Patricia Bata

1 Massey Square # 2714
Toronto, Ontario M4C 5L4

+639 17 526 8788
bata.patricia@gmail.com
github.com/patbata

Education

- **University of the Philippines - Diliman** — Quezon City, PH
  Bachelor of Science in Chemical Engineering (5-year program) — *Jun. 2013 - Jul. 2018*
  - Cumulative GWA: 1.82 / 1.00
  - Graduating Year GWA: 1.48 / 1.00
  - University Scholar[1] (1st Semester AY 2017-2018)
  - College Scholar[2] (1st Sem AY 2014-2015; 2nd Sem AY 2015-2016; 1st Sem AY 2016- 2017)
  - Award winning senior thesis: Chitosan-coated silica nanostructures from waste rice husk for carbon dioxide capture
  - Academic units: Mathematical Methods, Thermodynamics, Transport Phenomena, Chemical Reaction Engineering, Chemical Process Dynamics and Control, Project Management, Environmental Process Engineering, Chemical Engineering Research

[1]Merit award for students with a semestral general weighted average of at least 1.45 (magna cum laude standing).
[2]Merit award for students with a semestral general weighted average of at least 1.75 (cum laude standing).

Work Experience

- **OptumLabs, Inc.** — BGC Taguig City, PH
  Data Scientist — *Dec. 2019 - Present*
  - Use pertinent data and facts to identify and solve a range of problems within area of expertise
  - Applies statistics, machine learning, programming, data modeling, simulation, and advanced mathematics to recognize patterns, identify opportunities, pose business questions, and make valuable discoveries leading to prototype development and product improvement.

- **OptumLabs, Inc.** — BGC Taguig City, PH
  Data Science Apprentice — *Jul. 2019 - Dec. 2019*
  - Part of the BPI - UnitedHealth Group Data Science (BUDS) program, which is an intensive 6-month training program on different methods in data science, specifically statistics and machine learning algorithms to solve problems.
  - Practiced data science techniques and applications using real data from the health care and financial industries.
  - Trainings included Technical Challenges (TCs), which are one-day team-based activities that focus on a particular aspect of data science, and Case Studies, which are multi-week projects that use real-world data and aim to solve current projects from BPI and UHG R&D.

- **B&V People Business Co.** — Muntinlupa City, PH
  Admin Officer — *Feb. 2019 - May 2019*
  - Assisted leadership programs of B&V People Business Co., which include: Leadership Development Planning, Accountability Workshops, and Effective Communication Skills Workshops
  - Assisted during Global In-House Center Council (GICC) fellowship meetings, which are meetings between GICC's 60 members, predominantly Fortune 500 companies.

- **Polymer Research Laboratory** Quezon City, PH
  Student Researcher *Jun. 2017 - September 2018*
  - Worked closely with Dr. Bryan B. Pajarito on developing a novel carbon dioxide adsorbent from waste rice husk and chitosan.
  - Researched on the feasibility of the novel adsorbent and its possible application in industrial plants as an alternative to other absorbents/adsorbents used for carbon dioxide.
  - Experimented using a fixed-bed adsorption set-up and characterized adsorbent properties using OriginPro.

Conferences and Awards

- **Champion** Laguna, PH
  PIChE National Undergraduate Research Contest *Feb. 2018*
  - Organized by Philippine Institute of Chemical Engineers (PIChE)
  - Research based competition between all chemical engineering students across the country.
  - Contest theme: "Expanding Chemical Engineering Competencies for National Development"
- **2nd Runner Up - Research** Quezon City, PH
  UPD Engineering Undergraduate Project Contest *Jun. 2018*
  - Organized by the College of Engineering at University of the Philippines – Diliman
  - Design and research competition between all engineering students in UP Diliman
- **Best in Innovation & 1st Runner Up** Makati City, PH
  BPI-DOST Science Awards *Jun. 2018*
  - Organized by Bank of the Philippine Islands (BPI) and Department of Science and Technology (DOST)
  - Contest theme: "Forming a Sustainable Country through Science and Innovation"
  - A highly competitive nationwide science competition aimed to encourage and fund the country's top studies that support the United Nation's Sustainable Development Goals (SDGs).
  - Only the country's top 5 projects are invited to present and defend.
- **Champion** Quezon City, PH
  Magsaysay Future Engineers/Technologists Award (MFET) *Dec. 2018*
  - Organized by National Academy of Science and Technology, Philippines (NAST PHL) and Department of Science and Technology (DOST)
  - An award given by the country's highest advisory and recognition body to the government and the science community to encourage young Filipino students to pursue a career in science.

