AMAN PATHAK

Gainesville FL,USA | +1 352-507-5591 | pathakaman2909@gmail.com

in https://www.linkedin.com/in/amanpathak2997 | ⊙ https://github.com/pathak-aman

SUMMARY

Machine Learning Engineer with publications in Nature, ACM and AMIA, specializing in Natural Language Processing (NLP) solutions for healthcare and fintech with 3+ years of experience.

EXPERIENCE

Machine Learning Engineer | Bioinformatics - University of Florida   May 2023 - Present
- Led development of a GatorTron-based-LLM (F1: 0.95) automating thyroid analysis, saving clinicians 4000+ hours/yr.
- Published 3 high-impact clinical NLP papers in top-tier venues, including Nature (impact factor: 12.8) and AMIA.
- Developed BERT models to detect sleep disorders in 200K Alzheimer's patients, cutting care cost by $500K/yr.
- Deployed GPT-4 model automating 50+ monthly IRB reviews with 94% accuracy and reducing approval time by 60%.

Machine Learning Graduate Research Assistant | Bioinformatics - University of Florida   Jun 2022 - Apr 2023
- Outperformed state-of-the-art ML models, identifying 80% more undiagnosed high-risk suicide veterans (F1: 0.92).
- Engineered an Elasticsearch based solution to efficiently query 21M clinical reports, achieving <10ms search time.
- Created automated NLP pipelines for text de-identification of 100K+ reports monthly with an F1-score of 0.86.

Machine Learning Engineer | Monsoon CreditTech Pvt. Ltd   Aug 2020 - Jul 2021
- Built XGBoost and LGBM powered credit scorecards processing $100M+ in loans, reducing defaults by 25%.
- Deployed end-to-end ML pipeline on Azure with CI/CD, ensuring 99.9% uptime for serving 10K+ daily requests.
- Implemented spaCy-powered NLP system identifying high-risk borrowers, preventing $5M+ in potential defaults.

Data Science Intern | Hvantage Tech Solutions Pvt. Ltd   Feb 2020 - Aug 2020
- Designed an image pre-processing pipeline using OpenCV, boosting Google Vision API's OCR accuracy by 28%.
- Integrated LSTM-based NER capabilities to extract dates and timestamps, reducing extraction time by 45%.

TECHNICAL SKILLS

Languages: Python, R, Java, SQL, Bash/Shell, GO
NLP: Transformers, Langchain, Ollama, Open AI, SpaCy, BERT, GPT, Prompt Engineering
Machine Learning: NumPy, Pandas, Sklearn, Tensorflow, Pytorch, Time series analysis, Clustering, Regression
Data Science: Data Visualization, Statistical Modeling, Feature Engineering, A/B Testing, Hypothesis Testing
MLOps: MLflow, Docker, Kubernetes, AWS, Azure, CI/CD Pipelines, Version Control (Git, Bitbucket)
Frameworks & Databases: Django, Flask, Tableau, Excel, REST API, Elasticsearch, MongoDB, MySQL

PROJECTS

Duolingo Max Clone | github link | Python, LLM, Ollama, Llama3, DeepSeek-r1, Gen AI
- Developed an open-source AI-powered Spanish tutor, offering premium language learning features, saving $178/yr.
- Implemented context-aware conversations with vocabulary suggestions and real-time grammar correction.

GatorGuide | github link | Python, LLM, Langchain, Langsmith, Llama2, GPT-4, Gen AI
- Designed healthcare LLM system to simplify medical reports for patients and extract clinical insights for clinicians.
- Integrated Llama2 and GPT-4 to process EHR reports in <2 seconds with 91% accuracy.
- **Achievement**: Awarded first place in the UF HOBI Shark Tank 2024 for innovation and clinical efficiency.

Aggre-Gator | github link | Python, Flask, MongoDB, Elasticsearch, sklearn, Docker | Team: 5
- Launched a personalized news curator using item-based collaborative filtering for real-time content recommendation
- **Achievement**: Won Amazon Lightsail Hackathon 2021 for innovative approach to information overload.

EDUCATION

University of Florida *MS, Biomedical Informatics*, GPA: 4.00   Aug 2023 - Dec 2024
University of Florida *MS, Computer Science*, GPA: 3.77   Aug 2021 - May 2023

PUBLICATIONS

- Adekkanattu P, **Pathak A**, et al. Deep learning for ..., **Nature** npj Digital Medicine, 2024. **Impact factor: 12.4** | link
- **Pathak A**, et al., Extracting Thyroid Nodules Characteristics from ..., **AMIA** 2023. **Impact factor: 4.8** | link