Apuroop Kotha

Boston, MA | 6172388660 | kotha.a@northeastern.edu | LinkedIn | GitHub | Tableau

Education

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

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

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

- *Bachelor of Technology,*

Technical Skills

Programming Languages: Python, R Programming

Databases: SQL Server, Snowflake, MongoDB, MySQL, Oracle, MSSQL, PostgreSQL

Frameworks: Pandas, NumPy, Scikit-Learn, TensorFlow, PyTorch, Apache Hadoop, Apache Spark

Data Science: Artificial Intelligence, Machine learning, Deep Learning, OpenCV, Computer Vision, Supervised Learning, Unsupervised Learning, Model Evaluation, Hyperparameter Tuning, Natural Language Processing (NLP)

Data Engineering: Hadoop, Hive, Spark, ETL Processes, Database Management

Cloud: AWS (*S3, Redshift, EC2*), GCP - Google Cloud Platform (*BigQuery*), Azure (*Azure Machine Learning*)

Visualization Tools: Tableau, PowerBI, Plotly, Seaborn, Matplotlib

Work Experience

Data Scientist Intern | Bond.AI | Arkansas, United States

June 2024 – August 2024

- Conducted data collection by using the **API** from the several open-source datasets, cleaning, and analysis on large datasets using statistical techniques and **machine learning algorithms** to derive actionable insights.
- Created visually compelling **data visualizations** in **Tableau** to effectively communicate complex findings and insights to managers within the organization.
- Worked cross-functionally with software engineers, data engineers, and business analysts to integrate data-driven solutions into AI products and business initiatives.

Data Scientist | Latent View Analytics

September 2021 – July 2022

- **Machine Learning Architecture:** Implemented **Radial Basis Neural Network (RBNN)** for product categorization, integrating clustering and advanced training datasets.
- **Cloud Infrastructure and ETL**: Leveraged **AWS** for scalable cloud infrastructure, deploying **ETL processes** for data extraction and transformation, optimizing data workflows and enhancing efficiency.
- **Data Extraction and Analysis**: Conducted advanced data extraction and analysis from **Hadoop-based databases** using optimized SQL queries in **Hive and Presto**, achieving a **30% reduction** in data processing time.
- **Data Visualization**: Utilized **Tableau** to transform raw data into visually compelling dashboards, increasing data accessibility by 80%.

Data Analyst | Latent View Analytics

September 2020 – September 2021

- **Machine Learning and NLP**: Applied advanced quantitative techniques and **fine-tuned machine learning algorithms**, leveraging bigrams and trigrams within a **Custom Text Classifier model** to predict product categories.
- **Accuracy Enhancement**: Utilized **regex** patterns and **NLP techniques**, resulting in a substantial accuracy enhancement from **60% to 95% in product classification**.
- **Model Evaluation**: Evaluated model performance using statistical metrics such as **accuracy, precision, and F1 scores**, ensuring correct maintenance of product category hierarchy within the ML framework.
- **Data Storage and Visualization**: Employed **Power BI** for data storage, visualization, and insights dissemination, streamlining stakeholder access to model outcomes.
- **Model Validation and Data Engineering**: Validated models using advanced **keyword techniques** and developed Python fuzzy scripts for efficient client account mapping.

Data Analyst Intern | Latentview Analytics
April 2020 – September 2020

- Proficiency demonstrated in **SQL** queries on **MySQL** for extracting, manipulating, and analyzing data from relational databases, ensuring integrity and efficiency in **database management**.
- Utilized advanced techniques in SQL and Python Pandas for **cleaning**, **transforming**, and **validating** data, ensuring its high quality and consistency for analytical purposes.
- Collaborated with cross-functional teams to gather and analyze business requirements, design and develop data visualizations and reports, and deliver actionable insights and recommendations based on comprehensive data analysis.

Biomedical Sciences – Data analyst Mobility Programe | MAHSA University
April 2019 – September 2019

- Leveraged technical skills including **QlikView, Python** (utilizing **Pandas** and **Matplotlib**) to analyze biomedical data, enhancing insights into disease pathology, pharmacology, and medical technologies.
- Engaged in immersive biomedical science learning, focusing on **healthcare-data analytics integration**, contributing to improved patient care and treatment outcomes.
- Collaborated within multidisciplinary teams, applying data analytics techniques to interpret patient records, diagnostic tests, and experimental results, fostering innovation in personalized medicine and healthcare research.

Biotech Manufacturing and Data analytics Intern | Virchow Biotech
April 2018 – September 2018

- **Collecting data** from various sources such as **LIMS**, **API's,** quality control systems and flat files (csv, Excel, Json) by utilizing **Python Scripts** to **automate** the extraction process.
- Write the python scripts to **clean** and **preprocess** the collected data and integrate this data into centralized database or data warehouse by using the **SQL and ETL (Extract, Transform, Load)** processes.
- Apply **statistical methods** to analyze data and identify trends, patterns, and anomalies using Python libraries like **SciPy** and **Stats Models**.

Data Visualization Intern | Sathyabama University
April 2017 – September 2017

- Design and implement complex **Tableau visualizations** leveraging advanced features such as **LOD calculations**, table calculations, and parameter actions to provide dynamic and interactive data exploration experiences.
- Perform **data modeling** and optimization techniques including data blending, custom **SQL queries**, and efficient data source connections to ensure optimal performance and scalability of **Tableau dashboards** and reports.

Projects

- **Sneaker Resale Market Analysis,** 2024
  - Analyzed sneaker resale market data (SKU, condition, size, gender) to uncover trends and predict prices using ARIMA, SARIMA, and Gradient Boosting models.
  - Developed predictive models (ARIMA, SARIMA, Gradient Boosting) for accurate sneaker resale price forecasting, leveraging detailed market data analysis.
- **DocDigitizer : Analyzing and Enhancing Customer Interaction Through Data-Driven Insights**, 2023
  - Conducted extensive EDA on 6.1 million log entries, visualizing traffic patterns, server performance, and user behavior to optimize resource allocation and identify anomalies.
  - Developed predictive models with decision trees and random forests, achieving 99% accuracy, and created interactive dashboards using Tableau and Python for real-time data visualization and analysis.
- **Driving Data-Driven Innovation: Honda's Journey in the Age of Industry 4.0,** 2023 Tableau1, Tableau2
  - Conducted sentiment analysis and natural language processing (NLP) on customer reviews to extract the most common positive and negative sentiments, providing valuable insights into customer feedback.
  - Developed an interactive Tableau dashboard to visualize key sentiment trends and the most frequently mentioned words, enabling Honda to enhance customer experience and address critical issues efficiently.
- **Text Analysis of Customer Reviews: Uncovering Sentiment Trends and Key Topics,** 2023 Git
  - Performed sentiment analysis on customer reviews using TextBlob, categorizing sentiments as positive, negative, or neutral, and analyzed sentiment trends over various time periods to identify fluctuations.
  - Employed Latent Dirichlet Allocation (LDA) for topic modeling, revealing hidden topics and key themes within the reviews, and visualized frequent words associated with each topic to uncover prevalent themes.