Rohith Adhitya Chinnannan Rajkumar

chinnannanrajkumar.r@northeastern.edu | +1 (857) 701-8735 | LinkedIn | GitHub | Portfolio | Boston, MA

EDUCATION

Northeastern University **Boston, MA**
Master of Science in Data Analytics Engineering, (GPA – 3.91) May 2025
- Relevant Coursework: NLP, Data Mining, Neural Networks, Data Management for Analytics, and Computation & Visualization

Amrita Vishwa Vidyapeetham **Coimbatore, India**
Bachelor of Technology in Computer Science and Engineering, (GPA – 3.1) Jun 2023
- Relevant Coursework: Machine Learning, Computer Science fundamentals, and Foundations of Data Science

EXPERIENCE

Technocart **Tamil Nadu, India**
Business Data Analyst Apr 2022 –Jul 2023
- Leveraged Amazon Seller Central data to **extract** and **analyze** customer search terms, identifying high conversion keywords improving organic visibility and increasing sales up by 42.3%
- Managed Amazon Advertising campaigns by optimizing keyword bidding strategies and reallocating ad spend based on **ROAS** analysis, reducing **ACOS** by 18%
- Developed **Tableau dashboards** visualizing weekly sales, ad spend, and inventory KPIs, enabling stakeholders to drive decisions on pricing and stock allocation
- Automated product performance data pipelines using **PySpark** and **SQL**, reducing manual effort by 45% and improving data accuracy
- Conducted **Competitor analysis** to assess and optimize pricing, listing and design strategies by identifying pricing gaps, keyword trends and product feature opportunities
- Identified emerging customer trends through **statistical analysis** and **market research**, leading to a 15% increase in website engagement
- Collaborated with **Cross-Functional** teams to forecast demand and reduce stockouts, increasing product availability by 35% and decreasing returns by 12%

PROJECTS

Predicting age-related health risks using Machine Learning *(Airflow, ETL)* Sep 2024 – Dec 2024
- Optimized data pipelines using Apache **Airflow** to pre-process large-scale datasets, enabling seamless integration with machine learning models
- Built and deployed predictive ML models on distributed systems, achieving 82.5% accuracy in identifying age-related health risks
- Designed data models to support **ETL** workflows to analyze health metrics, driving insights into aging-related risks and enhancing predictive modeling

Account Payable Flow *(SQL, MongoDB)* Sep 2023 – Dec 2023
- Automated the Accounts Payable process using **SQL** and **MongoDB**, reducing fund release time by over 80% and increasing system reliability
- Formulated robust **SQL query models** to adaptively handle diverse scenarios with fault-tolerance, ensuring resilience in addressing real-world challenges
- Optimized **NoSQL** queries, achieving a **5-fold** throughput increase and **50% latency reduction**

Seizure Detection for Epilepsy Management *(Machine Learning, Python)* Sep 2023 – Dec 2023
- Designed and deployed **CNN** and **Decision Tree** models, achieving 88.9% accuracy and an F1 score of 0.89 in seizure detection from EEG data
- Improved seizure detection accuracy by rebalancing training data with **SMOTE**, enhancing model sensitivity in minority classes
- Boosted model performance by 10% using **ensemble techniques** and hyperparameter tuning

Analyzing multilingual news article similarities *(TensorFlow, SBERT)* Jan 2024 – Apr 2024
- Implemented a multilingual news similarity model using **SBERT** and **TensorFlow**, achieving high accuracy in cross-lingual content evaluation
- Built data preprocessing pipelines with Python (glob, re) and **Keras**, cleaning and transforming datasets across 18 languages
- Reduced validation error by optimizing regression models using **Adam optimizer** and **Huber loss**, improving model generalization

Identifying Influence behind LinkedIn posts using Topic Modeling and Sentiment Analysis Jul 2022 – May 2023
- Performed media **impact analysis**, achieving a 16% increase in post reach by incorporating media content
- Applied advanced topic modeling techniques, including **LDA**, **BERT**, **RoBERTa**, **NMF**, and a hybrid **LDA+BERT** approach, demonstrating superior performance with a coherence score of 0.59
- Evaluated and compared sentiment models (VADER, RoBERTa, SVM), with **VADER** achieving highest accuracy at 92.2%

TECHNICAL SKILLS

Programming Languages: Python (NumPy, Pandas, Matplotlib, Seaborn, Scikit-Learn, TensorFlow, NLTK), SQL, Java, R
Data Analytical Tools: Tableau, Power BI, MS Excel (VLOOKUP, Pivot Tables), Jupyter Notebooks, Google Analytics
Data Warehouse Tools: MySQL, MSSQL Server, GCP, AWS (S3, EC2), Snowflake, ETL, Airflow, Kubernetes, Docker
Data Science and ML: Regression, Classification, Decision Trees, SVM, Clustering, Neural Networks, NLP
Other Tools and OS: Jira, Git, Agile, MS PowerPoint, MS Word, Microsoft Office Suite, Windows, Linux, G Suite