KARTIK VASUDEV SHENOY

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

EDUCATION

University of Southern California — **Jan 2021 – Exp. Dec 2022**
Master of Science in Computer Science with Honors — GPA: **4 / 4**
Courses: Advanced Natural Language Processing, Advanced Computer Vision, Algorithms

University of Mumbai, India — **Aug 2014 – Jun 2018**
Bachelor of Technology in Computer Engineering; **Department Rank: 1** — GPA: **9.56 / 10**

EXPERIENCE

Machine Learning Engineer Intern – Bill.com | San Jose, CA (Remote) — **May 2022 – Aug 2022**

- 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.

Research Assistant – Information Sciences Institute, USC | Dr. Jon May | Los Angeles, CA — **Jan 2022 – May 2022**

- 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.

Research Assistant – Centre of Knowledge Graphs, USC | Dr. Filip Ilievski | Los Angeles, CA — **Feb 2021 – Jan 2022**

- Constructed a repository of scripts for identifying low quality statements in Wikidata knowledge graph amongst **1.1 billion** statements on the basis of deleted, deprecated statements and constraint violations using kgtk library.
- 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.

Software Developer – Barclays Global Service Centre | Pune, India — **Jul 2018 – Dec 2020**

- 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 Machine Learning models achieving a processing speed of **20ms per transaction**.
- Designed a real-time tweets **sentiment analysis** and classification engine for Barclays accounts to enable quick customer service response achieving an accuracy of around **90%** in pilot runs.
- Created a classifier application utilizing topics identified with the help of unsupervised algorithm **Latent Dirichlet Allocation** to extract insights by generating ontologies from iOS and Android application reviews and customer complaints.
- Implemented **dashboards** for automated generation of real-time delivery metrics of 30+ teams from Agile Central and Jira data sources which have been saving around **150 man-hours** annually. Bagged the **Barclays Award of Stewardship** for this initiative**.**

PROJECTS

- **Alexa Social Bot** - Contributed to the development of Alexa social bot by including conversational elements such as **Sentence BERT**, offensive speech classification, **SNIPS NLU** intent classifier, **FSMs**, **DialoGPT** and other neural text generation modules.
- **Generation Augmented Retrieval for Question Answering** achieving **92.6% top-1000 retrieval accuracy** and **43.9% EM accuracy** on Natural Questions dataset by fine-tuning BART-large model and using OTS Dense Passage Retriever models.
- **VeriSign** [Demo Link] [Github Link] - Deployed a web application to verify signatures and to detect forgery which 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.

PUBLICATIONS

- "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", Journal of Web Semantics, Elsevier, July 2021. [Paper Link] [Github 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] [Github Link]

SKILLS

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

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

Frameworks: OpenCV, Numpy, PySpark, Kafka

Web: LAMP stack, Angular JS, NodeJS, Flask

Database: MySQL, MongoDB, CassandraDB

Tools: AWS (EC2, RDS, S3, EKS, Comprehend, Athena, SageMaker), Github, Firebase, Matlab, Google Colab, Docker

ACHIEVEMENTS & INVOLVEMENTS

- **Semifinalist** of **Alexa Socialbot Grand Challenge 4** (2020-2021) as part of Team Viola
- '**Vice-President - Membership**' of Barclays Pune 2 Toastmasters Club (January 2019 - June 2019).
- Member of Junior Chambers International Club (2019) – contributing to **social service** causes such as afforestation, blood donation.