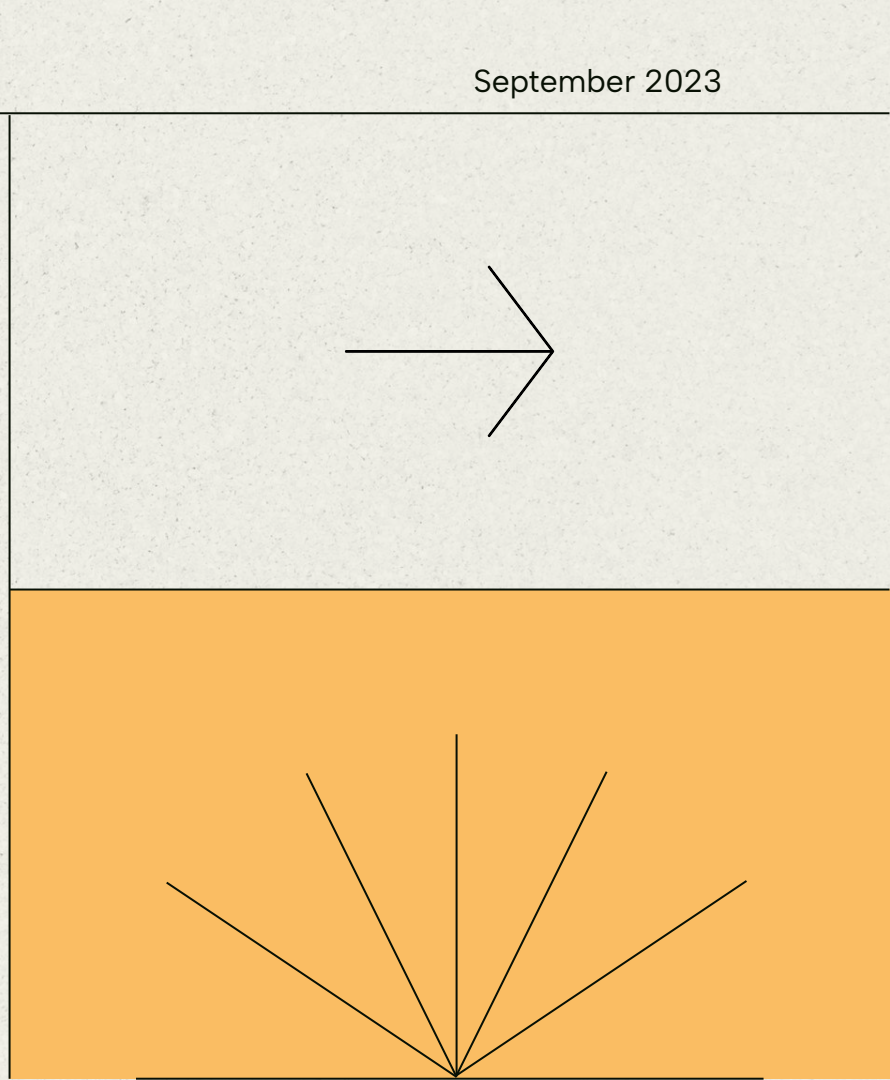
ALGER REMIRATA

September 2023

DATA SCIENTIST APPLICATION

Details of Tech Projects



Previous Companies

01 Neuroglee Therapeutics Pte Ltd

Neuroglee transitioned from digital therapeutics to virtual care company. Neuroglee provides scalable and evidence-based programs to manage cognitive care that leverages continuous monitoring, aiming for personalized care delivery using patient application developed inhouse, and monitored from a remote multidisciplinary clinical team.

Details: https://www.neuroglee.com/

02 Jewel Paymentech Pte Ltd

The company has been part of Advanced AI since 2022. Jewel Paymentech is a financial risk technology company that develops intelligent risk solutions for the banking and electronic payment industry.

Details: https://www.jewelpaymentech.com/

03 Savvysherpa Inc.

The company was acquired by Unitedhealth Group (UHG) in 2018 and now part of Optum Labs. Savvysherpa is a research-based company that combines healthcare economics and data science to produce valuable healthcare assets.

