Nicholas Lo

Analyst / Researcher / Data Scientist

https://www.linkedin.com/in/nicholasjlo/

Nicholas.lo@queensu.ca

About Me

I am an athlete at heart with an insatiable thirst for bettering performance. My experiences as a biomechanics researcher and coach allowed me to see the direct and tangible impacts of data-driven decisions.

I'm now looking to make a bigger splash on the journey of analytics application. Together with my MMA team, we were able to function as a cohesive unit at tackling machine learning projects, optimization problems and everything in between.

Aside from the projects listed in this portfolio, some of my interests include fantasy sports, sports/esports betting, baseball, bouldering, and ultimate frisbee

Education

Master of Management Analytics (MMA): Queen's University	2021-2022
Bachelor of Science – Kinesiology: University of Toronto	2012-2016

Portfolio Index

Tutoring Optimization

Optimization + Simulation
Team Lead

Project Goal

A colleague runs a tutoring franchise, "Study Buddy", and came across a hiring problem for the upcoming school year and wished to gain some MMA insight.

Through @risk and simulations, **which hiring strategy should Study-Buddy follow to maximize profits?**

Some stipulations include;

- Student demand varies per hour (younger students, earlier in afternoon)
- Potential tutors not available at all time slots
- Each class must have 1 tutor for every 4 students

Solution Design

- Calculate variable and fixed costs for the upcoming semester based on previous franchise reports;
  - Availabilities and wage for each tutor
  - Utilities, franchise fees, rent, supplies, insurance
  - Tuition
  - Marketing
- Establish an emergency tutoring company in case tutors on payroll cannot cover
- Use @Risk and simulation modeling to determine maximum profits

Outcomes

The two-tutor hiring scenario was able to **generate $1,000 more** per month than the next best available option.

Function	Graph	Minimum	Maximum
RiskOutput("Total Profit")		$ 3,547.99	$ 7,679.69
RiskOutput("Total Profit")		$ 3,150.99	$ 7,581.69
RiskOutput("Total Profit")		$ 3,392.19	$ 7,559.69
RiskOutput("Total Profit")		$ 3,072.99	$ 7,729.69

Next Steps

- While Covid-19 impacts max indoor capacity, may consider larger facility to accommodate demand
  - Attempt to project Covid-19 protocols
- Consider reputation damages for students being turned away for full capacity – discount system may remedy situation

Richter Predictor

Machine Learning
Team Lead

Project Goal

Nepal experienced a massive 7.8 magnitude earthquake in 2015, which equated to about 8 billion USD worth of initial damage, not including the many aftershock damage.

In order to reduce damage from the next earthquake, I lead a team of 8 to **create a model that could predict damage severity** (1, 2, or 3) of Nepalese buildings.

Solution Design

This DataDriven competition challenged us to properly categorize damage of Nepalese buildings. We did this by;

- Implemented FLAML, an auto machine learning, model selection, and tuning library to establish a baseline solution. Testing for random forest, neural networks then eventually landing on LGBM as the best model.
- Feature engineered geo_network variables to gather comparative weights, which were used to better identify class damage.

Outcomes

Using the LGBM mode, we were able to generate a **micro f1 score of 0.74**. This also achieved a **global ranking of 32/4801** on DataDriven.org.

Next Steps

- Attempt to run on a distributed server to reduce run time and allow for more attempts
- Coupled with above point, ensemble learning and try more combinations of hyper parameter tunning
- Explore possibility of mapping dataset onto latitude and longitude to locate earthquake epicenters

