Jan Michael Cayabyab Austria

datacamp.com/profile/janmichaelaustria
github.com/janmichael88

Email : jca88@cornell.edu
Mobile : 603-553-9327
204 North Mills Avenue, Claremont, CA 91711

Education

- **University of New Hampshire** — Durham, NH
  MS. Data Science; GPA: 4.00 — *May. 2019 – May. 2020*
  Thesis: Viral protein sequence creation and classification
- **Tufts University** — Medford, MA
  MS. Bioengineering (Withdrew); GPA: 4.0 — *Aug. 2017 – Dec. 2018*
  Project: Process Improvements of ultrafiltration membranes
- **Cornell University** — Ithaca, NY
  BS. Biochemistry: GPA:3.4 — *Aug. 2009 – May. 2013*
  Thesis: Auqaporin Membrane Desalination Systems

Skill Highlights

- **Programming Languages**: MATLAB, Python, R, JMP, SQL, SAS, C++
  Python Dependencies: numpy, scipy, tensorflow, keras, pytorch,pandas
  Technologies:Microsoft Power BI and Office, Tableau, Hadoop, Mapreduce, Spark
  Sectors: Data Analysis, Machine Learning, Deep Learning, Reinforcement Learning, Optimization, Biostatistics, Bioinformatics, Healthcare

Data Science Projects

University of New Hampshire

- **Cancer Recognition Detection**: Created a convolutional neural network using tensorflow/keras to identify severity of cancer from images of biopsies.
- **Novel virus creation and classification**: Developed classifier to predict protein type from protein sequence using both a CNN and RNN. Trained on 50000 different proteins. Used RNN to recreate novel viral proteins.
- **Hospital Ranking System**: Using python/sklearn and RandomForest/XG Boost, developed latent variable model to rank hospitals in area based on domains reported. Used to identify most efficient facilities.
- **Entity Resolution on Medical Claims Data**: Using python/sklearn and Logistic Regression models, identified individuals from medical claims data that were patients admitted to local brain trauma center after normalization of dataset.
- **Amazon comment rating classifier**: Created a convolutional neural network to identify star rating based only on text. Performed dimension reduction techniques such as UMAP and TSNE to group based on similarity.
- **Emergency Room Detection**: Using random forests and boosting algorithms, developed classifier to identify which patients were at risk of returning to emergency room within 6 months from medical claims data.

Work Experience

- **Cognia** — Portsmouth NH
  Data Scientist - Internship — *Oct.2019 – May.2020*
  - **Description**: Prototyped and designed Tableau dashboard with in house SQL integration for their employees to better understand and retrieve pertinent data in an efficient manner.
  - **Database Creation**: Extracted survey data from Cognia SQL server and ran exploratory data analysis in Python and R. Used the insights to create a data reporting system to summarize public/private surveys.
- **Manchester Boston Regional Airport** — Manchester NH
  Data Scientist - Internship — *Oct.2019 – May.2020*
  - **Description**: Used public surveys, flight data, and consumer market data to create a perceived value of time metric to educate travelers why choosing a certain airport (BOS or MHT) is more appropriate.
  - **Webscraping**: Used python and R to pull Uber, Lyft, and public transportation options to showcase fare vs time tradeoffs in southern NH and northern MA areas. Integrated with google API to create a user application for the airport. Prototyped chat bot to deliver cheapest fare to BOS or MHT given a user input.
- **Lonza** — Portsmouth NH
  Project Specialist — *Jan.2019 – Mar.2019*
  - **Description**: Created SQL databases to store in process manufacturing data specifications and parameters to increases efficiencies by 50% -70% allowing the company to allocate finances appropriately.
  - **Apache Spark**: Developed machine learning models using pyspark to identify which engineering and manufacturing projects would have a higher probability of succeeding within the next 6 months saving $3.5 million.
- **Pfizer** — Boston, MA
  Senior Associate Scientist — *Mar. 2018 – Jan. 2019*
  - **Description**: Performed analytical characterization of candidate molecules by using biophysical and biochemical characterization techniques (NMR, HPLC, capillary gel electrophoresis, and iCE). Developed dashboard to trend historical data from over 10 years to identify new potential molecules of interest.
  - **Keras and sklearn**: Used unsupervised learning techniques (TensorFlow, PyTorch, sci-kit learn) to identify patterns among monoclonal antibodies (bi-specific, tri-specific, etc). Model was deployed to classify incoming biologics and narrow down scope for testing.

- **Pfizer** Andover, MA
  Process Engineer *Mar. 2016 – Mar. 2018*
  - **Production Operations**: Supported upstream (cell culture) and downstream (purification) manufacturing operations for Pfizer's portfolio of monoclonal antibodies and vaccines for clinical and commercial campaigns.
  - **Continuous Improvements**: Trended data for processes to compare against other batches to identify manufacturing improvements. Tracked key performance indicators and created visualizations for metrics.
  - **Design Improvements**: Developed new impellors/agitators for 10000L centrifuge. Modeled flowrates and pressure curves using MATLAB scripts to optimize purification processes. Provided insight on scale up parameters of processes between research and development and manufacturing technology leads.
- **Pfizer** Andover, MA
  QC Scientist *Apr. 2015 – Mar. 2016*
  - **Description**: Performed microbiological assays of clinical products, commercial products, raw materials, production intermediate samples, bulk samples, direct drug substance, and drug substance intermediate samples. Updated and revised standard operating procedures for scientists such that current methods would be reflected and were in FDA specification.
- **Dana Farber Cancer Institute** Boston, MA
  Research Associate *Apr. 2014 – Apr. 2015*
  - **Desciption**: Core research/administrator for blood and tissue banks processing whole blood, plasma, serum, and tissue for breast cancer analytics. Queried databases for clinical, histological, and pathological data regarding cancer types for principal investigators. Responsible for database systems housing more than 100,000 clinical trial patient's samples.
- **Columbia University Medical Center** New York, NY
  Research Associate, Systems Biology *May.2013 – Apr. 2014*
  - **Research and Analytics**: Performed drug discovery, bioinformatics, and compound synergy experiments. Used accelrys pipeline pilot to analyze data for heat-map construction. Used automation and liquid handling machines for planned cell based experiments in microplate format. Performed cell culture and maintenance of multiple primary cell lines from patients at the hospital. Used microscopy techniques to determine physiological properties of stained cells.