Details: https://www.linkedin.com/company/savvysherpa-inc/

Highlights

Healthcare
Biotechnology
Financial Technology
Ecommerce
Payments

HIGHLIGHTS

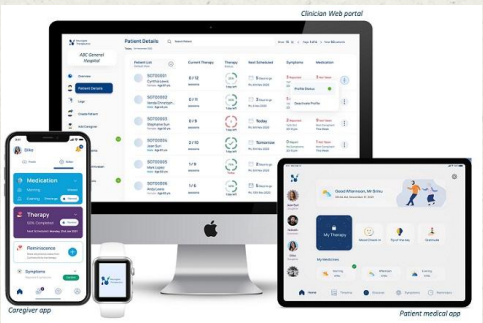
- More than nine years of experience in data science including end to end projects starting from Ideation and Research to Product Development and Product Maintenance
- Strong experience with Healthcare, Clinical, Payment, Ecommerce datasets
- Deep experience with machine learning (supervised, unsupervised, semi-supervised and self-supervised)
- Vast experience in deep learning (convolutional neural networks, recurrent neural networks, transformer models)
- Comprehensive experience with various databases (MSSQL/Postgres, MongoDB, Neo4j, AWS DynamoDB/Athena, Google Big Query, ElasticSearch/ Elastic Stack)
- Extensive work with big data analytics such as Apache Spark (PySpark/Scala), Hive, Hadoop
- Profound experience with visualization (Tableau, Bokeh, HoloView, MatplotLib, Seaborn)
- Substantial experience in Natural Language Processing (sentiment analysis, topic modelling, information retrieval)
- Significant experience in Speech Analytics and Audio Signal Processing
- Robust experience with Digital Signal Processing (accelerometer, continuous glucose monitor)
- Experience with Computer Vision (image processing, video analytics)
- Diverse and well-rounded skills in handling various datasets: transactional data, application interaction data, content data, healthcare claims data, clinical data, physiological data, wearable signal data, text data, audio data, images, video, etc.)

–

Neuroglee

Digital Therapeutics, Virtual Care,
Personalized Care Delivery,
Neuroscience

PROOF OF CONCEPT PROJECTS

	Project/Task Description
1. Analysis of Digital Biomarkers from NG001 and NG Connect 	NG001: https://alz.confex.com/alz/2023/meetingapp.cgi/Paper/75630 NG Connect: https://www.outsourcing-pharma.com/Article/2021/09/09/Neuroglee-Mayo-Clinic-partner-on-remote-care-for-Alzheimer-s - With consultation from cognitive scientist, we extract digital biomarkers based on patient's interaction data obtained from the applications developed by the company: NG001 and NG Connect - These digital biomarkers will be extracted, processed or computed based on the type of data collected. Example, techniques from digital signal processing will be used to extract features derived from wearable data upon performing wellness activity (either meditation or yoga). Techniques from natural language processing is used for analysis of textual and speech data derived from reminiscence activity.

NEUROGLEE THERAPEUTICS PTE LTD

November 2021 – Present

	Project/Task Description
2. Topic Modeling and Sentiment Analytics Pipeline	- Design and create topic modelling and sentiment analytics pipeline for text and speech data coming from feedback, patient outcomes, digital assessment and reminiscence therapy to be used for digital therapeutics and personalized virtual care delivery - Exploration on various topic modelling techniques and libraries such as Mallet's Latent Dirichlet Allocation, BERTopic, LDA2Vec - Exploration on various sentiment analytics libraries such as roBERTa - Visualization of topics and sentiments using Tableau
3. Recommendation System Framework for Personalized Care Delivery	- Collaborate with cognitive scientist, data engineers and cross-functional teams to design and create recommendation system framework for personalized care delivery - The recommendation system is a combination of rule-based engine based on domain knowledge from cognitive scientist, machine learning and deep learning techniques.

