Vatsal Kamlesh Khandor

(213) 245-5750 | khandor@usc.edu | linkedin.com/in/khandorvatsal/ | github.com/khandorvatsal | khandorvatsal.github.io

Education

University of Southern California **August 2022-May 2024**
Master of Science in Computer Science | ***CGPA: 3.83/4*** *Los Angeles, California*

Dwarkadas J. Sanghvi College of Engineering **August 2018-May 2022**
Bachelor's of Engineering in Computer Engineering | ***CGPA: 9.75/10*** *Mumbai, India*

Professional Experience

RadicalX | AI Engineer Intern | New York, New York **July 2023-August 2023**

- Applied data science techniques to optimize OpenAI's API calls, achieving remarkable 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.

Bombay Softwares | Machine Learning Engineer Intern | Ahmedabad, Gujarat **September 2021-December 2021**

- Implemented transient pose conditions using 33 skeleton keypoints and co-ordinates from BlazePose and re-framed other posture conditions. Detection of incomplete as well as 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.

Code13 Edutech Pvt. Ltd. | Machine Learning Engineer Intern | Nagpur, Maharashtra **May 2021-August 2021**

- Accomplished development of a chatbot with feature of "human handoff". Apart from handling general queries, bot was trained to assign issue for human interaction using NLP techniques on client-side and server-side with Botpress.
- Executed integration of chatbot suggesting courses based on user qualifications and preferences for company's website. Statistics indicate a 25% increase in resolved queries and engagement with chatbot on website.

Academic Projects

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

- Designed a crawler that retrieved 70,000+ GitHub Issue Reports in JSON format 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 | *BeautifulSoup, Feature Engineering*

- Developed a pipeline for asynchronous HTML sessions, automating data acquisition to extract 100 recent headlines for top 30 stocks from Yahoo Finance, and performed sentiment analysis using FinBERT model 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.

Eclectic Analysis of Classifiers for Fake News Detection | *NLP, Jupyter* | doi: 10.1201/9781003283249

- Performed research suggesting Long Short-Term Memory (LSTM) and Boosting Ensemble models like CatBoost and XGBoost works best for detecting and solving problem of Fake News. Leveraged five-fold cross-validation to ensure model stability and demonstrated classification report with an average accuracy of 91.4% for best working classifiers.

Credit Risk Assessment for Banking Systems | *Git, Github, Matplotlib, Ensemble Modeling, Machine Learning*

- Predicted loan defaults and probability on a dataset containing over 30000 rows with 11 variables. Trained, validated and tested XGBoost, k-nearest neighbors (KNN), and Logistic Regression models. RandomizedSearchCV and GridSearchCV is employed for hyperparameter optimization for all 3 models. Reported a recall of 75% while achieving 96% precision.

Cognitive Behavioral Therapy based Patient Analysis website | *HTML, CSS, Django, Tensorflow*

- Analyzed patients through Multiple Choice Question where sentiments are calculated using Recurrent 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.

Prediction of IPL matches using Machine Learning while tackling ambiguity in results | *Data Visualization, Seaborn*

- Rectified a key issue on data symmetry and inability of models to handle it explaining its extension to all types of classification models for comparing two or more classes using similar features for both classes (doi: 10.17485/IJST/v13i38.1649)
- Investigated Analytic Hierarchy Process (AHP) and Dream11 formula for score computation of each player in every team.

Technical Skills

Programming Languages: Python, SQL, Java, Javascript, C, C++ | **Databases**: MySQL, MongoDB, Postgres
Frameworks/Libraries: Flask, NLTK, OpenCV, Numpy, Pandas, scikit-learn, Keras, PyTorch, Spark, Hadoop, MapReduce
Other tools and services: BitBucket, AWS, Google Cloud Platform, MS Excel, Google Ads, PowerBI, Tableau

Leadership & Involvement

- President at GRIDS USC with 500+ active members. Teaching Staff Member for CSCI 585, DSCI 531, DSCI 250 and 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).