Skills

Languages: Python, Matlab, SQL, Tableau, RStudio, LaTeX

Operating Systems: Linux, Mac OS X, Windows UNIX

Modelling: Ordinary Linear Regression, Neural Networks (ANN, RNN, CNN), Boosting models (XGBoost, CatBoost, LightBoost), Transformer models (BERT, MedBert, GPT-3, Med7, NeuroNER)

Data Visualization : Sankey diagrams, SHAP diagrams, Plotly, Bokeh, Matplotlib, Seaborn

Data Expertise : Claims analysis, Health surveys, Free-text doctor's notes, Finanacial Data

Lab Skils: Sol-Gel process, UV-Vis Spectrometer, Fixed-bed adsorption analysis, High-temperature box furnace, Amination of polymers

Interpretation Skills: X-Ray Diffraction (XRD), X-ray fluorescence (XRF) spectrometer, Raman spectroscopy, Brunauer, Emmett and Teller (BET) surface area analysis, Fourier-transform infrared spectroscopy (FTIR), Field emission-scanning electron microscopy (FE-SEM), Transmission electron microscopy (TEM), Thermogravimetric analysis (TGA), Differential scanning calorimetry (DSC)

Miscellaneous: Intermediate video editing, Strong verbal and written communication skills, Excellent troubleshooting and debugging skills, Exceptional problem solving skills, Good leadership and team management skills, Good technical writing skills, Excellent at poster and manuscript preparation

Projects

- **Bank Customer Segmentation based on Customer Value** (Oct. 2019)
  - Aimed to segment the customers from a popular bank in the Philippines (BPI) based on banking activities.
  - Utilized Exploratory Factor Analysis (EFA) and k-means clustering to determine a customer value metric.
- **Quantifying Risk of Readmission for Diabetic Population** (Nov. 2019)
  - Aimed to determine features that are significant in patient readmission using data from 130 US hospitals
  - Developed 15 model configurations, and made use of linear and tree-based models using SMOTE and LASSO techniques
- **BPI ATM Forecasting with Unsupervised Clustering** (Dec. 2019)
  - Aimed to optimize tha bank's ATM reloading schedule to prevent premature loading and cash-out.
  - Developed forecasting models using architectures such as RNN - LSTM and FaceBook prophet.
  - Each model was assigned to an ATM cluster, developed using unsupervised techniques such as k-means, DBSCAN, and spectral clustering.
- **Creating Medical Embeddings for Lung Cancer Prediction** (Oct. 2020)
  - Developed medical embeddings using XLNet and GPT-3 from claims data to predict lung cancer.
- **Incorporating a Predictive Model in Micro-simulation Model (MSM) for Lung Cancer Costs** (Nov. 2020)
  - Incorporated a lung cancer prediction model into a microsimulation model to simulate costs and determine if the predictive model helps diagnose lung cancer earlier.
- **De-identification of Reliant Medical Data** (Aug. 2021)
  - Aimed to evaluate and refine the open-source Philter algorithm to de-identify free-text doctor's notes from Reliant Medical Group.
  - Made use of regular expressions and medical-based NLP models (Med7 and NeuroNER) to improve the algorithm.

- **Community Data Simulation and Modeling** (Sept. 2021)
  - Aimed to demonstrate that using community-level variables as predictors to models lead to confounding variables and bias.
  - Developed a Simulation and Model python class to demonstrate this bias.
  - A dashboard using these classes was developed to manipulate simulation parameters and model these in real-time.
- **Other projects**
  - Explored the association between activity and sleep patterns using Fitbit data. (Sept. 2019)
  - Developed a pipeline to analyze different CGM-derived metrics and adverse perinatal outcome for a pilot study with University of Pennsylvania. (Jan. 2020)
  - Explored UnitedHealth Group's claims data to identify common delays in lung cancer diagnosis and their associated cost. (Feb. 2020)
  - Conducted a retrospective study of lung module management, such as pathways to diagnosis and succeeding procedures, using medical claims. (May. 2020)
  - Analyzed Safegraph's mobility data to determine hospital patterns and surges before and after COVID-19 using time-series analysis. (Feb. 2021)
  - Created sankey diagrams to explore GLP-1 drop-offs in patients diagnosed with type 2 diabetes. (Jun. 2021)

Publications

1. Mehrotra S, Brantley H, Onglao P, Bata P, Romero R, Westman J, Bangerter L, Maity A. Divide-and-Conquer MCMC for Multivariate Binary Data. arXiv. Preprint posted online June 11, 2021. arxiv:2102.09008
   - Maintained GitHub repository using enterprise assets.
   - Validated R scripts and results
   - Point-person for formatting and submitting manuscripts to journal publications

Interests

Academic: Polymerization, nanotechnology, bioinformatics, genetics

Sports: Football, bouldering, and yoga

Computers: Neural networks, natural language processing, simulations, dashboarding

Musical: Playing piano and singing in chorale

Membership: Philippine Institute of Chemical Engineers (PIChE), UP Alumni Engineers Association

Other: Reading novels, foreign language learning