**Education**

**Masters of Computer Science Sept 2022 – June 2024 (Expected)**

University of California, San Diego (UCSD) CGPA: 3.97/ 4

*Relevant Courses*: Deep Learning, Scalable Data/ML Systems, Recommender Systems, Computer Vision, Advance NLP – LLMS

**Bachelor Of Engineering (Computer Engineering) August 2018 – July 2022**

Vivekanand Education Society’s Institute of Technology (VESIT) CGPA: 9.013/ 10

*Relevant Courses*: Machine Learning, Software Engineering, Database Management, Big Data Analytics, Advance Algorithms

**Internship Experience**

*Machine Learning CV, NLP Researcher,* **Ujima S&P Lab**  **March 2023 – Present**

* Building a “Smart Mirror” on Raspberry Pi using VGG-Face TF-Lite Model, trained on 100,000 images, to detect ethnicity, address the existing biases in CV datasets, and contribute to the public domain via user feedback.
* Developed a rule-based security chat assistant using Rasa in an unstructured, fast-paced startup-like environment.
* Led 4 undergraduates through the Early Research Student Program in analyzing privacy data for MMO AR/VR games using TF-IDF and K-means Clustering.

*Machine Learning & Cyber Security Intern,* **Legendary Entertainment**  **June 2023 – Dec 2023**

* Created a shortcut dashboard that reduced the total time taken by the SOC analyst from 80 minutes to about 4 minutes daily in tracking user-anomaly events by integrating Splunk Dashboard and Azure MSGraphs APIs via Python Scripting.
* Maintained a high level of security while streamlining a user’s authentication process by integrating a FIDO Alliance product into the existing SAML/OIDC SSO workflow.
* Assisted the VP of security in laying the foundation of a next-gen universal anomaly and user-behavior detection platform based on LLMs in collaboration with Sky High Security by building a prototype in Splunk.

*Full Stack Development Intern,* **Makos Infotech (Startup)**  **June 2021 – August 2021**

* Integrated and developed server-side code using JQuery, PHP, and MySQL for an early startup targeting the automation of the On-campus placement process agile methodologies.
* Created and merged relational databases using MySQL workbench and deployed it on AWS RDS to develop a college-student-company social network inspired by Facebook’s friend system.
* Established a mentorship-onboarding program for new undergraduate interns, aligning them with the existing codebase and processes, saving the company at least 1 week of time and effort.

*Data Analyst Intern*, **Leadingindia.ai** **May 2020 – July 2020**

* Collaborated with a team of four to develop a vaccine prediction model for H1N1 and seasonal flu vaccines, accurately predicting public acceptance trends (41%) for the COVID-19 vaccine, securing first place among 85 intercollege peer groups.
* Published a [research paper](https://doi.org/10.1007/978-981-16-0401-0_11) in Springer & authored a [blog](https://medium.com/@jjhaveri1906/pandemics-a-harsh-reality-7c05254e907b) highlighting the correlation between H1N1 and COVID-19 pandemics.

**Projects**

**[Alt Bot for Mastadon: An automatic image alt generation bot](https://docs.google.com/presentation/d/1smZzOd8u-NhgbotJRkn2Eqw5WmXVxA-XSB0afzxaNWE/edit?usp=sharing)** [(Github)](https://github.com/CSE210-Fall23-Team2/AltBot) **Sep 2023 – Dec 2023**

* Developed a REST-API based Chrome extension in JS to help visually impaired people browse decentralized social media feeds by leveraging hugging face image captioning models to generate alternative image descriptions.
* Deployed 3 levels of custom cache system to ensure efficient performance with minimal lag complemented with testing scripts.

[**MedLM: Exploring Language Models for Medical QnA Systems**](https://arxiv.org/abs/2401.11389)[(Github)](https://github.com/JayJhaveri1906/CSE291_MedLM) **March 2023 – Aug 2023**

* Led a team of 4 to fine-tune language models (e.g., Bloom, T5, GPT-2) on the MedQuad dataset, in collaboration with Microsoft researcher Dr. Asma Ben Abacha, comparing performance against larger models (GPT-3.5, GPT-4) through dynamic prompting.
* Implemented ROUGE and BLEU metrics and conducted human surveys for doctors and patients to evaluate the models.

[**GrooveGenie: A copyright-free music generator**](https://github.com/JayJhaveri1906/GrooveGenie-A-copyright-free-music-generator) **March 2023 – June 2023**

* Created an open-source music generation model, utilizing Facebook’s EnCodec Transformer model to compress audio wav files into embeddings for model interpretation.
* Trained a conditioned GAN network to generate music based on user-provided genre inputs embedded using the BERT model, aimed at producing only copyright and royalty-free music, trained on the FMA dataset.

[**Game Genre and Recommendation Classification using Steam Reviews**](https://github.com/JayJhaveri1906/Game-Genre-and-Recommendation-Prediction) **Nov 2022 – Dec 2022**

* Designed data pipelines to preprocess and apply machine learning techniques for classifying game genres, analyzing user sentiment, and curating a personalized game recommendation system using user reviews.
* Achieved 90.53% accuracy with Random Forest, balanced data & TF-IDF, outperforming N-Gram, Multinomial NB, Linear SVC.

**[Divya-Drishti: An Independent Aid for the Visually Impaired](https://github.com/JayJhaveri1906/Divya-Drishti) Aug 2020 – May 2021**

* Achieved a 400% net cost reduction by creating a real-time Voice-activated AI-IoT android application to help Visually Impaired People (VIPs) comparable to state-of-the-art OrCam in detecting currency, objects, and scenes.
* Published a [research paper](https://dx.doi.org/10.2139/ssrn.3867707) highlighting the needs of VIPs, funded by the Mumbai University Minor Research Grant.

**Selected Research Publications**

Inampudi S., **Jhaveri J.** et al., (2021) **Machine Learning Based Prediction of H1N1 and Seasonal Flu Vaccination**. Advanced Computing. IACC 2020. Communications in CIS, vol 1367. Springer, Singapore. [DOI.org Link](https://doi.org/10.1007/978-981-16-0401-0_11)

**Skills:** Python, PyTorch, Tensorflow, OpenCV2, Hugging Face, Pandas, Scikit-Learn, Splunk, Linux, Git, AWS, Azure, GCP, Firebase