NEUROGLEE THERAPEUTICS PTE LTD

November 2021 – Present

	Project/Task Description
4. Speech Analytics Pipeline	- Design and create machine learning based speech analytics pipeline to determine presence of cognitive impairment with features computed based on techniques from audio signal analysis, digital signal processing (spectral analysis, fast fourier transform and wavelet transform), natural language processing and deep learning - Explored various speech analytics library such as librosa, pyaudio, pydub and deep learning frameworks such as HuggingFace, wav2vec, HuBert
5. Multiprocessing and Multithreading in Feature Extraction for Wearable	- Experience in using Apache PySpark and Dask for multiprocessing and multithreading of feature extraction functionalities on wearable data

NEUROGLEE THERAPEUTICS PTE LTD

November 2021 – Present

PRODUCTION-BASED PROJECTS

	Project/Task Description
1. Data Analytics and Predictive Analytics Pipeline for NG Connect and NG001	- Design and create data and predictive analytics pipeline/framework for the company's main products: NG Connect and NG001. - Create reporting, data analytics and predictive analytics dashboard with use cases across multiple departments, including senior leadership team, clinical care and operations, product and technology teams - These dashboards has been initially created using Tableau, Bokeh and Holoview. Data extraction techniques involves data processing using SQL and pandas.
2. Data Collection Schema for Learn, Wellness and Calendar Modules	- Create data collection schema and data validation scripts which serve as guides for the engineering team. Use of "jsonschema" library to generate validation scripts. - Highlight observations and bugs needed to be fixed based on data generated by the application validated against proposed schema
3. Analysis of Cognitive Assessment Data	- Experience in analysis and interpretation of several cognitive assessment such as NTB, MoCA, GAD-7, PROMIS, MMSE, MBI-C, DASS, etc.

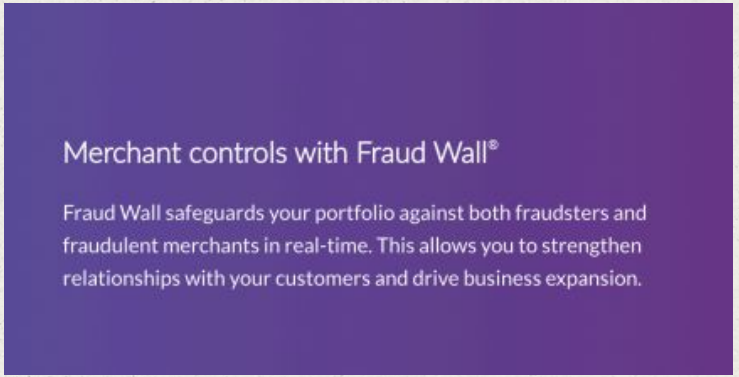
Jewel Paymentech

An Advanced AI company

Financial Risk and
Compliance Technology
Company

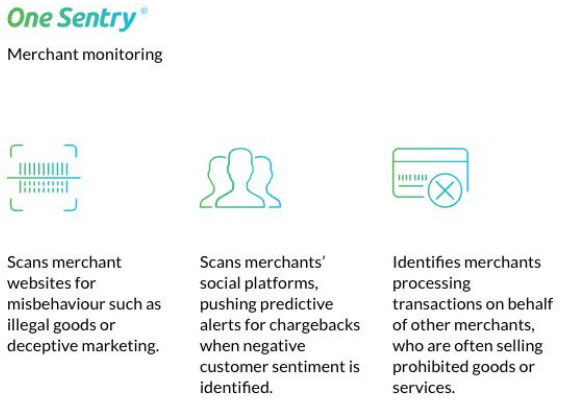
JEWEL PAYMENTECH PTE LTD – (END TO END PRODUCTION PROJECTS)

November 2018 – November 2021



Fraudwall

One Sentry



JEWEL PAYMENTECH PTE LTD – (END TO END PRODUCTION PROJECTS)

