

## Industry Projects

- Working on user behavior fraud detection, user identification tracking for multiple Fortune 500 banks and e-commerce clients using **statistics, machine learning** and **cost sensitive evaluation** methods, achieved 70% ~ 80% AUC of ROC
- Through **cognitive research** in fraudsters' behavior, creatively **developed typing and mouse movement features** for machine learning, the features are ranking high importance across clients
- Developed **automated feature generation pipeline** and **machine learning pipeline**, reusable across clients
- Generate aggregated analysis and biometrics behavior analysis through **data visualization** using **Tableau, Python & R**, deliver business reports
- Write **advanced SQL queries** to extract facts, insights and convert to knowledge for customer and business delivery
- Apply **statistics, data preprocessing, feature engineering & machine learning methods** in detailed biometrics behavior analysis, attack analysis and model evaluation
- **Drove the implementation** of a mathematics model for MasterCard New York Demo about AI, **experimented** on various data sources, devices and typing behaviors
- **Cross team collaboration** with Architect, Software Developers in **production deployment**, successfully built the technical and business understanding between teams
- Collaborated with Data Engineer in **pre-production big data simulation**, evaluated the deployment risks

### Skills & Tools

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

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

2017 May -  
Now

<ul style="list-style-type: none"> <li>• <b>Advanced SQL queries</b> to generate weekly report for Fair &amp; Fast Loan, providing <b>insights for loan decision makers</b> to help people with good credit to get rid of poverty easier</li> <li>• Interviewed the loan decision makers from <b>all types of Vancity loans</b>, got familiar with the data and the business</li> <li>• Applied <b>Natural Language Processing (NLP)</b> methods in mortgage loans to find patterns in customer requirements</li> <li>• Built a light weighted <b>search engine with real time neural network ranking</b> to help customer search Vancity business merchants fast and accurate</li> <li>• Applied <b>machine learning</b> methods on finding leading factors that influence loan pay back</li> <li>• Machine learning, software development and advanced SQL query on other confidential projects</li> <li>• Face-to-face <b>customer communication</b></li> </ul> <p><b>Skills &amp; Tools</b></p> <ul style="list-style-type: none"> <li>• Customer Requirements Collection</li> <li>• Data Collection - SQL</li> <li>• Data Preprocessing - SQL, Python, R</li> <li>• Data Analysis - SQL, Python, R</li> <li>• Feature Engineering, Machine Learning, Model Evaluation - R, Python, Spark</li> <li>• Data Visualization - Python, R, PowerBI</li> <li>• Software Development - Python, Spark</li> </ul>	<p><b>Vancity Savings and Credit Union</b> Vancouver, Canada</p>	<p>2016 May - 2016 Dec.</p>
<ul style="list-style-type: none"> <li>• Collected business background and requirements on HP financial systems from HP internal 5+ departments</li> <li>• Understanding on new HP internal financial systems changes and provide training to 200+ HP internal employees</li> <li>• Plan making on multiple training design, training delivery timeline</li> <li>• Frequent communication with business contacts, stakeholders, technical experts and customer representatives</li> </ul> <p><b>Skills</b></p> <ul style="list-style-type: none"> <li>• Customer Requirements Collection</li> <li>• Business Understanding</li> <li>• Project management</li> <li>• Training Design, Story Telling, Presentation</li> </ul>	<p><b>HP</b> Shanghai, China</p>	<p>2015 Mar. - 2015 Aug.</p>