Geo Level ID 1

Geo Level ID 2

Geo Level ID 3
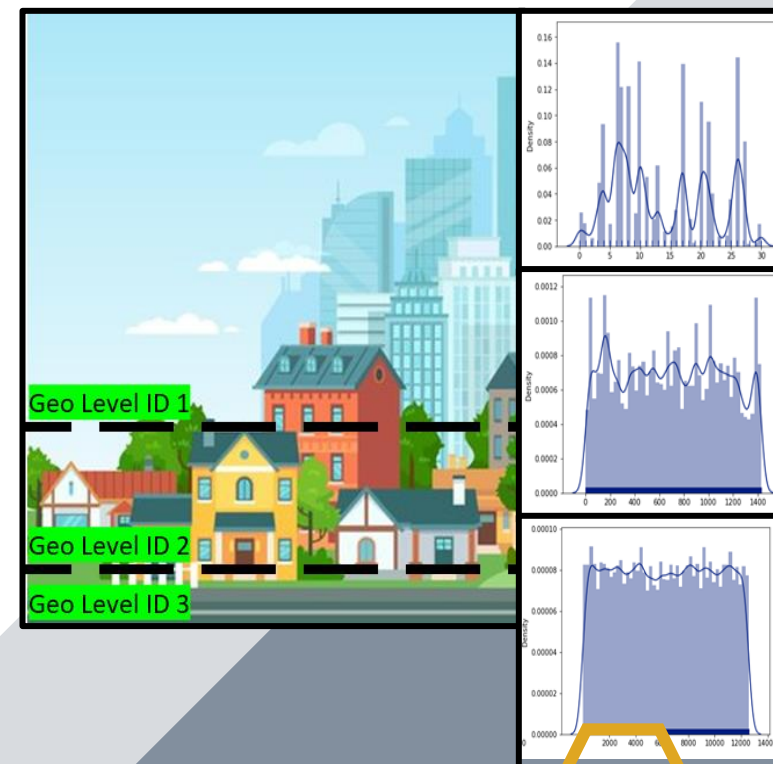
Amazon Reviews

Natural Language Processing
Machine Learning + Big Data
Team Lead

Project Goal

Amazon is an ecommerce empire that ships around 1.6 million orders per day. They've revolutionized the shopping experience and impacted every consumer business, making customer reviews an industry standard.

With that in mind, the Amazon review ecosystem had some interesting trends. We created a model to **determine whether a review was helpful or not** and then provide potential solutions to address the insights.
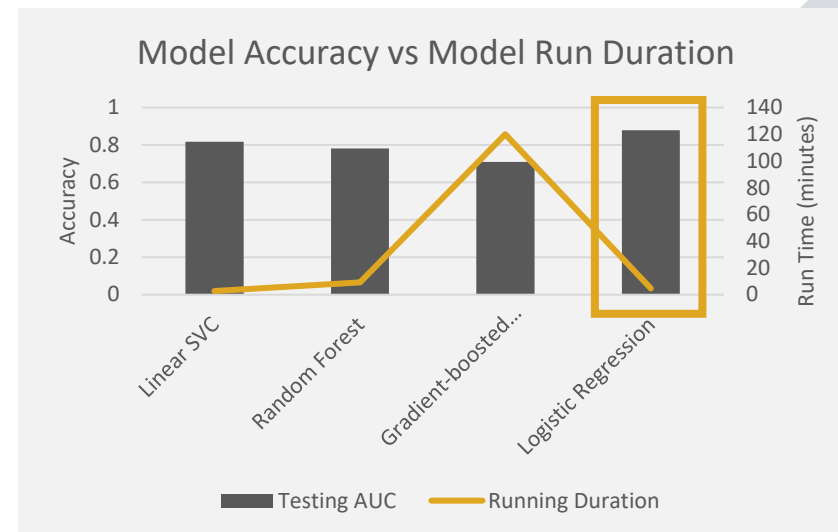
Solution Design

The challenge was privately hosted for our program (MMA865 – Big Data Analytics) on Kaggle. We competed amongst our cohort to climb the leaderboard.

- In total, dealing with 3+ million product reviews, therefore used **Databricks Spark** for parallel processing to reduce individual system demands and processing times
- Traditional data cleaning (e.g., dealing with NA, duplicates, unixtime, test/train splits) as well as NLP preprocessing (e.g., tokenization, stop words, vectorization, n-grams) were used to analyze the data
- TF-IDF, Feature Assembler, one hot encoding of vectors, feature hasher and logistic regression were the key tools used to model the data

Outcomes

While using Logistic Regression, we scored an **accuracy of 0.865, rank #5 of 16**, at determining whether a review would be helpful or not.

Model Accuracy vs Model Run Duration

Accuracy (left axis): Linear SVC, Random Forest, Gradient-boosted…, Logistic Regression
Run Time (minutes) (right axis)

Legend: ■ Testing AUC — Running Duration

Next Steps

- From business perspective:
  - Revamp the quality customer reviewer badges for better reliability and confidence
  - Improve cluttered UX design
- From model perspective:
  - Balance the data by collecting more data, huge spike of negative reviews during 2012-2017
  - Hyperparameter tuning