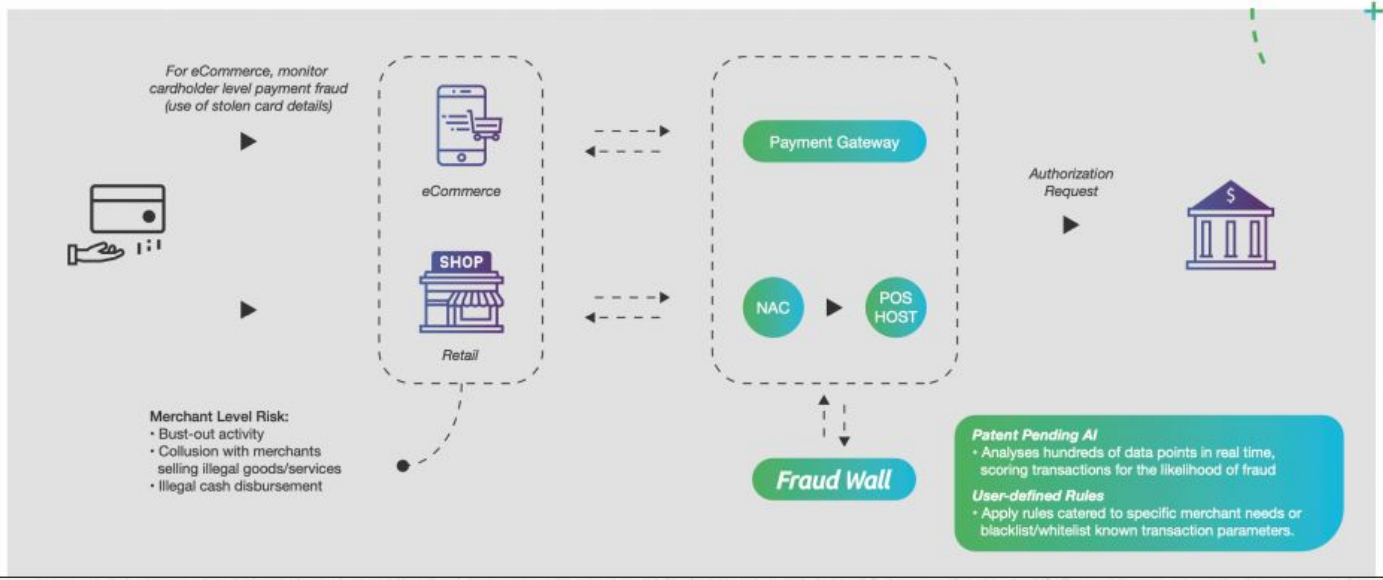
November 2018 – November 2021

Fraud Wall®

Protect yourself from every attack

Fraudsters are getting smarter and more adaptive. Fraud Wall is a comprehensive and dynamic risk management engine that shuts down fraud before it can happen. By combining AI with your fraud expertise, it lets you to effectively manage merchant level fraud with a holistic view of activity across your entire portfolio.

By taking an active approach to combatting fraud, Fraud Wall allows your merchants to provide better customer experiences. Through the reduction of both fraud losses and false decline rates, you can significantly increase cardholder spend and improve customer retention.



JEWEL PAYMENTECH PTE LTD – (END TO END PRODUCTION PROJECTS)

November 2018 – November 2021

One Sentry® Marketplace

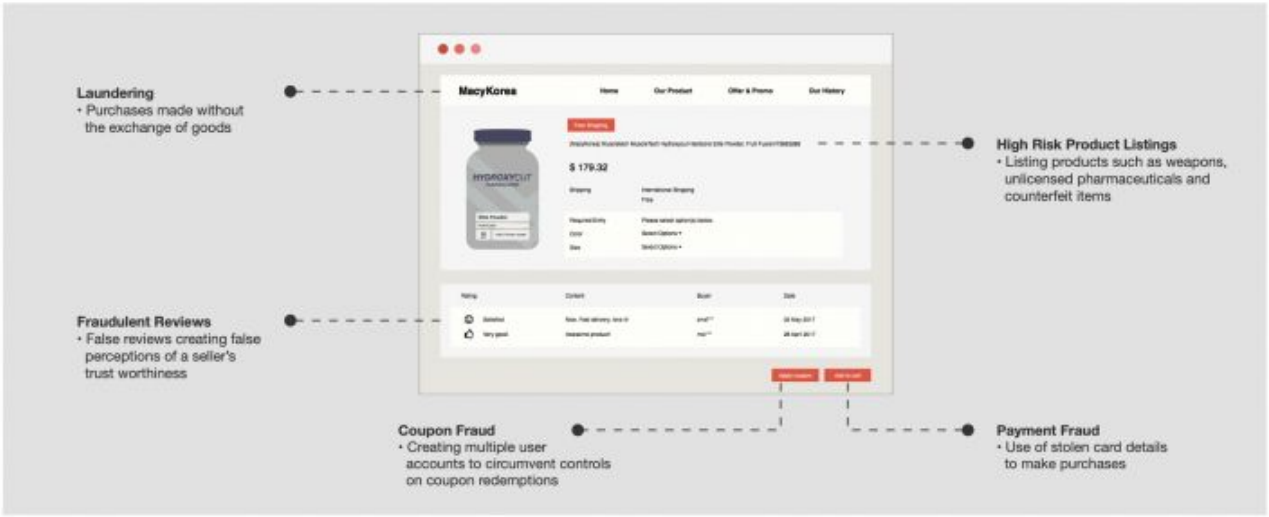
A safer marketplace for all

The online marketplace can be an attractive place for fraudsters. One Sentry Marketplace analyses elements of the member lifecycle to detect and prevent fraudulent activity. Whether it's analysing purchase patterns, online reviews or your membership portfolio, this intelligent platform can report on a range of illegal activities in real time.

