

Sixth Annual Kick-Off Event



The University of Toronto Sixth Annual 2024-2025 IMI *BIGData***AIHUB** Big Data and Artificial Intelligence Competition

All U of T graduate and undergraduate students with
big data/AI experience are welcome to compete

Register @ uoft.me/BDAIHComp



\$30,000
IN PRIZE MONEY





Professor Kevin Yousie

Chair, IMI Big Data & Artificial Intelligence Competition

Associate Professor, Teaching Stream, Institute for Management & Innovation

The University of Toronto

Welcome Comments

Today's Agenda

Welcome Comments

- **Professor Shauna Brail**, Director, Institute for Management & Innovation
- **Louis Rosario** – Senior Vice President, Compliance & AML- Governance, Strategy, and Enterprise Programs, Scotiabank
- **Carol Hall** – Vice President, Global Head - AML/TF Risk Detection, Scotiabank

Overview, Logistics & Next Steps

- **Professor Kevin Yousie**, Chair, IMI *BIGDataAIHUB Big Data & Artificial Intelligence Competition*

Technology in This Competition

- **Professor Gerhard Trippen**, CTO, IMI *BIGDataAIHUB Big Data & Artificial Intelligence Competition*

Overview of the Use Case and Data

- **Dr. Duncan Smith-Halverson**, Director, AML/ATF Models and Analytics, Scotiabank
- **Parsa Vafaie**, Data Scientist, Scotiabank

Adjourn (By 1:00 pm)

Norms for this Meeting

Please:

1. Keep your video and audio muted.
2. To ask a question or make a comment, put your “hand up”.
3. When called upon, please unmute both audio and video.
4. Keep your comments relatively brief and to the point.
5. After speaking please mute your audio and video again.
6. Honour timelines.
7. Thank you.



Professor Shauna Brail

Associate Professor

Director, Institute for Management & Innovation

The University of Toronto

Land Acknowledgement

We wish to acknowledge this land on which the University of Toronto operates. For thousands of years it has been the traditional land of the Huron-Wendat, the Seneca, and the Mississaugas of the Credit.

Today, this meeting place is still the home to many Indigenous people from across Turtle Island and we are grateful to have the opportunity to work on this land.

Louis Rosario

Senior Vice President, Compliance & AML- Governance, Strategy, and Enterprise Programs, Scotiabank

Carol Hall

Vice President, Global Head - AML/TF Risk Detection,
Scotiabank



Professor Kevin Yousie

Chair, IMI Big Data & Artificial Intelligence Competition

Associate Professor, Teaching Stream, Institute for Management & Innovation

The University of Toronto

Overview, Logistics & Next Steps

Thank you to the team!



Recognizing Members of the Team

IMI Faculty & Staff

- **Professor Kevin Yousie** – Chair, IMI Big Data & AI Competition, *BIGDataAIHUB*
- **Professor Shauna Brail** – Director, Institute for Management & Innovation (IMI)
- **Professor Gerhard Trippen** – Rotman Master of Management Analytics Program
- **Professor Lisa Zhang** – Department of Mathematical and Computational Sciences
- **Professor Bree McEwan** – ICCIT and the Digital Science Institute
- **Claire Westgate** – Marketing, Institute for Management & Innovation
- **Ryan Cerrudo** – Marketing, Institute for Management & Innovation
- **Sarah Ogbiti** – Marketing, Institute for Management & Innovation
- **Abdullah Qaisar** – Marketing, Institute for Management & Innovation
- **Jesse Stamplecski** – Information and Instructional Technology
- **Brian Hamilton** – Information and Instructional Technology
- **Kenji Le** – Information and Instructional Technology
- **Muhammad Atif Rasheed** – Information and Instructional Technology
- **Simreet Aulakh** – Administration, Institute for Management & Innovation
- **Maria Andrade** – Administration, Institute for Management & Innovation
- **Clerissa Albores** – Administration, Institute for Management & Innovation
- **Ignacio Mongrell** – ICUBE, Institute for Management & Innovation
- **Saima Zulqarnain** – IMIx, Institute for Management & Innovation
- **Heather Tang** – *BIGDataAIHUB*, Institute for Management & Innovation

Scotiabank

- **Louis Rosario** – Senior Vice President, Compliance & AML- Governance, Strategy, and Enterprise Programs
- **Carol Hall** – Vice President, Global Head - AML/TF Risk Detection
- **Dr. Duncan Smith-Halverson** – Director, AML/ATF Models and Analytics
- **Parsa Vafaie** – Data Scientist
- **The AML/TF Data Science Team**

STEM Fellowship

- **Dr. Sacha Noukhovitch** – President and Editor-in-Chief, Student Research and Innovation Network

