# Curriculum Vitae


## Manish Kumar

Email : sihagmanish36@gmail.com
github.com/sihagmnis36 Mobile: +91 9167856885


## Education

**Indian Institute of Technology, Bombay** **[July'14 – July'18]**

*Bachelor of Technology in Chemical Engineering; CPI: 6.33/10*

## Professional Experience

**Data Scientist**, *Aganitha Cognitive Solutions, Hyderabad* **[Sep'18 – Present]**

**Research Browser** *based on Neo4j and ReactJS*

- **Usecase:** With millions research publications already published and thousands being added every day, managing and staying up to date with the latest research is very crucial for lab scientists, even more so in the times such as the coronavirus pandemic
- **Research:** Tried different databases including postgres and neo4j as a backend; Solved the challenge of a free text search in real-time; Researched about methods for author and affiliation name disambiguation and mapping them to right publications
- **Results:** Developed a database pipeline to efficiently store ~250 million publications; Developed topical research hubs for COVID-19, Alzheimer's Disease and Cancer to fetch relevant and up-to-date publications

**Pharmacokinetic Properties Prediction**

- **Usecase:** In drug discovery, beforehand knowledge of pharmacokinetic properties like Bioavailability and Metabolism rate can save a significant amount of time and money resources for researchers
- **Research:** Learned how a drug process in a body and different factors it depends on; Researched, processed, and analyzed different datasets available on the internet; Replicated then state of the art solutions for building baseline models
- **Models:** Developed different variations of Graph Convolutional Networks like Gated-Attention Networks
- **Results:** The model was able to identify drugs with higher oral Bioavailability with 90% accuracy; Achieved an overall RMSE of 0.18 and $R^2$ value of 0.58 which became the next **State-of-the-art** solution

**Service Request Analysis** *using Natural Language Processing*

- **Usecase:** Within a large company, redirecting thousands of emails from customers to respective divisions requires significant time and manpower
- **Research:** Developed tools to clean the data with significant noise; Implemented a pipeline to generate synthetic text representative of the ground truth classes
- **Models:** Trained the state-of-the-art models like RNNs, LSTMs, GRUs and Self-Attention networks
- **Results:** Automation of this process saved hundreds of work-hours for the client weekly

**Chem Data Lake Solution**

- **Usecase:** Chemical and medical experimental data storage pipelines often struggle with uniformity and standardization across different teams and labs. It can become increasingly tough to replicate lab experiments and keeping track of the results
- **Research:** Explored different pipelines used by pharma companies and developed a standardized data model

- **Results:** Developed a graph-based data storage pipeline and an integrated data model to store the experiment results in an easy to retrieve way; Developed a webapp and a website for the ease of lab scientists from non-technical backgrounds

---

## Technical Skills

| | |
|---|---|
| **Programming languages** | Python, R, SQL, JavaScript, HTML, CSS, GraphQL, Bash |
| **Familiar with** | Java, C, C++, Octave, MATLAB |
| **Tools & OS** | Postgres, Neo4j, Docker, Linux, Git, AWS, VMs |
| **Libraries** | ReactJS, Tensorflow, RdKit, Sklearn, Pandas, NumPy |

---

## Internship

**Machine Learning Intern**, *Eli Research India, Faridabad* **[May'17 – July'17]**

**Chatbot** *using Natural Language Processing*

- **Usecase:** To any customer-oriented company, a large number of queries from customers are often repetitive and can be answered by an automatic chatbot machine
- **Solution:** Built a Chatbot to answer common queries related to the healthcare domain; Built tools to improve medical text search algorithms; Added support for context-sensitive search by leveraging Natural Language Processing techniques

---

## Relevant Courses

- Introduction to C++, *IIT Bombay, On Campus, Aug'14 – Nov'14*
- Introduction to Computer Science, *Harvard, edX, Jul'16 – Nov'16*
- R Programming, *Johns Hopkins University, Coursera, Aug'16 – Sep'16* [Certificate]
- Algorithms, *University of California, San Diego, Coursera, Dec'16 – Jan'17* [Certificate]
- Differential Equations I & II, *IIT Bombay, On Campus, Aug'15 – Nov'15*
- Linear Algebra, *IIT Bombay, On Campus, Jan'15 – Mar'15*
- Calculus, *IIT Bombay, On Campus, Aug'14 – Nov'14*
- The Data Analytics Edge, *MITx Courseware, edX, Jan'17 – Feb'17*
- Machine Learning, *Stanford University, Coursera, Mar'17 – Jun'17*
- Natural Language Processing, *Stanford University, YouTube, May'17 – Jun'17*
- Deep Learning Specialization, *deeplearning.ai, Coursera, Aug'17 – Feb'18* [Certificate][Certificate][Certificate]

---

## Extra-Curricular Activities

- Volunteered for an education camp for village school children organized by Group for Rural Activities(GRA)
- Educated underprivileged children about potential opportunities in their future as part of the camp
- Key event organizer in Asia's largest college cultural festival Mood Indigo
- Organized and executed five cultural competitions for over 500 participants from all over the world
- Won a Gold Medal in Kho-Kho College General Championship
- Awardee for building best Remote-Controlled model aeroplane as part of Aeromodelling club, IIT Bombay