NIKITA HEMANT KODKANY

Los Angeles | kodkany@usc.edu | linkedin.com/in/nikitakodkany/ | nikitakodkany.github.io/ | 213-691-9301

EDUCATION

University of Southern California **JAN 2023 – DEC 2024**
Master of Science, Computer Science Los Angeles, CA
***[Coursework**: Machine Learning, Natural Language Processing, Deep Learning, Algorithms]*

KLS Gogte Institute of Technology **AUG 2018 – JUL 2022**
Bachelor of Engineer, Computer Science India

TECHNICALS SKILLS

Programming: Python (Flask, Django), C/C++, Java, SQL, R
Frameworks: scikit-learn, TensorFlow, PyTorch, NumPy, Pandas, Seaborn, Matplotlib, PySpark, Streamlit, nltk
Tools & Tech: Apache (Hadoop, Spark, HIVE, Airflow), Snowflake, AWS (Redshift, Sagemaker, S3), Microsoft (Azure, Power BI), QlikSense, Tableau, Hue, Docker, Terraform, Git, UNIX, LINUX

EXPERIENCE

Keck Medicine of USC **APR 2023 – Present**
Data Science & Analytics Intern | *QlikSense, Tableau, PowerBi, SQL, Hue, AWS, Python* Los Angeles
- Converted 3 dashboards (QlikSense to Tableau), achieving 20% reduction in processing time across all visualizations
- Engineered health indicators (including SEDoH) from 40M+ EHR data; developed ML models for predicting length of stay and mortality using AWS Redshift
- Crafted SQL queries for validation across databases (QlikSense, Hue) and dashboards (Snowflake, PowerBI)
- Built ML models to forecast business financial trends using demographic and procedure code data from EHR system

Mercedes Benz Research And Development India **JUL 2022 – DEC 2022**
Data Engineer | *HIVE, PySpark, Databricks, PyTorch, HuggingFace, Azure, Python* Bangalore, India
- Resulted in 5% improvement in efficiency for car engine data pre-existing ETL pipeline using PySpark on HIVE tables
- Automated manual effort by engineering Machine Learning pipeline employing LLMs for patent classification; integrating CI/CD pipeline for deployment and monitoring
- Presented insights on Signal Processing in Machine Learning for the Automobile domain to over 50 attendees

Eyesec Cyber Security Solutions **DEC 2021 – JUL 2022**
Machine Learning Engineer| *nltk, Spacy, PyTorch, TensorFlow, NumPy* Belagavi, India
- Spearheaded Machine Learning pipeline for industry-aligned university syllabus generation exploiting 50,000+ web-scraped job descriptions, analyzing 100+ universities, improved job relevancy by 30%
- Authored Machine Learning course, featuring 7 projects and 5 internships, guiding 80% participants engagement

Cognius.ai **SEP 2021 – MAY 2022**
Machine Learning Engineer| *Python, PyTorch, NLG, NLU, Optuna, GCP AI Platform* Remote
- Led research for Natural Language Generation pipeline using Google's T5, implementing custom named entity (NER); conducted A/B testing on model configurations, resulting in a 14% improvement in pipeline accuracy
- Enhanced pipeline accuracy by 14% through fine-tuning IBM Watson's API with hyperparameter optimization for Natural Language Understanding tasks

PROJECTS

Prediction of Individuals' Influence on Community Dynamics | *Python, nltk, PyTorch, NLP*
- Developed large language model to predict social media engagement metrics injecting tweets and Reddit discussions, outperforming MLP and LSTM models by 30% to predict influence of individual on large audience
- Fine-tuned BERT for sentiment analysis and embeddings; obtained 20% reduction in prediction error

Sunglass Hub Dataflow |*Azure, Terraform, Airflow, Docker, Python, Insomnia, PostgreSQL*
- Implemented end to end ETL pipeline to scrape, process, and store data from Sunglasseshut API
- Delivered validated data to Azure Synapse using Terraform, Airflow, Docker, Insomnia, PostgreSQL; enabled data visualization, providing insights into sales trends, customer preferences, and inventory management

Urban Area Reconstruction using GAN |*Python, TensorFlow, PyTorch, GAN, CycleGAN, OpenStreetMap*
- Architected CycleGAN model to generate high-resolution urban reconstructions from satellite imagery and OpenStreetMap data; enabled visualization of city growth and infrastructure expansion
- Augmented reconstruction accuracy by integrating building density, roads, and green spaces; used transfer learning and data augmentation