For example, One Sentry Marketplace combats the sale of unlawful products. From illegal items (such as unlicensed pharmaceuticals or weapons) to counterfeit goods, it allows you to mitigate the risk of facing hefty regulatory fines and legal action.

Other concerns surrounding new merchant activity are coupon fraud (where buyers and sellers monetise promotion incentives) and fraudulent product reviews – both of which can negatively impact your business.

Whichever issue you are facing, One Sentry Marketplace allows you to keep ahead of fraudsters, even as they adapt their methods of attack.

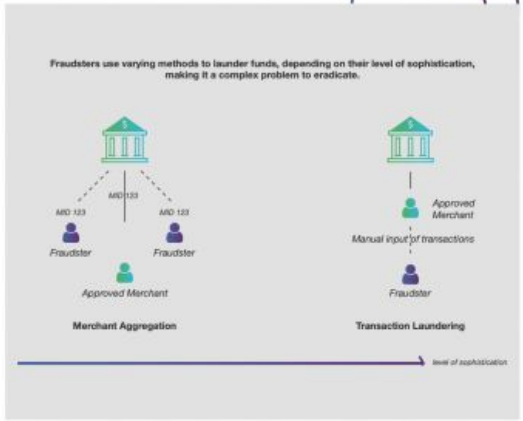


One Sentry®

Welcoming eCommerce with confidence

One Sentry is an advanced engine that monitors your merchants' web, social and transactional behavior, alerting you in the event of unusual activity. This means you can implement preemptive controls if a new merchant starts behaving suspiciously after onboarding. One Sentry can also detect and eradicate cases of merchant laundering – where legal businesses are used as a front for the sale of illegal products.

With a detailed view of your portfolio activity, you can accept more eCommerce merchant applications with confidence. If any new merchants that you acquire turn out to be unlawful, you can significantly reduce or prevent the impact of their activities on your business.



JEWEL PAYMENTECH PTE LTD – (END TO END PRODUCTION PROJECTS)

November 2018 – November 2021

	Project/Task Description
Fraudwall AI Engine	**Dynamic Real Time Fraud Prevention System** - combination of both rule based method (plugins) and machine learning method that score each transaction based on its fraud risk in a span of milliseconds **Static Deep Learning Real Time Fraud Prevention System** - use of deep learning methodology to score each transaction based on fraud patterns learned from historical data
Responsibilities within Fraudwall AI	- Responsible for anything involve with Fraudwall AI from ideation to production deployment (end to end development) - Work with fraud analyst and business intelligence team to identify patterns used by fraudsters based on chargeback, issuer response data and domain knowledge - Responsible with code development of all needed functionalities, bug fixes, feature enhancement, server maintenance and components upgrade. Also frequent model retraining to ensure efficiency of the machine learning models. - Point person to investigate Fraud Wall AI production problems identified by the fraud analysts teams, development team, products team, platform team or the client itself. - Part of the responsibility is maintenance of the server where AI engines has been deployed like maintenance and upgrade of mongoDB, elasticsearch and neo4j.

JEWEL PAYMENTECH PTE LTD – (END TO END PRODUCTION PROJECTS)

November 2018 – November 2021

	Project/Task Description
One Sentry AI Engine	One Sentry AI uses techniques from natural language processing to classify website contents into one of the 15 categories. One of the categories is a valid website while the other 14 categories are high risk website which involves MLM (multi level marketing), selling alcohol, etc.
Responsibilities within One Sentry AI	- Responsible for anything involve with One Sentry AI from ideation to production deployment (end to end development) - Exploratory project on creating a recommendation engine for merchant category code based on its website's contents. - Project in production involves classification of merchant website based on 15 categories including normal and those that are of high risk such as alcohol, MLM, gambling, prostitution, etc. - Point person to investigate One Sentry AI production problems identified by the fraud analysts teams, development team, products team, platform team or the client itself. - Responsible with code development of all needed functionalities, bug fixes, feature enhancement, server maintenance and components upgrade. Also frequent model retraining to ensure efficiency of the machine learning models.

