Utkarsh Singh

647-482-1523 | u25singh@uwaterloo.ca | LinkedIn | GitHub

EDUCATION

University of Waterloo **Waterloo, ON**
M.S. in Management Science *Sep 2025*

- **Relevant Coursework:** Operations Research, Advanced Machine Learning, Statistics for Data Analytics, Text Analytics

Purdue University **West Lafayette, USA**
B.S. in Business Analytics *Dec 2023*

- **Academic Honors:** Dean's List (2021)
- **Relevant Coursework:** Financial Data Modeling, Database Management, Econometrics, Data Mining, Data Visualization

SKILLS

Programming Languages: Python, SQL, R

Machine Learning Frameworks: Scikit-learn, TensorFlow, PyTorch, XGBoost, Natural Language Processing (NLP), Neural Networks (RNNs, CNNs, Transformers)

Data Analysis Libraries: Pandas, NumPy, SciPy, Statsmodels, BeautifulSoup

Data Visualization: Power BI, Tableau, Excel, Matplotlib, Seaborn

Database & ETL: MySQL, PostgreSQL, MongoDB

Concepts: ETL/ELT Pipelines, Data Modeling, Big Data Processing, Statistical Analysis, Text Analytics, Sequence-to-Sequence Models, Prompt Engineering

Agile Tools: Atlassian Jira, Trello

Collaboration Tools: Slack, Microsoft Teams

Soft Skills: Analytical thinking, problem-solving, communication (technical/non-technical), teamwork, independent work

RELEVANT WORK EXPERIENCE

Extern, Amazon Fulfillment Center **Remote**
Operations Strategy & People Analytics Extern *June 2025 – Aug 2025*

- Analyzing unstructured employee feedback to identify workforce attrition risks, productivity blockers, and key operational pain points within Amazon Fulfillment Center
- Creating segmented cohort profiles using qualitative data analysis and thematic coding to surface role specific challenges and recommend targeted interventions.
- Developing business ready briefing notes and strategy decks synthesizing frontline insights and actionable operational recommendations.

Purdue University, Data Mine Learning Community **West Lafayette, USA**
Research Analyst Intern *Aug 2021 – Dec 2021*

- Conducted bias analysis on Chicago's public health ML model, processing 20,000 daily Yelp API data rows to uncover demographic disparities in inspection patterns.
- Improved inspection coverage by 25% in underserved areas through statistical insights, demonstrating regulatory impact assessment skills.
- Collaborated with peers to present findings, enhancing team communication and stakeholder engagement.

PROJECTS

Parkinson's Disease Progression Prediction **Jan 2025 – Mar 2025**

- Developed regression and classification models using Python, Scikit-learn, and TensorFlow to predict UPDRS scores for Parkinson's disease severity on a Kaggle dataset (4,838 entries).

- Implemented Random Forest Classifier for severity categorization, addressing class imbalance and identifying medication status as a key predictor (feature importance: 0.352).
- Designed a Feed Forward Neural Network with Batch Normalization, improving R^2 from 0.056 to 0.166 and reducing MAE from 5.16 to 4.64, enhancing clinical prediction accuracy.

Analyzing Substance Use Pattern Among Canadian Youth **Jan 2025 – April 2025**

- Developed 5 predictive models (Bayesian, Lasso, Multilevel, PSM, Random Forest) using Python, boosting substance use prediction accuracy up to 0.79 AUC to inform youth health strategies.
- Analyzed 62,850-student dataset with pandas/seaborn, revealing bullying-smoking link (chi-square $p < 0.001$) to guide prevention programs
- Validated models with cross-validation/LOO, uncovering social influence as key driver, enhancing targeted intervention planning.

Crop Yield Prediction **May 2024 – July 2024**

- Developed a crop yield prediction model using Random Forest and LightGBM algorithms, achieving R^2 scores above 90% for both, demonstrating high predictive accuracy on a Kaggle dataset from World Bank data.
- Conducted comprehensive exploratory data analysis (EDA) on a dataset of 28,242 entries, identifying key patterns in crop yield factors such as rainfall, pesticide usage, and temperature, enhancing data-driven insights for agricultural optimization.
- Implemented data preprocessing and feature engineering, including label encoding and standard scaling, resulting in a 30% reduction in model training time while maintaining robust performance metrics (MAE, MSE, R^2) for crop yield predictions.

NGO Funding Data Analysis **May 2024 – July 2024**

- Developed a K-means clustering model with PCA on 167 countries' socioeconomic data, identifying high- and low-priority clusters to guide NGO funding, improving allocation efficiency by 40%.
- Performed EDA on 9 features using Pandas and NumPy, preprocessing 167 records to enable accurate clustering, reducing data preparation time by 25%.
- Visualized high- and low-priority clusters with a Matplotlib choropleth map for 167 countries, accelerating NGO decision-making by 50% for targeted interventions.

Healthcare Data Mining **Aug 2022 – Dec 2022**

- Achieved 97.59% accuracy in predicting health plan purchases using the K-Nearest Neighbors (K-NN) model, outperforming the baseline model.
- Discovered critical insights through logistic regression, finding that education (+22.54%) and marital status (+89.43% for single individuals) significantly impact purchase likelihood.
- Implemented decision tree modeling, revealing that females with income >$22,000 and unmarriedmales with higher education are ideal target customers.

Pfizer Database Analysis **Aug 2022 – Dec 2022**

- Identified Hepatitis Aas the most-ordered vaccine with 23.6 million units, helping optimize production and supply chain planning.
- Uncovered a 3.36x variation inmaterial costs across vaccines, highlighting opportunities for reducing manufacturing expenses.
- Optimized vaccine distribution by analyzing manufacturing order trends, revealing peak production periods and potential supply chain bottlenecks.