

Industry Projects

- **Data evaluation** on billions of data records for clients from giant banks, e-commerce companies and financial institutions, to understand fraud detection requirements and provide suggestions for later model development.
- **Research** on academic papers and industry competitors, to creatively provide solutions for fraud detection, including feature generation, model building & evaluation.
- **Feature generation** has contributed 50+% important features successfully deployed onto production, leading to 70+% AUC and 70+% AVP in multiple fraud detection projects.
- **Model development** using statistics or machine learning methods with thorough experiments and project specific evaluation metrics, operating on millions of data records on average.
- **Software development** in data science pipeline building, including feature generation pipelines, machine learning pipelines and statistical model pipelines. Participating in frequent code refactor and code review.
- **Data visualization** for customer reports, model performance analysis, data science pipeline monitoring, etc.
- **Cross team collaboration** to deploy data science models to production successfully and bridge the understanding cross teams.

Skills & Tools

- Business Understanding, Information Collection
- Data Preprocessing, Feature Engineering, Machine Learning, Model Evaluation - PSQL, Python
- Data Analysis - PSQL, Python
- Data Visualization - Tableau, Python
- Software Development - Python

**NuData Security, A
MasterCard
Company**
Vancouver, Canada

2017 May -
Now

- **Interviewed** business experts in business loans, individual loans, credit card groups and gained valuable knowledge about Vancity business, in order to provide data science suggestions.
- **Model development for Individual Loan Payback Prediction** with supervised machine learning methods, experimenting on 100+ billion data records, reaching to 96+% balanced accuracy.
- **Model development for Mortgage Loan Purpose Understanding** with Natural Language Processing (NLP) methods to help business experts understand main reasons for clients to borrow mortgage loans.
- **Model development for Customer Complaints Auto Processing in Vancity System Immigration Project**, to help Vancity decision makers quickly understand daily customer complaints when Vancity was immigrating to a new financial system.
- **Model development for Merchant Search Engine Project** to help Vancity clients search for preferred merchants with customized neural network searching ranking system.
- **Data engineering and research** support to bridge Vancity and Simon Fraser University (SFU) data mining lab on **Merchant Entity Recognition Project**.
- **Data preprocessing** on daily **Fair & Fast Loan Reports**, merging data from 30+ tables stored in 3 data warehouses with 10+ billion records, achieved 100% merging accuracy.
- **Tools assessment** on Azure and open source machine learning tools, to provide suggestions to the data science team.

Skills & Tools

- Customer Requirements Collection
- Data Collection - SQL
- Data Preprocessing - SQL, Python, R
- Data Analysis - SQL, Python, R
- Feature Engineering, Machine Learning, Model Evaluation - R, Python, Spark
- Data Visualization - Python, R, PowerBI
- Software Development - Python, Spark

**Vancity Savings
and Credit Union**
Vancouver, Canada

2016 May
- 2016 Dec.

<ul style="list-style-type: none"> • Collected business background and requirements on HP financial systems from HP internal 5+ departments • Understanding on new HP internal financial systems changes and provide training to 200+ HP internal employees • Plan making on multiple training design, training delivery timeline • Frequent communication with business contacts, stakeholders, technical experts and customer representatives <p>Skills</p> <ul style="list-style-type: none"> • Customer Requirements Collection • Business Understanding • Project management • Training Design, Story Telling, Presentation 	<p>HP Shanghai, China</p>	<p>2015 Mar. - 2015 Aug.</p>
<ul style="list-style-type: none"> • Got familiar with all the PowerBI v1 products and attended Power Pivot Language M development • Designed & Implemented SQL intelligent spinning system, to allow SQL Server automatically kill and wake up jobs based on system resources usage, improved system efficiency 25+% • Data Engineering work on Hadoop System, such as ETL work between SQL Server and Hadoop HDFS, advanced Hive Query on big data analysis • Collaborated with data scientist and data architects on Microsoft servers anomalies detection, implemented automatic anomalies detection logic in Hive and applied on 3TB+/day data, found 2 anomalies over 3000+ servers • Implemented automatic data rescue system, prevented the customer from losing 3TB+ data when servers were down • Implemented Windows 8.1 digital app with real time data visualization on teams' performance and data analysis on products usage report, providing decision makers with insights on product marketing campaign and teams' sprint plans <p>Skills & Tools</p> <ul style="list-style-type: none"> • Customer Requirements Collection and Analysis • Relational Database - SQL, Azure Storage, TFS • Big Data & Machine Learning - Hadoop, Hive • Software Development - C#, PowerShell • Production Deployment • Data Visualization - C#, XAML, PowerBI 	<p>Microsoft Redmond, US</p>	<p>2013 Jul. - 2015 Jan.</p>

