Pratish Mashankar

Radford, VA | (571) 507-3388 | pratishsmashankar@gmail.com | linkedin.com/in/pratish-mashankar/ | github.com/PratishMashankar

PROFILE SUMMARY

Data-driven professional with a Master's in Computer Science (Machine Learning) and 3+ years of experience in data analysis, GenAI, and project management. Skilled in deploying ML and NLP models, designing data pipelines, and developing interactive dashboards to translate complex data into actionable insights for strategic decision-making.

EDUCATION

George Mason University | Master of Science in Computer Science (**Machine Learning**) | Fairfax, VA | GPA: 3.7/4.0   **May 2024**

Jawaharlal Nehru Technological University | Master of Science in Information Technology | Hyderabad, India | GPA: 3.3/4.0   **July 2021**

EXPERIENCE

Data Analyst | Radford University   **Radford, VA | June 2024 – Present**

- Pioneered a data collection strategy increasing knowledge rate by 101.5% in 2 months through data integration, survey, and LinkedIn outreach to 3000+ graduates, establishing the first comprehensive post-graduation dataset for Radford University.
- Executed the duties of a DA Project Manager liaising 4 Lightcast products to capture graduation outcomes and ensure effective data integration into the University's data ecosystem for strategic planning and decision-making.
- Designed and deployed interactive Power BI dashboards integrating 5 data sources, writing DAX queries, and creating calculated measures that translated over 15 complex outcomes into actionable insights for university leadership.
- Conducted extensive User Acceptance Tests on a new CRM platform identifying over 15 abnormalities and helping 7+ universities.

Data Scientist | George Mason University   **Fairfax, VA | October 2022 – May 2024**

- Developed a text summarization model by fine-tuning the PEGASUS transformer from HuggingFace trained on 16,000 conversations and deployed the application as a Fast API webpage on AWS EC2 through GitHub actions, ECR and Docker. [GitHub]
- Conducted Python exploratory data analysis (EDA) on 9,000 VoIP data for GMU's Telecom System Replacement and generated reports outlining migration strategies and ROI projections, ensuring smooth transitions for 5,000 faculty and staff members.
- Devised 16 SPARQL templates automating the dataset generation of over 10,000 Wiki data queries for fine-tuning LLMs to forecast an increase in the accuracy of query generation using natural language prompts beyond 6%.

Software Engineer | Virtusa Corporation   **Hyderabad, India | July 2020 – August 2022**

- Implemented and operationalized a sentiment analysis and risk classification NLP solution for detecting and mitigating GxP compliance risks with 85% accuracy by constructing custom AWS Comprehend models trained on 4000 clinical audit reports
- Established data pipelines for risk prediction of audit summaries hosted on AWS EC2 Nginx server by integrating Lambda triggers on S3 for file upload monitoring and incorporated real-time prediction insights visualization using Amazon Quick Sight.
- Developed Autosampler, a stratified data sampling framework on AWS EC2 for extracting subsets from clinical audit reports, enabling auditors to control randomness by allocating specific weights to risk factors like Non-Compliance or Delayed Submission.
- Collaborated on UI development of Autosampler by establishing Django RESTful APIs for data handling, achieving an 80% reduction in audit time.

ACADEMIC PROJECTS

Multiclass Hate Speech Detection by introducing Noah BERT | Python, Transformer, BERT, Classification   **December 2023**
- Aggregated 3 data sources using pandas to introduce a 3-class hate speech dataset with 7000 documents to fine-tune BERT with 75.9% baseline accuracy for hate speech classification and introducing a novel transformer NoahBERT with 78.9%.

Multilingual Intent Text Classification and Text Revision Model | BLEU, GenAI, Transformer   **May 2023**
- Spearheaded the reproduction of Grammarly's framework for iterative text revision on 8000 multi-domain multilingual annotated corpora, fine-tuned 6 LLMs achieving 71% accuracy for intent classification and a 0.87 BLEU score for text generation.

Machine Learning Bias Auditing | Machine Learning, Classification, Bias   **May 2023**
- Evaluated 9 machine learning models to predict the risk for recidivism in 7000 convicts achieving 69.4% accuracy using gradient boosting and identified biased 0.42 false positive rate against African Americans to recidivate twice than Caucasians.

SKILLS

Programming	Python (Pandas, NumPy, Django, Scikit-learn), Java (Android), C++, C, MySQL RDBMS, Apache Spark
Data Science	Generative AI (LLMs, Transformers), Text Summarization, Machine Translation, Classification, Regression, Ensemble, Unsupervised Learning, Feature Engineering, Statistical Modelling, Neural Networks, CNN, RNN
ML Framework	TensorFlow, Keras, PyTorch, HuggingFace, MLOps (AWS SageMaker, Comprehend, Bedrock), Power BI, Tableau
DevOps	Amazon Web Services (S3, EC2, ECR, boto3), Kubernetes, Git (GitHub actions, workflows, CI/CD), Docker, Terraform

PUBLICATION

International Conference on Computational Vision and Bio-Inspired Computing   **November 2021**
- Introduced an ML paper in 2021 Springer focusing on Twitter API data collection of 60,000 microblogs on the Indian Agricultural Ministry for sentiment analysis using Random Forest and KNN to identify the societal impact of Farm Laws with 90% accuracy.

LEADERSHIP AND COMMUNITY SERVICE

- Trained over 2000 students in public speaking, theatre and presentation by founding Vachan Speaker's Club.
- Tutored high school students from marginalized backgrounds on Mathematics and Science in vernacular language.
- Volunteered at the SV Lotus Temple Mega Kitchen overseeing the preparation and distribution of food to 1000+ devotees per day.
- Mentored 6 authors as the editor of the kMITRA international magazine published monthly with 10,000+ readers in 7 countries.