Uma Sankar Sahoo

Data Scientist

Total Experience : 6+ years

I am a life long learner, fascinated about finding meaning in data using my expertise in artificial intelligence technologies. I am passionate about programming to deliver data science and analytical solutions to business

umasankar641@gmail.com

+91 9304954739 / 8984749723

India

LinkedIn

Github

WORK EXPERIENCE

Data Scientist
Tata Steel Ltd
01/2020 - Present *Jamshedpur*

Data Scientist
Scalers and Victors(S&V)
09/2018 - 12/2019 *Bangalore*

Technology Analyst
Infosys Ltd
05/2014 - 09/2018 *Bhubaneswar*

PROJECTS

Equipment Reliability using predictive maintenance
Client : Tata Steel In-house
Potential savings : 1 billion INR on 50 medium scale critical equipment
Problem Statement : High breakdowns, more downtime, reduced production and less safety in plant operation

- I am Leading this vertical across Tata Steel intending to digitize entire plant under smart plant initiative
- We have developed one patented product known as the "Multi-Model Equipment Reliability System". This solution is backed-up by advanced self-customized algorithms, dynamically programmable configurations, adaptive learning (no-retuning) mechanism giving prescriptive solutions. This is a closed feedback system with SAP integrated messaging services and actions.
- My responsibility includes core architecture design, data science pipeline, establishment, model sustenance, and ensuring continued stable performance on-cloud(GCP)/on-premise.
- We organize monthly workshops for newer opportunities, understand existing problems of business and helping business understand the power of AI is revolutionizing manufacturing

AUTOCATS : AI taxonomy for data quality management and data cataloging
Client : ADNOC (Abu Dhabi National Oil Company)
Potential savings : 674k AED (0.9 Cr INR every month + 0.5 cr. man-hour/month)
Problem Statement : Mismanaged inventory data and man hour effort

- Built a centralized AI for the material management system by creating a smart inventory system.
- With the help of advanced text analytics using NLP, we came up with AUTOCAT - A DQM and Data Cataloging solution that helped the business automate the entire process starting from material creation to inventory management
- A master data dictionary was built to handle categorize the materials to their corresponding noun, modifier, attribute, and values. A ten year's historic data was scanned to build this massive data catalog
- A sequence of NLP models was pipelined for the data classification task i.e TF-IDF vectorization, LDA, guided LDA, topic modeling

iPhone defects Prediction
Client : Apple
Potential savings : Reduced waiting time of 72 hours per customer
Problem Statement : Higher waiting time leading to customer dissatisfaction

- The Project aims to predict a defect before the actual defect during a diagnosis phase of an apple finished product.
- A customer coming to a retail store is validated with its product originality and a detailed check is carried out on the damaged product.
- The service period is usually time-consuming and the customer has to submit his device after taking a backup of data and involves discomfort.
- We offered our clients a solution to detect the possible defects category and severity even before the starting of the actual diagnosis in the backend.
- This early categorization speeds up the notification generation process at GCRM and reduces the actual service time

AppleCare refunds and detecting fraudulent transaction in AppleCare space
Client : Apple
Potential savings : 0.12 Million USD per year
Problem Statement : Unnoticed frauds and problems during GL

- The application aims to calculate refund amounts to be made to customers undergoing a repair or when they decide to discontinue apple's agreement of the AC+ plan.
- We designed a solution to avoid the tiring activities of detecting anomalous transactions using a machine learning model.
- The model makes a prediction and estimates the refund cost that defines a set of contributing parameters. Once this cost is not adjusted by the third party selling insurance to customers it flags this to be unusual behavior.
- This helped reduce the erroneous transactions and saved revenue that went unnoticed while refunding customers.
- My responsibilities include ideation of the problem statement and providing a machine learning-based solution over a traditional threshold-based solution.

TECHNICAL SKILLS

Coding
Machine learning and Deep learning using python
Framework
Scikit-Learn, Tensorflow, Keras, Pyspark,OpenCV,NLTK
Database
Teradata, Oracle,MongoDB Atlas,Google Bigtable
Production
Google cloud platform and on premise deployments
Performance monitoring
Adaptive learning with integrated SAP Z2 feedbacks
Visualisation
Tableau

EDUCATION

Bachelors in Technology(B.Tech)
BPUT, Silicon Institute Of Technology
Year of Passing *2014*
CGPA *9.17*

All India Senior School Certificate Examination(AISSCE)
CBSE, Kendriya Vidyalaya Sangathan
Year of Passing *2010*
Percentage *94.8*

All India Secondary School Certificate Examination(AISSC)
CBSE, Kendriya Vidyalaya Sangathan
Year of Passing *2008*
Percentage *91.4*

ACHIEVEMENTS

- Winner of Data hack summit organized all over India by Analytics Vidya and LinkedIn influencer Tarry Singh at Bangalore
- Winner of Innovista 2020 award organized by Tata Steel
- Awarded as best student award(2010-2014 batch) amongst all technical institute in the eastern zone by Indian Society for Technical Education(ISTE)
- All India CBSE rank 13th in AISSCE
- Two times national champion (Gold-2009 and Bronze-2007) organized by CBSE