

Sai Venkata Kaushik Pillalamarri

984-895-7844 | spillal2@ncsu.edu | [LinkedIn](#) | [GitHub](#) | Raleigh, NC

EDUCATION

North Carolina State University, Raleigh, NC || PhD in Computer Science August 2024 - May 2027
Courses: Advanced Machine Learning, Database Management Systems
North Carolina State University, Raleigh, NC || Master of Computer Science August 2022 - May 2024
Courses: Neural Networks, Automated Learning and Data Analysis, Natural Language Processing, DevOps GPA:4.0/4.0
National Institute of Technology AP, India || Bachelor of Technology in Electrical Engineering August 2018 - May 2022
Courses: Data Structures and Algorithms, Problem solving using Computer Programming GPA:7.61/10

SKILLS

Languages: C++, Python

Libraries/Frameworks: TensorFlow, PyTorch, Keras, Scikit-learn, Seaborn, NumPy, Pandas, BeautifulSoup.

Tools and Frameworks: Git, GitHub, MATLAB, MySQL, Bash, Ansible, Kubernetes, Docker, Github Actions CI/CD, Azure, Prometheus.

WORK EXPERIENCE

Graduate Student Researcher, Christopher G.Healey, Raleigh, NC June 2023 – Aug 2023

- Devised a system based on Clustering and Large Language Models(LLMs) on news articles using extractive summarization to provide different ideological perspectives on the given news topic, for readers who are interested in seeing different perspectives.
- Implemented web scraping techniques to collect data from various online news sources, utilizing BeautifulSoup.

Research Intern, International Institute of Information Technology, Hyderabad, India December 2021 - January 2022

- Performed exploratory data analysis, statistical analysis, data cleaning, aggregation and feature engineering using pandas.
- Implemented traditional Machine Learning techniques-based models like RandomForest, SVM and Naive Bayes and observed and compared the results.

PROJECTS

LOLgarithm: Humor Classification [GitHub](#)

- Defined humor recognition as a classification task, distinguishing between humorous and non-humorous instances and leveraged Lexicons for sentiments and Statistics of Structural Elements (SSE) for statistical insights in sentence structure.
- Explored the semantic layers of humor using Word2Vec embeddings, and analyzed the incongruity, ambiguity, and phonetic structures within sentences and incorporated contextual information through ColBERT embeddings resulting in an increase in the accuracy for the SOTA Colbert model from 50% to 62%, and achieved a 26% improvement in F1-Score on unseen data.

CI/CD Pipeline for Coffee Ordering System [GitHub](#)

- Engineered a robust CI/CD pipeline for the Coffee-project, a Node.js application, automating key stages of software development.
- Leveraged Docker for efficient containerization, ensuring consistency across various environments and orchestrated deployment and rollback using Kubernetes, optimizing application scalability and resilience.
- Integrated GitHub Actions to automate continuous integration and delivery, enabling seamless and reliable production deployments and employed Ansible for automation, streamlining repetitive tasks and enhancing overall workflow efficiency.

Terrain Identification based on Time-series Data [GitHub](#)

- Built a Deep Learning based classifier employing Pytorch and TensorFlow to distinguish the action of a person using accelerometer and gyroscope data.
- The baseline models were CNN and LSTM based networks using TensorFlow, which achieved an F1 score of 73% and 77%.
- Built a CNN Bi-directional LSTM model utilizing Pytorch and fine-tuned it, which attained an F1 score of 91%.

Credit Card Fraud Detection [GitHub](#)

- Devised a system in Python with TensorFlow to detect fraudulent credit card transactions on a highly imbalanced dataset.
- Boosted the Recall of each model at least 20% by employing sampling techniques like ADASYN, SMOTE and Centroid Clustering to resolve the class imbalance in the data set.
- Implemented a variant of Random Forest using Numpy and Pandas and Sci-Kit Learn and achieved PRAUC of 75%.
- Developed CNN and ANN networks on the resampled data and compared their performances with Random Forest, Voted Ensemble baseline models. Finetuned the architectures utilizing Keras-tuner to achieve higher PRAUC of 76% and 80%.

Hierarchical K-Prototypes based Explanation System [GitHub](#)

- Developed a semi-supervised explanatory system and tested it on various datasets and observed the results.
- Implemented the clustering technique, K-Prototypes, to group data instances using numerical and categorical data, resulting in a hierarchical structure of these groups.
- Demonstrated the effectiveness of our approach on several datasets and showed that our system can provide meaningful explanations for numerical and categorical data.
- Implemented GitHub Actions CI workflow for the project, ensuring seamless integration and deployment of code changes.