3

Pitch Mound Heights
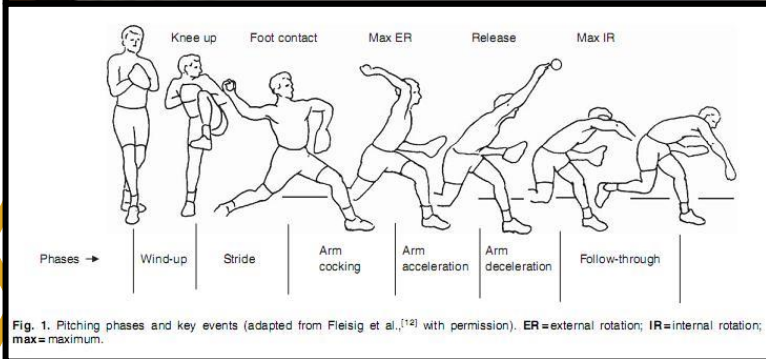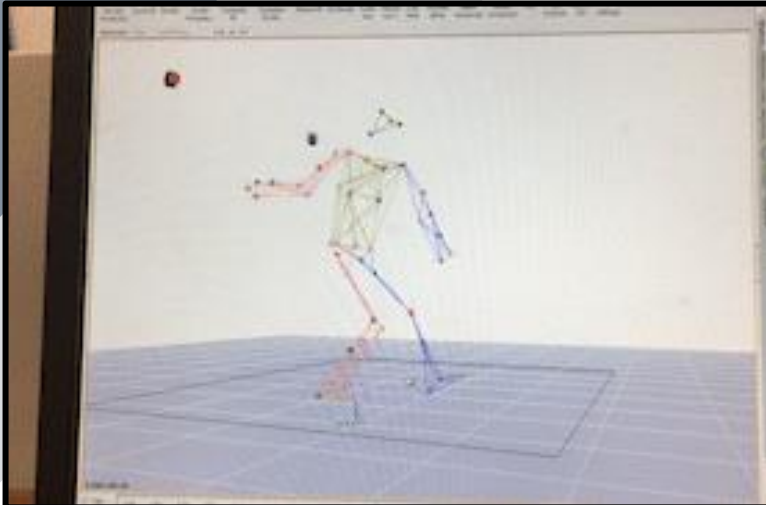
Biomechanics Research
Research Intern

Fig. 1. Pitching phases and key events (adapted from Fleisig et al.,[12] with permission). ER=external rotation; IR=internal rotation; max=maximum.

Project Goal

A constant concern for pitchers in the MLB all the way down to Little League is longevity of arm health. We seek to determine if the height of the mound is inherently a risk factor to ulnar collateral ligament injuries, colloquially called the Tommy John Injury.

Funded by MLB and USA Baseball, the American Sports Medicine Institute and Motus Global set to collect data of pitchers from 4 mounds of varying heights (15, 10 being regular, 8, and 6 inches). The research question is: **does mound height influence the biomechanics of the pitcher?**

Solution Design

- Invited local college teams from division 1-3 to USA Baseball Headquarters in Raleigh, North Carolina to conduct an outdoor data collection
- Partnered with experts from Motus Global to navigate concerns with outdoor data collection
- Mound expert consulted to construct mound heights with traditional 1 inch decline per foot of runway
- 5 full effort fastballs and 5 effort curve balls thrown from each mound (40 pitches total)
- Repeated measures analysis of variance was used to detect significant differences

Outcomes

Lowering the mound may not affect the pitcher's ball movement, but may slightly reduce shoulder and elbow kinetics, **possibly** reducing risk of injury.

Published here
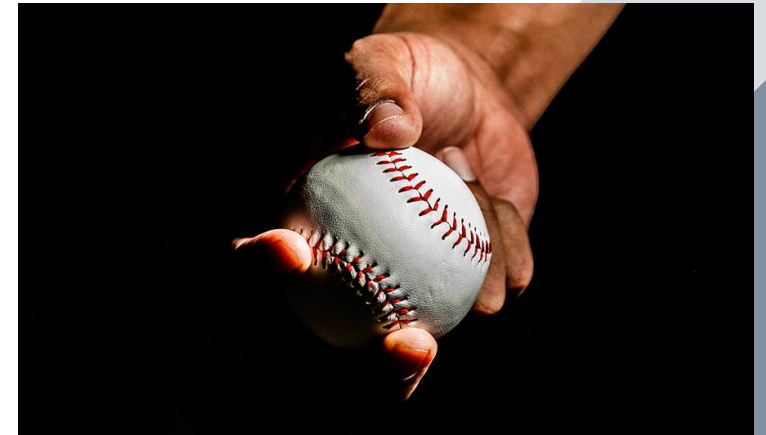
Next Steps

Since there were minimal changes in pitcher kinetics and ball kinematics, investigating mound distance from plate or size of strike zone may be the next steps for reducing overall arm stress.

