DINABANDHU BEHERA

Data Scientist

@ beheradinabandhu50@gmail.com ☎ +91-78964 92396 in linkedin.com/in/dinabandhu50 ⚲ github.com/dinabandhu50

EXPERIENCE

Data Scientist
AIRA MATRIX

📅 Oct 2020 – present 📍 Thane, India

Drug Induces Liver Injury Prediction | POC
- Data has been scrapped from Pubchem website which contains SMILES and their corresponding dili level.
- The task is to predict dili level (0-6) of the drug like molecules.
- Mainly we are applying Graph Neural Network architecture to predict the classes. (Still working on it)

Liver Toxicity Prediction pre-clinical | POC
- The animal study data has been collected from TG-GATES database and NTP website. In total 400 labeled molecules were there in dataset.
- It was a multi-label classification problem, where we have to classify each molecules to 10 different liver lesions. classes.
- We have used XGB, RF, Catboost, ANN, Graph Neural Network for modeling and used Optuna for hyperparameter optimization. Also implemented a web demo using Streamlit for client presentation.

Skin Toxicity Prediction | POC
- The data has been collected from open source websites for this binary classification problem.
- We have used multi-modal approach using features from tabular data, image data and graph data to predict the skin toxicity.
- Also implemented a Flask based web application for the demo of the project.

- -

Machine Learning Engineering
Virtusa, xLabs

📅 Aug 2018 – Aug 2020 📍 Bengaluru, India

Retail recommendation system based on click-stream data | POC
- The goal was to develop and deploy recommendation system for retail website using Google Cloud Platform (GCP) services and user click-stream data.
- We have used open source divolte collector, flask, app engine, cloud bigquery, google cloud dataprep services from GCP to implement the solution.

Determining frequent opiate prescriber | Client IBM Watson
- Our objective was to reduce fatalities due to drug overdose.
- We classified the prescription data based on how frequent an opiate-drug had been prescribed in an year.
- We applied tree-based algorithms such as Random Forest and XGboost for classification and developed a flask API for the same.

Precision Medicine for Cancer | Client IBM Watson
- Developed a binary classifier to classify blood cell cancer types (ALL/ AML), such that appropriate treatment could be provided to the patients.
- The problem belongs to HDLSS problem class, with dimension (rows:38 + 34, cols:7129) ($p >> N$). We used techniques and algorithms such as PCA, Lasso, HSIC Lasso , SVM and XGboost and flask API for interaction.

EDUCATION

IIT Bombay - M.Tech
📅 Aug 2016 – Aug 2018 🎓 Biomedical (Imaging)

IIT Guwahati - B.Tech
📅 July 2012 – May 2016 🎓 Electrical and Electronics

SKILLS

Technologies - `AI/ML` `Cloud` `Health care`

Languages - `Python` `Matlab` `C++ (basic)`

Web - `Web-Scrapping` `Streamlit` `Flask` `Git`

Non-Tech - `Team-work` `Leadership` `Research`

ACHIEVEMENTS / PUBLICATIONS

- Certified as an Associate Google Cloud Engineer.
- Contributed to open-source soft-wares on github e.g. https://github.com/ecrl/padelpy
- " Multi-modal approach for prediction of skin toxicity in small molecules " at EUROTOX 2021, year 2021, Poster presentation
- " Multi-label classification model for the prediction of Mechanism-of-Action of small molecules " at Pharmacology 2021, British Pharmacological Society, year 2021, Oral presentation

OTHERS

Mechanism of Action prediction | Kaggle Competition
- Secured a rank of 707 out of 4373 total participants with a single model prediction.
- It is a multi-class classification problem. We used Pytorch to implement a 3 layered neural network along with clever preprocessing to predict the outputs.

Bristol-Myers Squibb – Molecular Translation | Kaggle Competition
- Problem here is to predict InChI string representation from scanned molecular structure.
- We are approaching this problem as an image captioning problem. And we are still working on this competition.

Vehicle Routing Optimization Problem (IoT) | Nestle
- Used google OR-Tool to find out the optimal vehicle route.
- Developed a web interface to intake warehouse position details and display the optimal route on website.

Real Time vehicle location tracking (IoT) | Nestle
- In this project, I have used NodeRED for real time vehicle location tracking simulation.
- I used open-source "CloudMQTT website" as the pub/sub broker and displayed real-time location, vehicle speed on NodeRED dashboard.