ANURAG SANGEM

+1 (551) 220-7098 || ansangem@iu.edu || LinkedIn || Website

EDUCATION

Indiana University, Bloomington. (Aug,2022 – July,2024) Bloomington, US

Master of Science in Data Science

Coursework: Data/Social Media Mining, Statistics and Random Variables, Elements of Artificial Intelligence, Machine Learning.

Vellore Institute of Technology (Jul,2016 – Jun,2020) Chennai, India

Bachelor of Technology in Electronics and Communication Engineering – GPA: 8.72/10

WORK EXPERIENCE

Data Scientist (Aug,2020 – Jul,2022) Bangalore, India

Oracle

- Focused primarily on building propensity to buy models using algorithms like Random Forest, SVM, etc.
- Modelled and scored more than 40 Oracle owned products like ACONEX, NETSUITE etc. on the historical data and updated the 'potential customers to target' to the sales teams this resulted in 1.8x lift in opportunity win rate for top ranked accounts.
- 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.
- Designed dynamic attribution framework using Markov chain algorithm to measure the impact of the marketing campaign on pipeline creation
- Developed three ranking algorithms namely Sequential Ranking, Weighted Ranking and Ranking using generalized linear model (GLM) on the archival data of the Tech Cloud products and presented the top 50 potential customers in each region of the LAD (LAD Territory Account Ranking).
- Leveraged the inbuilt functions of OML4py to improve the efficiency of python scripts by more than 35% (Based on the time taken to train the models on the same data) in OML notebooks.
- 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.
- Performed statistical analysis on the Oracle CX Sales Data which is being advanced to a new tool called Datafox (acquired by Oracle).
- Accelerated the CXD to Datafox matches by 17% by using Jaro-Winkler similarity.

Data Analyst Intern (Jun,2019 – Jul,2019) Chennai, India

Appyhub Technology Solutions

- Created illustrative dashboards using Tableau, SQL, MS Excel, 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 business campaigns, user responses and provided weekly reports with the insights and outcomes.
- Conducted market research for sales and procurement data which resulted in 2% increase in the contracts in 1 month.

PERSONAL PROJECTS

Covid-19 Vaccine - Tweets Sentiment Analysis (Deployed in AWS)

- Worked on the Restaurant Reviews data set from Kaggle to train a supervised model with sentences of positive and negative emotions, used several NLP techniques/libraries such as stop words, PorterStemmer, TfidfVectorizer for data preprocessing.
- Leveraged the open source MLFlow platform to keep a track of my model performance on various parameters and had all the instances logged to mlruns.
- Pushed the docker container image of the model and loaded it to S3 bucket in AWS, used the mlflow. sagemaker module to provide an API to deploy my MLFlow models.
- Using BOTO3 as the AWS SDK the endpoint of my model in Amazon Sage Maker is invoked to make predictions on the recent tweets related to 'Covid-19 Vaccine' which were extracted using tweepy.

Age Detection Using CNN and Handcrafted Features

- Modeled the estimator using traditional Machine Learning algorithms trained on Handcrafted Features as well as Deep Learning Based Convolutional Neural Network (CNN).
- Extracted features such as Local Binary Patterns, Gabor Filter Responses and Landmark Facial Ratios.
- In addition to ML algorithms such as Random Forest, and Gradient Boosted Trees, CNN is used in developing the estimator model which resulted in reducing the MAE by 85%.

PROFESSIONAL CERTIFICATION

- Oracle Machine Learning using Autonomous Database 2021 Certified Specialist

TECHNICAL SKILLS

Programming Languages: Python, C++, R, PL/SQL.

Developer Tools: Oracle machine learning notebooks, Docker, MLflow, Kuberflow, GCP, Kubernetes, AWS Sage maker, Tableau, GIT.

Data science /Model Deployment skills: Statistical Data Analytics, Hypothesis Testing, EDA, Predictive Modelling, Machine Learning, Deep Learning, Natural Language Processing, Neural networks, Flask, YAML, Boto3 and Prometheus.