Sik-Faan Pitch

Entrepreneurial Pitch
Individual

Project Goal

Cuisine is huge in Chinese cultures, acting as uniting vehicles for families. Unfortunately, Chinese-Born-Canadians have been culturally detached from their roots, causing cultural drift. Therefore, **this service aims to reconnect Canadians with their Chinese heritage.**

Solution Design

In a Dragon's Den-esque competition, I pitched an entrepreneurial idea in order to gain angel investor funding to take the project off the ground.

- Preserve cultural heritage with authentic, crowd sourced Chinese food recipes
- A service motivated by current food delivery services (e.g., GoodFood, HelloFresh). Provide food delivery kit with all the necessary ingredients
- A platform fueled by community of customers to share recipes and traditions
- Scalability in reproducibility for using framework on other cultural cuisines

Outcomes

[Role playing] Currently product testing with 1 year into business expecting to deliver 10,000 meals per month. Asking for **$750,000 for 10% of the company**.
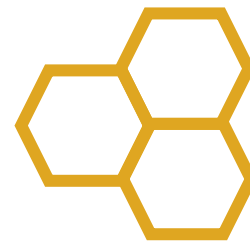

Next Steps

- Build out the recipe list through 1on1 interviews and PO box submissions
- Explore storage and refrigeration options for produce to-be delivered as well as delivery logistics

BCB Branch Optimization

Optimization + Simulation
Individual

Project Goal

This BCB Bank case study was looking to **optimize the number of managers they need to hire** to oversee all 35 banking branches in the Edmonton area. The goal of this extensive project was 3-fold.

1. Descriptive analytics. Gain intuition through visualization
2. Predictive analytics. Obtain data we don't know
3. Prescriptive analytics. Optimize how many branches should be considered "home branches"

Solution Design

- Descriptive: Using Tableau, we were able to extract branch locations from postal codes, then appropriately size branches based on annual revenue
- Predictive: By using GoogleMaps API, commute times from one branch to another was vectorized, then compared within matrixes
- Prescriptive: Setting objective = max profit, variable = branch assignments, and constraints = exclusivity and travel time for linear programming model

Outcomes

From a financial standpoint, **five BCB home branches was best** to accommodate the demand of each of it's 35 locations and maximize total profits.
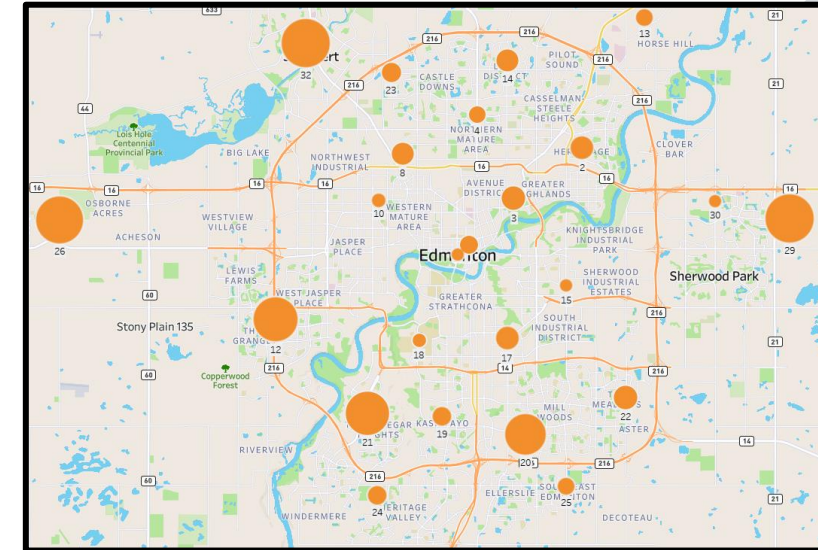
Next Steps

The five largest branches were considered as home branches for this optimization problem.
- However, this may not be the five geographically spaced-out branches. Further analysis into geo-spatial distancing could potentially have similar, but more convenient solutions

Extracurricular Activities

American Baseball Biomechanics Society Member (2019-Present)

Staying up to date with the current research within baseball and biomechanics. Allows me continue with the baseball passion as a side project and continue to push out baseball-related content that would potentially fall onto a community that can learn, refine, or reproduce similar insights.

Smith's MMA Student Executive Committee VP of Operations: (2021-2022)

While working closely with the Career Advancement Center, my main role was to connect students with the resources needed to make that next jump in their career. Hosting alumni panel discussions to bring perspective or directing students to portfolio building seminars.