Manish SIHAG

Data Scientist | Natural Language Processing | Al in Drug Discovery

♀ Hyderabad, Telangana, India



Indian Institute of Technology, Bombay

Bachelor of Technology in Chemical Engineering

JULY'14 – JULY'18



PROFESSIONAL EXPERIENCE

Data Scientist, Aganitha Cognitive Solutions, Hyderabad

SEP'19 - CURRENT

Apr'21 - Current | Drug Solubility Prediction

- > Designed a database schema and setup a Postgres instance for storing the data for the project
- > Led a team of 3 members and guided the team with various aspects of the project
- > Trained ML models for predicting the solubility of molecules in different assays and solvents
- > Trained deep learning models to predict quantum descriptors of molecules
- > Developed a webapp to interactively train new models for scientists from non-technical backgrounds
- > The models are currently used by researchers from a \$B Pharma based out of the US

Python PostgreSQL Tensorflow Docker Git

Jul'20 - Apr'21

Disease Hub and Research Browser

🗹 COVID Research Browser 💢 Rare Disease Research Browser 🧭 Cancer Research Browser Developing a near real-time knowledge graph to collate and structure the information about diseases like COVID-19. Also, developing a research browser app that can fetch latest publications related to any topic

- > Wrote ETL scripts for fetching financial data for any pharma company
- > Developed a master data management pipeline to disambiguate author names and their affiliations
- > Setup an ETL pipeline to fetch, process, and store research papers related to any disease
- > Designed a graph-based database schema for storing data in neo4j
- > Built a ReactJS based website to efficiently search millions of research papers in real-time
- > The hub is currently being used by marketing research teams as a competitive intelligence platform Python | Neo4j | GraphQL | ReactJS | JavaScript | HTML | CSS | Docker |

Mar'19 - Dec'19

Suzuki Reaction Yield Prediction

- > Developed python modules for featurizing chemical reactions and molecules
- > Setup an ETL pipeline for performing quantum computation for chemical molecule descriptors
- > Trained multiple variations of Graph Convolutional Neural Network models
- > Helped client in 30% reduction of their lab expenditure by identifying low yielding reactions
- > A research paper for joint publication is underway in collaboration with a \$B Pharma company

Python Postgres Tensorflow Docker RdKit GraphCNN Airflow

Junior Data Scientist, Aganitha Cognitive Solutions, Hyderabad

SEP'18 - SEP'19

Jan'19 - Mar'19

Pharmacokinetic Properties Prediction

- > Learned how a drug process in a body and different factors it depends on
- > Researched and analyzed different open-source datasets and replicated state of the art solutions
- > Trained graph-based deep learning models to identify drugs with higher oral Bioavailability
- > Achieved state-of-the-art results for the trained models

Python | GraphCNN | Tensorflow | RdKit | Docker

Sep'18 - Dec'18

Email Request Analysis

Classifying customer emails into 70+ categories to save resources needed for manual ticket generation

- > Developed a python package for converting Microsoft Outlook's PST files to a dataframe
- > Implemented a pipeline to generate synthetic text representative of the underrepresented classes
- > Trained various NLP models like RNNs and Self-Attention networks to classify incoming emails
- > Improved customer experience by auto-generating ticket numbers for automated tasks and reduced the waiting time from 2 days to 5 minutes

Python Postgres NLP Tensorflow Docker

May'17 – July'17

QA Chatbot

Developing a QA chatbot to answer common and repetitive queries by customers

- > Carried experiments for full sentence generation using RNNs and GRUs given initial words
- > Developed deep learning models to classify customer queries into different categories
- > Helped the company in reducing the number of customer calls by 10% by setting up a QA chatbot

Python NLP Tensorflow Postgres

TECHNICAL SKILLS

Programming Python, SQL, GraphQL, HTML, Javascript, CSS, Bash

Databases Postgres, Neo4j

Libraries Tensorflow, RdKit, ReactJS, Keras, Scikit-learn, pandas, numpy

Others Docker, Linux, Git

Familiar with R, Java, C, C++, Octave, MATLAB

RELEVANT COURSEWORK

Aug'16 – Sep'16	R Programming, Johns Hopkins University, Coursera 🗹 Certificate	
Dec'16 – Jan'17	Algorithms, University of California, San Diego, Coursera 🖸 Certificate	
Jan'17 – Feb'17	The Data Analytics Edge, MITx Courseware, edX	
Mar'17 – Jun'17	Machine Learning, Stanford University, Coursera	
May'17 – Jun'17	Natural Language Processing, Stanford University, YouTube	
Aug'17 – Feb'18	Deep Learning Specialization, deeplearning.ai, Coursera C Certificate	Certificate
	☑ Certificate	