Our Values

IMI Big Data & Artificial Intelligence Competitors Are

Ethical

Respectful

Collegial

Focused on Personal / Professional Development

Environmentally Friendly

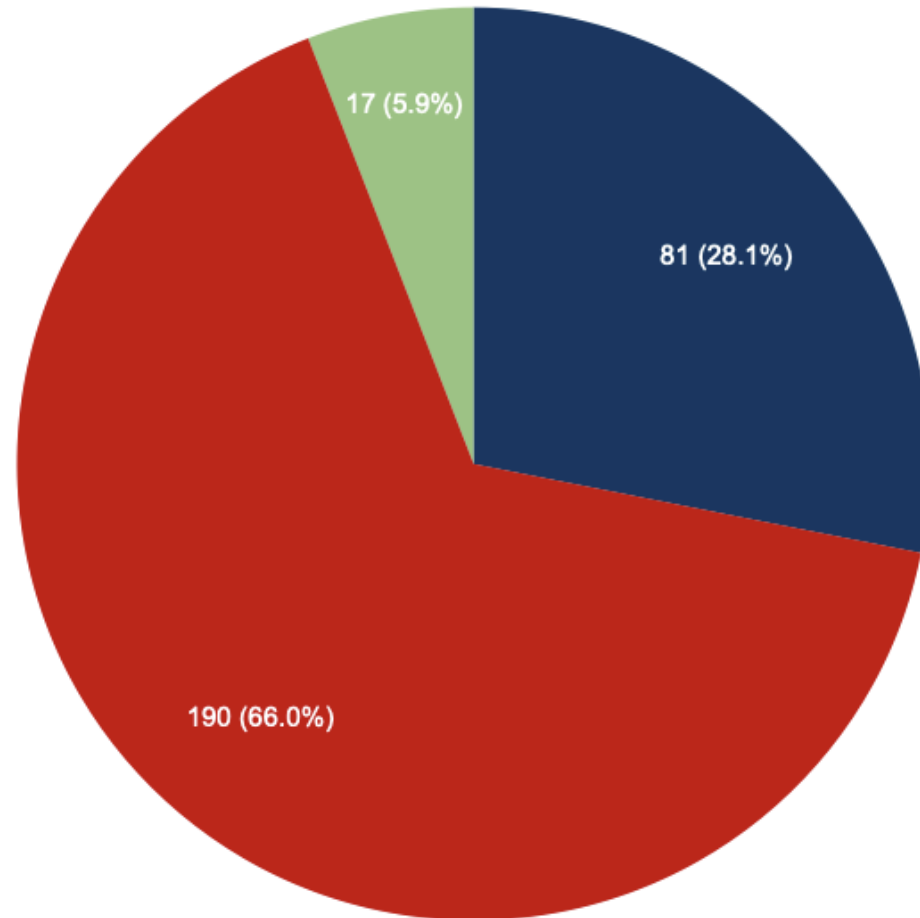
Failure to behave in this manner could potentially result in elimination from the competition.

