Arun Gaonkar

919.349.6169 | arun.rg37@gmail.com | linkedin.com/in/arun-gaonkar | github.com/ArunGaonkar | Portfolio | Raleigh

EDUCATION

Master of Computer Science Aug 2021 – **May 2023**
North Carolina State University, GPA: 3.96/4.00 Raleigh, NC, USA
Courses: Neural Networks, Natural Language Processing, Automated Learning & Data Analysis, Artificial Intelligence, Software Engineering, Design and Analysis of Algorithms, Algorithms for Data Guided Business Intelligence

Bachelor of Technology, Electronics and Communications Engineering Aug 2016 - Jun 2020
PES University, *GPA: 3.65/4.00* Bangalore, India

EXPERIENCE

Social AI Lab, North Carolina State University, Raleigh, NC | Research Assistant Jan 2023 – Present
- 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, DistilBERT for identifying prejudiced sentences.

IEC Lab, North Carolina State University, Raleigh, NC | Research Assistant Sept 2022 – Present
- 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 | Data Science Intern [repo, blog] May 2022 – Aug 2022
- Discovered causal relations using Bayesian networks, conditionalization and drew causal models with 9+ variables.
- Leveraged discretization technique to increased speed of probabilistic dependence tests by 10-fold.
- Analyzed behavior of the HPCC_Causality toolkit on synthetic and real-world datasets for causal discovery.

SKILLS

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

PROJECTS

Context-based Sarcasm Detection: *Python, Colab, Keras, TensorFlow, GLoVe, Bi-LSTM, RoBERTa.* [repo, report]
- Curated a news dataset of 28,000+ entries using BeautifulSoup. Pre-processed using pandas and numpy (ETL).
- Investigated context dependency for sarcasm detection by employing embedding & tokenization by NLTK libraries.
- Developed and trained Bi-LSTM & RoBERTa models for sarcasm detection, achieved 96% classification accuracy.

Example-Bot: Personalized Code Assistant: *Nodejs, JavaScript, REST API, Ansible, MongoDB, Unit testing*
- Designed a chatbot to assist developers to create, store and retrieve personalized code snippets and API examples.
- Deployed server-based chatbot for CRUD operations on MongoDB by leveraging Ansible, Git bash & CI/CD pipeline.
- Improved test coverage to 96% by utilizing unit testing with Chai & Mocha following scrum & agile methodologies.

Wildfire Data Analysis and Cause Prediction: *Python, Jupyter, Keras, RFC, KNN, Bi-LSTM, CNN.* [repo, report]
- Led a team of 3 to build an end-to-end machine learning solution for wildfire cause prediction.
- Analyzed 1.88 million records using pipelined ETL, data visualization techniques such as matplotlib and seaborn.
- Employed models like RFC, KNN, Bi-LSTM, CNN to predict wildfire reason. Best accuracy of 93% was achieved by CNN.

Brain Tumor Image Classification: *Python, Keras, TensorFlow, PyTorch, Computer Vision, Bi-LSTM, CNN.* [repo, report]
- Developed MRI image classification models by training deep learning models like Bi-LSTM & CNN using TensorFlow.
- Improved classification accuracy to 92% by optimizing hyperparameters.

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