Chieh-An Chang

North York, ON, Canada | chiehan.chang.job@gmail.com
LinkedIn | GitHub | Personal Website

SUMMARY

Incoming Master of Data Science & AI student (University of Waterloo, Vector Institute–recognized) with a B.Sc. in Computer Science & Statistics (University of Toronto, CGPA 3.81). Skilled in Python, SQL, Spark, AWS, and ML. Achieved a top-6 Kaggle ranking using CatBoost. AWS and Microsoft Power BI certified; seeking a **Data Science Intern starting May 2026.**

EDUCATION

University of Waterloo ON, Canada | April 2027 (Expected)
*Master of **Data Science and Artificial Intelligence** (Co-op)*

- Recognized by **the Vector Institute** as an AI master's program.

University of Toronto ON, Canada | June 2025
*Bachelor of Science, Double Major in **Computer Science and Statistics**, with **High Distinction (CGPA 3.81**/4.00)*

- **Awards**: **Ranked 6th** in a Kaggle competition; **Dean's List Scholar** (2021, 2022, 2024); **Entrance Award** recipient.

PROJECTS

Travel Prediction Model Development | Python, Scikit-learn, Pandas, NumPy ON, Canada | January 2024 - April 2024

- Cleaned 1,000+ records from a raw CSV dataset and split them into **68% training, 12% validation, and 20% test sets** using **Pandas** and **NumPy** for efficient data wrangling.
- Created 6+ **boxplots** to visualize key variables, identifying and **replacing outliers with mean** values to maintain data consistency.
- Implemented 4 supervised machine learning algorithms—**Random Forest, K-Nearest Neighbors, Multinomial Naive Bayes, and Logistic Regression**—**using Scikit-learn** to efficiently explore potential models.
- Optimized Random Forest hyperparameters (e.g., number of estimators, maximum tree depth) via **grid search** on the validation set to enhance model validation accuracy.
- Constructed 4+ **contingency tables** (confusion matrices) to analyze **precision** and **recall**, achieving a good balance between **sensitivity** and **specificity**.
- Attained **~80% accuracy** on the **20% unseen test set**, demonstrating robust generalization and solidifying model reliability.

Student's Preferences of Learning Methods | R, Data Cleaning & Analyzing ON, Canada | September 2024 - December 2024
Main Topic: Does the student's preferred learning method affect their academic achievement?
Instructor: Luai Al Labadi, Assistant Professor, University of Toronto

- Investigated the impact of Generative AI on study habits by **conducting surveys** of 60+ students, leading to informed recommendations for improving learning efficiency.
- Collected data from 60+ participants with diverse language backgrounds using **stratified random sampling**, ensuring a balanced dataset and enhancing the validity of statistical analyses.
- Performed a **Chi-square test for independence in R** by categorizing participants based on their first language, uncovering significant variations in learning preferences.
- Conducted a **one-way ANOVA in R** to compare academic performance across multiple learning methods (ChatGPT, textbooks, online platforms), identifying which method yielded the highest CGPA.

SKILLS

- **Programming Languages:** Python, R, Java, Scala, SQL, NoSQL (Neo4j, MongoDB)
- **ML & Analytics Tools:** Spark/PySpark Scikit-learn, Pandas, NumPy, PyTorch, LangChain (LLM), Matplotlib, Tableau, Power BI
- **DevOps & Platforms:** AWS**,** Docker, Kubernetes, Git/GitHub, CI/CD, Linux/Unix
- **Soft Skills:** Problem Solving, Collaboration, Effective Communication Skills, Analytical Thinking

EXPERIENCES

University of Toronto, *Teaching Assistant* ON, Canada | January 2024 - Present
Courses: CSC343H (**Introduction to Databases**), STA256H (Probability I), STA260H (Probability II), STA360H (Introduction to Bayesian Statistics)

- Delivered tutorials to 40+ students and provided feedback on assignments, midterms, and final exams for 300+ students, clarifying concepts in probability, hypothesis testing, regression, and Bayesian inference.