Manish **Sihag**

Data Scientist | Natural Language Processing | AI in Drug Discovery

chemicbook.com linkedin.com/in/manish-sihag github.com/sihagmnis36
+91 9167856885 sihagmanish36@gmail.com

Education

Indian Institute of Technology, Bombay July'14 – July'18
Bachelor of Technology in Chemical Engineering

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

Professional Experience

Data Scientist, Aganitha Cognitive Solutions, Hyderabad Sep'18 - Current

Feb'19 - Dec'19 — **Suzuki Reaction Yield Prediction**
Given a chemical reaction, predicting the amount of product that will form
- Developed python modules for featurizing chemical reactions and molecules
- Setup an ETL pipeline for performing quantum computation for chemical molecule descriptors
- Developed a jupyter notebook based webapp to interactively train new models and make predictions
- The model was able to identify 3 least yielding reactions out of the given 10 paths with 95% confidence
- A research paper for joint publication is underway in collaboration with a $B Pharma company

Python | Postgres | Tensorflow | Docker | RdKit | GraphCNN

Jul'20 - Aug'20 — **Research Browser**
COVID Research Browser *Rare Disease Research Browser* *Cancer Research Browser*
Developing a research browser app that can fetch latest publications related to any topic like COVID-19
- Developed modules for disambiguating authors and organizations across research papers
- Experimented with different databases like Postgres and neo4j as a backend
- 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

Python | Neo4j | GraphQL | ReactJS | JavaScript | HTML | CSS | Docker

Sep'18 - Dec'18 — **Email Request Analysis**
Classifying customer emails into 70+ categories to save time and manpower needed for manual work
- Developed a python package for converting Microsoft Outlook's PST files to a dataframe
- Wrote a Regex based Python module to identify dates from a free-text email
- 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

Python | Postgres | NLP | Tensorflow | Docker

Feb'19 - Mar'19 — **Pharmacokinetic Properties Prediction**
Predicting pharmacokinetic properties like Bioavailability for drug-like molecules
- Learned how a drug process in a body and different factors it depends on
- Researched and analyzed different datasets available openly and replicated state of the art solutions as baseline models
- Trained graph-based deep learning models to identify drugs with higher oral Bioavailability
- Achieved an overall RMSE of 0.18 and R^2 value of 0.58 which was a 3% improvement over then State-of-the-art solution

Python | GraphCNN | Tensorflow | RdKit | Docker

Machine Learning Intern, Eli Research India, Faridabad MAY'17 – JULY'17

May'17 – July'17 | **Medical Query Chatbot**

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

- Trained deep learning models to generate full sentences given initial words
- Setup a Postgres database for storing text data
- Developed deep learning models to classify a customer query into various categories

Python | NLP | Tensorflow | Postgres

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 Certificate Certificate Certificate