**Pavan Sabnaveesu**

**832 274-8181|** [**sabnaveesuprasanth@gmail.com**](mailto:sabnaveesuprasanth@gmail.com) **|** [**GitHub**](https://github.com/prasanthsabnaveesu) **|** [**LinkedIn**](https://www.linkedin.com/in/prasanth-sabnaveesu/)

**Professional Summary**

Machine learning engineer with 5+ years of experience in deploying scalable AI solutions across NLP, computer vision, and recommendation systems, skilled in full-stack development, cloud platforms, and data visualization for impactful insights

**Education** *Texas A&M University* Master of Science, Computer Science  ***CGPA: 3.81***

**Technical Skills**

Programming & Databases : Python, R, SQL, PostgreSQL, DSA, C, C++, Pinecone

Web Development : Django, HTML, CSS, jQuery, Bootstrap, REST API, Fast API, Flask

Data Visualization : Tableau, Matplotlib, Seaborn, Plotly, Power BI

Machine Learning : NumPy, Pandas, Scikit-learn, TensorFlow, Keras, PyTorch, OpenCV

Cloud & DevOps : AWS, Azure, GCP, Docker, Kubernetes, CI/CD (Jenkins)

Data Processing : Apache Spark, Hadoop, Kafka, A/B Testing

Machine Learning : Regression, XGBoost, Random Forest, LSTM, CNN ,GRU

Advanced ML : PEFT (LoRA, QLoRA), RAG, GPT, BERT, Transformers

**Work Experience**

***Graduate Research Assistant, Texas A&M University* *Feb 2023 – Present***

* Designed and deployed a wind turbine blade segmentation system using Mask R-CNN, YOLOv7, and YOLOv8 achieving 98.7% mAP and a 25% reduction in training time
* Developed real-time object and lane detection for self-driving cars, boosting lane accuracy by 21% with U-Net ResNet 101 and enhancing traffic sign/pedestrian tracking via YOLOv8 and ByteTrack
* Researched and implemented state-of-the-art complex neural network architectures and algorithms
* Optimized LLM performance by 20% using Parameter-Efficient Fine-Tuning techniques (LoRA, QLoRA)
* Developed a RAG model using Hugging Face Transformers and Pinecone, enhancing enterprise document search accuracy by 35% and integrating seamlessly with existing knowledge bases to extract business insights resulting in a 15% increase in revenue

***Software Developer – AI, NEXT ROW Private Limited* *July 2021 – Dec 2022***

* Created prototypes for machine translation, speech recognition, and face recognition, resulting in a 15% increase in accuracy and a 20% decrease in latency
* Collaborated with cross functional teams to design, build, and deploy AI models for various domains, ensuring robust solutions to business problems
* Translated Chinese to English using NLTK and wubi, cleansing and tokenizing sentences for input. Employed GRU-based encoder-decoder architecture for accurate language translation.
* Optimized and troubleshooted SageMaker model deployments, ensuring seamless performance and adaptability through continuous monitoring, validation, and automated model updates
* Led the development of a recommendation system integrated with a graph database (AWS Neptune), which increased user engagement by 20% through personalized recommendations
* Web scraped preprocess data from various sources, ensuring data quality and integrity through rigorous cleaning and validation processes
* Developed CNNs from scratch using NumPy and Pandas, achieving over 99.1% accuracy on image classification
* Enhanced predictive model achieving a 25% increase in accuracy through advanced feature engineering and hyperparameter tuning

***Software Developer – AI, Meslova Systems Private Limited* *Sept 2018 - June 2021***

* Developed full-stack applications using Python and Django, integrating ML and AI models resulting in a 30% increase in performance and response time
* Enhanced ETL processes with Hadoop and SQL which increased efficiency by 25%, and built real-time pipelines with Kafka and SageMaker to automate ML retraining for model accuracy improvement
* Streamlined CI/CD processes with Jenkins and Docker, boosting model accuracy by 20% and deployed AI applications on Docker and Kubernetes for seamless scaling and continuous performance optimization