Dhanush Binu

Phone Number: (412) 390-9237 | binudhanush0@gmail.com | Personal Website | Linkedin | Medium

Motivated Data Scientist with a broad range of technical expertise and research experience in machine learning and big data. Actively seeking roles in generative AI and NLP to further explore and contribute to these evolving areas of interest.

Education

University of Pittsburgh, Pittsburgh PA — August 2023 - Present
Computer Science | Masters in Science

Indian Institute of Information Technology, Kottayam — August 2019 - May 2023
Computer Science and Engineering | Bachelor of Technology (Hons.) | CGPA: 8.45

WORK EXPERIENCE

CleanTech Solar — **Mumbai, India**
Data Analyst Intern — *June 2022 - August 2022*

- Created eight custom visualizations, allowing for monitoring complex metrics across 10+ project sites.
- Managed central database using SQL, implementing constraints and triggers for data integrity.
- Implemented a database automation system for geographic data, resulting in a 30% reduction in manual data processing time.
- Resolved bugs in email and PV system data, leading to an estimated 15% reduction in system downtime.

PROJECTS

Hate Speech Target Group Identification using RoBERTa — **August 2023 - Ongoing**

- Curated a dataset by integrating Civil Comments, Social Bias Inference, Kennedy et al. 2020, and HateXplain.
- Enhanced model accuracy by over 10% through preprocessing and fine-tuning of a RoBERTa model.
- Adapted RoBERTa model for implicit target prediction, achieving 77% weighted F1-score across 10 classes.

Distributed Data Parallel LLM training with Apache Spark — **August 2023 - December 2023**

- Built and managed a Docker-based Apache Spark environment consisting of several workers and one master node.
- Trained DistilBERT across multiple worker nodes resulting in a substantial 45% reduction in training time.
- Achieved an additional 5-8% reduction in training time by tuning parameters such as worker node counts and core allocations.

Crime Trend Prediction and Analysis using Big Data — **August 2022 - March 2023**

- Designed a Spark-based data pipeline on Databricks for processing and analyzing large-scale crime data.
- Employed a generalized additive model to identify patterns in crime within Atlanta from 1997 to 2020.
- Analyzed multiple parameters in dataset using an ensemble of clustering and time series models to identify recurring patterns in distribution and frequency of criminal incidents.

Cheating Detection in Exams using Pose Detection and Machine Learning — **January 2022 - December 2022**

- Modeled a robust motion detection algorithm using OpenCV resulting in a 90% accuracy in detecting cheating behavior in exam videos.
- Designed an accurate face recognition system for identifying individuals in exam footage.
- Implemented human pose extraction from video frames for vector conversion and classified them using an ExtraTree ensemble classifier.

PUBLICATIONS

Academic Dishonesty Detection in Exams Using Pose Extraction | SpringerLink — Published
International Conference on Inventive Systems and Control ICISC 2023, Springer
Conference Date: 31th January 2023

An Efficient Framework for Crime Type Prediction using Feature Engineering and Machine learning | SpringerLink — Published
International Conference on Data & Information Sciences ICDIS 2023, Springer
Conference Date: 17th June 2023

TECHNICAL SKILLS

Data Science : Python,Numpy, Pandas, Sklearn, TensorFlow, PyTorch, Keras, OpenCV, Excel, Matplotlib, PowerBI.
Data Engineering and Mining: Docker,Apache Hadoop, Apache Pig, Apache Spark, Apache Kafka, Selenium, Beautiful Soup.
Cloud Technologies: Azure ML, Azure SQL, Google BigQuery, Kubernetes.