Savvysherpa Inc.

Merged with United Health Group and became UHG Research and Development
Now part of Optum Labs

Healthcare,
US Medical Insurance

SAVVYSHERPA INC – (RESEARCH-BASED PROJECTS)

May 2013 – October 2018

	Project/Task Description (Savvysherpa Research Projects)
2013 - 2014	Started working in Savvysherpa as researcher. Initial work involves being familiar with healthcare claims data and creating SQL queries, subqueries and stored procedures to answer business and research problems such as medical adherence, medical risk adjustment factor computation, healthcare utilization, insurance claims cost analysis, medical claims management
June - November 2014	Savvysherpa acquired Red Brain Labs, a machine learning tech company based in Draper Utah. Together with them, we work on a project involving optimization of call center operations and telemarketing in one of our clients. Created supervised machine learning model to predict sales, predict call duration, predict sale of add-ons. Also work with optimising matching between caller and agent within a telemarketing campaign (both inbound and outbound).
November 2014 - May 2015	Savvysherpa partnered with a telecommunications company. Through their call logs data, we infer demographic information of our prepaid subscribers based on the location logs of the calls made by their subscribers. Also, we tried to infer which of their subscribers are travelers, home buddy, living near commercial areas and working on night shift bases.
May 2015 - Feb 2016	Create own implementation of block principal pivoting method (BPP), a state of the art method of matrix factorization and tensor factorization. Implement the algorithm in Scala (the backend language of Apache Spark) and inserted the algorithm to Apache Spark. Through this project, skill sets on Scala, Apache Spark (and PySpark), Hadoop and Hive has been developed. Involve with configuration setting and installation of Apache Spark in the company

SAVVYSHERPA INC – (RESEARCH-BASED PROJECTS)

May 2013 – October 2018

	Project/Task Description (Savvysherpa Research Projects)
Feb 2016 - Dec 2016	Data mining projects using factorization methods on terabytes of data from Unitedhealth Group by the use of Apache Spark. Discover research studies on comorbidities of diseases, fraudulent claims submission, fraudulent procedures conducted by specific providers and disease progression. Also involve in algorithm creation for bioidentity verification of the company's motion program (https://unitedhealthcaremotion.com/) using techniques of digital signal processing on accelerometer data.
January 2017 - May 2017	Applied deep learning methods such as convolutional neural network (CNN) and recurrent neural network (such as long short term memory networks (LSTM)) on identifying disease progression of type-2 diabetes patients. Use LSTM to identify the diagnosis for the next doctor visit of a patient, based on the paper "Doctor AI: Predicting Clinical Events via Recurrent Neural Networks"
June - July 2017	Undergo extensive training in data science, machine learning and deep learning through IMPACT (Interdisciplinary Mentoring Program in Analysis, Computation and Theory) in Brigham Young University in Utah
July 2017 - Sept 2017	Attempted to use CNN to identify activities within a 7 minute workout video but abandoned the idea of solving the problem using video and image data. Instead we make use of accelerometer data and LSTM to classify signal data within the 7 minute workout accelerometer data.

