Apuroop Kotha

Boston, MA | 6172388660 | apuroopkotha2@gmail.com | Portfolio | LinkedIn | GitHub | Tableau

Technical Skills

Data Visualization: Tableau, Power BI(DAX), Quick sight, Seaborn, Matplotlib, Plotly
Programming: Python, R Programming, SQL
Databases: SQL Server, Snowflake, PostgreSQL, MySQL, Oracle, MondoDB
Cloud & Data Warehousing: AWS (S3, EC2), GCP (BigQuery), Snowflake
Data Engineering: ETL Processes, Database Management, Data Pipeline Automation, Alteryx, Hadoop, Spark, Hive, Presto
Machine Learning: Pandas, NumPy, Scikit-learn, TensorFlow, PyTorch, NLP, Supervised/Unsupervised Learning, Model Evaluation, Time Series Analysis
Soft Skills: Analytical Thinking, Problem-Solving, Cross-Functional Collaboration, Project Management, Agile Methodologies

Education

Northeastern University, Boston, MA **Sept 2022 - March 2024**

- *Master of Science in Data Analytics – Data Science, Applied Machine Intelligence, Risk analytics*

Sathyabama University, India **June 2016 - March 2020**

- *Bachelor of Technology,*

Work Experience

Data Scientist Intern | Bond.AI | Arkansas, United States ***June 2024 – August 2024***

- **Conducted A/B testing and model evaluation** on budgeting data using Scikit-learn to optimize predictive models for consumer expenditure based on spending patterns and transaction datasets.
- **Automated the extraction, transformation, and loading (ETL)** of housing, census, and consumer data using APIs and, streamlining the data pull process by python for efficient analysis and ensuring large-scale data integration.
- Integrated consumer spending **models into production**, focusing on scalability and performance using TensorFlow and GCP, facilitating the analysis in relation to the CPI index and other state-level indicators.

Data Scientist | LatentView Analytics ***April 2020 – September 2022***

- Designed and deployed scalable **AWS infrastructure, integrating S3, RDS, and Redshift** to optimize large-scale data pipelines for machine learning operations.
- Developed and implemented **Radial Basis Neural Networks (RBNN)** and **Custom Text Classifiers**, enhancing automated data categorization accuracy by leveraging **unsupervised clustering algorithms**.
- Optimized data preprocessing using **SQL, Python, Hive, and Presto**, reducing data query processing time by 30% and ensuring high-quality data for ML models.
- Improved categorization accuracy by 35% through advanced feature engineering with **regex patterns, Bigrams & Trigrams**, and **NLP techniques** within preprocessing workflows.
- Deployed and monitored ML models in cloud environments, ensuring real-time performance with AWS and automating data mapping and validation processes using Python.
- Created interactive **Tableau and Power BI dashboards**, increasing data accessibility and enabling real-time decision-making across teams by 80%.
- Collaborated with **cross-functional teams** to gather requirements, design scalable ML architectures, and deliver data-driven insights for business decision-making.

Data analyst Intern | MAHSA University, Malaysia **April 2019 - *September 2019***

- **Conducted time series** analysis on biomedical data using Python, improving disease prediction models by 30% through the application of machine learning.
- **Developed interactive Tableau dashboards** to visualize health data trends, enhancing stakeholders' ability to explore complex datasets dynamically.

Data analyst Intern | Virchow Biotech ***April 2018 - September 2018***

- **Automated Data Collection & Integration** using Python from LIMS, APIs, and quality control systems. Streamlined data integration into centralized SQL databases.
- **Applied statistical methods using SciPy and Stats Models** to detect trends and anomalies in biomedical data, supporting operational decision-making.

Projects

- **Sneaker Resale Market Analysis,** 2024 - Analyzed resale market data using ARIMA, SARIMA, and Gradient Boosting models to predict sneaker prices with high accuracy. Built interactive dashboards using Tableau to visualize price trends.
- **DocDigitizer : Analyzing and Enhancing Customer Interaction Through Data-Driven Insights**, 2023 - Analyzed 6.1M log entries at DocDigitizer, identifying server performance issues. Developed decision tree models, 91% accuracy, presenting insights in Tableau.
- **Driving Data-Driven Innovation: Honda's Journey in the Age of Industry 4.0,** 2023 Tableau1, Tableau2 - Performed sentiment analysis on customer reviews using TextBlob, extracting insights to enhance the customer experience. Created an interactive Tableau