This Year – Final Registration Numbers

356 Competitors

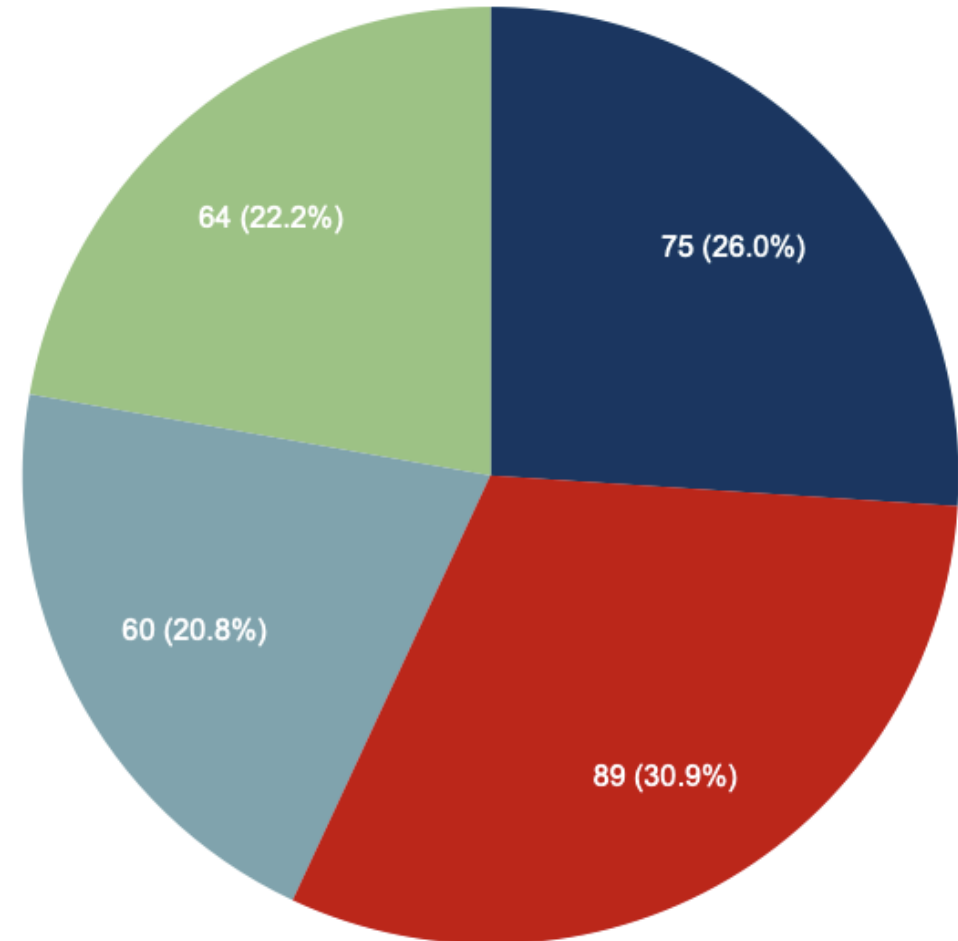
LEVEL OF STUDY

● Masters ● Undergrad student with advanced big data & AI skills ● Ph.D.

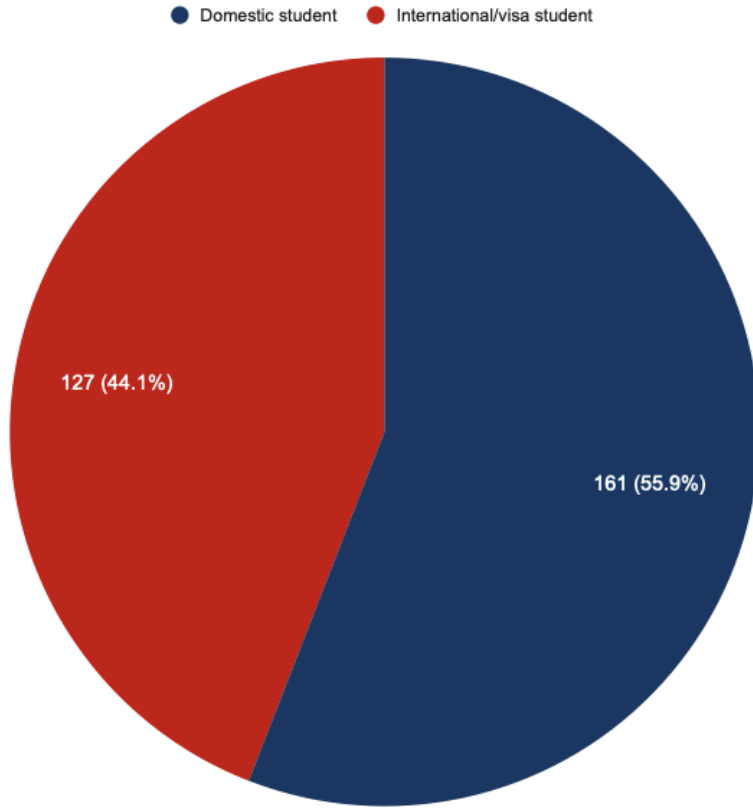


YEAR OF STUDY

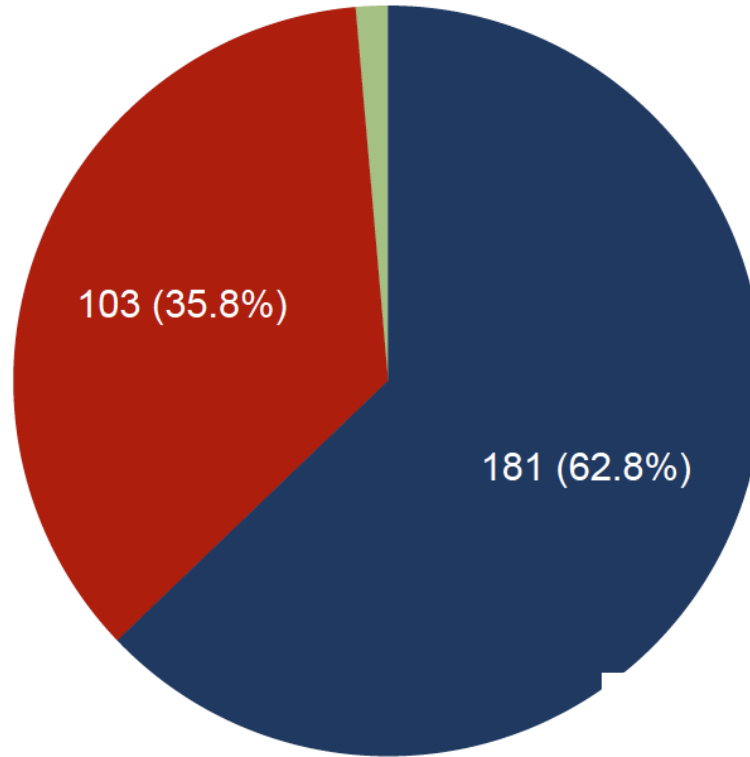
● 4th year or higher ● 1st year ● 2nd year ● 3rd year



