

Liam Bui

📍 Vancouver, BC

☎ (604) 352-1568

✉ liam_bui@outlook.com

🌐 <https://liambll.github.io/blog/>

TECHNICAL SKILLS

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

PROFESSIONAL EXPERIENCE

Data Scientist (Coop) – 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 signal and image 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
- Developed term partitioned index mechanism in Java to enable fast document search in Accumulo (HBase version with cell level security)

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
- Managed vendors to deliver Qlikview dashboards for marketing metrics tracking

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
- Led infrastructure design and deployment for an enterprise data management system

Software Engineer (Intern) - Hutcab Consulting, Singapore

May 2011 – Aug 2011

- Implemented inventory management and demand forecasting functions in a Decision Support System with Java Servlet to facilitate optimal inventory decision making
- Reduced data processing time by automating Data Extract, Transform and Load (ETL) tasks with Microsoft SQL Server Integration Services (SSIS)

EDUCATION

MSc. Computer Science (Big Data, Machine Learning) - GPA 4.12/4.33

Sep 2016 – Dec 2017

Simon Fraser University, Burnaby, BC, Canada

- *Big Data Systems, Big Data Analytics, Machine Learning, Deep Learning, Data Mining, Regression Analysis, Statistical Learning*

Liam Bui

📍 Vancouver, BC

☎ (604) 352-1568

✉ liam_bui@outlook.com

🌐 <https://liambll.github.io/blog/>

MSc. Management of Technology

National University of Singapore, Singapore

Aug 2012 – Dec 2013

BEng. Computer Science, BSc. Business Analytics

Nanyang Technological University, Singapore

Aug 2008 – Jun 2012

DATA SCIENCE PROJECTS

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 and conferences
- 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

All State Claim, Kaggle Data Science Competition [[Kaggle](#)]

Sep 2016 – Dec 2016

- Predicted car accident's insurance claim using Linear Regression, Neural Network and Boosted Tree
- Ranked top 20% in the Kaggle competition
- Technologies: Python, scikit-learn, Keras, XGBoost

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: Scala, Spark, 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' attitudes towards movies
- Technologies: Matlab, Matconvnet, Word2Vec