<ul style="list-style-type: none"> <li>• Got familiar with all the <b>PowerBI v1 products</b> and attended Power Pivot Language M development</li> <li>• Designed &amp; Implemented <b>SQL intelligent spinning system</b>, to allow SQL Server automatically kill and wake up jobs based on system resources usage, improved system efficiency 25+%</li> <li>• <b>Data Engineering</b> work on <b>Hadoop</b> System, such as <b>ETL</b> work between SQL Server and Hadoop HDFS, <b>advanced Hive Query</b> on big data analysis</li> <li>• Collaborated with data scientist and data architects on Microsoft servers <b>anomalies detection</b>, implemented <b>automatic anomalies detection logic</b> in Hive and applied on <b>3TB+/day</b> data, found 2 anomalies over 3000+ servers</li> <li>• Implemented <b>automatic data rescue system</b>, prevented the customer from losing 3TB+ data when servers were down</li> <li>• Implemented Windows 8.1 digital app with <b>real time data visualization</b> on teams' performance and data analysis on products usage report, providing decision makers with insights on product marketing campaign and teams' sprint plans</li> </ul> <p><b>Skills &amp; Tools</b></p> <ul style="list-style-type: none"> <li>• Customer Requirements Collection and Analysis</li> <li>• Relational Database - SQL, Azure Storage, TFS</li> <li>• Big Data &amp; Machine Learning - Hadoop, Hive</li> <li>• Software Development - C#, PowerShell</li> <li>• Production Deployment</li> <li>• Data Visualization - C#, XAML, PowerBI</li> </ul>	<p><b>Microsoft</b> Redmond, US</p>	<p>2013 Jul. - 2015 Jan.</p>
<ul style="list-style-type: none"> <li>• <b>Data Collection and ETL</b> on Amazon merchants and products data</li> <li>• Backend development on merchants and products <b>search feature</b>, reduced searching time from 20+ minutes to less then 5 minutes</li> <li>• Frontend development on merchants and products search feature</li> </ul> <p><b>Skills &amp; Tools</b></p> <ul style="list-style-type: none"> <li>• Relational Database - SQL</li> <li>• Software Development - Java, Javascript</li> </ul>	<p><b>Amazon</b> Seattle, US</p>	<p>2012 May - 2012 Aug.</p>

## Research Projects

- 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

### Skills & Tools

- 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

### Published Open Source

- **SFU Comment Extractor:** <http://bit.ly/2pESH6V>
- **SO-CAL Sentiment Analysis:** <http://bit.ly/2BGx9bR>

**SFU**  
**Web Scraping**  
Burnaby, Canada

2016 Oct.  
- 2017 Apr.

- 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

### Skills & Tools

- Front-End development - jQuery, jQuery Mobile
- Biometrics Data Capture - C#
- Data Preprocessing - Java

**University of Arizona**  
**Human Computer Interaction**  
Seattle, US

2013 Mar.  
- 2013 Jun.

- 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**

### Skills & Tools

- Software Development - Java

**University of Washington**  
**Development in NLP**  
Seattle, US

2013 Apr.  
- 2013 Jun.

<ul style="list-style-type: none"> <li>• <b>Web Scraping</b> billions of Mid-East web forums to help detect potential terrorism comments</li> </ul> <p><b>Skills &amp; Tools</b></p> <ul style="list-style-type: none"> <li>• Software Development - Java, SQL</li> </ul>	<p><b>University of Arizona</b>  <b>Web Scraping</b>  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"> <li>• <b>Web Scraping</b> to extract Questions, Answers and Comments from 2000+ posts that has 3+ answers, <b>found API vulnerability</b> that will leak private data</li> <li>• <b>Data Preprocessing and Feature Generation</b> on extracted data, applied <b>program analysis</b> methods on the code pieces in Questions, Answers and Comments</li> <li>• <b>Machine Learning</b> to predict underrated answers using both supervised methods and unsupervised methods, achieved 90%+ accuracy</li> </ul> <p><b>Skills &amp; Tools</b></p> <ul style="list-style-type: none"> <li>• Web Scraping - Python</li> <li>• Data Preprocessing - Python, R</li> <li>• Program Analysis - Radon, Pylint</li> <li>• Sentiment Analysis - Stanford Core NLP</li> <li>• Feature Engineering &amp; Machine Learning - R</li> </ul> <p><b>Project Link</b>  <a href="http://bit.ly/2DWrNuh">http://bit.ly/2DWrNuh</a></p>	<p><b>SFU</b>  <b>OS, Program Analysis &amp; Cyber Security</b>  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"> <li>• <b>Data Collection</b> through face-to-face interviews</li> <li>• Using <b>social science</b> and <b>qualitative analysis</b> methods to analyze the collected data and <b>extract information</b> as well as <b>hidden requirements</b></li> <li>• <b>Provide solutions</b> based on extracted information</li> </ul> <p><b>Skills &amp; Tools</b></p> <ul style="list-style-type: none"> <li>• Research - qualitative analysis</li> <li>• Face-to-face interview, information collection, hidden requirements analysis</li> <li>• User Experience Design</li> </ul> <p><b>Research Paper Link</b>  <a href="http://bit.ly/2lez5Bn">http://bit.ly/2lez5Bn</a></p>	<p><b>SFU</b>  <b>Human Computer Interaction</b>  Burnaby, Canada</p>	<p>2017 Feb.  - 2017 Apr.</p>