<ul style="list-style-type: none"> • Data Collection and ETL on Amazon merchants and products data • Backend development on merchants and products search feature, reduced searching time from 20+ minutes to less than 5 minutes • Frontend development on merchants and products search feature <p>Skills & Tools</p> <ul style="list-style-type: none"> • Relational Database - SQL • Software Development - Java, Javascript 	<p>Amazon Seattle, US</p>	<p>2012 May - 2012 Aug.</p>
<h2>Research Projects</h2>		
<ul style="list-style-type: none"> • Implemented & published SFU Comment Extractor, a web scraping tool • Modified & published the code of sentiment analysis tool SO-CAL • Web Scraping & built large scale text corpus from 10,000+ News Articles and 4+ million News Comments • Research & Development to find best solutions for web mining, automatic web searching • Research on Natural Language Processing (NLP) tools • Attended weekly meetings to discuss NLP research papers • Participated crowd sourcing work • Presentation on Spark and NLP <p>Skills & Tools</p> <ul style="list-style-type: none"> • Web Scraping - Python, Open Sources such as Scrapy, Gigya, Selenium, etc. • Web Searching - Google CSE, Bing, XGoogle, Yahoo! Query Language, pygoogle, etc. • Research & Development • NLP - Stanford Core NLP, NLTK, Spacy, etc. • Presentation - Spark Cluster <p>Published Open Source</p> <ul style="list-style-type: none"> • SFU Comment Extractor: http://bit.ly/2pESH6V • SO-CAL Sentiment Analysis: http://bit.ly/2BGx9bR 	<p>SFU Web Scraping Burnaby, Canada</p>	<p>2016 Oct. - 2017 Apr.</p>

<ul style="list-style-type: none"> Developed the front-end for both Desktop version and Mobile version Lie Detection test Captured real time user keyboard typing, mouse movement data, and preprocessed the data <p>Skills & Tools</p> <ul style="list-style-type: none"> Front-End development - jQuery, jQuery Mobile Biometrics Data Capture - C# Data Preprocessing - Java 	<p>University of Arizona Human Computer Interaction Seattle, US</p>	<p>2013 Mar. - 2013 Jun.</p>
<ul style="list-style-type: none"> Executed UW Ollie and Reverb on millions of comments to extract relational nouns, paralyzed phrases and N-ary Generated summarized reports Attended weekly meeting to learn AI in NLP <p>Skills & Tools</p> <ul style="list-style-type: none"> Software Development - Java 	<p>University of Washington Development in NLP Seattle, US</p>	<p>2013 Apr. - 2013 Jun.</p>
<ul style="list-style-type: none"> Web Scraping billions of Mid-East web forums to help detect potential terrorism comments <p>Skills & Tools</p> <ul style="list-style-type: none"> Software Development - Java, SQL 	<p>University of Arizona Web Scraping Tucson, US</p>	<p>2011 Oct. 2012 May</p>
<h2>Academic Projects</h2>		
<p>Detect Underrated StackOverflow Answers Automatically.</p> <ul style="list-style-type: none"> Web Scraping to extract Questions, Answers and Comments from 2000+ posts that has 3+ answers, found API vulnerability that will leak private data Data Preprocessing and Feature Generation on extracted data, applied program analysis methods on the code pieces in Questions, Answers and Comments Machine Learning to predict underrated answers using both supervised methods and unsupervised methods, achieved 90%+ accuracy <p>Skills & Tools</p> <ul style="list-style-type: none"> Web Scraping - Python Data Preprocessing - Python, R Program Analysis - Radon, Pylint Sentiment Analysis - Stanford Core NLP Feature Engineering & Machine Learning - R <p>Project Link http://bit.ly/2DWrNuh</p>	<p>SFU OS, Program Analysis & Cyber Security Burnaby, Canada</p>	<p>2017 Mar. - 2017 Apr.</p>