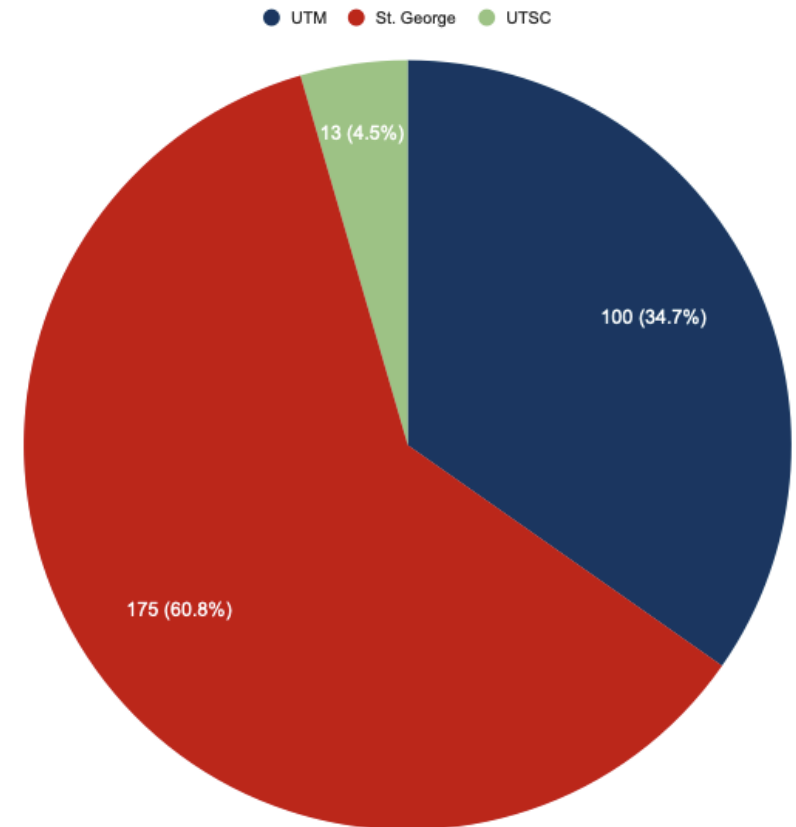
STATUS



Male Female Other

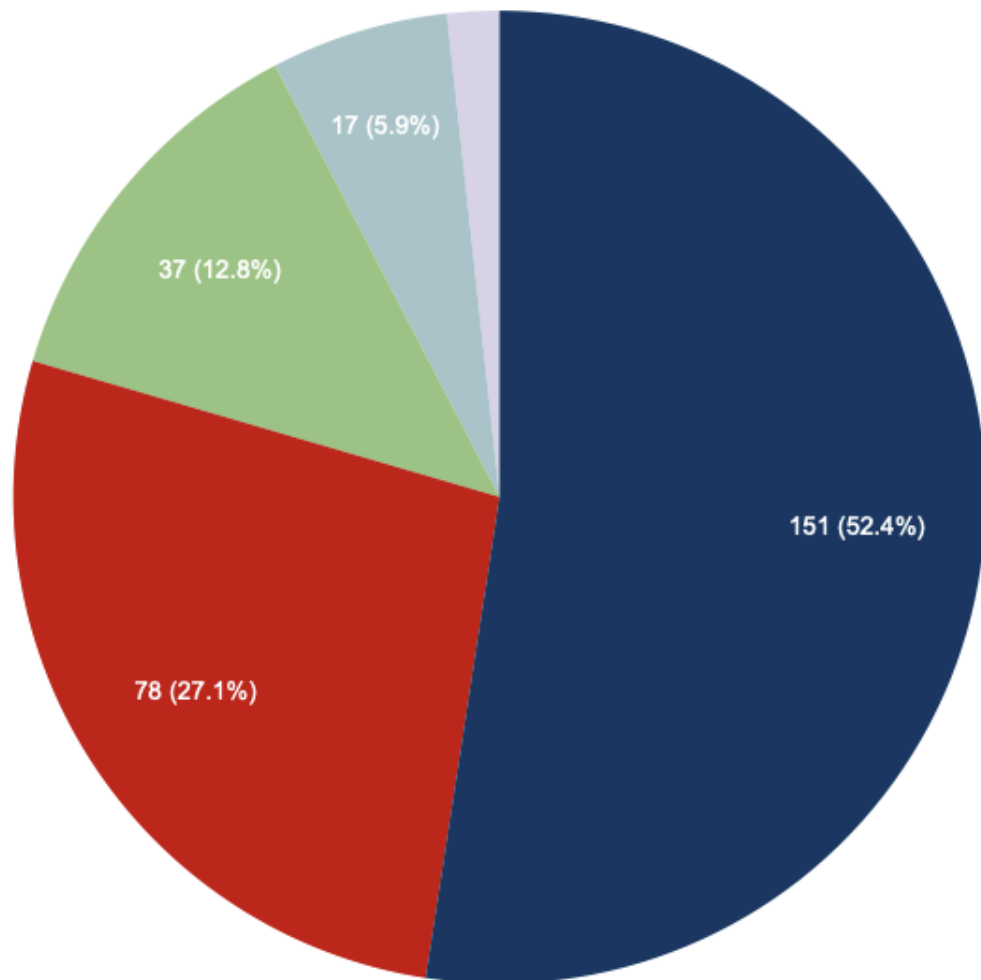


CAMPUS



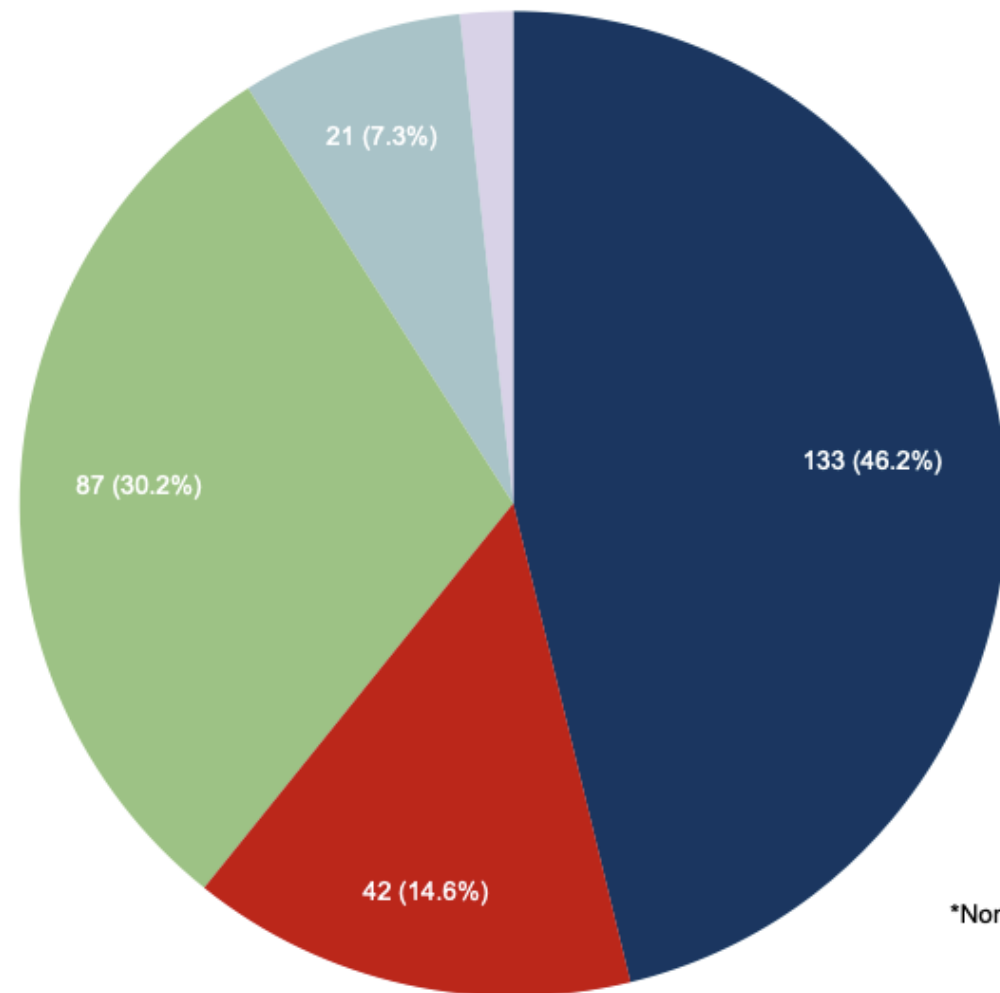
BIG DATA KNOWLEDGE

● Average ● High ● Low ● Very High ● None



AI KNOWLEDGE

● Average ● Low ● High ● Very High ● None



*None: 5 (1.7%)

*None: 5 (1.7%)

Next Steps & Key Dates

1. Ensure you **accept your Quercus invitations** to the Competition.
2. **Form your teams** in Quercus (People then “Official Teams”). Anyone not in a team **in a week** will be placed in one.
3. You will be receiving a very brief **computing capacity quiz / survey**.
4. **Sources of additional information** are now available in the Modules section of the competition Quercus site (e.g. Technical training resources).
5. Instructions with respect to **accessing the data & optional JupyterHub technology** will follow later.
6. **The competition data** will be released in December. You will be advised more about that later.

