Rohan Singh

Linkedin: https://www.linkedin.com/in/singhrhn
Github: https://github.com/ygivenx

Email : singhrohan@outlook.com
Mobile : +1-312-885-0056

Skills Summary

- **Languages**: Python, R, SQL, Java, Scala, Go, C++
- **Databases**: Postgres, Redcap, DynamoDB, ElasticSearch, MySQL, MS-SQL, Redis
- **Cloud**: AWS, GCP, Azure
- **Tools & Skills**: Pytorch, DVC, MLFlow, Data Science, Machine Learning, Analytics, Data Viz, Deep Learning, Statistics, Transformers, NLP, Data Mining, Terraform, Spacy, Spark, Tableau, Keras, GIT, MS-Excel

Experience

- **Memorial Sloan Kettering Cancer Center** — New York City
  Data Scientist III — *April 2021 - Present*
  - **Predict Time-to-event of Venous Thromboembolism** : Conceptualized and developed a deep learning model to predict the risk for venous thromboembolism in patients who had MSK IMPACT sequencing panel performed on solid tumors from 2014-2019 with a **c-index of .72** for the internal cohort and **.69** for the external cohort. (**Provisional Patent Issued**)
  - **Predicting Impeding Fracture from Radiology Report and CT Scans**: Created a data pipeline to curate a cohort of patients(n= 3000) with metastasis in the pelvic region using NLP and developed models based on clinical and textual features to predict the risk of an impending fracture. Further work is being performed on this cohort by adding CT images to improve the model metrics using multi-modal analysis.
  - **Predicting Airway Obstruction using Radiology Reports**: Developed a front end to curate data and labels for prospective patients who require intervention for airway obstruction to be able to do such predictions automatically.
  - **Programmatic Data Labelling**: Led the initiative for programmatic labeling of clinical data in collaboration with snorkel.ai. Enabled several data teams to leverage weak supervision to label data and improve AI/ML workflows
- **IBM** — San Francisco, CA
  Senior Data Scientist & Managing Consultant — *July 2018 - April 2021*
  - **Covid Recovery Index**: Lead a team of 4 to develop a COVID impact planning strategy for one of the top automobile insurers in the US by predicting and explaining the impact of COVID on driving at a county level using ping data from cellphones and other socio-economic and epidemiological variables. **The results helped the executive team to issue a refund on auto insurance policies during COVID.**
  - **Advanced Automation Platform**: Conceptualized an AWS Serverless architecture to automate claims processing for one of the top insurers in the United States. Lead and worked with a team of developers to implement and roll out to production within 4 months. Improved straight-through processing by **30%**.
  - **Legal Billing Anomaly Detection**: Developed a proof of concept and later helped implement a full product for a Fortune 10 company to help them find anomalies in legal billing data using NLP and multi-class classification. Helped the client save over **20%** in legal bills and reduce the review times by **90%** per invoice for a particular subset of legal billing category.
- **Carnegie Mellon University** — Pittsburgh, PA
  Research/Teaching Assistant - Alessandro Acquisti/Janusz Szczypula — *May 2017 - May 2018*
  - **Behavioral Economics & DBMS**: As a Behavioral Economics RA, developed a data gathering tool to analyze the effects of advertisement and their placements in Google search results - organic vs sponsored. As a DBMS TA, conducted labs and graded assignments for Database Management Systems
- **Citadel LLC.** — Chicago, IL/ London, GB
  Software Engineer — *Feb 2012 - Apr 2017*
  - **Data Quality Monitoring System**: Developed a flexible monitoring system to check the quality of external data from vendors using state-of-art statistical data analysis, resulting in 200% decrease in turn-around time.
  - **Commodity Trading Modelling**: Architected a front-end interface, enabling the trading team to run models on-demand and make quick decisions. Revamped the legacy MS-Excel analysis system to use python along with modern scientific packages, reducing the model development and testing time exponentially.
  - **Future Pricing Model**: Implemented a Future Value model to generate trading signals for soft commodity futures.

Education

- **Carnegie Mellon University** — Pittsburgh, PA
  Master of Information Systems Management; GPA: 3.87 (Highest Distinction, Dean's List) — *May - 2018*
  Courses: Machine Learning, Distributed Systems, Database Management, NoSQL, Statistics, Data Structures, Optimization, Economic Analysis, Practical Data Science, Managing Disruptive Technologies, Data Mining
- **Birla Institute of Technology** — Mesra, India
  Bachelor of Computer Science; GPA: 3.5 (7.70/10.0 - First Class with Distinction) — *June - 2011*