<p>Study a group of conversational data miners and provide solutions for online education programs.</p> <ul style="list-style-type: none"> • Data Collection through face-to-face interviews • Using social science and qualitative analysis methods to analyze the collected data and extract information as well as hidden requirements • Provide solutions based on extracted information <p>Skills & Tools</p> <ul style="list-style-type: none"> • Research - qualitative analysis • Face-to-face interview, information collection, hidden requirements analysis • User Experience Design <p>Research Paper Link http://bit.ly/2lez5Bn</p>	<p>SFU Human Computer Interaction Burnaby, Canada</p>	<p>2017 Feb. - 2017 Apr.</p>
<p>Analyze Vancouver crime data through visualization.</p> <p>Skills & Tools</p> <ul style="list-style-type: none"> • Visual Design Principles • Data Preprocessing - Python • Data Visualization - Tableau <p>Visualization Link http://tabsoft.co/2pBjhOc</p>	<p>SFU Data Visualization Burnaby, Canada</p>	<p>2017 Feb. - 2017 Mar.</p>
<p>Applied NLP, Web Mining, Search & Ranking, Data Visualization methods in traveling service.</p> <ul style="list-style-type: none"> • Web Mining social media such as Instagram, Flickr, Twitter to generate traveling trends, real popular tourism spots and real time traveling topics • Applied search & ranking and NLP methods to build a traveling focused search engine with neural network real time ranking adjustment, optimized BST, PageRank performance with Spark GraphFrame • Data Visualization on traveling photo post trends <p>Skills & Tools</p> <ul style="list-style-type: none"> • Web Mining - Python, SQLite • Search Engine Creation- Spark, graph algorithms • Traveling features development - Spark Cluster • Data Visualization - d3 <p>Project Link http://bit.ly/2zBvVg3</p>	<p>SFU Applied Big Data Science II Burnaby, Canada</p>	<p>2016 Mar. - 2016 Apr.</p>

<p>Implemented a new e-commerce recommendation systems based on daily news text mining.</p> <ul style="list-style-type: none"> • Web Mining daily news to collect and preprocess text data, generated features for later machine learning analysis • Evaluated multiple machine learning methods that good for text classification, and implemented Non-negative Matrix Factorization (NMF), fisher classifier based on better statistical evidence <p>Skills & Tools</p> <ul style="list-style-type: none"> • Web Mining, Text Mining - Python, MySQL • Data Preprocessing - Python • Machine Learning & Statistics - Python <p>Project Link http://bit.ly/2zBKPCN</p>	<p>SFU Applied Big Data Science I Burnaby, Canada</p>	<p>2015 Dec.</p>
<p>Video movement detection by using LSTM (Long Short-Term Memory) with changed classifier.</p> <ul style="list-style-type: none"> • Replaced the last layer classifier of LSTM with Random Forests, reaching to accuracy 82% and improved the accuracy 15% • Experimented on using unsupervised learning to create clusters as a new feature in supervised learning, reaching to accuracy 85% • Evaluated multiple classifiers such as SVM, neural network, KNN, decision tree, etc. <p>Skills & Tools</p> <ul style="list-style-type: none"> • Data Preprocessing - R, Python • Feature Engineering & Machine Learning - R <p>Project Link http://bit.ly/2CebSAQ</p>	<p>SFU Machine Learning Burnaby, Canada</p>	<p>2015 Dec.</p>
<p>Apply machine learning in hiking trails recommendation system.</p> <ul style="list-style-type: none"> • Web Mining from hiking websites and social media to collect 15+ features for recommendation system • Implemented Jaccard Similarity to recommend hiking trails based on weather, location, user's preference, etc. 15+ features <p>Skills & Tools</p> <ul style="list-style-type: none"> • Data Collection & Data Preprocessing - SQL • Machine Learning - Java 	<p>University of Arizona Data Mining Tucson, US</p>	<p>2012 Apr. - 2012 May</p>

<p>Design and implement a call center system with software design patterns.</p> <ul style="list-style-type: none"> • Designed a call center system with automatic customer services, customer queuing and caller assignment features • Applied 5+ software design patterns and implemented the call center successfully, the design patterns include Abstract Factory, Prototype, Singleton, Factory Method, etc. <p>Skills & Tools</p> <ul style="list-style-type: none"> • System Design • Software Development & Design Patterns - Java 	<p>University of Arizona Software Design Patterns Tucson, US</p>	<p>2012 Nov. - 2012 Dec.</p>
<p>Provide database solutions for a primary school in Tucson, help them manage students information.</p> <ul style="list-style-type: none"> • Back and forth customer communication to understand the school's current system and database design, digging out the hidden customer requirements • Redesigned and implemented the database system by connecting separated tables into relational database <p>Skills & Tools</p> <ul style="list-style-type: none"> • Customer Communication & Information Collection • Business Requirements Understanding • System Design • Relational Database - SQL • Story Telling & Presentation 	<p>University of Arizona Database Systems Tucson, US</p>	<p>2011 Nov. - 2011 Nov.</p>
<p>Provide business solutions to 20+ 2011-2012 business cases in industry with professional story telling and business presentation. Face-to-face customer communication and provide business solutions for 2 network companies.</p> <p>Skills</p> <ul style="list-style-type: none"> • Customer Communication & Information Collection • Business Requirements Understanding • System Design • Story Telling & Presentation 	<p>University of Arizona Business Communication Tucson, US</p>	<p>2011 Sep. 2012 Dec.</p>
<ul style="list-style-type: none"> • Presentation & Lead Discussion • Machine Learning • Data Visualization • Natural Language Processing (NLP) • Data Visualization • Human Computer Interaction • Cyber Security & Program Analysis 	<p>Data Science Presentations</p> <p>http://bit.ly/2BO1KUZ</p>	<p>2015 Dec. - 2017 Apr.</p>

