Banu Priya R K

Columbus, OH | +1-614-483-7875| banupriya2351@gmail.com | linkedin/banupriya-rk | github.com/banupriya-15

Summary

Data Scientist with expertise in machine learning, and statistical analysis, transforming raw data into meaningful insights through processing, visualization, and storytelling. Passionate about solving practical challenges in business and healthcare through data interpretation, predictive modeling, and impactful decision-making.

Education

Master of Science in Computer Science and Engineering – Ohio State University – CGPA: 3.58 Dec 2024
Coursework: Neural Networks, Data Mining, Fairness in Artificial Intelligence and Databases, Data Visualization, Parallel Computing, Advanced OS, Cybersecurity

Bachelor of Technology in Computer Science and Engineering – Anna University – CGPA: 4 June 2021
Coursework: Problem-Solving, Object-Oriented Design, Data Structures & Algorithms, Databases, Operating Systems, System Design, Networking, Machine Learning

Skills

- **Languages** : Python, R, Java, D3.js, HTML/CSS, JavaScript
- **Tools** : Tableau, Git, GitLab, Jira, Docker, Kubernetes, JUnit, Postman, RabbitMQ, OIPA, Microsoft Excel
- **Frameworks/Libraries** : PyTorch, TensorFlow, Keras, Scikit-Learn, H2O, Flask, Node.js, Express.js, React.js, Angular.js, TextBlob, OpenCV, RestAPI, Pandas, Matplotlib, Apache Spark, MLlib, Airflow, Hadoop, AWS, Microsoft Azure
- **Databases** : MySQL, PostgreSQL, Oracle SQL, NoSQL, DynamoDB, Elasticsearch, Neo4j
- **Certifications** : *Microsoft*-Azure Cloud Fundamentals, Azure AI Fundamentals, *Google*- Python for Data Science Certification, CyberArk Trustee Certification, *NPTEL* - Java, Machine Learning

Experience

The Ohio State University Wexner Medical Center Columbus, Ohio, USA
Data Research Scientist – Emergency Medicine *Jan 2025 – Present*

- Implemented the **BERT LLM model** to extract medical entities and classify cardiac arrest types from complex patient records, enhancing the accuracy of patient data categorization and improving analysis efficiency by 30%.
- Developed an algorithm to calculate medication response time and applied **Mann-Whitney U tests**, **IQR analysis**, and stratified analysis using Matplotlib and MLlib to assess differences in patient responses, discovering breathing patterns linked to survival probability.

Data Science Intern *May 2023 – Dec 2024*

- Applied **random forest model** to predict survival rates with 92.4% accuracy, identifying critical survival determinants based on resuscitation efforts, and performed time series analysis to assess how these determinants influence patient outcomes over time.
- Developed a Python module with **Apache Airflow, Pandas, and NumPy** to automate processing of 5,500+ longitudinal capnography records, improving efficiency by 40% through data pipeline optimization and structured transformation.

Cognizant Technology Services Chennai, India
Software Engineer – Data & Migration *Feb 2022- Dec 2022*

- Led a **data migration** project for Nationwide Insurance, transitioning 200K+ Universal Life Insurance contracts to the **Oracle Insurance (OIPA)** platform using **SQL** and ETL pipelines, ensuring data accuracy and seamless integration.
- Developed automated migration strategies with **Apache Spark and ETL pipelines**, improving data transfer efficiency by **30%** and ensuring scalability with a **PIT**(Point-In-Time) migration strategy.
- Reduced data discrepancies by 50% through **SQL debugging, data reconciliation, and validation** in **PostgreSQL**, resolving financial and non-financial mismatches and improving system reliability.

Programmer Analyst *Sep 2021- Jan 2022*

- Configured product parameters and rates in OIPA using **SQL, XML, and XSLT**, improving setup efficiency by **20%** and enabling faster feature deployment. Performed **root cause analysis** post- migration in OIPA, leading to reduced downtime and improving data integrity.

Programmer Analyst Intern *May 2021- Aug 2021*

- Developed a retail banking application using **microservices architecture** and Spring Boot, integrated with PostgreSQL for database management, and deployed on AWS, enhancing the scalability and efficiency of financial services.

Projects

Worded Bar Chart Visualization: Developed a Python script to visualize the geographical spread of non-profit organizations across the US using worded bar charts, improving data interpretability and accessibility.
Newspaper Sentiment Analyser: Designed an NLP tool using TextBlob and Web Scraping to extract, summarize, and analyze sentiment from NYTimes technology articles, providing insights into public opinion trends.
Smart Mobility: Built microservices using Bing Maps API and Open Street Maps for real-time data extraction, enhanced with Graph Convolutional Networks (GCN) to improve predictive analytics and accurately predict accident zones.
Pokémon Go Behavioral Analysis: Utilized Python and SQL for preprocessing and analyzing Pokémon Go player sensor and survey data, uncovering psychological insights by correlating in-game movements with player feedback.
Early Detection of Plant Pathology: Created a machine learning app using CNNs to detect plant diseases with 92% accuracy, offering early predictions and treatment suggestions to enhance agricultural disease management.
Face-X: Developed Face-X, a facial recognition-based attendance system using OpenCV and Haar Cascade, achieving 98% accuracy and significantly improving attendance tracking precision.

Publications

Nassal, M. M. J., **Banu R. K., et al.** (2024). *Temporal trends in end-tidal capnography and outcomes in out-of-hospital cardiac arrest. JAMA Network Open,* DOI:10.1001/jamanetworkopen.2024.19274. Contributed by analyzing end-tidal CO_2 trends, developing visualizations, and calculating **slope** values using **regression** analysis to assess its role in predicting cardiac function recovery.