VATSAL KAMLESH KHANDOR

San Jose, CA (Open to Relocation) | khandorvatsal203@gmail.com | +1 (213) 245-5750
linkedin.com/in/khandorvatsal | github.com/khandorvatsal | khandorvatsal.github.io

EDUCATION

University of Southern California — **Aug 2022 – May 2024**
Master of Science in Computer Science; GPA: 3.85/4.0 — Los Angeles, California

Dwarkadas J. Sanghvi College of Engineering — **Aug 2018 – May 2022**
Bachelor of Engineering in Computer Engineering; GPA: 9.75/10.0 — Mumbai, India

SKILLS

Languages and Databases: Python, R, Java, JavaScript, C, C++, SQLite, MySQL, PostgreSQL, MongoDB
Frameworks/Libraries: Flask, Selenium, Numpy, Pandas, Scikit-Learn, TensorFlow, PyTorch, PySpark, Keras, Hadoop
Other tools and services: Git, Github, Bitbucket, Postman, Docker, AWS, Google Cloud Platform, PowerBI, Tableau

EXPERIENCE

Software Engineer | University of Southern California | Los Angeles, California — **Jan 2024 – May 2024**

- Developed full-stack web app leveraging frontend and backend technologies for Dr. Moats at Children's Hospital LA.
- Implemented Role-Based Access Control (RBAC) containing data repository, ensuring secure access for diverse roles (Doctor, Analyst, Administrator) and compliance with healthcare privacy regulations by excluding all PHI.

Machine Learning Engineer Intern | Bombay Softwares | Ahmedabad, India — **Sept 2021 – Dec 2021**

- Implemented transient pose conditions using **33** skeleton keypoints and co-ordinates from BlazePose and re-framed other posture conditions. Detection of incomplete/inaccurate reps increased by approximately **25%**.
- Designed metrics for A/B testing. Improvised existing metadata by carrying out **23+** exercises, including redesigning global API conventions and code refactoring. Eliminated **57%** redundant code per exercise, reducing technical debt.

Machine Learning Engineer Intern | Code13 Edutech Pvt. Ltd. | Nagpur, India — **May 2021 – Aug 2021**

- Engineered chatbot with "human handoff" addressing general queries and course recommendations. Leveraged NLP techniques across client/server side with Botpress, enhancing query resolution and engagement by **25%**.

PROJECTS

Automated Calculation of Underwriting for Multifamily Property Investment *(Hackathon Winner – 1st place)*

- Crafted Flask app, scraping **5000+** data points with Selenium, testing APIs with Postman, and applied RandomForest Regression for assisting buyers in determining fair value of listings. Containerized with Docker for deployment.
- Built RAG system using Gemini-1.5-Pro LLM and Streamlit for real-time property insights (NOI, DSCR, COC, CAP Rate). Attained **65%** faster query responses and utilized Plotly for dynamic visualizations for historical county data.

Efficient API Usage for Document Querying (LLM's) *(Langchain, GPT)*

- Applied data science techniques to optimize OpenAI's API calls, achieving **92%** reduction in token utilization through innovative PDF chunking method involving document similarity. Reduced response time of user query by **18%**.
- Conducted fine-tuning of Davinci and Curie models for custom knowledge base while also exploring open-source Large Language Models (LLMs) like Llama2 for development of personalized onboarding manager chatbot.

Automated Extraction of Reproduction Steps from Android Bug Reports *(Deep Learning, Google Colab)*

- Designed crawler that retrieved **70,000+** GitHub Issue Reports in JSON for ETL based on heuristics using markdown parser and NLTK. Utilized BIO labeling technique for named entity recognition to form dependent variable.
- Achieved F1 score of **97.65%** deploying a combination of BERT and CRF for predicting sequences in bug reports.

Algorithmic Trading using Reinforcement Learning & Sentimental Analysis *(Feature Engineering)*

- Devised end-to-end pipeline with asynchronous HTML session to extract **100** recent relevant headlines for **top 30** stocks from Yahoo Finance, performing sentiment analysis using FinBERT based on pre-trained financial corpus.
- Evaluated performance of 5 RL algorithms for training agent in developing profit-maximizing strategies, resulting in a **119% increase** in profitability by incorporating sentiment analysis compared to standard index returns.

Analysis of Classifiers for Fake News Detection *(NLP, Machine Learning, Jupyter) | doi: 10.1201/9781003283249*

- Conducted research utilizing LSTM, CatBoost and XGBoost ensemble classifiers, reaching a **91.4%** accuracy.
- Methods incorporated five-fold cross-validation, hyperparameter tuning to balance precision and recall.

Cognitive Behavioral Therapy based Patient Analysis website *(HTML, CSS, Django, TensorFlow)*

- Analyzed patients through Multiple Choice Question where sentiments are calculated using Neural Networks.
- Built website for patients suffering with depression and allots Psychiatrists to patients based on health condition. Secured a position at final round of hackathon of **top 14** teams from 300+ teams participating.

LEADERSHIP & INVOLVEMENT

- Ex-President leading GRIDS USC (Graduates Rising in Data Science) with **16** E-Board Members and **500+** active members. Ex-Lead Teaching Staff Member for CSCI 585, DSCI 531, DSCI 250, CSCI 455.
- Led a group of 6 students for presenting a working prototype of a "Smart Mirror" for state level "Strike" competition developed over a period of 1 year after getting shortlisted from 59 teams (ISBN: 978-93-5437-776-1).