Manish Prajapati

manishpjt1998@gmail.com | +917905210526 | Gorakhpur, U.P | LinkedIn | YouTube

SUMMARY

Machine Learning Engineer with 3 years of expertise across data science and MLOps. Successfully bridged the gap between model development and production deployment by implementing robust CI/CD pipelines and automating model training workflows. Experienced in creating automated workflows and maintaining reliable ML pipelines that runs smoothly 99% of the time. Demonstrated ability to translate business requirements into practical ML solutions that deliver measurable results.

EXPERIENCE

MLOPS Engineer, Wysa *Sep 2022 - Present*

- Led end-to-end MLOps initiatives implementing robust **CI/CD** pipelines via **GitHub Actions,** automated **model deployment** across development and staging environments. Engineered automated testing and evaluation pipelines for seamless model weight updates, ensuring consistent quality assurance. Improved **productivity a**nd speed of team by **2x a**nd cutting manual work by **100%**.
- Successfully led a cross-functional team and directing project scope and **feature implementation** while maintaining clear **communication** channels. Demonstrated strategic leadership by initiating the project independently and seamlessly transitioning it to team ownership.
- Conceptualized and initiated development of a comprehensive **LLM Playground platform** from scratch that cut testing time by **80%** enabling **prompt engineering, document processing**, and **RAG** testing capabilities. Led a cross-functional team in expanding platform features, incorporating cost estimation tools and document query functionalities while maintaining seamless integration with multiple **LLM providers**.
- Planned and initiated the implementation of **RAG** architecture for **GenAI** applications, integrating both managed and open-source LLMs.
- Developed innovative **LLM a**pplications including personalized user interaction systems and emotion tracking analytics. Engineered comprehensive solutions for automated weekly **report generation** and user **sentiment analysis im**proving **user engagement b**y **40%** and **cost reduction** of roughly **60%**.
- Transformed manual model aptness evaluation processes through **end-to-end automation**, reducing assessment time from **months** to **days** through sophisticated **bash scripting** and **Python** implementation. Architected large-scale data processing infrastructure utilizing **orchestration tools** and **AWS** ecosystem for optimizing **data pipeline** efficiency.
- Implemented distributed model training pipeline for **transformer-based** model and large-scale dataset processing using **PyTorch**.
- Collaborated with **Microsoft Research** to establish pioneering evaluation metrics for **AI** systems in **Indic languages**.

Data Scientist, Willow AI *Jul 2021 - Jul 2022*

Client: Audible Health AI

- Optimized existing prediction system up to 3x by hosting **ML** models separately using **Triton Inference Server** and using **GRPC** requests instead of **HTTP** to get the lower latency.
- Developed a real-time GPU monitoring tool using **Nvidia DCGM**, **Prometheus**, and **Grafana**.
- Developed model benchmarking and load testing tools using **Python** and **shell script** to automate logging model's performance with keeping track of versions.
- Implemented **RabbitMQ** messaging queue for more **reliable**, **persistent**, and **high availability**.
- Led a team of 3 members. Managed complete **client communications**, and **expectations**, and ensure on-time delivery of results.

Client: Dr. Fossums's Pet Care

- Created an **NLP pipeline** to read, extract and save text data from reports to **MYSQL database**.
- Developed an **OCR** using **Paddle OCR**. Collected 10000+ samples of symptoms and disease from patient reports.
- Trained **XGBoost m**odel to automate prediction of disease using symptoms in reports and achieved **87% accuracy** verified by human validation.

Data Science Intern, Willow AI *Feb 2021 - Jul 2021*

Client: Marshal Marketing

- Performed **EDA** and identified key drivers for **high sales** growth. This was used to calibrate instore experiences and management decisions for **200+ stores** in **U.S**
- Implemented single-click automation of legacy dashboard, utilized **Tableau** reducing production time by **50%**

EDUCATION

B.Tech, Computer Science and Engineering *Aug 2017 - Jul 2021*
Lovely Professional University GPA: 7.1

Created with **resumepuppy.com**

SKILLS

Languages: Java, Python, Bash scripting
Core skills: Deep Learning, Predictive Modelling, Image Processing, NLP, CI/CD, LLM, RAG, Model deployment and Automation, Technical Leadership
Frameworks: Tensorflow, PyTorch, FastAPI, ChromaDB, Langchain, Streamlit
Tools: Docker, Kubernetes, Git, Tensorflow serving, Triton Inference Server, AWS (EC2, Lambda, S3, CloudFormation), Ollama, Monitoring tools

Publications

Action Replication in GTA5 using Posenet Architecture with LSTM Cells

Certificates

Advanced Machine Learning With TensorFlow On GCP, Tensorflow In Practice Specialization

Created with **resumepuppy.com**