Data Science Side Projects

<ul style="list-style-type: none"> • Statistical Data Preprocessing • Feature Engineering, Dimensional Reduction • Applied Machine Learning Algorithms in detail • Advanced Machine Learning Tools • Multiple Languages - R, Python, Java • Multiple Data Sources 	Data Preprocessing & Machine Learning http://bit.ly/2CONTi4	2015 Aug. - Now
<ul style="list-style-type: none"> • Time Series • Stock Prediction • Natural Language Processing & Neural Network 	Sequential Analysis http://bit.ly/2C2f8sH	2015 Aug. - Now
<ul style="list-style-type: none"> • Digit Recognition • Image Classification • Audio Classification • Neural Network Architect • Deep Learning Learning Notes 	Deep Learning & Neural Network http://bit.ly/2Cfqy9e	2015 Aug. - Now
<ul style="list-style-type: none"> • Web Mining • Text Preprocessing • Search & Rank • Information Retrieval • Sentiment Analysis • Topic Modeling • Entity Extraction • Research & Open Source 	Natural Language Processing (NLP) http://bit.ly/2zDVIEu	2015 Aug. - Now
<ul style="list-style-type: none"> • I mined all the social media you have heard of • NLP & Machine Learning • Business Decision Support • Data Visualization 	Mining Social Media http://bit.ly/2zFLPpv	2015 Aug. - Now
<ul style="list-style-type: none"> • Core Python RDD & DataFrame • Spark Machine Learning 	Spark http://bit.ly/2llrzVp	2015 Aug. - Now
MapReduce in Java: http://bit.ly/2BN4SAa NoSQL HBase: http://bit.ly/2BLh6t8	Hadoop MapReduce & Hadoop HBase	2015 Aug. - 2016 Jan.
<ul style="list-style-type: none"> • Detailed notes & urls in Statistics, Data Preprocessing, Feature Engineering, Machine Learning, Model Evaluation, Data Visualization, etc. Resource 1: http://bit.ly/2Dmvk4b Resource 2: http://bit.ly/2liPpAU	Data Science Resources	2015 Aug. - Now

<ul style="list-style-type: none"> Implemented supervised & unsupervised learning algorithms such as Collaborative Filtering, KNN, Neural Network, K-Means, Decision Tree, etc. Implemented optimization methods such as Genetic Algorithm, Simulated Annealing, etc. Applied to Recommendation System, Search Engine, Optimization Problems, etc. 	Machine Learning Algorithms Implementation http://bit.ly/2BMX5ST	2015 Aug. - Now
Other Experience		
<p>Help kids with cancer to have a wish come true.</p> <ul style="list-style-type: none"> Public Speaking to 300+ people in Space Needle, University Washington, recruited 10+ new volunteers and raised \$2000+ donation Visit sick kids to understand their wishes Design and organize wish events 	Make-A-Wish Volunteer work Seattle, US	2013 Mar. - 2013 Jul.
<ul style="list-style-type: none"> Face-to-face present Karaoke discounts and events to 100+ customers per day Understand customer requirements and suggest suitable Karaoke activities 	Nanjing Melody KTV Sales Promotion Nanjing, China	2010 Dec. - 2011 Feb.
<ul style="list-style-type: none"> Familiar with all the prices, pros & cons of different clothing brands such as CK, Levis, etc. Recommend customers with suitable clothing collocation 	Nanjing Laidi Shopping Mall Sales Promotion Nanjing, China	2009 Dec. - 2009 Dec.

Visit My GitHub for More: <https://github.com/hanhanwu>



2009 - Now