<p>Analyze Vancouver crime data through visualization.</p> <p><b>Skills &amp; Tools</b></p> <ul style="list-style-type: none"> <li>• Visual Design Principles</li> <li>• Data Preprocessing - Python</li> <li>• Data Visualization - Tableau</li> </ul> <p><b>Visualization Link</b>  <a href="http://tabsoft.co/2pBjhOc">http://tabsoft.co/2pBjhOc</a></p>	<p><b>SFU</b>  <b>Data Visualization</b>  Burnaby, Canada</p>	<p>2017 Feb.  - 2017 Mar.</p>
<p>Applied NLP, Web Mining, Search &amp; Ranking, Data Visualization methods in traveling service.</p> <ul style="list-style-type: none"> <li>• <b>Web Mining social media</b> such as Instagram, Flickr, Twitter to generate traveling trends, real popular tourism spots and real time traveling topics</li> <li>• Applied search &amp; ranking and NLP methods to build a <b>traveling focused search engine</b> with <b>neural network real time ranking</b> adjustment, <b>optimized BST, PageRank performance</b> with Spark GraphFrame</li> <li>• <b>Data Visualization</b> on traveling photo post trends</li> </ul> <p><b>Skills &amp; Tools</b></p> <ul style="list-style-type: none"> <li>• Web Mining - Python, SQLite</li> <li>• Search Engine Creation- Spark, graph algorithms</li> <li>• Traveling features development - Spark Cluster</li> <li>• Data Visualization - d3</li> </ul> <p><b>Project Link</b>  <a href="http://bit.ly/2zBvVg3">http://bit.ly/2zBvVg3</a></p>	<p><b>SFU</b>  <b>Applied Big Data Science II</b>  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"> <li>• <b>Web Mining</b> daily news to collect and preprocess text data, <b>generated features</b> for later <b>machine learning</b> analysis</li> <li>• <b>Evaluated</b> multiple <b>machine learning methods</b> that good for text classification, and implemented Non-negative Matrix Factorization (NMF), fisher classifier based on better statistical evidence</li> </ul> <p><b>Skills &amp; Tools</b></p> <ul style="list-style-type: none"> <li>• Web Mining, Text Mining - Python, MySQL</li> <li>• Data Preprocessing - Python</li> <li>• Machine Learning &amp; Statistics - Python</li> </ul> <p><b>Project Link</b>  <a href="http://bit.ly/2zBKPCN">http://bit.ly/2zBKPCN</a></p>	<p><b>SFU</b>  <b>Applied Big Data Science I</b>  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"> <li>Replaced the last layer classifier of LSTM with Random Forests, reaching to accuracy 82% and improved the accuracy 15%</li> <li>Experimented on using <b>unsupervised learning</b> to create clusters as a new feature in <b>supervised learning</b>, reaching to accuracy 85%</li> <li>Evaluated multiple classifiers such as SVM, neural network, KNN, decision tree, etc.</li> </ul> <p><b>Skills &amp; Tools</b></p> <ul style="list-style-type: none"> <li>Data Preprocessing - R, Python</li> <li>Feature Engineering &amp; Machine Learning - R</li> </ul> <p><b>Project Link</b>  <a href="http://bit.ly/2CebsAQ">http://bit.ly/2CebsAQ</a></p>	<p><b>SFU</b>  <b>Machine Learning</b>  Burnaby, Canada</p>	<p>2015 Dec.</p>
