

# Liam Bui

📍 Vancouver, BC, Canada

✉ [liam\\_bui@outlook.com](mailto:liam_bui@outlook.com)

🌐 <https://linkedin.com/in/liambui>

## TECHNICAL SKILLS

- **Machine Learning:** Scikit-learn, Tensorflow, OpenCV
- **Big Data:** Hadoop, Spark, Hive, Cassandra, MongoDB
- **Programming:** Python, Scala, Java, JavaScript
- **Statistics/Operations Research:** R, SAS
- **Visualization:** Qlikview, Tableau, D3.js
- **Others:** SQL, AWS, Cloudera, GIT, JIRA

## PROFESSIONAL EXPERIENCE

**Machine Learning Engineer – Terramera, Vancouver, BC** Jan 2018 – Present

- Research machine learning and computer vision algorithms for image analysis, bioinformatics, and cheminformatics to accelerate and support data-driven bioscience R&D projects using Python, OpenCV, Scikit-learn, Tensorflow.
- Define data requirement and implement end-to-end pipeline to collect, process, and analyze data, leveraging on imaging, sensor, and robotic platforms.

**Machine Learning Engineer (Intern) – Armourgel, Burnaby, BC** Sep 2017 – Dec 2017

- Researched signal processing and machine learning algorithms for fall detection and activity classification based on biomedical signal data collected from wearable sensors
- Visualized, pre-processed and annotated the signal data to create a labelled dataset ready for analysis
- Implemented data science pipeline to extract time-domain & frequency-domain features, train machine learning models, and evaluate the models' performance with Python and Scikit-learn

**Data Scientist (Intern) – PHEMI, Vancouver, BC** May 2017 – Aug 2017

- Developed distributed data processing and analytics prototypes using Spark (Scala) and Hive to demonstrate fast query and analytics on terabytes of clinical data
- Proposed machine learning and deep learning demos using Python, Scikit-learn and Tensorflow to show how medical imaging data can be analyzed to support diagnosis
- Implemented natural language processing pipeline in Java and cTAKES, a library with both rule-based and machine learning techniques, to extract clinical information from unstructured medical text

**Business Analytics Specialist - DBS Bank, Singapore** Jul 2015 – Jul 2016

- Developed SAS code to extract data from Teradata SQL databases and perform statistical analysis for Card & Unsecured Lending sales and marketing
- Liaised with data science team to deploy predictive models (Recommender System, Location Analytics) for targeted marketing, leading to 2x lift in customer response rate in digital campaigns
- Proposed experiment design and hypothesis testing on different content factors to improve customer engagement in email marketing

**Data Analyst - DBS Bank, Singapore** Jul 2012 – Jun 2015

- Developed Java analysis reports and Qlikview dashboards to provide technology and operations teams with insight on process improvement and risk control
- Performed process mapping and simulation modelling to optimize business processes, resulting in 10% reduction in operating cost
- Managed modelling team to document enterprise-wide data assets using industry standard notations (ER, UML, BPMN diagrams), leading to consistency and standardization

## EDUCATION

**MSc. Big Data, Machine Learning, GPA 4.0/4.0** Sep 2016 – Dec 2017  
Simon Fraser University, Burnaby, BC, Canada

# Liam Bui

📍 Vancouver, BC, Canada

✉ [liam\\_bui@outlook.com](mailto:liam_bui@outlook.com)

in <https://linkedin.com/in/liambui>

**BEng. Computer Science, BSc. Business Analytics, GPA 4.45/5.0**

Aug 2008 – Jun 2012

Nanyang Technological University, Singapore

## DATA SCIENCE PROJECTS

**Diabetes prediction based on clinical measurements** [[Github](#)] Sep 2017 – Dec 2017

- Analyzed Pima Indians Diabetes dataset to understand relationship between diabetes diagnosis and clinical measurements using generalized additive models and tree-based methods.
- Implemented an ROC curve interactive dashboard to show the optimal cut-off threshold based on a specific True Positive Rate or False Positive Rate criterion
- Technologies: R, shiny, glmnet, gam, gbm

**Topic Model and Network Analysis on research publications** [[Github](#)] Jan 2017 – Apr 2017

- Developed web scraping function to retrieval publications from major machine learning journals
- Analyzed topic models (Latent Dirichlet Allocation) to explore underlying topics and performed graph analytics to visualize interesting relationship among publication topics
- Technologies: Scala, Spark ML, Spark GraphX, MongoDB, Gephi

**Question-Answer Style Chatbot** [[Github](#)] Jan 2017 – Apr 2017

- Developed a chatbot using deep recurrent neural network (Sequence to Sequence LSTM) for language models, together with several improvement techniques (Beam Search, Attention Mechanism)
- Developed a web-based user interface for user-chatbot conversation
- Technologies: Python, NLTK, Tensorflow, Flask

**Longitudinal study: Effect of tobacco use on mortality** [[Github](#)] Jan 2017 – Apr 2017

- Analyzed National Longitudinal Mortality dataset to understand the effect of tobacco use on mortality using mixed model logistic regression and survival analysis (Kaplan-Meier estimator)
- Technologies: R, ggplot2, glm, lme4, survival

**Online Restaurant Recommender System** [[Github](#)] Sep 2016 – Dec 2016

- Implemented Collaborative Filtering Recommender System (Matrix Factorization, Item-Similarity) based on restaurant detail and user rating datasets from Yelp
- Developed a web-based Proof of Concept to demonstrate and visualize the recommender system's result
- Technologies: pySpark, MLlib, Cassandra, Hadoop Cloudera Distribution, Google Map API, D3.js

**Facial Expression Recognition in Real-Time** [[Github](#)] Sep 2016 – Dec 2016

- Implemented and Benchmarked different models (Principle Component Analysis, Linear Discrimination Analysis, Convolutional Neural Network) to extract features from facial expression images and predict emotion in real-time
- Technologies: Python, OpenCV, Keras, Amazon AWS

**Sentiment Analysis on Movie Comments** [[Github](#)] Sep 2016 – Dec 2016

- Performed natural language processing and text mining on movie comments using Convolutional Neural Network in Matlab to analyze users' sentiment towards movies
- Technologies: Matlab, Matconvnet, Word2Vec