SURAJ GATTANI — surajgattani.g@gmail.com • LinkedIn • GitHub • +1 (504) 434-9462

Work Experience

- **Schrodinger** — Los Angeles, CA
  Senior Scientist II, ML — *June 2022 – Present*
  - **Architected novel unified graph-text ML model** achieving state-of-the-art molecular property predictions across **2M+** molecules, matching or exceeding specialized single-property models.
  - **Led the platform development for LiveDesign Learning** architecting scalable **Argo/Kubernetes** ML pipelines generating **$500K+** revenue while enabling automated training and real-time inference for pharma clients and 15 internal therapeutics projects.
  - **Built and deployed LLM-based search tool** achieving **85%+** accuracy for clinical trial alerts and internal document retrieval.
  - **Led Material Science initiatives** developing ML pipelines for formulation prediction and integrating multiple property prediction models into production.
  - **Implemented Graph Neural Networks** yielding **8-10%** r-squared score improvements across molecular property predictions at scale.

  Senior Scientist I, ML
  - **Developed smart ensemble methodology** enhancing model performance by **6%** through qualified data selection.
  - **Enhanced ML dashboard capabilities** with comprehensive backtesting features for MedChemists and Material Scientists.

- **GV20 Therapeutics** — Cambridge, MA
  Machine Learning Engineer — *Oct 2019 – Oct 2020, Oct 2021 – June 2022*
  - **Implemented** deep learning model architecture for cancer type classification **(MCC: 0.97)**, target gene expression **(PCC: 0.67)**, and survival prediction **(Cidx: 0.75)**.
  - **Applied Transfer Learning and Multitask Learning** to improve the expressed target gene correlation by **up to 40%.**
  - **Used** AWS Sagemaker to train and optimize the model parameters with hyperparameter optimization search.
  - **Created** a **Python package** containing various models from recent antibody-related research.

- **Machine Learning and Bioinformatics Laboratory, University of New Orleans** — New Orleans, LA
  Research Engineer — *July 2019 – Oct 2019*
  - **Data preprocessing** to analyze the risks and mishaps and build a forecasting model with **97% accuracy**.
  - **Developed a tagging** based system **(94%)** to forecast the risks according to the location and type.
  - Created a **software plugin** to automate the risk forecasting and model update with new data.

Publications

- StackCBPred: A Stacking based Prediction of Protein-Carbohydrate Binding Sites from Sequence.
- Leveraging High-throughput Molecular Simulations and Machine Learning for Formulation Design. In Review
- IGSF8 is a Natural Killer Cell Checkpoint and Cancer Immunotherapy Target.
- Machine Learning based Prediction of Protein-Peptide Binding Sites from sequences.

Generative AI Projects

- **Multilingual Conversational Agent for Government services** — July 2024 – Oct 2024
  - **Built** end-to-end multilingual conversational system using **large language models** and WhatsApp.
  - **Developed** context-aware knowledge retrieval engine with **95% accuracy**.
  - **Integrated** ASR/TTS capabilities supporting **8 languages**.
- **Agent for E-Commerce Product Recommendation & Search** — March 2024 – Sept 2024
  - **Architectured** hybrid search system improving accuracy by **40%**.
  - **Implemented** LLM reasoning engine reducing irrelevant recommendations by **65%**.
  - **Designed** scalable architecture handling 100k queries at sub-200ms.

Programming Languages & Frameworks:

Python, C, TensorFlow, PyTorch, Cuda, SQL, Linux, Git, AWS, GCS, Docker, Argo, Kubernetes, Knative, Helm, Vector Database, Streamlit

Education

- **Master of Science, Computer Science** Aug 2017 – May 2019
  The University of New Orleans *New Orleans, LA*
  Courses, Skills: Machine learning, deep learning, computer vision, bioinformatics, natural language processing, agile development, SCRUM, algorithms, data analysis.

- **Bachelor of Engineering, Electronics and Telecommunication Engineering** Aug 2013 – May 2017
  Savitribai Phule Pune University *Pune, India*

Research Projects

- **NASA Patent and Software Classifier Tool for text classification:** Applied machine learning to classify the patent and software into 15 categories each. Implemented filtering and vectorization (Word2Vec, GloVe) to create an SVM based classifier. The tool created for the test purpose improved the runtime to 1-2 seconds through serialization.
- **NASA mishap analysis & risk forecasting using NLP:** Implementing deep learning model to analyze the text data and forecast the risk and mishap at NASA Stennis Space Center. Created a software plugin for the test, report, and further investigations.
- **A Stacking based Prediction of Protein-Carbohydrate Binding Sites from Sequence (thesis) -** A balanced predictor, STACKCBPred using **stacking** based approach with **accuracy** & **sensitivity** of **86% & 86.1%** resp. Utilizes features extracted from evolution-driven sequence profile and several predicted structural properties of residues.

Certifications

- Building LLM Applications from scratch into Production
- Building better Generative Adversarial Networks (GANs)
- Natural Language Processing with Sequential and Probabilistic Models

Achievements and Awards

Robotics Competitions: 1st Place - Sinhgad Institutes, Pune - Sociobotics 2016, 2nd Place - Veermata Jijabai Technical Institute, Mumbai - VJTI Robotics 2015. **Awards:** Privateer Graduate Award, Research Assistantship, etc.