

# Harpal Dhillon

Raritan, NJ • 908-523-8035

Email: [hd227@scarletmail.rutgers.edu](mailto:hd227@scarletmail.rutgers.edu)

Github: <https://github.com/hpdhillon>

Linkedin: <http://bit.ly/harpallinkedin>

## PROFESSIONAL EXPERIENCE

### Data Scientist, Komodo Health

Nov 2021 – Feb 2024

- Partnered with client to scope analysis, defined metrics (MCC) and analyzed patient journey data (**Python, SQL**) to ID critical points affecting patient journey, leading to new revenue & retention of \$6M client.
- Consolidated patient cost data, conducted exploratory data analysis (**Python**) to assess data quality and sparsity, documented critical data context and built missing data imputation to create model ready dataset.
- Developed ML model (**Python, GBM**) to predict patient costs from incomplete patient records, model demonstrated 50% MAE improvement over baseline, launched as new product for customers (like Aetna).
- Analyzed **GBM** model outputs to highlight feature impact to patient cost, with the goal of enabling cost reductions and competitive healthcare plan price adjustments, improving profit.
- Created proposal for new mathematical representation of patient journeys and applying **LLMs** to improve patient segmentation, used to accelerate clinical trial candidate sourcing and speed up drug development.
- Built new reporting product (**Pyspark, SQL**) by creating ingestion pipelines and implementing statistical imputation, enabling customers insights into cost and utilization metrics based on user-defined filters.
- Consulted with customers to determine needs and translate to technical requirements and delivery plan.

### Decision Science Analyst, Axtria

Feb 2021 – Nov 2021

- Developed negative exponential omni-channel marketing mix model in R to generate optimal spending insights and recommendations for large pharma prostate cancer product (\$10,000,000+ spending budget).
- Implemented clustering algorithms (optimized kmeans++) to improve customer segmentation, created best practices that were adopted across teams and generated 9.6% greater model accuracy.
- Helped an exogenous team with an ad-hoc geospatial project by co-authoring *Driva*, thereby reducing processing time by >90% and MAPS API costs by >\$3000 for that project and each subsequent project.

### Data Science Intern, Schneider Electric/ASCO

Jun 2019 - Aug 2019

- Discovered opportunities for supply chain enhancements with data visualizations, created data asset (SQL Server) and dashboards (Tableau) for critical part tracking, reduced part sourcing time by >90%.
- Analyzed data and utilized DBSCAN clustering (Python) to find optimal locations for new technician hires

## PROJECTS

### OPTOAI - [LINK](#)

Oct 2020 - Jan 2021

*OptoAI is a journaling app that detects how optimistic/pessimistic/neutral a user is feeling and makes recommendations.*

- Gathered and curated training data (**Prolific**), engineered NLP features (Sentence-BERT, Cosine Similarity) and achieved a 91% LOOCV accuracy with a GBM.
- Developed and deployed web application (Streamlit) as user interface for journaling insights.

### Co-Founder – Rutgers Data Science Bootcamp - [LINK](#)

Aug 2019 - Aug 2020

- Co-founded free bootcamp, taught 45+ undergrads data science from the ground up.

## EDUCATION

### Rutgers University – B.A in Statistics (Honors Program)

May 2020

**Major:** Statistics, **Minors:** Mathematics, Computer Science, Economics, **GPA:** 3.4 / 4.0

Economics Honor Society | Vice President Rutgers Statistics Club | Honors Program Peer Mentor

## SKILLS

**Languages:** Python (Pandas, Scipy, Seaborn, StatsModels, Huggingface, scikit-learn), PySpark, R (Forecast, devtools, distGeo, Driva [author] ), SQL (MS SQL Server, MySQL, Snowflake)

**Technical:** Tableau, Dataiku, Airflow, Streamlit, Heroku, AWS

**Certifications:** Intro to Healthcare (Stanford School of Medicine), Dataiku Core Designer + Developer