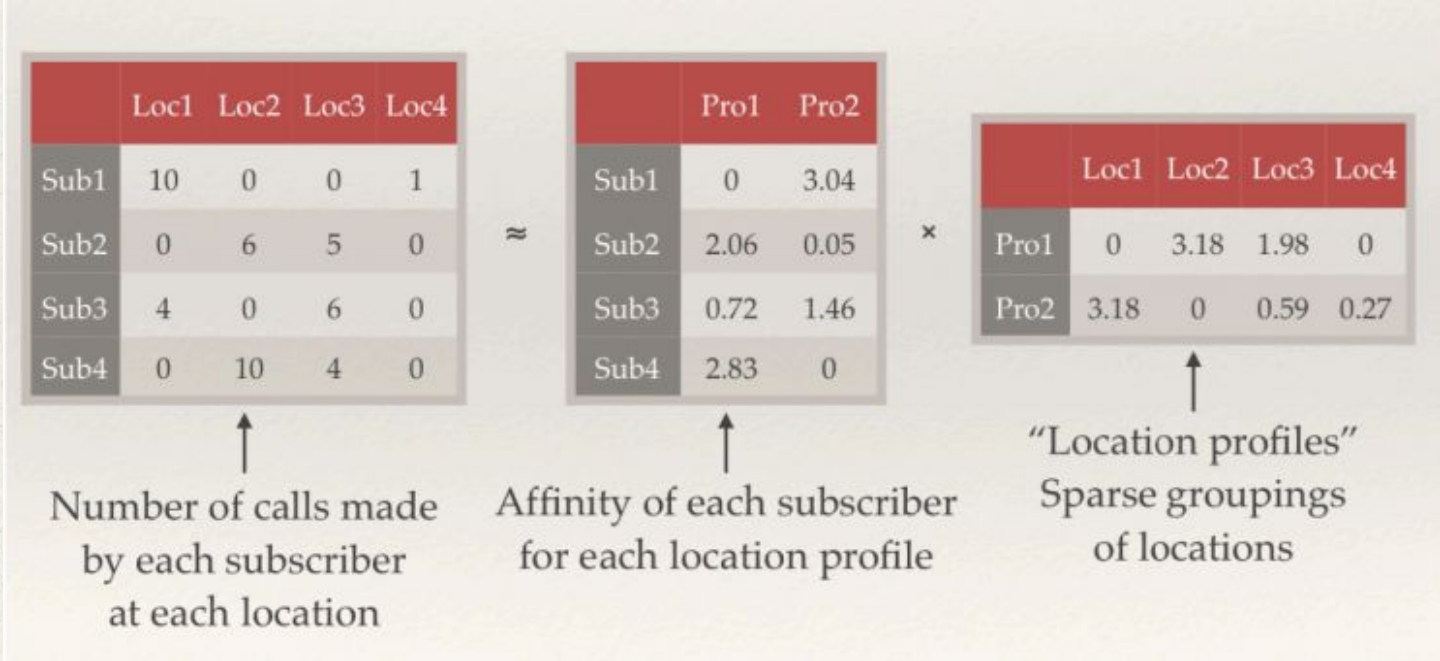
SAVVYSHERPA INC – (RESEARCH-BASED PROJECTS)

May 2013 – October 2018

	Project/Task Description (Savvysherpa Research Projects)
Oct 2017 - Dec 2017	Create in-house medical word embedding model trained on Unitedhealth Group Insurance claims data to discover disease comorbidities in disease progression projects and to be used as embedding layer for the deep learning models for healthcare applications. Also create customised word embedding models trained on 2016 PubMed dataset which is used for internal document query tool.
January - March 2018	Apply word embedding to discover off label drugs used for some chronic diseases. Also create a regression model to be used as proxy for risk adjustment factor (RAF) cost model.
April 2018 - June 2018	Part of the team working on creating algorithm for meal detection using continuous glucose monitor (CGM) data and accelerometer data. Detecting meal has been part of diabetes management program of the company in order for members to perform physical activities few minutes after the meal.
June 2018 - August 2018	One of the mentors for the new data scientist on training hired by the company. As one of the participants of IMPACT in 2017, I was assigned to handle optimization lessons and administer laboratory exercises that revolves around optimization especially convex optimization.
August 2018 - October 2018	Continue with creation of meal detection algorithm using CGM data and accelerometer data. Also work with project that study progression of neuropathy and its comorbidities (on its initial stage)