Next Steps & Key Dates

7. Direct email address to the Scotiabank Data Science Team.
8. After the data release in December – First Q & A Session.
9. By 5:00 pm Monday, February 24th, 2025 – Initial video and code submission is due.
10. Top five finalist teams will be announced.
11. By 5:00 pm Monday, March 17th, 2025 – Top five teams PowerPoint slides and final code due.
12. Saturday, March 22nd, 2025 – in-person *Finalist PowerPoint Presentations* in front of the judges as well as the *Fighting Crime With Big Data Conference* in the Rotunda at the Institute for Management & Innovation, Mississauga.



Professor Gerhard Trippen

CTO, IMI Big Data & Artificial Intelligence Competition

Associate Professor, Teaching Stream and Academic Director, Master of Management Analytics, Rotman School of Management

The University of Toronto

A Few Words About the Technology In This Competition

Dr. Duncan Smith-Halverson

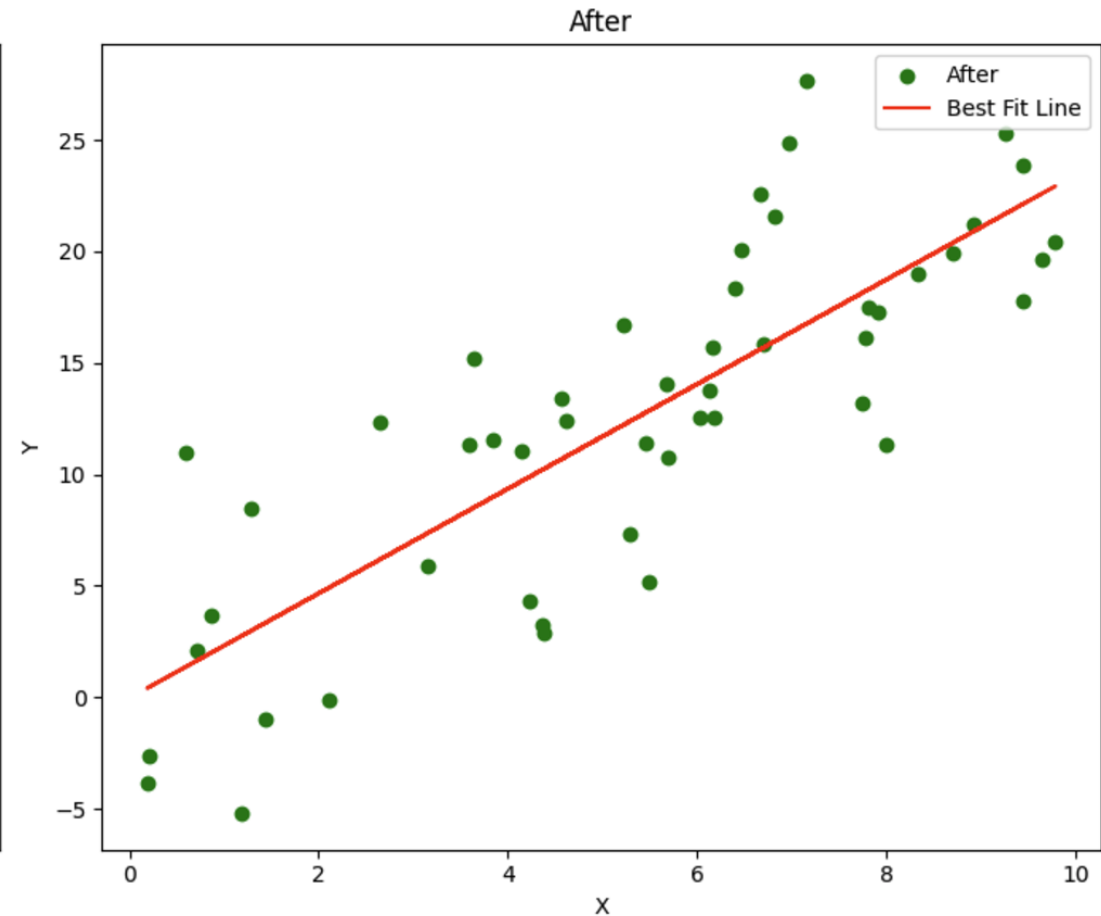
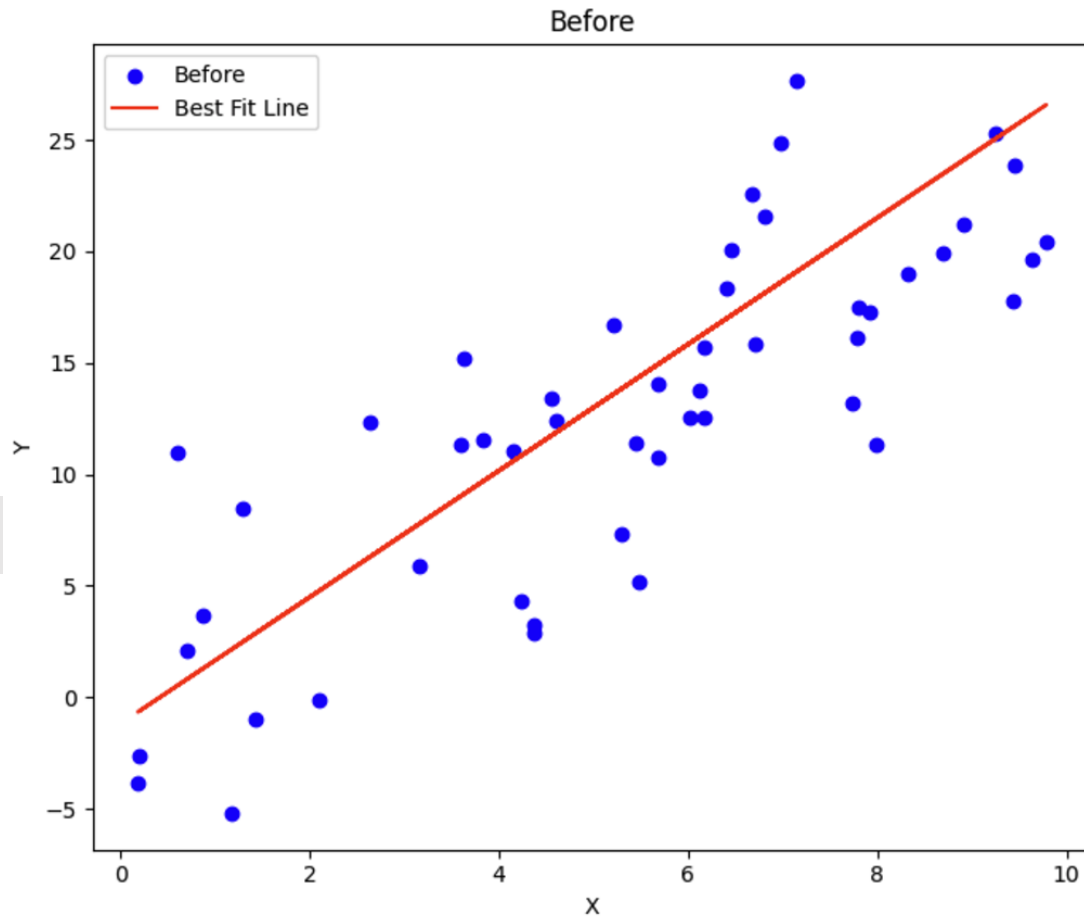
Director, AML/ATF Models and Analytics,
Scotiabank

