Chinmayi Hegde

San Jose, CA | +1 669-649-3847 | chinmayilokeshwar.hegde@sjsu.edu

Data Scientist/ Junior Developer

GitHub | LinkedIn | Portfolio

MS in Computer Science	GPA: 3.8/4.0	**San Jose State University, CA**	Expected May 2024
BE in Computer Science	GPA: 7.9/10.0	**PES University, India**	Aug 2020

Coursework - Data Mining, Machine Learning, Deep Learning, LLM & Generative AI, Cloud Computing
Relevant Certifications - Architecting with Kubernetes, Google Cloud Infrastructure — View all certifications
Publications - Vehicle Trajectory Prediction using GANs, Autonomous Defense Device with GSM and NLU

TECHNICAL SKILLS

Programming	Python, Java, R Programming
Databases	SQL, NoSQL - MongoDB, Cassandra
Cloud Platforms	Google Cloud (GCP), Microsoft Azure, Docker
ML/AI	Deep learning, Natural Language Processing, Large Language Models, Time series forecasting, Natural Language Generation, Recommendation system, Image + Video processing, Predictive/statistical analysis
Tools	Git, Hadoop, Spark, Tableau, Looker, Airflow, Vertex AI
Packages	TensorFlow, LangChain, Hugging Face, Spacy, Sci-kit learn, Keras, OpenCV, PySpark

PROFESSIONAL EXPERIENCE

WebMD – NY, USA (Remote) Jun 2023 – Aug 2023
Software Development Intern

- Engineered **machine learning pipelines** for recommendation systems using **BigQuery,** Python on **Vertex AI,** and **Airflow**, to accurately target potential customers. Classification and regression were used.
- Finetuned **PaLM 2 (LLM)** on Google Cloud Platform Vertex AI to create an information extraction solution for pharmaceutical data in **BigQuery,** resulting in an extraction accuracy of 95%. Exploratory analysis on **Tableau** and **SQL.**
- Conducted **clustering** and **time series forecasting** using **Python** and **AutoML** for customer segmentation for a propensity model to identify customer behavior trend analysis, increasing market penetration by 15%.

Merkle – Bangalore, India Sep 2020 – Jun 2022
Data Scientist

- Used ARIMA **time series forecasting** with SQL, and **PySpark** for an **anomaly detection** feature in an in-house tool.
- Developed ETL data pipeline with **GCP Airflow** and Python for over 2 million records across 60 tables from BigQuery (SQL), and Adobe Analytics to ensure data quality.
- Built an automated Tableau dashboard for product performance from transaction data on Google BigQuery using **Natural Language Generation** and **NLP Python and R** scripts.
- Automated a Python script for **data analysis**, and prioritization in data products, for the development of a data dictionary for the data governance team, enabling data alignment activities.
- Owned the **User Acceptance Testing (UAT)** for features and bug resolutions of the data product. Strategized, and executed meticulous testing protocols, ensuring seamless integration.

Merkle – Bangalore, India Jan 2020 – Mar 2020
Software Development Intern

- Built a logo tracking application POC with ResNet & YOLOv3 for **video processing** with 89% accuracy, using **Keras** and **Streamlit.**
- Conducted fake review detection for client's brand analytics using **Selenium** web scraping to gather customer reviews and utilized **Natural Language Processing (NLP),** and neural networks (CNNs) with **PySpark** for large-scale classification in **Python.**
- Received training in R, SQL, GCP, problem analysis, and PySpark, developing skills in data analysis.

PROJECTS

Multi-Lingual Image Description Assistant - *Hugging Face, LLM, Image processing, NLP* - **[GitHub]** Spring 2023

- The application takes an image input to produce its description in audio output in the selected language. Employed Keras and Hugging Face's image-to-text model, LangChain (ChatGPT API) for language translation, and Hugging Face's text-to-speech model to create a multilingual audio description from user-uploaded images.

Deep Learning-based Inference for Cyberbullying - *Transformers, NLP, Neural networks* - **[Ongoing]** Fall 2023

- Utilized BERT, GloVe, fastText, and deep learning models (LSTM, RNN, CNN) with TensorFlow for proactive cyberbullying detection. Named Entity Recognition (NER) and Topic Modelling for context mining with Flask for real-time blocking, with Python. Achieved 92.7% accuracy, prototyping on Streamlit. Distributed pipeline using Spark

Website Ranking with Keyword Extraction & PageRank - *Python, MongoDB, NLP* - **[GitHub]** Spring 2023

- Implemented query optimization techniques on MongoDB for a news article ranking system, leveraging PageRank and TF-IDF keyword extraction algorithms in Python

Finance Analytics with Llama2 - *Large Language Model (LLM), Python* Fall 2023

- An automated financial analysis tool utilizing Python and Llama 2 to classify bank statement transactions into predefined categories and provide Tableau data visualizations, to save time and achieve efficiency in managing finances

PUBLICATIONS/ HACKATHONS

OTTO Multi-Objective Recommendation System - *Python, AWS* - **[Kaggle Competition]** Spring 2023

- Built an e-commerce (OTTO dataset) recommendation engine using Amazon SageMaker and Kinesis streaming, handling data preparation, model training and deployment, and real-time inference. Designed and implemented secure and scalable data pipelines with Kinesis Firehose, and AWS S3.

Vehicle Trajectory Prediction using GANs - *Image/Video Processing, Deep learning* - **[GitHub] [Article]** Fall 2019

- Devised a collision avoidance model for vehicles by path inference using video processing with TensorFlow YOLOv3 for object detection and tracking, and Generative Adversarial Networks to predict vehicle trajectories. Achieved 88-93% accuracy.

Autonomous Defense Device with GSM and NLU - *IoT, Machine Learning* - **[Article]** Spring 2018

- Top 10 at Kludge Intercollege Hackathon in 2018 and presented at the National RISE Conference.
- Designed the safety device using IoT technology, including features such as GSM, DialogFlow and cloud-based SOS call/SMS using machine learning for user analysis, and GPS location tracking.

LEADERSHIP AND VOLUNTEERING

- Student organizer - IEEE ICACCI, Genesis DevCon
- Head of Design - TEDxPESU (2018, 2019), IEEE GirlsGeekHack and inGenius
- Website Optimization Student Assistant - Department of Computer Science