Venkata Satya Sri Ram Giduthuri

(680) 216-3364 | vgiduthu@syr.edu | github.com/GvsSriRam/ | linkedin.com/in/venkata-satya-sri-ram-giduthuri/

EDUCATION

May, 2025	**Syracuse University**	**Master of Science in Computer Science**	**CGPA: 4.0**
June, 2020	**Indian Institute of Technology Kharagpur**	**Bachelors & Master of Technology (Honors)**	**CGPA: 8.0 / 10**

SKILLS

Languages	Python · SQL · C++ · HTML · CSS · JavaScript · Java · R
Tool & Frameworks	TensorFlow · Keras · Sklearn · Numpy · Pandas · Flask · Django · Spring Boot · ReactJs · NodeJS · Git · Docker
Cloud Technologies	Google Cloud Platform, AWS, Azure, Databricks
Data Technologies	Kafka · Dataflow · DBT · FEAST · REDIS · SQL Server · PostgreSQL · MongoDb
Miscellaneous	Linux · Object Oriented Programming · Design Patterns · Web Scrapping · Prompt Engineering · Kubernetes

CERTIFICATIONS & COURSEWORK

Machine Learning	Data Analytics	Natural Language Processing
Design & Analysis of Algorithms	GCP – Professional Data Engineer Training	DBMS
Object Oriented Design	Social Media & Data Mining	Probability & Statistics

EXPERIENCE

HCL Software, India   **Senior Software Engineer**   **Sep, 2020 – July, 2023**

- Worked as a Machine Learning and Data Engineer in HCL Commerce - Data & AI - R&D division

PRODUCT RECOMMENDATION SYSTEMS

- Created a *Collaborative Filtering model*, to predict user preferences with 95% accuracy based on implicit feedback in a large product catalogue
- Built a *Content-Based Recommendation model* using NLP techniques (TF-IDF, word embeddings) increasing customer interaction by 10%
- Enhanced *Frequently Bought Together Model* recommendations based on Association Rule Mining algorithms resulting in 15% more cart size
- Devised training, fine-tuning, prediction pipelines and utilised Redis store to cache and retrieve recommendations within 10 ms

DYNAMIC SEGMENTATION MODEL

- Designed a multi-label classification algorithm to classify online shoppers into customer segments for precision marketing with 95% accuracy
- Applied feature selection methods to improve interpretability and achieve an explainable AI, with 18% increase in classification accuracy
- Employed Deep Auto-Encoders for dimensionality reduction, enhancing computational efficiency by 85% without a significant loss of data

DYNAMIC PROMOTION OPTIMIZATION

- Developed a Multi-Armed Bandit architecture with Epsilon Greedy, UCB, and Thompson Sampling algorithms for dynamic promotions
- Setup real-time ETL pipelines on Trino with DBT to monitor promotions performance and devised custom metric to maximize conversions

DATA ENGINEERING

- Utilized DBT TRINO views to optimized feature engineering for ML models and reduced prediction latency from 1.6 sec to 600 ms
- Built macros in DBT to streamline complex SQL operations - one-hot encoding, incremental materialization and improving code reusability
- Established online & offline feature stores by leveraging FEAST, DBT & TRINO to minimize real-time data retrieval latency to 100 ms

BACKEND APPLICATION DEVELOPMENT

- Engineered generic and configurable Python Fast-API based APIs for ML model training and Node JS APIs to load and serve predictions
- Setup continuous deployment on Cloud Run, ensuring seamless upgrades from git, horizontal scalability and response times of 20-30 ms

Skill-set: Deep Learning, Recommendation Systems, CI/CD, Jira, Agile, Best SDLC practices, Testing, OOP, SQL

PROJECTS

Balance AI: AI for Business Analytics   **Feb, 2024 – Present**

- Engineered an LLM chatbot driving actionable insights for business optimization, marketing strategy definition, and dynamic graph plotting
- Innovated a robust bi-directional LSTM and HMM-Viterbi model for tailored query responses, enhancing user engagement
- Automated monthly report generation for clear financial insights through meticulous time series analysis with an accuracy over 90%

Sign Language Recognition   **Summer, 2024**

- Built a 97% accurate ASL recognition web app (trained on 223,075 images) using MediaPipe, OpenCV, and TensorFlow Decision Forests
- Integrated context-aware autocorrect, translation, text-to-speech, Google search integration and virtual assistant messaging
- Optimized on-device hand landmark detection to achieve latency below 200 ms, resulting in 35% faster response compared to similar solutions

Visual Similarity Recommendations Model   **June, 2021 – July, 2021**

- Inspired by Visual Similarity Recommendation feature of Adobe Sensei for Adobe Commerce, 2021 release
- Customized pre-trained keras VGG-16 model's input and output layers while keeping weights undisturbed, resulting in accurate predictions
- Pre-processed and normalized a dataset of 3000 product images to extract visual features and computed similarity to offer recommendations

Whisper Wallet: An AI Enhanced Expenses Recognition Model   **Dec, 2023 – Present**

- Developed and tested an Open CV-based expense detection model for generic bills, handling both scanned and normal images
- Integrated the Gemini Pro Vision LLM model into the API to accurately parse expenses, quantities, and amounts in generic bills
- Created a user-friendly web application interface for uploading bills, selecting individuals to split expenses, and displaying the expense share
- Reduced latency by 25% during bill uploads and parsing in the web app, improving overall user experience for large-scale uploads