Parsa Vafaie

Data Scientist,
Scotiabank

Overview of the Use Case and Data

Asking Questions



A Brief History of Statistical Modelling

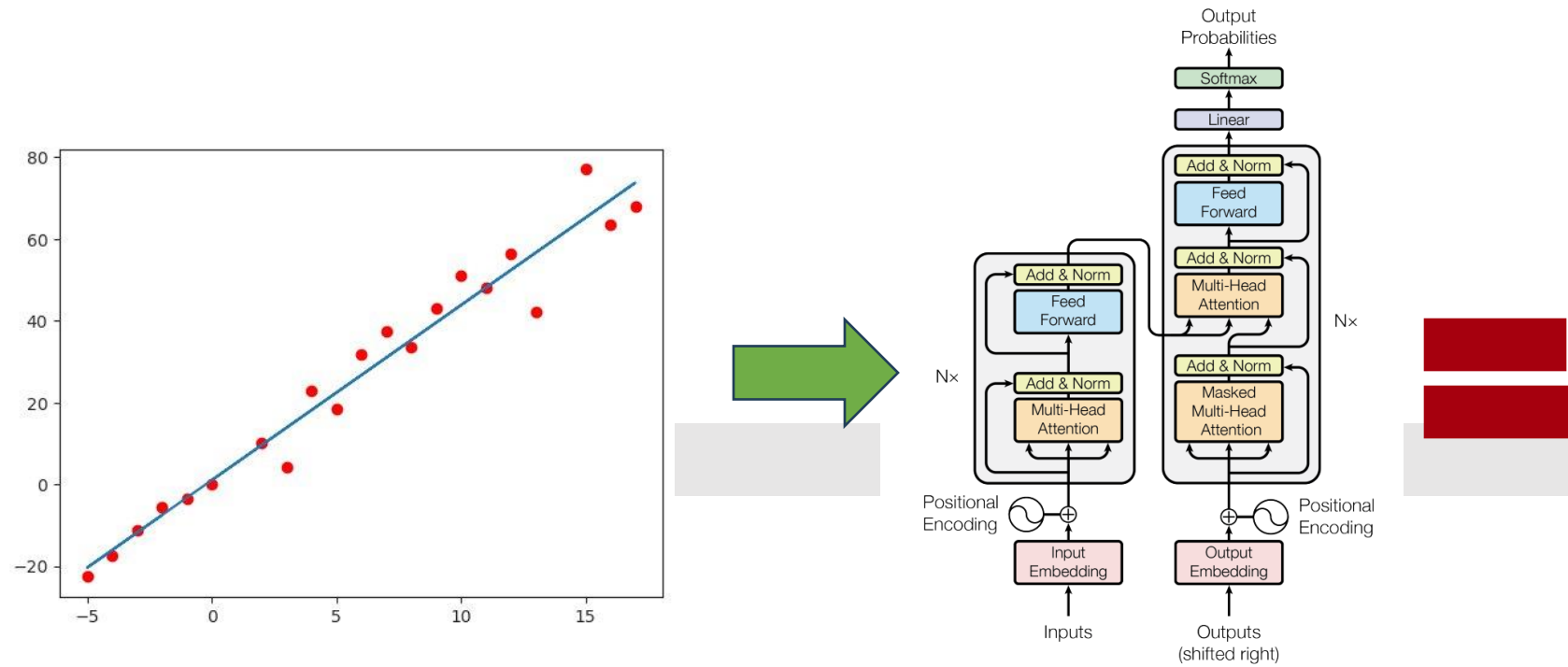
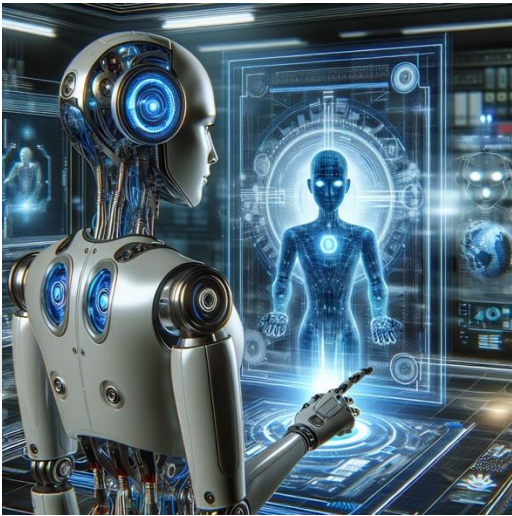


Figure 1: The Transformer - model architecture.



?

What Makes a Model Model?



$$F = G \frac{m_1 m_2}{r^2}$$

How can Models Help Fight Crime?



Well-represented representation... Deviations from which may be of interest



Anti-Money Laundering involves matching known patterns against client data, as well as identifying unknown patterns. Typically, networks involving large flows of funds are targeted



Anti-Terrorist Financing involves matching almost exclusively unknown patterns against client data. Networks may not be identifiable, and funds may be small.

Overview of tasks

Theme: Finding bad actors through anomalous behaviour

Introductory task: Unsupervised learning



Build unsupervised learning models, such as clustering techniques and anomaly detection, for customer segmentation pertinent to AML risk.

Advanced task: Customer foundation model



Can we create embeddings out of customer data by learning a representation from their transactions and customer profile?
Can a bad actor search in embedding space improve detection of unknown patterns?

Moderate task: Production pipeline



It's time to use your solution! You are tasked with delivering your code to us with clear pipelines and in a Docker format so that we can use it for evaluation.

Overview of data



Transactions: ATM, Cheque, EMT, Electronic transfers, Card transactions, wires



Customer: Established date, Tenure, Industry, City, Income, Number of employees

Learning opportunities



Unsupervised learning

- Learn to work with multiple data sources.
- Tackle real problems with missing labels.
- Utilize unsupervised learning algorithms.
- Work with synthetic data obtained from a state-of-the-art GenAI process.



Production pipeline

- Improve your ability to write quality code and pipelines.
- Learn how to work with Docker.
- Develop the ability to create well-documented scripts for streamlined execution.



Customer foundation model

- Research and work on state-of-the-art problems.
- Use challenging models such as contrastive learning.
- Visualize your results and identify patterns in the data.

Closing Comments

Sixth Annual Kick-Off Event



The University of Toronto Sixth Annual 2024-2025 IMI *BIGData***AI**HUB Big Data and Artificial Intelligence Competition

All U of T graduate and undergraduate students with
big data/AI experience are welcome to compete

Register @ uoft.me/BDAIHComp



\$30,000
IN PRIZE MONEY

