Raj Bhowmik

(812)606-2478 | rbhowmik@iu.edu | https://www.linkedin.com/in/rajbhowmik **|** https://github.com/RajBhowmik

Experienced data scientist with a demonstrated history of developing and applying statistical analysis and modeling as well as machine learning techniques to solve real-world problems. Highly skilled in Machine Learning, Deep Learning, Statistical Modeling, Quantitative Analytics, Python and SQL.

EXPERIENCE

Data Scientist Intern at **Ascentt Business Systems, Fremont, California**   **May 2022 – December 2022**

- Built a state-of-the art ML model (predicting health index of vehicle parts) by integrating vehicles' sensor and parts data. The model generated a performance improvement of **30%** over the current production model and was responsible for a reduction in equipment breakdown by **60%. ~** Survival Regression, Autoencoder.
- Developed an ML boosting forecasting model for vehicle sales forecast in major global markets to improve operational efficiency and profitability. Solution space includes Prophet, Random Forest, and Deep learning LSTM-RNN models significantly improving over current models by **20%**.
- Designed an anomaly detection system (Isolation Trees) that ensures safety and reliability by detecting hidden risks in machinery parts.

Data Scientist at **Infosys Limited, Bengaluru, India**   **August 2016 - April 2021**

- Built, validated, tested, and deployed a predictive model (XGBoost with an accuracy of **88%**) for a major automobile client's Pricing System focused on improving business outcomes. Conceptualized, built, and continually improved on vehicle sales forecasting models by **20%**. This enabled informed decision-making on planning, purchasing, and allocation.
- Formulated different NLP techniques like Bert Multilingual, Zero-Shot TM, information extraction and sentiment analysis using pre-trained transformers for tele-critical care data from Twitter, which resulted in the identification of critical insights and trends in the data.
- Developed a predictive maintenance model for a commercial airline using AWS, leveraging sensor data to predict when aircraft components are likely to fail. Utilized AWS Sagemaker for model training and deployment, and AWS Lambda for real-time inference. This reduced unplanned downtime and maintenance costs by **25%**. Implemented multi-threading to reduce runtime by **60%**
- Optimized client's data pipeline architecture by identifying, designing, and implementing process improvements such as automating manual processes, optimizing data delivery, and re-designing infrastructure for greater scalability.

TECHNICAL SKILLS

- **Statistics**: Descriptive statistics, EDA, Bayesian, Correlation, Causal inference, Experimental design, Hypothesis testing
- **Machine Learning**: Regression, Classification, Clustering,NLP, Time-Series, Deep Learning, PyTorch, TensorFlow, scikit-learn.
- **Computing**: Python, R, SQL, Shell Scripting, Tableau, AWS, Postgres, MongoDB, Spark, Hadoop, Hive, Airflow, Kafka

ACADEMIC PROJECTS

- **Text summarizer**: Developed a text summarizer that leverages cosine similarity and OpenAI's GPT-3 language model for text summarization. Conducted comparative analysis of the two approaches and identified that GPT-3 outperformed cosine similarity in generating summaries that are more coherent and flexible.
- **Object and License plate detection**: Constructed a robust automatic number plate recognition system with the help of OpenCV libraries. By leveraging the use of CNN with VGG16 backbone we composed a model that can detect the type of vehicles and extract the license plate of the vehicles.
- **Machine Translation**: Crafted a scalable machine translation model using RNN with LSTM. The model was improved using RNN with embedding, along with GRU and tweaking encoding layers, which outperformed LSTM by **60% ~ TensorFlow, RNN**

EDUCATION

Indiana University Bloomington   **August 2021-May 2023**
Master of Science in Data Science

Satyabhama Institute of Science and Technology   **August 2012-April 2016**
Bachelor of Engineering- Computer Science