<p>Apply machine learning in hiking trails recommendation system.</p> <ul style="list-style-type: none"> <li><b>Web Mining</b> from hiking websites and social media to collect 15+ features for recommendation system</li> <li>Implemented <b>Jaccard Similarity</b> to recommend hiking trails based on weather, location, user's preference, etc. 15+ features</li> </ul> <p><b>Skills &amp; Tools</b></p> <ul style="list-style-type: none"> <li>Data Collection &amp; Data Preprocessing - SQL</li> <li>Machine Learning - Java</li> </ul>	<p><b>University of Arizona</b>  <b>Data Mining</b>  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"> <li><b>Designed a call center system</b> with automatic customer services, customer queuing and caller assignment features</li> <li>Applied 5+ <b>software design patterns</b> and implemented the call center successfully, the design patterns include Abstract Factory, Prototype, Singleton, Factory Method, etc.</li> </ul> <p><b>Skills &amp; Tools</b></p> <ul style="list-style-type: none"> <li>System Design</li> <li>Software Development &amp; Design Patterns - Java</li> </ul>	<p><b>University of Arizona</b>  <b>Software Design Patterns</b>  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"> <li>• Back and forth <b>customer communication</b> to understand the school's current system and database design, <b>digging out the hidden customer requirements</b></li> <li>• Redesigned and implemented the database system by connecting separated tables into relational database</li> </ul> <p><b>Skills &amp; Tools</b></p> <ul style="list-style-type: none"> <li>• Customer Communication &amp; Information Collection</li> <li>• Business Requirements Understanding</li> <li>• System Design</li> <li>• Relational Database - SQL</li> <li>• Story Telling &amp; Presentation</li> </ul>	<p><b>University of Arizona Database Systems</b> 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><b>Skills</b></p> <ul style="list-style-type: none"> <li>• Customer Communication &amp; Information Collection</li> <li>• Business Requirements Understanding</li> <li>• System Design</li> <li>• Story Telling &amp; Presentation</li> </ul>	<p><b>University of Arizona Business Communication</b> Tucson, US</p>	<p>2011 Sep. 2012 Dec.</p>
<ul style="list-style-type: none"> <li>• Presentation &amp; Lead Discussion</li> <li>• Machine Learning</li> <li>• Data Visualization</li> <li>• Natural Language Processing (NLP)</li> <li>• Data Visualization</li> <li>• Human Computer Interaction</li> <li>• Cyber Security &amp; Program Analysis</li> </ul>	<p><b>Data Science Presentations</b></p> <p><a href="http://bit.ly/2BO1KUZ">http://bit.ly/2BO1KUZ</a></p>	<p>2015 Dec. - 2017 Apr.</p>
<p><b>Data Science Side Projects</b></p>		
<ul style="list-style-type: none"> <li>• Statistical Data Preprocessing</li> <li>• Feature Engineering, Dimensional Reduction</li> <li>• Applied Machine Learning Algorithms in detail</li> <li>• Advanced Machine Learning Tools</li> <li>• Multiple Languages - R, Python, Java</li> <li>• Multiple Data Sources</li> </ul>	<p><b>Data Preprocessing &amp; Machine Learning</b></p> <p><a href="http://bit.ly/2C0NTi4">http://bit.ly/2C0NTi4</a></p>	<p>2015 Aug. - Now</p>
<ul style="list-style-type: none"> <li>• Time Series</li> <li>• Stock Prediction</li> <li>• Natural Language Processing &amp; Neural Network</li> </ul>	<p><b>Sequential Analysis</b></p> <p><a href="http://bit.ly/2C2f8sH">http://bit.ly/2C2f8sH</a></p>	<p>2015 Aug. - Now</p>



