DINABANDHU BEHERA

Data Scientist

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

EXPERIENCE

Data Scientist

AIRA MATRIX

Oct 2020 – present Thane, India

Toxicity Prediction | POC

- Working on a multi-modal approach to predict toxicity of drug molecules.
- In the multi modal approach we are applying deep learning to graph data, image data and tabular data simultaneously to predict the toxicity of a drug molecule.

Mechanism of Action prediction | Kaggle Competition

- Secured a rank of 707 out of 4373 total participant with 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.

Machine Learning Engineering

Virtusa, xLabs

Aug 2018 – Aug 2020 Bengaluru, India

Retail recommendation system based on click-stream data | POC

- Developed and deployed recommendation system for retail website using Google Cloud Platform (GCP) services.
- Collected click-stream data using open source divolte collector software from the website which was developed using flask and bootstrap technology and we deployed the website using App engine to GCP.
- The click-stream data was stored in cloud bigquery, and the data analysis was done using google cloud dataprep service

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.
- Developed a custom sklearn compatible python library for preprocesing and uploaded it to pypi repository using poetry.

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

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

Web - Web-Scrapping | Flask - API | Git

Non-Tech - Team-work | Leadership | Research

Technologies - AI/ML | G.C.P | Linux

ACHIEVEMENTS

- Certified as an Associate Google Cloud Engineer.
- Successfully managed to get an investment worth 15 Million USD from IBM

PROJECTS

Dog Breed Identification

- This dataset comprises 120 breeds of dogs and the goal of the project is to create a classifier capable of determining a dog's breed from a photo
- Developed custom data pipeline for image pre-procesing and augmentation using pytorch and opencv.
- Pre-trained classification models such as VGG16 and Inception v3 were used to classify the dog breed using pytorch.

Extracting Entity from Abstract of Research Papers

- Data collected from online resources using beautifulsoup4 library. This data contained 300+ scientific paper abstracts, out of which 200 were used for labeling process.
- The annotation of data was done using BRAT rapid annotation tool and then a script was developed to convert BRAT standoff format to conll (IOBE) format.
- Implemented a named entity recognition (NER) system to extract author name, year of publication and keywords from the abstract of scientific research paper.

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.