Praveen Kumar Sridhar

Boston, MA • (857)-424-8758 • sridharpraveenkumar@gmail.com
•linkedin.com/in/praveenks23 • Portfolio • github.com/PraveenKumarSridhar

Professional Experience

Data Scientist Intern **May 2022 - Aug 2022**

Meta(FaceBook) - Ads & Business - Seattle, Washington

- **Redefined** how the organization looks at **advertiser churn**. Created a strong framework to predict churn and identify levers that reduce churn.
- Designed an **exhaustive advertising lifecycle model** and utilized it to construct a **churn taxonomy.**
- Developed numerous ML models to forecast churn (with an 85% ACC) and used **calibration curves** to demonstrate the models' dependability. The critical features that are likely to generate churn, were extracted using **SHAP values**.
- **Built friction vectors** to identify and isolate levers the organization can control to reduce friction(obstacles that lead to churn). Wrote highly optimized SQL queries to create these features.
- Implemented **causal inference models** like **X learners, & Causal trees** to measure the impact of each friction to cause churn.
- **Presented the findings to the organization and leadership.** The churn model & framework developed will be used in the future as a part of the broader **lifetime model** for advertisers.

Data Scientist **Jun 2018 - Aug 2021**

Intellect Design Arena Ltd. - R&D team, Chennai, India

- Designed, developed, and shipped Deep learning models: **CNNs, LSTMs**, and **Bidirectional LSTMs with attention**.
- These models achieved accuracy upward of 90% in the production environment. Also designed and implemented a module to capture feedback from users.
- Designed, built, and shipped a complex ensemble classifier, built using **BERT** and **ROBERTA**.
- Experimented with the best OCRs **tesseract, easyOCR, paddleOCR, aOCR**, etc. Combined **CRAFT** with tesseract which produced a 5% increase in accuracy.
- Used image processing tech. with **Tesseract** & **CRAFT** to read data from Machine Readable Zone in passports.
- Developed an entire NLP pipeline using **RabbitMQ** (from tokenization to spell checking) which runs on multiple servers which are completely customizable wrt the number of workers/consumers and flow.

Education

Masters, Data Science **Aug 2021 - May 2023**

Northeastern University, **Khoury College,** Boston, MA **GPA**: 4.0/4.0

- **Coursework**: NLP, Supervised Machine Learning, Algorithms, Deep learning
- **Positions held**: Teaching Assistant, Research Assistant

B.Tech, Computer Science **May 2014 - Apr 2018**

VIT University, Chennai, TN **GPA**: 8.93/10

Technical Knowledge

- **Languages**: Python, R, Scala, C/C++, Java
- **ML & DL Packages**: TensorFlow, Keras, PyTorch, sklearn, Plotly, Matplotlib, NEAT, OpenCV, tesseract, EasyOCR.
- **Databases**: MongoDB, Redis, SQL Server, PostgreSQL.
- **Technical Skills**: NLP, Image Processing, Deep Learning, Machine learning, Tableau, Data Cleaning & Interpretation, A/B testing, Causal inference techniques, git.
- **Certifications**: Natural Language Processing in TensorFlow(Coursera), Deep learning specialization(Coursera).

Projects

- **Poetry Generator**: Trained **Bidirectional LSTM** neural networks to generate poems in 3 languages (English, Hindi, Tamil) last 2 being regional South Indian languages.
- **Question Answering model**: Created a question answering model with the help of transformer-based models like **BERT, DistilBERT, ALBERT,** etc. These models achieved an **F1 score of 81%**.
- **AI Flappy bird**: Built the traditional flappy bird game using pygame and further trained an AI using NEAT (NeuroEvolution of Augmenting Topologies) to play the game. The AI trained quickly and has achieved **a high score of 1000** and plays the game flawlessly.
- **Art generation**: Generated new art using a pre-trained **VGG-19** given a content image and a style image. The generated image has the content from one and the style from the other image.