**Suman Kalapatapu**

12805 Shaker Blvd, Cleveland, OH 44120| Graduating Dec 2020|Open to relocation

kalapasn@mail.uc.edu | (513) 307 - 6877 | <https://www.linkedin.com/in/skin/>| <https://github.com/suman500bn>

**Education**

**University of Cincinnati** **Fall 2019 - Dec 2020**

Master's in Computer Engineering, GPA: 3.7

Graduate Coursework in Intelligent Data Analysis, Machine learning, Big data analytics, Cloud computing, Deep Learning

**CVR College of Engineering** **Fall 2012 - May 2016**

Bachelor's in Electronics and Communication Engineering GPA: 3.8

Undergraduate Coursework in Object Oriented Programming, Operating Systems, Algorithms and Data Structures

**Areas of Expertise**

|  |  |
| --- | --- |
| Programming Languages | Python, Java, JavaScript, Linux shell, SQL, C-Programming |
| Big Data and Cloud | Hadoop, PySpark, Hive, Informatica, AWS EC2, Lambda, S3, Docker, Azure |
| Framework and Database | Node.js, Express, Django, Flask, Git, MS SQL server, MySQL, Postgres, MongoDB |
| Data Science Libraries | Pandas, NumPy, Scikit-Learn, Seaborn, Matplotlib, PyTorch, NLTK, Tableau |

**Professional Experience**

**Backend Developer and ML Engineer Y apps Lab May 2020**

* Lead team building media intensive social networking app for foodies using node.js handling millions of requests per day.
* Developed ML algorithm to generate newsfeed based on personalized interests and advertisement system to target specific users based on interests. Also developed batch script which records transactions for these promotions in Stripe and Paypal.
* Developed search autocomplete API and integrated with Lambda, S3, API Gateway, CDN to support concurrent users.

**Data Science Internship Cincinnati’s Children Hospital Jan 2020**

* Implemented proof of concept to find impact of chemotherapy in cancer patients at different phases using Machine Learning
* Visualized survival analysis of patients having different side effects through Cox Proportional Hazard in R
* Created a novel approach to predict survival for different patients (age, gender, ethnicity) from initial phases which can help physicians to change treatment procedure.

**Software Development Analyst NTT DATA Global Delivery Services Ltd. Oct 2016 – Jul 2019**

* Trained and deployed predictive models on sales across different regions which impacted critical business decisions
* Built, tested, and maintained ETL data pipelines integrating data from multiple servers into a cohesive data warehouse and other repositories for reporting and analysis.
* Created dashboards in Tableau showing sales across different regions which made marketing to strategize their promos.
* Reduced 30% human efforts creating Spark scripts which cleaned, extracted, merged data across disparate systems.
* Developed automatic backup, restore linux scripts and deployed at 1300+ retail stores which reduced 40% data loss issues

**Projects**

* **Recommender system**: Built 5 types of recommender system based on popularity, classification, collaborative filtering, content filtering approach using Pandas, Scikit-learn on movie, bank, restaurant, housing datasets.
* **Hosted Docker application in Cloud**: Built CRUD application with upload and download functionality using Python Flask, MySQL for Backend and HTML, CSS for front end, containerized application using Docker and deployed it in AWS EC2.
* **Weather Data analysis**: Cleaned and collated TBs of datasets using SparkQL, Python and developed solutions for efficient analysis of weather data of Cincinnati region for last 10 years on Hadoop cluster.
* **Crime data analysis using Azure Databricks:** Explored parquet files data using optimized PySpark queries and visualized through Tableau integration. Extracted time Series data and forecasted crime count through ARIMA model
* **Car Purchase Evaluation**: Performed data analysis and performed classification and clustering techniques to make the model successfully classify which class car belong to.

**Certification**

* Deep Learning Specialization by Nvidia| Machine Learning certification by Stanford university
* Microsoft Azure AZ900 certification| Tableau certified Data Scientist | PySpark certification from DataCamp

**Awards and Achievements**

* 1st Prize for paper presentation on topic E-paper at Mahatma Gandhi institute of technology over 200 participants.
* Spot Award of excellence in one year at NTT Data and got promoted to grade 6 within one and half year of joining.
* Recipient of Graduate Incentive Scholarship