<ul style="list-style-type: none"> <li>Digit Recognition</li> <li>Image Classification</li> <li>Audio Classification</li> <li>Neural Network Architect</li> <li>Deep Learning Learning Notes</li> </ul>	<b>Deep Learning &amp; Neural Network</b> <a href="http://bit.ly/2Cfqy9e">http://bit.ly/2Cfqy9e</a>	2015 Aug. - Now
<ul style="list-style-type: none"> <li>Web Mining</li> <li>Text Preprocessing</li> <li>Search &amp; Rank</li> <li>Information Retrieval</li> <li>Sentiment Analysis</li> <li>Topic Modeling</li> <li>Entity Extraction</li> <li>Research &amp; Open Source</li> </ul>	<b>Natural Language Processing (NLP)</b> <a href="http://bit.ly/2zDVIEu">http://bit.ly/2zDVIEu</a>	2015 Aug. - Now
<ul style="list-style-type: none"> <li>I mined all the social media you have heard of</li> <li>NLP &amp; Machine Learning</li> <li>Business Decision Support</li> <li>Data Visualization</li> </ul>	<b>Mining Social Media</b> <a href="http://bit.ly/2zFLPpv">http://bit.ly/2zFLPpv</a>	2015 Aug. - Now
<ul style="list-style-type: none"> <li>Core Python RDD &amp; DataFrame</li> <li>Spark Machine Learning</li> </ul>	<b>Spark</b> <a href="http://bit.ly/2llrzVp">http://bit.ly/2llrzVp</a>	2015 Aug. - Now
MapReduce in Java: <a href="http://bit.ly/2BN4SAa">http://bit.ly/2BN4SAa</a> NoSQL HBase: <a href="http://bit.ly/2BLh6t8">http://bit.ly/2BLh6t8</a>	<b>Hadoop MapReduce &amp; Hadoop HBase</b>	2015 Aug. - 2016 Jan.
<ul style="list-style-type: none"> <li>Detailed notes &amp; urls in Statistics, Data Preprocessing, Feature Engineering, Machine Learning, Model Evaluation, Data Visualization, etc.</li> </ul> Resource 1: <a href="http://bit.ly/2Dmvk4b">http://bit.ly/2Dmvk4b</a> Resource 2: <a href="http://bit.ly/2liPpAU">http://bit.ly/2liPpAU</a>	<b>Data Science Resources</b>	2015 Aug. - Now
<ul style="list-style-type: none"> <li>Implemented supervised &amp; unsupervised learning algorithms such as Collaborative Filtering, KNN, Neural Network, K-Means, Decision Tree, etc.</li> <li>Implemented optimization methods such as Genetic Algorithm, Simulated Annealing, etc.</li> <li>Applied to Recommendation System, Search Engine, Optimization Problems, etc.</li> </ul>	<b>Machine Learning Algorithms Implementation</b> <a href="http://bit.ly/2BMX5ST">http://bit.ly/2BMX5ST</a>	2015 Aug. - Now
<b>Other Experience</b>		
Help kids with cancer to have a wish come true. <ul style="list-style-type: none"> <li><b>Public Speaking</b> to 300+ people in Space Needle, University Washington, recruited 10+ new volunteers and raised \$2000+ donation</li> <li>Visit sick kids to understand their wishes</li> <li><b>Design and organize wish events</b></li> </ul>	<b>Make-A-Wish Volunteer work</b> Seattle, US	2013 Mar. - 2013 Jul.

<ul style="list-style-type: none"> <li>• <b>Face-to-face present</b> Karaoke discounts and events to 100+ customers per day</li> <li>• <b>Understand customer requirements</b> and suggest suitable Karaoke activities</li> </ul>	<b>Nanjing Melody KTV Sales Promotion</b> Nanjing, China	2010 Dec. - 2011 Feb.
<ul style="list-style-type: none"> <li>• Familiar with all the prices, pros &amp; cons of different clothing brands such as CK, Levis, etc.</li> <li>• <b>Recommend customers</b> with suitable clothing collocation</li> </ul>	<b>Nanjing Laidi Shopping Mall Sales Promotion</b> Nanjing, China	2009 Dec. - 2009 Dec.

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



2009 - Now

