Manish Kumar

chemicbook.com | mnis@uchicago.edu | +1 (872) 904-5284 | 5339-1 S. Kimbark Ave, Chicago, IL 60615

Summary

A data science professional with 4 years of experience in developing AI solutions in the Drug Discovery domain. Experienced in collaborating with multi-billion dollar clients in a high-paced environment and delivering high-quality end-to-end solutions for Machine Learning projects.

Education

The University of Chicago Sep '22 – Dec '23

Master of Science in Analytics

Indian Institute of Technology, Bombay Jul '14 – Jul '18

Bachelor of Technology in Chemical Engineering

Experience

Full Stack Data Scientist, *Aganitha, Hyderabad, India* Sep '20 – Jun '22

- Trained and deployed Graph Neural Network models to predict quantum descriptors of molecules and identify high-yield chemical reactions, resulting in reduced cost and experimentation time for scientists | RMSE 0.21
- Collaborated with a leading global pharmaceutical company to research and predict the solubility of molecules using machine learning; Work to be published in a top journal soon | RMSE 0.33
- Created a ReactJS-based web application with neo4j as the backend, adding the efficient and real-time search of millions of research papers related to any topic, to enable a faster intelligence solution for R&D scientists
- Managed and mentored a 5-member team by writing tutorial blog posts, designing quizzes, and collating study materials for both technical and domain scientist roles, shortening the onboarding time from 12 to 4 weeks

Jr Data Scientist, *Aganitha, Hyderabad, India* Sep '18 – Sep '20

- Spearheaded the company's transition to AI in the Drug Discovery domain by researching and implementing tools needed, training team members, and improving overall efficiency
- Improved client customer service by automating the classification of incoming customer emails using self-attention neural networks, reducing response time from 2 days to 5 minutes
- Simulated and modeled a Biochemical plant using polynomial function forecasting and encoder-decoder neural networks to help optimize multivariate time series process control parameters
- Built an Airflow-based framework to visualize and monitor the ETL pipeline with 30+ docker containers, resulting in a streamlined process and optimized pipeline management

Machine Learning Intern, *Global Growth IOC, Faridabad, India* May '17 – Jul '17

- Added a word-prompt feature for users while typing, resulting in better user experience and productivity
- Trained initial models for a QA chatbot that led to a 10% reduction in the number of customer calls down the line

Projects

COVID-19 Literature Analysis, *MSCA31013 Big Data Platforms* Oct '22 – Dec '22

- Leveraged Hadoop and PySpark to manage and process 70Gb+ text data
- Performed topic modeling using LDA to cluster 1M+ research papers, resulting in better insights discovery

Skills

Experienced: Python, R, SQL, Docker, Tensorflow, PySpark, Hadoop, Neo4j, Airflow, JavaScript, Linux, Git, Bash, Keras, RdKit, DeepChem, Data Science, Machine Learning, Deep Learning, NLP, RNN, LSTM

Familiar with: ReactJS, Kubernetes, MLFlow, Hive, GraphQL, NWChem