RAHUL AVADHOOT

DATA SCIENCE & CHEMICAL ENGINEERING

CONTACT

rahul.avadhoot@gmail.com
(206) 954-7943
Los Angeles, CA
linkedin.com/in/rahulavadhoot
github.com/rahulavadhoot

PROFILE

Data Scientist with 2+ years of experience building large-scale predictive-modeling and analytics solutions to reduce the cost and time for developing new drugs with a background in chemical engineering.

EDUCATION

2018
UNIVERSITY OF WASHINGTON [SEATTLE, WA]
MS in Chemical Engineering
Specialization in Data Science

2015
NATIONAL INSTITUTE OF TECH. [NAGPUR, INDIA]
BTech in Chemical Engineering

MANAGEMENT SKILLS

- Agile Software Development
- Project and Team Management
- Communication and Presentation

TECHNICAL SKILLS

- **Programming**: Python (NumPy, SciPy, Pandas, Scikit-Learn, Keras, Matplotlib, Seaborn), R, SQL
- **Databases**: SQLite, MongoDB
- **Machine Learning**: Regression, Classification, Clustering, Neural Networks (MLP, CNN, RNN)
- **Statistics**: Distributions, Hypothesis Testing, Statistical Inference
- **Software & Tools**: Linux, Git, Jupyter, Docker, Tableau, AWS, Azure

EXPERIENCE

2019 - PRESENT
Data Science Resident | Qulab Inc.

Leadership

- **Driving strategy and vision** for cheminformatics applications by translating research, and available data into innovative solutions for stakeholders.
- Actively collaborating with Chemistry, and AI teams to **develop products** that measurably and efficiently improve the drug discovery and design process.

Projects

- Trained state-of-the-art models using **Random Forests, Support Vector Machines, and Graph Convolutional Networks** to predict ADMET molecular properties like bioavailability, hERG inhibition, CYP450 inhibition, BBB permeability, etc. which outperformed existing commercially available models by **20%**.
- Developed highly scalable **drug discovery and patent databases (20,000,000 molecules)** on the **MongoDB API for Azure Cosmos DB**; Delivered a **Command Line Interface** to upload, download, and **query molecular data** for the Chemistry team.
- Created a **Chemistry Bot for Slack** that looks up compound names or SMILES and renders it along with its molecular properties. The bot runs on an **RDKit-based API** on **Heroku**.

2016–2018
Graduate Student Researcher | University of Washington

- Developed **PyMolSAR**, a Python package to compute 760 2D molecular descriptors and train the most appropriate **machine learning** model to **predict molecular properties**.
- Trained machine learning models on a highly imbalanced dataset of **300,000 molecules** to predict USP1 inhibition with a best F1-score of **0.25**.

PROJECTS

E-Commerce Product Classifier

- Developed an ensemble model for classifying **60,000 e-commerce products** into **36 categories** using a **bag-of-words** model for product descriptions and a **pre-trained VGG-16** model for product images. Achieved an overall accuracy of **60%**.

Current State of Nuclear Energy Research

- Delivered web-scraping, and a **Tableau dashboard** to **track collaborations in nuclear energy research** from 2000-2016 by web-scraping meta-data of **30,000 research papers** from Web of Science using Selenium.