Computer Skills:** Python, C++, Spark, SQL, SAS, AWS, Microsoft SSIS, SPSS, Power BI, Tableau, Excel, GitHub, HIVE, Sqoop.
- **Technical Skills:** Statistical Analytics, Big data, Machine Learning, Deep Learning, NLP, Data Visualization, Decision Modelling.
- **Statistical Tools:** Hypothesis testing, A/B Testing, Time Series Forecasting, Regression Analysis, Design of Experiment
- **Key Skills:** Storytelling, Analytical Reasoning, Problem Solving, Project Management.
- **Extracurricular Activities:** College Spokesperson at Green Revolution Program, Captain of College Basketball Team.