

Zeeshan Hooda

DATA SCIENTIST

📞 +1 (587) 568-6239 | ✉️ zhooda@protonmail.com | 🌐 zhooda | in zeeshanhooda

Education

University of California, Berkeley

DOCTOR OF PHILOSOPHY (PHD) IN COMPUTER SCIENCE

Berkeley, USA

May 2022 - June 2026 (Expected)

- **Specialization** — Learning and Probabilistic Inference
- Combined four-year MS/PhD program with one dissertation

University of California, Berkeley

BACHELOR OF SCIENCE (BSc) IN COMPUTER SCIENCE, GPA: 3.98/4.0

Berkeley, USA

September 2019 - April 2022

- **Specialization** — Data Science & Statistics

Skills

Languages Python, Java, JavaScript, HTML, CSS, Go, Rust, C, C++, Assembly (x86, ARM)

Frameworks AWS, GCP, Docker, Flask, FastAPI, GraphQL, Flutter, Svelte

Data Science SQL (Postgres, MySQL), noSQL (MongoDB, CouchDB), Data Warehousing (BigQuery, ETLs), Apache Spark (PySpark, Spark SQL), PrestoDB, Analytics (Superset, Plotly, Tableau)

Experience

Smart Access

Edmonton, Canada

LEAD DATA SCIENTIST

April 2021 - Present

- Orchestrated and carried out a production-scale migration of event-driven analytics schemas to ensure minimal uncertainty in analytics product. This migration reduced the total dataset uncertainty from 15.6% to 1.9% and ensured the company's ability to make accurate data-based decisions
- Implemented version control and environment isolation (dev, SIT, UAT, prod) to increase robustness of ETL code development
- Modeled and transformed raw user event data to infer the usage of certain features for use in business intelligence applications
- Developed a normalized DCG metric to measure relevance of search results
- Revamped the search engine with lemmatization and semantic search, resulting in a 70% increase in relevant results and a 30% increase in search usage

SensorUp

Calgary, Canada

DATA SCIENCE INTERN

May 2020 - September 2020

- Developed an Airflow pipeline to combine fluid transportation facility and vehicle data from over 20 private and public datasets to be used in a fluid hauling extension to the SensorUp platform
- Created a production ready visual recognition service to identify hazards on remote sites with limited internet accessibility, which is now deployed on multiple well sites providing hazard recognition to large producers. Used Nvidia edge devices, DeepStream, OpenCV, Python, & RabbitMQ

Redman Technologies

Edmonton, Canada

SOFTWARE ENGINEERING INTERN

May 2017 - September 2018

- Created a tool used by Realtors across Canada to visualize neighborhood statistics (education, crime, etc.) gathered from public and private data sources. Used CouchDB, Express.js, Node.js, & Mapbox
- Collaborated on all stages of the development lifecycle, from requirements gathering to production releases

Projects

Reddit Sentiment Analysis

DATA SCIENCE

- A sentiment analysis project on 200+ million Reddit comments from popular subreddits using PySpark, PyTorch & pandas