Robins Yadav

● robiny@arizona.edu ● (757) 831-8753 ● in/robinyuarizona ● datascienceportfol.io/robinyuarizona ● US Green Card Holder

SUMMARY

Data professional with a master's degree specializing in Data Science and Machine Learning. Proficient in Python, SQL, AWS, and Big Data, with expertise in healthcare analytics, NLP, and LLM. Skilled in model development and MLOps for deploying scalable machine learning solutions.

EXPERIENCE

Software Engineer - Data Science
Cotiviti, Inc. (Healthcare) **Sep 2022 - May 2024, Nepal**

- Led a team to victory at the Cotiviti Hackathon 2023 by automating the manual process of reviewing and labeling document pages (text mining) using machine learning and natural language processing in the Mavis application, driving strategic initiatives through efficient page classification.
- Conducted research on advancements (technology adoption) in generative AI and machine learning, integrating innovative techniques with PySpark for data exploration and visualization to enhance U.S. healthcare data for member and provider.
- Designed and implemented a unified master record system for members and providers, applying advanced data processing using DocumentDB, and optimal software engineering strategies including version control software like Git.
- Developed and deployed Python-based tools for **JSON** parsing, and REST **API** integration, optimizing **data interaction** and API capabilities, which facilitated efficient exception management and across organizational data stewards to reduce risk management.

Graduate Researcher (PhD) - Data Science & Machine Learning
Arizona State University & Boise State University **Aug 2021 - July 2022, USA**

- Conducted advanced research on multivariate **time series analysis** to improve building energy technology and **fault detection**, enhancing building manufacturing performance through the application of specialized algorithms using statistical modeling package using scripting language python.
- Developed and implemented algorithms to detect resilient multivariate temporal (RMT) features, significantly improving the precision and efficiency for quantitative analytics such data models, data mining, and anomaly detection in heavy industry environment.
- Created and optimized PySpark modules for predictive analytics and machine learning in a Databricks environment, utilizing Python libraries and Spark tools to develop data engineering ETL pipelines and job scheduling.

Graduate Researcher (MS) - Data Science & Machine Learning
University of Arizona **Aug 2018 - May 2020, USA**

- Conducted in-depth evaluation of quantization-optimization algorithms, including SignSGD and TernGrad, establishing mathematical modeling within machine learning **distributed systems** to optimize data transmission and enhance training efficiency, significantly boosting performance.
- Utilized data acquisition via Snowflake to manage data integration, data manipulation, data cleansing, and analyze large datasets (unstructured data structures), employing data analytical techniques to enhance production system efficiency and capabilities by ensuring data quality.
- Designed and implemented an interactive dashboard for data visualization software such as Tableau, and performed statistical testing using statistical packages like R

PROJECTS

Generative AI, Large Language Models (LLMs), Recommender System, and Data Analytics; GitHub: github.com/robinyUArizona/

- **T-shirt Store (Retail Industry):** Developed an **LLM** system using Google Palm (PaLM2) within the **LangChain** framework. This system could interpret user queries in natural language, convert them into **SQL** queries, and execute them on a MySQL database, facilitating user interaction via **Streamlit** and other development tools including engineering principles.
- **Hotel Booking Demand:** Developed a **predictive modeling** to forecast hotel booking cancellations, utilizing exploratory data analysis (EDA), **feature selection**, and **oversampling techniques** to address class imbalance; evaluated five classification ML models including deep learning, logistic regression, and achieved an 87% precision as performance metrics using best model after **hyperparameter** refining**.**
- **Anti-Money Laundering Fraud Detection:** Built a fraud detection financial system using **Decision Trees, Random Forest**, and **XGBoost** with **feature engineering**, **over sampling** to address class imbalance, and **hyperparameters** tunning. **Deployed** the best model on cloud service AWS (**Docker**, **ECR**, **EC2**), integrated with **CI/CD** via **GitHub Actions**, and monitored using **MLflow**. Developed a **Streamlit** interface for financial services entity.
- **Lending Club Datasets:** Performed comprehensive univariate and multivariate analysis, including statistical testing in **Databricks'** Spark Cluster to evaluate borrower (consumer lending) repayment likelihood through regression analysis. Extracted, cleaned, and prepared large e-commerce datasets stored in Amazon S3, enhancing data transformation and enabling more solutions using recommender systems.

EDUCATION

M.S. in Electrical and Computer Engineering; GPA 4.0/4.0; **University of Arizona** **May 2020, Tucson, AZ**

B.S. in Electrical Engineering; GPA 3.7/4.0; **Boise State University** **May 2018, Boise, ID**

TECHNICAL SKILLS

Python, R, Java, MongoDB, SAS, Perl, C++, Go, Matlab, NumPy, Pandas, SQL, NoSQL, Seaborn, Scikit-learn, PyTorch, Keras, Tensorflow, NLTK, Spark (PySpark), MLLib, Hadoop, Hive, Pig, Django, Data Analysis, Computer Science, Programming languages, Linux, Data Management, Transformers (BERT, GPT), RAG, Docker, Jenkins, GitHub Action, AWS (S3, EC2, IAM, SNS, Lambda, SageMaker), Microsoft Office, GCP, Microsoft Azure, MLflow, Flask, Streamlit, Tableau, Power BI, Project Management