Manohar Lakshmana Vadaga

mxv5801@g.rit.edu|+1-585-300-1213
https://www.linkedin.com/in/manohar-vadaga/

EDUCATION

Rochester Institute of Technology, Rochester, NY   **Expected: May 2021**
Master of Science in Computer Science **GPA: 3.6/4**
Course Work: Advanced Object-Oriented Programming, Data Structures and Algorithms, Distributed Systems, Statistical Machine Learning, Big Data Analytics, Deep Learning, Computer Vision
University of Mumbai, Mumbai, India   **May2017**
Bachelor of Engineering in Computer Engineering with *Distinction* **GPA: 8.25/10**

SKILLS

Programming Languages: Java, Python, SQL
Framework and Tools: Docker, Docker Compose, Kubernetes, Airflow
Big Data: Hadoop, Hive, Spark, HBase, Pig. **Cloud Platform:** Microsoft Azure Cloud, AWS S3, AWS Lambda, AWS Kinesis
Web Technology: HTML5, JavaScript, CSS3, PHP, Selenium, Scrapy for Web Scraping, Django
Technical Expertise: Database Management, Data Analytics, Data Visualization, Big Data Processing, Software Engineering

PROFESSIONAL EXPERIENCE

GlobalFoundries, Malta, NY – *Machine Learning Co-op*   **August 2019 – December 2019**
(Jupyter Notebook, Python, Selenium, Scikit-Learn, Pandas, Numpy, MS SQL)
- Developed a machine learning model which would help to predict the DBF points on the silicon wafer after the lithography process based on the levelling data decoded from the mdl files and achieved an RMSE of 6nm.
- Developed a Beta version of machine learning model using Bayesian optimization to predict the failure of the vacuum pumps attached to high end tools inside the FAB based on the sensor data and newly generated features which could save **$1 million annually**.

West Pharmaceutical Services, Exton, PA – *Data Science Co-op*   **June 2018 – December 2018**
(Azure Databricks, Power BI, Python, Selenium, Scikit-Learn, Pandas, Numpy, NetworkX, MS SQL, Gephi)
- Worked in Agile Methodology (team) environment and collaborated with different teams across the organization and created a pipeline to scrape the newly added trials using Microsoft Azure Data Factory and performed data transformation and Exploratory Data Analysis from different data sources using Microsoft Azure Databricks and visualized it in Power BI and reported and presented the results to the stakeholders.
- Performed Network Analysis on the government clinical trials data using Python NetworkX and visualized it using Gephi tool.
- Performed text mining and Topic Modelling to cluster the research articles based on the abstracts from PubMed Database using LDA model and then used Tf-Idf and RAKE for each cluster to extract keywords for each document.

PROJECTS

Implement Model Compression Techniques for CNNs with Model Interpretability (Pytorch, Python, ImageNet dataset)
- Implemented Pruning, Layer Ablation and Quantization techniques for pruning popular CNN models to achieve similar accuracy as baseline model on ImageNet Dataset by reducing model weights by over 20% and ran the model on multiple GPUs in parallel.
- Implemented Class Activation Maps (CAMs) to understand model interpretability for CNN models.

Implemented End to End Neural Machine Translation Model (Kubernetes, Docker, GCP, Flask, Keras, Pandas, Numpy)
- Designed a Neural Machine Translation Model using Keras LSTM layers and used Glove vectors for word vector representation in Word Embeddings Layer of the model and designed a Python Web App using Flask to interact with the model in real time.
- Deployed the model in Docker container and hosted it in Kubernetes in Google Cloud for scalability and loadbalancing.

Apache Airflow and Docker for Machine Learning Pipeline (Python, Docker, Airflow)
- Implemented Decision Tree and Adaboost Classifier from scratch to classify English and French sentences.
- Used Airflow to create Airflow DAGs to create tasks for data preprocessing, training and testing of decision tree, Adaboost models and implemented the above pipeline using docker compose.

Publisher Subscriber System (Java, Docker, Docker Compose)
- Designed a time and space decoupled system where a subscriber can subscribe to the topics of their choice.
- Multithreaded Publisher, Subscriber and Event manager modules were deployed on separate Docker containers.

Peer to Peer Distributed System Architecture (Java, Docker, Docker Compose, Kademlia Protocol)
- Designed a decentralized peer to peer system where a client can store and look-up for files based on the unique hashcode.
- Designed with Docker compose and used JSON RPC for communication and Kademlia Protocol for scalability to decide where to store the files. Trie data structure was employed for building routing table in each node.

Microsoft Azure Big Data-Capstone (Azure HDInsight, Azure Databricks, Azure Data Factory, Power BI)
- Created a partitioned Hive table for the inventory retails data stored in Azure Blob Storage and performed data cleaning and transformation using the Python user defined functions and exported the Hive table to an Azure SQL Database using SQOOP.
- Collected the real time data from Azure Event Hubs and captured it using Spark Structured Streaming on Azure Databricks and stored the clean data on Azure Blob Storage and visualized it in Power BI.