

SUMANTH MANDURU

@smanduru@gmu.edu 📞: +1 571-585-9446 📧: sumanthmanduru 📄: github.com/Sumanth197

EDUCATION

- **George Mason University** Virginia, USA
Master of Science - Computer Science, GPA: 3.9 *August 2022 - May 2024*
- **Indian Institute of Information Technology, Sri City** Chittoor, India
Bachelor of Technology - Computer Science *August 2015 - May 2019*

SKILLS SUMMARY

- **Programming** Python, C/C++, R, Javascript, NodeJS
- **Frameworks** Numpy, Pandas, Scikit Learn, PyTorch, Tensorflow, Hadoop, Apache Spark, SQL, NoSQL - MongoDB, Neo4J
- **Software Development** FastAPI, Flask, Postman, Streamlit, Docker, CI/CD Pipeline, HTML, CSS
- **Cloud Technologies** AWS-S3, AWS-SageMaker, GCP, GitHub, Hugging Face
- **Machine Learning** Kernel Machines, ANN, CNN, RNN, LSTM, Transformers, Computer Vision (OpenCV), NLP (LLM)

EXPERIENCE

- **George Mason University, C4i and Cyber Center** Fairfax, USA
Graduate Research Assistant *Sep 2022 - Present*
 - Built a reinforcement learning agent to select maneuvers for a high speed aerospace vehicle to conduct an emergency descent. Executed the trained agent with varying initial conditions for altitude and velocity, covering a wide range from $\pm 10\%$ to $\pm 40\%$.
 - Analyzed diverse datasets with neural and factorization models, and demonstrated the pivotal role of context features in recommendation systems to enhance accuracy by 5% and optimize performance.
- **Jio Platforms Limited** Hyderabad, India
Deputy Manager - Data Science *Sep 2021 - Aug 2022*
 - Developed an Interpretation Engine encompassing Data Interpretability and Model Interpretability. Led Model Interpretation efforts, integrating 20 Explainable AI (XAI) methods for enhanced Model Explainability.
 - Contributed to the AJIO business project on Demand Planning and Discount Optimization, constructing validation APIs for Self Causality, Forecasting, and Clustering Pipelines, leading to an 18% increase in sales.
 - Fabricated an Ensemble Engine incorporating 3 Base Ensemble techniques such as Max Voting, Average, and Weighted Average, along with 3 advanced methods including Stacking, Blending, Boosting, and Bagging.
- **Stride.ai** Remote
Jr Data Scientist *July 2020 - Sep 2021*
 - Achieved 93% accuracy in document classification through logistic regression and neural networks, resulting in \$50,000 OpEx reduction and saving 200-man hours annually. Assembled BERT for NER, incorporating 40 Financial entities with Python, PyTorch, and SpaCy.
- **Nikulsan - Adohm** Bangalore, India
Data Science Engineer *May 2019 - June 2020*
 - Calculated similarity measures to develop a predictive model to classify real-time leads to block irrelevant placements, keywords for campaigns and generating client savings of \$100 per lead.
 - Constructed an LSTM Model training on ad copy descriptions and headlines, resulting in ad-copies generation with a seed-text of efficient search keywords, giving rise to increase of 74% impressions, and 29% conversions.

PUBLICATIONS

- **Relative Switching Point based Dynamic Positional Representation for Code-Mixed Sentiment Analysis** M Ali, S T Kandukuri, Sumanth Manduru, P Patwa, A Das. AAAI 2022 Student Abstract and Poster Program.
- **DFW-PP: Dynamic Feature Weighting based Popularity Prediction for Social Media Content.** Viswanatha R G, Chaitanya B S N V, Prathyush P, Sumanth M, Mrinalini C, Dileep K P, Snehasis M. The Journal of Supercomputing 2023.

RESEARCH PROJECTS

- **IDIA Summer Fellowship with 22nd Century Technologies, Inc. (Python, HuggingFace, FastAPI, AWS)**
 - Designed an efficient model inference pipeline for keyword extraction, leveraging KeyBERT, and further integrated BART for advanced summarization capabilities.
 - Effectively optimized document processing, S3 storage, and Llama indexing, elevating the chatbot's responsiveness for accurate user query responses. Resulted in a 30% enhancement in search efficiency.
- **Neural Network(s) Pruning: One and Ensembles (Python, PyTorch)**[Project Link](#)
 - Engineered a compression technique for Deep convolutional neural networks (DCNN) architectures, minimizing size and parameters while maintaining equivalent accuracy to the original model, yielding an 80% cost reduction and a 5% accuracy improvement. Orchestrated the experiments on ARGO and Hopper Clusters, managing SLURM jobs for efficient resource utilization.
- **Hierarchical Attentional Graphs: Scaling NLP Beyond Sequence Length Constraints (PyTorch, Graph Neural Networks (GNNs), HuggingFace)** [Project Link](#)
 - Reproduced HiPool for long document classification, examined multilinguality with generated 500K (512-token) and 10K (2048-token) multi-lingual datasets. Evaluated robustness with GNNs, applying techniques like Word Dropping, Jumbling, and Misspelling.
 - Conducted a thorough comparative analysis of HiPool BERT against GraphSage, GCN, and GAN architectures. Secured 53% accuracy on multilingual datasets using BERT Multilingual and XLM Roberta models.
- **Unraveling Insights for January 2016 - Predictive Analysis of NYC Yellow Taxi Services (PySpark, SHAP)**[Video Link](#)
 - Implemented an extensive Exploratory Data Analysis (EDA) on a 1.7GB dataset with 10 million rows and 19 columns, exploiting PySpark scalability for seamless processing. Employed outlier handling, feature engineering, and machine learning regression models to predict the target variable.
 - Furthermore, the utilization of SHAP provided valuable insights into the individual contributions of features, allowing for the extraction of numerical feature importance.