Monu Kumar Pathak

Industrial Engineering and Operation Research
Indian Institute of Technology Bombay

M.Tech
23m1526@iitb.ac.in
+91-9971088936

PROFESSIONAL EXPERIENCE

Cognitio Analytics India PVT LTD | Data Science Intern *(May'24-Jul'24)*

- Reviewed over **10 research papers** on Generative AI and documented their use case into the medical domain
- Preprocessed medical database containing **58,000** hospital admissions, **38,000+** adult patients and their diagnosis
- Developed **LSTM** and **Transformers** based models to predict patient next visit embedding from medical history
- Classified diseases into **5 classes**, achieved **19%** higher **AUROC** with **LightGBM** over end-to-end NN models

ACADEMIC WORK AND PROJECTS

A.I. for Advance Data Analysis and Simulation | *Master's Thesis Project* *(May'24-Present)*
Guide: Prof. Jayendran Venkateswaran

Objective: To develop a system that analyzes simulation data & generate meaningful insights using **Generative A.I.**

- Implemented **Data Farming techniques**, integrating the **explainable AI (XAI)** to analyze large simulation data
- Performed **1M+ simulations** & clustered the output data using **K- Means, DBSCAN & Hierarchical** algorithms
- Extracted insights & improved model transparency with explainability in **Neural Networks** and **Random Forest**
- **Future work**: Integration of the **LLM** technologies such as **Microsoft's InsightPilot** and **JarviX** into the system

Differentially Private Movie Recommendation System | *Course Project* | *Machine Learning* *(Jan'24-April'24)*
Guide: Prof. Balamurugan Palaniappan

- Applied **3 privacy-enhancing** algorithms by injecting **Laplace noise** at the input, processing & output stages
- Developed **Matrix Factorization** algorithms from scratch, including **Alternating Least Squares(ALS)** & **SGD**
- Achieved **95**% accurate prediction while balancing the trade-off between **privacy parameter** & model **RMSE**

LLM for Automated Commonsense Knowledge Graph Construction | *Seminar* *(Jan'24-April'24)*
Guide: Prof. N. Hemachandra

- Implemented **COMET** by **fine-tuning GPT-1** to generate and enhance the commonsense knowledge inferences
- Leveraged the **ConceptNet** and **ATOMIC** knowledge graphs with over **40M** and **877K** commonsense concepts
- Achieved **perplexity** of **11.14** & **BLEU-2** score of **15.10**, reflecting quality & novelty in commonsense generation

Simulation & Optimization of the Operations in Rocktron | *Course Project* | *Simulation* *(Jan'24-April'24)*
Guide: Prof. Jayendran Venkateswaran | *IISE Rockwell Simulation Competition*

- Analyzed process, input, and layout data, fitting valid distributions using **KS Test** for the accurate simulation
- Utilized **DES** to model 5 different scenarios using **Arena** & analyzed worker & machine utilizations as **KPIs**
- Developed a plan to **reduce the workforce** by **51%** while ensuring it could handle **1.5x** of the current capacity

Optimizing Flight Delays using Heuristics | *Course Project* | *Application of OR in Service Sectors* *(Jan'24-April'24)*
Guide: Prof. Narayan Rangaraj

- Developed a **heuristic** based solution for a complex **multi-objective** problem to effectively **reduce flight delays**
- Utilized **Gurobi** Solver to generate **Pareto optimal solutions** & created a pool of efficient options for selection
- Improved Pareto optimal solutions using a **threshold-based approach**, iteratively minimizing the flight delays

Airline Ticket Sales Prediction using Time Series Forecasting | *Self Project* *(Jan'24-April'24)*

- Analyzed the given data to check for **stationarity** & decomposed it to get **level, trend, seasonality**, and **residue**
- Performed **ADF test** for stationarity & used **ARIMA, SARIMA, Prophet** and **XGBoosT** to predict future sales
- Achieved **14.27% MAPE** using Facebook's Prophet model, surpassing SARIMA's **19.23%** in forecasting accuracy

Vehicle Routing using Metaheuristics | *Course Project* | *Optimization Techniques* *(Aug'24-Dec'24)*
Guide: Prof. Vishnu Narayanan

- Utilized a dense network of **36 nodes**, including a depot, with X and Y coordinates & customer demand values
- Applied **Genetic Algorithm** & **Genetic K-means** metaheuristics to approximate optimal solution for the VRP
- Achieved an optimal cost of **1,702** units with **3 vehicles** using GA, and **957 units** with **5 vehicles** using GKM

Retrieval Augmented Gereration(RAG) Based Chatbot | *Course Project* | *Deep Learning for NLP* *(Jan'24-May'24)*
Guide: Prof. Pushpak Bhattacharyya

- Developed an **AI-powered chatbot** using **Retrieval-Augmented Generation (RAG)** to minimize hallucinations
- Integrated the **OpenAI's GPT-3.5** for language modeling via LangChain for better conversational capabilities
- Managed a **30-book** vector space for domain-specific knowledge integration and built a UI using Streamlit

KEY COURSES & TECHNICAL SKILLS

- **Machine Learning**: Principles and Techniques
- **Deep Learning** for **NLP**
- Modelling and Computation Lab
- Mathematical **Optimization** Techniques
- Engineering **Statistics**
- Simulation Modelling and **Analysis**

- **Languages**: Python, C++, Java, SQL, Kotlin,
- **Libraries**: Scikit-Learn, Pytorch, NLTK, Pandas, Matplotlib
- **Softwares & Tools**: Hugging Face, Power BI, Matlab, OR Tools, OpenCV, Android Studio, Azure ML, Gurobipy

SCHOLASTIC ACHIEVEMENTS & EXTRACURRICULARS

- Represented **IIT Bombay** on a **global level** in the **IISE Rockwell Student Simulation Competition** *(2024)*
- Achieved **AIR 23** in GATE 2023 and **AIR 196** in GATE 2022 Textile Engineering and Fiber Science. *(2023)*
- Developed an **IoT device** for hostel power conservation and secured **3rd place** in Jarvis Technical GC *(2023)*