Vyshnev S

Bengaluru, India | +91-9400380840 | vyshnevnandanam@gmail.com | linkedin.com/in/vyshnev | github.com/vyshnev

Experience

LTIMINDTREE March 2022 – May 2025

ENGINEER October 2022 – May 2025

- Contributed to a customer churn prediction project by experimenting with models ranging from **logistic regression to XGBoost**. Improved churn identification by 7%, leading to **cost savings of £1.2M** over 3 years.
- Expanded the feature set from an initial 22 to over 40 by engineering features from customer behavior logs and basic sentiment analysis of feedback texts, resulting in **an AUC-ROC improvement from 0.71 to 0.83**.
- Collaborated with the DevOps team to deploy the churn model via a batch prediction pipeline on AWS, **processing roughly 5,000 predictions per hour** and integrating it into a CRM dashboard.
- Participated in a data quality check to identify bias in model outcomes. An adjustment reduced disparity across two key customer segments by 2%.

JUNIOR ENGINEER March 2022 – October 2022

- Used **New Relic to monitor key performance** metrics and identify anomalies in the production environment.
- Helped maintain data accuracy in Salesforce pipelines, supporting internal dashboards used by the business.
- Logged recurring technical issues and created structured **Jira tickets, accelerating resolution time**.
- Worked with analysts to optimize reports using SQL and Python, **improving report refresh time by 15%**.

Projects

End-to-End Used Car Price Prediction:(Github) | *Python, Streamlit, MLFlow, DVC, Docker, AWS*

- Developed a price prediction model by training algorithms like XGBoost on web-scraped data stored in AWS S3.
- Managed the ML lifecycle by tracking experiments with MLflow and versioning data and models with DVC.
- Built and containerized a user-friendly Streamlit application with Docker for consistent and isolated deployment.
- Deployed the application to an AWS EC2 instance, making the predictive model accessible via a web interface.

Transformer Dialogue Chatbot(Github, Huggingface) | *Python, PyTorch, Gradio, Hugging Face Hub*

- Designed and implemented a deep learning sequence-to-sequence Transformer chatbot from scratch in PyTorch, using multi-head self-attention, positional embeddings, and encoder-decoder stacks.
- Developed an end-to-end NLP pipeline for Cornell Movie-Dialogs Corpus, creating a 15k+ token vocabulary and training a model to generate dialogue, addressing challenges like mode collapse.
- Deployed the chatbot using Gradio and managed the 280MB model artifact on Hugging Face Hub, enabling on-demand model loading and maintaining a lightweight, production-ready code repository.

Model Optimization: Distillation and Quantization(Github, Huggingface) | *Python, PyTorch, Optimum, ONNX*

- Architected a model optimization pipeline, reducing a BERT-base model's on-disk size by 85% (439 MB to 64 MB) and boosting CPU inference speed by over 5x.
- Implemented knowledge distillation by subclassing the Hugging Face Trainer to transfer knowledge from a BERT teacher to a compact DistilBERT student, retaining over 99% of the performance with 92.5% accuracy.
- Applied post-training quantization using ONNX Runtime, compressing the distilled model by an additional 75% and achieving a 4.3x speedup with a negligible 0.6% drop in accuracy.
- Established a reproducible MLOps workflow by versioning and publishing both distilled and quantized model artifacts to the Hugging Face Hub, enabling seamless deployment and collaboration.

Education

SRI MANAKULA VINAYAKA ENGINEERING COLLEGE PUDUCHERRY, IN

B.Tech in Mechanical Engineering *August 2017 – October 2021*

Technical Skills

Languages: Python, SQL, HTML/CSS, C++
Libraries and Frameworks: Seaborn, Matplotlib, Pandas, NumPy, Scikit-learn, Keras, PyTorch, Flask, FastAPI, Streamlit, Gradio, Hugging Face, Selenium, Docker
Developer Tools: Git, Postman, Jupyter Notebooks for Data Science and Machine Learning, Jira
Databases: PostgreSQL, MySQL, SQLite
Cloud Technologies: AWS (EC2, S3, ECR, RDS, Beanstalk, CI/CD (Github Actions)), Azure