Contact

rjvora98@gmail.com

www.linkedin.com/in/rijul-
vohra-367764126 (LinkedIn)
rijulvohra.github.io/rijulvohra/
(Portfolio)
github.com/rijulvohra (Other)

Top Skills

Apache Spark

PySpark

Scala

Certifications

Neo4j Fundamentals

pandas Foundations

LangChain for LLM Application
Development

Python for Data Science and
Machine Learning Bootcamp

Rijul Vohra

Knowledge Graphs @ Jefferies | MS Data Science @ USC
New York City Metropolitan Area

Summary

Skilled working on Linked Data, Machine Learning, building large
scale spark pipelines with industrial and academic experience.
Experienced working on Linked Data / Data Science domain for the
Financial and Pharmaceutical industry.

I currently work at Jefferies, building and maintaining a Knowledge
Graph for the bankers to achieve their business goals.

I graduated from University of Southern California with a Masters in
Applied Data Science. My interests range from Machine Learning,
Natural Language Processing, and creating a Knowledge Base
through Information Retrieval, Entity Resolution and Ontology
Mapping. I am interested in working towards making data linked and
accessible using Knowledge Graphs.

Previously, I have worked as a Research Assistant at USC
Information Sciences Institute at the Center on Knowledge Graphs
where I worked on building an entity linking system that links tables
to a Knowledge Base. I also have worked on building a linked data
source on Drugs using RxNorm, Wikidata, etc. Previously, I have
also worked as a Data Science intern at Novartis working towards
FAIRification of Biomedical data using Knowledge Graphs.

Webpage: http://rijulvohra.github.io/rijulvohra/
Source Code for some my very interesting work can be found at:
https://github.com/rijulvohra

Skills

Languages and Databases: Python, C++, SQL, MATLAB, Firebase,
MongoDB, MySQL, RDF, Blazegraph, Neo4j, SPARQL
Data Science Skills: Supervised Learning, Unsupervised Learning,
Natural Language Processing, Embeddings, Seq2Seq models,
Knowledge Graphs, Linked Data, Semantic Web

Significance Testing: Hypothesis Testing, A/B Testing

Experience

Jefferies
Knowledge Graph | Assistant Vice President(Data Science & Analytics)
May 2021 - Present (4 years 9 months)
New York City Metropolitan Area

Information Sciences Institute
Graduate Research Assistant
August 2020 - May 2021 (10 months)
Los Angeles, California, United States

◦ Table-Linker: Entity Linking System for linking tables to Wikidata
◦ Developed algorithm for candidate generation for entities in a table.
Achieved a recall of 0.986
◦ Trained and tuned Pairwise Contrastive Loss Neural Network for Candidate
Ranking
◦ Achieved top 1 score of 0.902 on T2DV2 tables and 0.87 on Semtab Round
4 tables

◦ Harmonize: Python, KGTK, Wikidata dump, RxNorm data dump, Blazegraph,
ElasticSearch
◦ Created a linked data source for drugs with Wikidata and RxNorm
◦ Developed a pipeline to generate triples and load them to Blazegraph
◦ Indexed 92 million data items from wikidata, wikipedia using ElasticSearch
◦ Developing algorithm for candidate generation for entities in a table by
linking those entities to wikidata
◦ Contributing to open source projects: KGTK(https://github.com/usc-isi-i2/
kgtk) and table-linker(https://github.com/usc-isi-i2/table-linker)

Novartis
Data Strategy Intern
June 2020 - August 2020 (3 months)
East Hanover, New Jersey, United States

◦ FAIRification of Data: Python, Wikidata, SPARQL, RDF, AWS EC2, S3, Git,
Docker, Jenkins
◦ Cleaned dirty transaction data by developing Linked Master Data
Management using Wikibase infrastructure
◦ Linked bioportal entities to Qnodes in Wikidata, overall precision 85%

◦ Streamlined reconciliation process for data curators by integrating entity linking webservice with OpenRefine UI
◦ Devised pipeline to validate shape of new entities added to satellite Wikibase allowing manual curation

University of Southern California - Marshall School of Business
Graduate Research Assistant
November 2019 - May 2020 (7 months)
Los Angeles, California, United States

◦ Business Open Knowledge Network: Python, Snorkel, Scrapy, Spacy, BeautifulSoup, Spacy, FLAIR NLP
◦ Developed broad crawler to crawl 100,000 company's webpage extracting Mergers and Acquisitions
◦ Achieved recall of 69% on extracting names of target and acquirer companies
◦ Extracted customer relations from Capital IQ database for 20 years(2000 - 2020) using Snorkel with an F1 score of 68.1%

Moscow Institute of Physics and Technology (State University) - MIPT, Phystech
Intern
December 2018 - May 2019 (6 months)
Moscow, Russian Federation

Worked with time series data to predict the wheat yield in 5 states of the US. Used statistical techniques to generate new features. Analyzed the importance of different features and compared the performance of Random Forest Regressor, Gradient Boosting Machine, Neural Network.

———

Education

University of Southern California
Master's degree, Applied Data Science · (2019 - 2021)

Thapar Institute of Engineering and Technology
Bachelor of Technology - BTech, Electronic and Communications Engineering Technology · (2015 - 2019)

Sanskriti School, Chanakyapuri, New Delhi
Higher Secondary, Science · (2013 - 2015)