Arun Gaonkar

Email : arun.rg37@gmail.com
Linkedin: https://www.linkedin.com/in/arun-gaonkar/
Mobile : +1-919-349-6169
Website: https://arungaonkar.github.io/

Education

- **North Carolina State University** — Raleigh, NC
  Masters in Computer Science; GPA: 3.96 — *Aug 2021 - May 2023*
- **PES University** — Bangalore, India
  Bachelor of Technology, ECE; GPA: 3.64 (8.57/10.0) — *Aug 2016 - Jun 2020*

Skills Summary

- **Languages**: Python, R, C/C++, NodeJS, Shell Scripting
- **Frameworks**: HTML, JavaScript, CSS, REST, Express, Apache, Jenkins, Ansible, Chai, Mocha
- **Tools & Technologies**: Keras, TensorFlow, PyTorch, Git, GitHub, VS Code, Jupyter Notebook
- **Databases**: MySQL, SQL, PostgreSQL, MongoDB, AWS DynamoDB, PowerBI, Tableau
- **ML & NLP Models**: CNN, RNN, LSTM, BERT, Word2Vec, GLoVe, ELMo, Tf-Idf
- **Data Science**: Pandas, NumPy, NLTK, SpaCy, Scikit-Learn, Causality

Experience

- **North Carolina State University** — Raleigh, NC
  Research Assistant - Social AI Lab — *Jan 2023 - Current*
  - Scraped around 26000 reddit posts from r/Asianparents pushshiftio API.
  - Labelling 5000 sentences (belonging to -600 posts) for prejudice, type of prejudice and prejudice topic.
  - Planning to leverage multiple transformer-based models like RoBERTa for identifying prejudiced sentences.

- **North Carolina State University** — Raleigh, NC
  Research Assistant - IEC Lab — *Sept 2022 - Current*
  - Utilized interactive learner to teach AI agents on how to solve algebra equations.
  - Conducted 15+ sessions and spearheaded data collection procedures to understand learning patterns.
  - Assisted in data wrangling and data labeling to improve its performance.

- **LexisNexis Risk Solutions** — Alpharetta, GA
  Intern - HPCC Systems — *May 2022 - Aug 2022*
  - Discovered causal relations using Bayesian networks, conditionalization and drew causal models with 9+ variables.
  - Leveraged discretization technique to increase the speed of probabilistic dependence tests by 10-fold.
  - Analyzed behavior of the HPCC_Causality toolkit on synthetic and real-world datasets for causal discovery.

Academic Projects

- **Context Based Sarcasm Detection (Text Processing)**: Created a dataset of 28,000+ news entries, utilized NLP techniques including embedding, tokenization, and Bi-LSTM & RoBERTa models to achieve 96% classification accuracy for context-based sarcasm detection, resulting in improved understanding of sarcasm in language.
- **Example-Bot: Personalized Code Assistant (App Building)**: Built a server-based chatbot using Ansible and Git bash for CRUD operations on MongoDB, designed to help developers create, store, and retrieve customized code snippets and API examples, while following Scrum and Agile methodologies. Achieved 96% test coverage by implementing unit testing with Chai and Mocha, and deployed a seamless CI/CD pipeline.
- **Wildfire Data Analysis and Cause Prediction (Data Analysis & Data Visualization)**: Maximized wildfire cause prediction accuracy to 93% by leading a team of 3 in building an end-to-end machine learning solution and analyzing 1.88 million records using ETL and data visualization techniques, potentially enabling more effective and efficient fire prevention and response measures.
- **Brain Tumor Image Classification (Image Processing)**: Achieved 92% accuracy on MRI image classification for brain tumors by developing deep learning models with Bi-LSTM and CNN using TensorFlow and optimizing hyperparameters, leading to potential improvements in early detection and treatment
- **Large-size Matrix Inversion Using Recurrent Neural network**: Developed Hopfield Neural Network by constructing a matrix representation with op-amps for analog realization and reducing the computation complexity of matrix inversion from O(N$\hat{3}$) to O(N$\hat{1}$.5), improving the efficiency and speed of the neural network training process.

Honors and Awards

- Won prestigious Prof. CNR Rao award for being a consistent top 10% performer in PES University.