Anurag Sangem

+1 551-220-7098 ansangem@iu.edu anurag-sangem Portfolio

Education

Indiana University, Bloomington, US | Master of Science in Data Science **Aug 2022 - May 2024**

Coursework: Elements of Artificial Intelligence, Machine Learning, Deep Learning Systems, Statistics and Random Variables, Data/Social Media Mining, Cloud Computing, Data/Scientific Visualization, and Big Data Applications

VIT, Chennai, India | B.Tech in Electronics and Communication Engineering **Jul 2016 - Jun 2020**

Work Experience

Oracle | **Data Scientist** | *Bangalore, India* **Aug 2020 - Jun 2022**

- Focused primarily on building propensity to buy models using algorithms like Random Forest, SVM, etc.
- Modeled and scored more than 40 Oracle-owned products like ACONEX, NETSUITE, etc. on the historical data that resulted in a 1.8x lift in opportunity win rate.
- Utilized multi-channel attributions along with an Attention-based RNN and a fully connected neural network, which aims to find out the right marketing channels that ultimately lead to a sale.
- Leveraged the inbuilt functions of OML4py to improve the efficiency of Python scripts by more than 35%.
- Performed Hypothesis Testing for Fusion ERP, EPM, SCM & HCM: Impact of Marketing Touches on Win Rate and Average Won Pipe occurring at different times in the B2B sales funnel.
- Transitioned the CX Sales Data to the Datafox platform resulting in a 17% acceleration of matches through the utilization of Jaro-Winkler similarity for the advanced analytics-marketing team.

Appyhub | **Data Analyst Intern** | *Chennai, India* **Jun 2019 - Jul 2019**

- Created illustrative dashboards using Tableau, SQL, MS Excel, and Power Query based on the requests from the BA teams, saving approximately 8 hours of manual reporting work per week.
- Interpreted and visualized data from 6 marketing campaigns, and user responses and provided weekly reports.
- Conducted market research for the sales data which resulted in an 8% increase in contracts.

Academic Projects

TechPath Navigator | *TensorFlow, NLP, BERT, CNN, LSTM, Fine-tuning, Flask, Pyngrok* **Aug 2023 - Dec 2023**

- Orchestrated rigorous experiments to develop a job recommendation system on 5,000 job descriptions and titles.
- Implement BERT-based models (BERT + CNN, BERT + LSTM), achieving testing accuracies of 82.28% and 80.45%.
- Led the implementation of model deployment setup using Pyngrok and Flask, ensuring seamless access to the BERT-based hybrid models for real-time job recommendation and enhancing user experience.

IU Grad Admit Predictor | *Beautiful Soup, Scrapy, Knn, JavaScript, Flask, AWS, Heroku, GIT* **Jan 2023 - May 2023**

- Created a university recommendation system by analyzing admission data from multiple sources, utilizing a weighted KNN algorithm for personalized recommendations with feature-weighting.
- Achieved accuracy rates of 68.83% for MIS and 54.57% for Computer Science. Deployed the system on Heroku with a CI/CD pipeline for scalability and accurate deployment.

Technical Skills

Programming Languages: Python, C++, R, PL/SQL
Developer Tools: Docker, MLflow, Kuberflow, AWS Sage maker, GCP, Microsoft Azure, Heroku, PowerBI, Tableau
Data science Skills: Statistical modeling, Hypothesis Testing, A/B testing, Deep Learning, Statistics, Postgres, Oracle, Pandas, NumPy, TensorFlow, PyTorch, Keras, Natural Language Processing, GPT, ETL pipelines, Data Visualizations
Big Data: Map Reduce, Hadoop, Apache Spark, PySpark, NoSQL

Certifications and Courses

Stanford NLP By Prof. Chris Manning (CS224N-Spring'19) **Aug 2023 - Jan 2024**

- Word2Vec, N-Gram, Attention, Transformers, BERT, GloVE, Recurrent Neural Networks and Language Models, Hugging Face, Transformers, Large Language Models (LLMs), Generative Adversarial Networks (GAN)

Oracle Machine Learning using Autonomous Database 2021 Certified Specialist **Jan 2022 - Feb 2022**

- Oracle Machine Learning, Autonomous Database, SQL PL/SQL, OML4PY, OML4SQL, Predictive Modeling, Data Mining, Feature Selection, Model Evaluation, Clustering, Classification, Regression, Time series analysis