Chinmayi Hegde

San Jose, CA 95126 | +1 669-649-3847 | chinmayilokeshwar.hegde@sjsu.edu | 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

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, Javascript, HTML, CSS
Databases	SQL, BigQuery, NoSQL - MongoDB, Cassandra, Neo4j
Cloud Technologies	Google Cloud (GCP), Microsoft Azure, Docker
Tools	Git, Flask, Hadoop, Spark, Tableau, Looker, Airflow, Vertex AI,
Packages	TensorFlow, LangChain, Hugging Face, Spacy, Sci-kit learn, Keras, OpenCV, PySpark
ML/AI	Deep learning, Natural Language Processing, LLM finetuning, Time series forecasting, Recommendation system, Image + Video processing

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.**
- 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
- Orchestrated the **migration** of critical ETL scripts from Analytics Workbench to the **Google Cloud Platform (GCP)** for scalability
- Pipelined **Natural Language Generation** and **NLP Python** scripts for automated KPI reports in a **Tableau** dashboard
- **Automated** a Python script for data analysis, and prioritization in data products, for the development of a data dictionary
- Developed **ETL data pipeline** with GCP **Airflow** and **Python** for 2 million records across 60 tables from BigQuery (SQL), and Adobe Analytics to ensure data quality

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),** neural networks (CNNs) with **PySpark** for large-scale classification in **Python**
- Received training in R, SQL, GCP + Kubernetes problem analysis, and PySpark, developing skills in data analysis.

PROJECTS AND PUBLICATIONS

Multi-Lingual Image Description Assistant | *Hugging Face, LLM, Image processing, NLP* GitHub

- The application takes an image input to produce its description in audio output in the selected language
- Employed 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, Flask* GitHub

- Utilized BERT, GloVe, and fastText and deep learning models (LSTM, RNN, CNN) for proactive cyberbullying detection. Named Entity Recognition (NER) and Topic Modelling (LDA) 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, Network Analysis* GitHub

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

Vehicle Trajectory Prediction using GANs | *Image/Video Processing, Deep learning, TensorFlow* GitHub

- Devised a collision avoidance model 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.

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