KARTIK VASUDEV SHENOY

(323) 986-9136 | kartikshenoy.com | kshenoy@usc.edu | Linkedin: kartik2112 | Github: kartik2112 | Los Angeles, CA

EDUCATION

University of Southern California
Master of Science in Computer Science (Artificial Intelligence) with Honors
University of Mumbai, India
Bachelor of Technology in Computer Engineering; **Department Rank: 1**

January 2021 – Exp. Dec 2022
GPA: **4 / 4**
August 2014 – June 2018
GPA: **9.56 / 10**

EXPERIENCE

Bill.com
Machine Learning Engineer Intern

May 2022 – August 2022
San Jose, CA (Remote)

- Generated a dataset of **1 million** invoices and statements from user data after sanity checks for improving the amount OCR model.
- Leveraged an ensemble of document classification models to fix the imbalanced data distribution of statements in training dataset.
- Improved the amount prediction test accuracy from 75.5% to **77.67%** with 50% confidence threshold.

Information Sciences Institute, USC
Research Assistant, May Team

January 2022 – May 2022
Los Angeles, CA

- Worked on designing bots capable of understanding communication and strategizing for the board game – diplomacy.
- **Increased the win rates by an average of 5%** for the powers in the game by generating and understanding DAIDE messages and using rules on top of the pre-trained reinforcement learning based Dipnet bot.

Information Sciences Institute, USC
Research Assistant, Centre of Knowledge Graphs

February 2021 – Present
Los Angeles, CA

- Constructed a framework for identifying low quality statements in Wikidata knowledge graph amongst **1.1 billion** statements on the basis of deleted, deprecated statements and constraint violations.
- Enhanced the **graph embeddings** of nodes using retrofitting based on **BERT** embeddings and structural, textual properties extracted from **Wikidata, Probase and DBPedia** datasets increasing **Spearman correlation from 0.66 to 0.73** on WordSim353 benchmark.

Barclays Global Service Centre
Software Developer, Barclaycard UK

July 2018 – December 2020
Pune, India

- Devised a prototype fraud detection system to identify mule accounts by building a pipeline of **Kafka** queues, **Cassandra DB** and **PySpark** servers having an ensemble of ML models.
- Designed a tweets **sentiment analysis** engine which fetches tweets where Barclays accounts were tagged in real-time to enable quick customer service response achieving an accuracy of around **90 %** in pilot runs.
- Created a classifier application utilizing **ML algorithm Latent Dirichlet Allocation** to extract insights by generating ontologies from iOS and Android application reviews and customer complaints.
- Spearheaded the design of a system that helps in connecting the colleagues with available bandwidth and skillsets with the colleagues needing assistance in their work, using **AngularJS, Java, MySQL**, saving more than **900 man-hours** annually.
- Translated VISA, Mastercard scheme mandates into system code which processes millions of transactions every day.

Virtual Labs, Indian Institute of Technology, Bombay
Web Development Intern, Team Leader

March 2017 – August 2017
Mumbai, India

- Led a team of three to develop a Virtual Lab for the online demonstration of machine learning concepts such as neural networks, learning rules and optical character recognition. This lab has won the Global Online Laboratory Consortium International Lab Award.

PROJECTS

Alexa Social Bot
- Contributed to the development of Alexa social bot including conversational elements such as **Sentence BERT**, offensive speech classification, **SNIPS NLU** intent classifier, **FSMs**, **DialoGPT** and neural text generation modules.

Generation Augmented Retrieval for Question Answering
- Addressed the Question Answering task by fine-tuning BART-large model and using OTS Dense Passage Retriever models achieving **92.6% top-1000 retrieval accuracy** and **43.9% EM accuracy** on Natural Questions dataset.

VeriSign [Demo Link] [Github Link]
- Deployed an application to verify signatures by comparing with the original one and detect forgery.
- Achieved an accuracy of 96 % by training a Convolutional Siamese Network using the concept of One-Shot Learning.

