VENKATA SAI MUKTEVI

Versatile Data Scientist, (224) 249-0303, Seattle, WA
linkedin.com/in/smuktevi|github.com/smuktevi|smuktevi.github.io|muktevi.venkatasai@outlook.com

EXPERIENCE

Microsoft, Data Scientist II, Redmond, WA — Apr 2022-Present

Microsoft AI, Microsoft Edge Growth Data Science team.

- Developed an embedded representation of raw browsing history user data (252 TB), identifying user behavior and patterns across websites to train a neural network for predicting the next most likely URL. Achieved 66% precision on the first recommended site, enhancing Windows Taskbar recommendations.
- Led multiple personalization experiments as the primary Data Scientist for Edge Feature Delighter upsells, New Tab Page variants, and Bing Rewards. Improved retention metrics by 2% through the upsell of a shopping coupon feature, impacting millions of users globally.
- Designed and launched the Churn Labeling and Metric Decomposition Dashboard, empowering leadership with data-driven insights to track and mitigate user churn. The dashboard became the cornerstone of weekly Product Growth Retention meetings, significantly reducing churn mitigation turnaround time.
- Ensured DMA privacy compliance by transitioning to a compliant telemetry signal, analyzing data to establish an alternative KPI. This preserved continuous metric monitoring and experimentation without disrupting business operations.
- Created the Most Valuable User (MVU) metric through server-side analysis and implemented it as a client-side segment on the product in C++ using our model deployment infrastructure, enabling personalized experiments for high-value users across the product.
- Developed user segments using a URL categorization model and client-side signals, leading to improved personalization. Integrated ONNX models to run segmentation efficiently on the client-side.
- Engineered a horizontal solution for experiment owners to measure Total Addressable Market (TAM) and ROI targeting, pivoting product telemetry on various relevant signals for more effective opportunity sizing in experiments.
- Aligned server-side telemetry with client-side signals to ensure accurate measurement and utilization of client behavior metrics, improving overall product analytics.
- **Intern (Jun-Sep 2021):** Designed a scalable method to analyze user behavior using unsupervised clustering, identifying quick churn conditions. Leveraged SHAP and XGBoost for feature interpretability, providing actionable insights on "Quick Churners."

IBM, Cognitive Business Decision Science, Bangalore, India — Jul 2019-2020

- Spearheaded and composed an automated cross-database validation system that monitored consistency of vital product data being transferred across different data resources.
- Automated data transformation, processing and other data-related business tasks using Python scripts saving the time and manual efforts of 8-10 employees on the same tasks.

PROJECTS

Topic-Sentiment Analysis and Relevance System (TSAR) [https://tinyurl.com/y3rwcyxb]

UW Capstone Project **sponsored by Meta**.

- Designed and built the TSAR system, an automated pipeline for detecting conflict, bias, and relevance in social media discussions using BERT embeddings.
- Applied TSAR to Reddit data, analyzing 50 subreddits and over 10,000 comments and 1,000 posts per subreddit to create the "Map of Reddit."
- Presented to the Reddit Data Science Team to improve API development and defined subreddit health metrics.
- Developed BEReddiT, an interactive Python DASH app for visualizing and exploring the analysis results.

Plant Maintenance Recommendation System for Ntwist (Startup)

Python Machine Learning Student Associate, **incubator at CIE IIIT Hyderabad, India**

- Developed a prescriptive analytics system for an industrial plant customer to optimize plant operations, reduce maintenance costs, and increase profits by 42%, while bringing 17% of operational hours back within limits.
- Modeled sensor and lab data from a High-Pressure Acid Leaching plant (Ni/Co extraction) using Deep Belief Networks, RBMs, Feed-Forward Networks, PCR, and PLSR, delivering a high-impact solution.

"Unveiling Collective Wisdom" – Topics on Microsoft Teams Search – Microsoft Global Hackathon 2023

- Designed and implemented a feature for Microsoft Teams that automatically categorizes chats and channel discussions trained using BERTopic and GPT-3.5 Turbo, with GPT-4 for labeling validation sets. Developed UX enhancements to tag relevant contacts via Teams Telemetry.
- This innovation was adopted by the Microsoft Teams Search, Assistant, and Intelligence teams for integration into the product.

EDUCATION

University of Washington, Seattle, Washington — Mar 2022

Master of Science in Data Science | GPA: 3.99/4.0

Some Relevant Coursework: Statistical Machine Learning, Large Scale Data Systems, Machine Learning with Big Data

Ramaiah Institute of Technology, Bangalore, Karnataka, India — Jul 2019

Bachelor of Engineering in Computer Science and Engineering | GPA: 9.42/10.0 | First Class with Distinction

SKILLS

Data Science:	Statistical Analysis, Machine Learning, Analytics, Programming, Visualization, A/B Experimentation, Classification, Regression, Tree-based Learning, Feature Engineering, Data Pipeline Architecture,
Technical Skills:	C++, Python, SQL, AWS Certified Solutions Architect-Associate (2020-2023), Web Development [MEAN], Git