Pneumonia Detection from Chest X-Ray Scans [Github Link]

- Trained a CNN on Chest X-Ray Scans with histogram equalization achieving 94.56 % accuracy and a recall score of 0.97.

Sign Language Translator [Github link]

- Managed a team of four to develop an Android application for recognizing Indian Sign Language used by the hearing- and speech-impaired. This is connected to a remote Python server.
- Extracted centroid contour distance curve of hand poses followed by normalization with FFT for feature-set generation.
- Trained **HMMs** for Gesture Recognition from gesture videos with 97.23 % accuracy and **k-NN** for Hand Pose Recognition with 99.7 % accuracy. Results published in a paper indexed by IEEE Xplore.

Feed-Forward Neural Network Implementation in NumPy [Github link]

- Implemented Neural Network Training using Back Propagation with Mini-Batch Gradient Descent, regularisation, variable momentum and learning rate annealing. Used this flexible implementation to train MNIST dataset achieving 96.81 % accuracy.

PUBLICATIONS & ARTICLES

- "Viola: A Topic Agnostic Generate-and-Rank Dialogue System", 4th Proceedings of Alexa Prize (Alexa Prize 2020). [paper link]
- "A Study of the Quality of Wikidata", Preprint submitted to Elsevier, July 2021. [paper link]
- "Keystroke Dynamics Analysis and Prediction", Towards Data Science, January 2021. [Part 1 (EDA), Part 2 (Model Training)]
- "LSTM Back-Propagation - the Math Behind the Scenes", Medium, July 2020. [Article Link]
- "Real-time Indian Sign Language (ISL) Translation", 9th International Conference on Computing, Communication and Networking *Technologies (ICCCNT)*. Published in IEEE Xplore, October 2018. [paper link]
- "An Effective Pixel-Wise Approach for Skin Colour Segmentation - Using Pixel Neighbourhood Technique", International Journal *on Recent and Innovation Trends in Computing and Communication (IJRITCC)*, 2018. [paper link]

SKILLS

Programming Languages: Python, C++, Java, JS

Machine Learning: TensorFlow, Pandas, Keras, Scikit-learn, NLTK, Spacy

Frameworks: OpenCV, Numpy, Matplotlib, Kafka

Web: LAMP stack, Angular4, NodeJS, Flask

Database: MySQL, MongoDB, CassandraDB

Tools: AWS (EC2, RDS, S3, EKS, Comprehend), Git, Firebase, Matlab, Google Colab, Heroku

CERTIFICATIONS

- Natural Language Processing with Classification and Vector Spaces
- Natural Language Processing with Probabilistic Models
- Natural Language Processing with Sequence Models
- Deep Learning Specialization (Coursera)
- Machine Learning (Stanford Online)
- Angular 2 (Udemy)

ACHIEVEMENTS & INVOLVEMENTS

- **Semifinalist** of Alexa Socialbot Grand Challenge 4 (2020 – 2021) as part of Team Viola.
- **Guided 900+ students** as a **Teaching Assistant** for the courses Artificial Intelligence and Algorithms (Jan 2022 – Present) at USC
- Bagged the **Barclays Award of Stewardship** for automating the generation of real-time delivery metrics of more than 30 teams from Agile Central and Jira data sources. These dashboards have been saving around 150 man-hours annually.
- **Sir Dorabji Tata Trust scholarship** for securing 1st Rank in a class of 120 students during undergraduate studies.
- '**Vice-President - Membership**' of Barclays Pune 2 Toastmasters Club (January 2019 – June 2019).
- Active member of Junior Chambers International Club – contributing to **social service** activities such as creating seed balls for afforestation in areas devastated by floods and blood donation.
- **Taught basic computer skills** of Microsoft Office to underprivileged students over a span of 10 days.
- **Placement Coordinator** during under-graduation, managing logistics for around 40 companies and 800 students.
- Indian Classical Music Level 3 certified singer and Harmonium player and a beginner guitar player.