Discovering Latent Location Profiles Application to Telecommunications Data

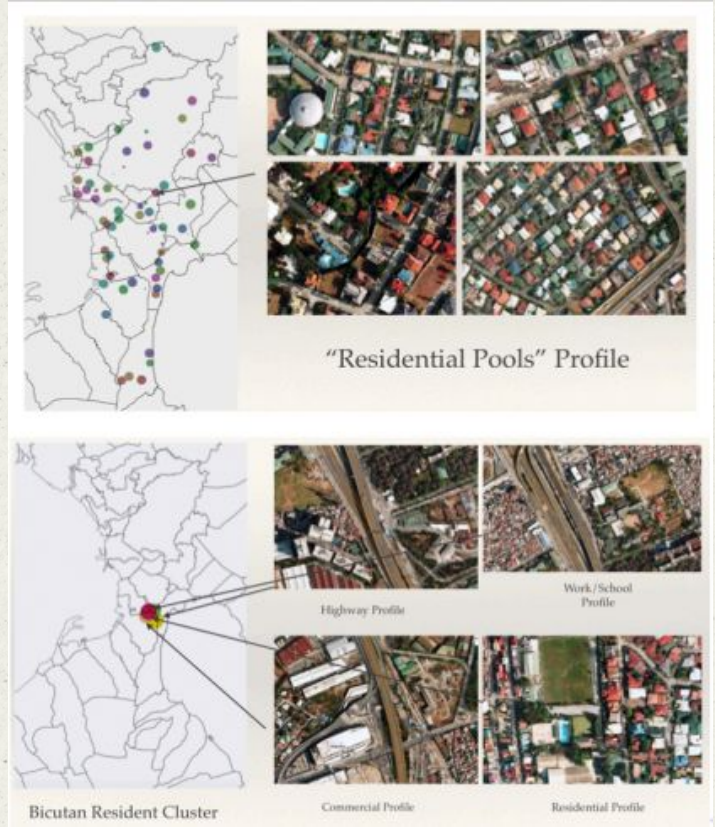
Problem Statement: Using prepaid call logs data, can we discover latent demographic information of prepaid subscribers based on location and duration of calls they have made?



SAVVYSHERPA INC – (RESEARCH-BASED PROJECTS)

May 2013 – October 2018

Discovering Latent Location Profiles Application to Telecommunications Data



SAVVYSHERPA INC – (RESEARCH-BASED PROJECTS)

May 2013 – October 2018

Extension of Using Matrix Factorization to Tensor Factorization Application to Telecommunication and Healthcare Data

Problem 1: Using call logs data, can we describe mobility of our prepaid subscribers, i.e, we want to identify who among our subscribers are home buddy, traveler, night-shift workers, etc.

Problem 2: With terabytes of claims data we have in Unitedhealth Group serving approximately 20% of the US population, how can we mine this data, discover patterns within our members?

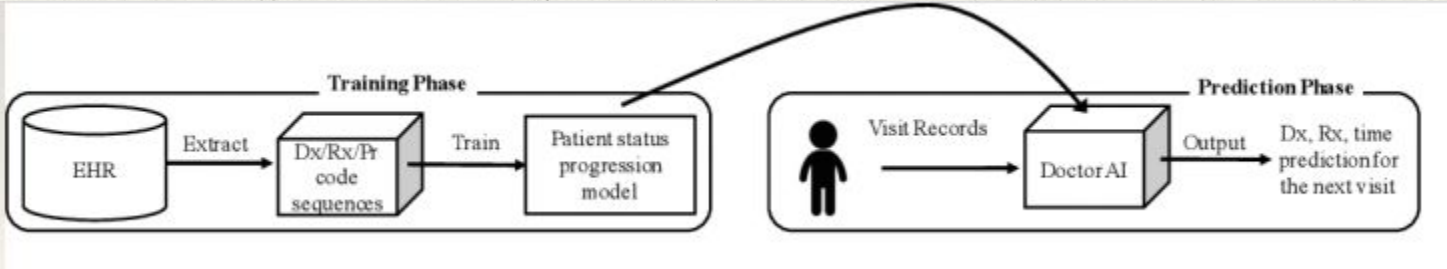
Solution:

- A better algorithm for factorization: Block principal pivoting method
- An algorithm that will handle more than 2 data entities: Tensor factorization (instead of matrix factorization)
- A platform to process terabytes of data: Hadoop and Apache Spark

Natural Language Processing Application to Healthcare

Problem Statement: Predict future diagnosis and cost of some chronic diseases using historical claims data.

Solution: Use deep learning LSTM to predict future diagnosis by storing each of the procedure codes into time-arranged sequences. Since the embedding layer of each of these deep learning architectures are random by default, we trained an in-house medical word embedding trained on historical claims data to represent an embedding layer in our LSTM models. This is based on the paper "Doctor AI: Predicting Clinical Events via Recurrent Neural Network" which predict the diagnosis of a patient in the next visit given historical medical codes.



Doctor AI: Predicting Clinical Events via Recurrent Neural Networks

SAVVYSHERPA INC – (RESEARCH-BASED PROJECTS)

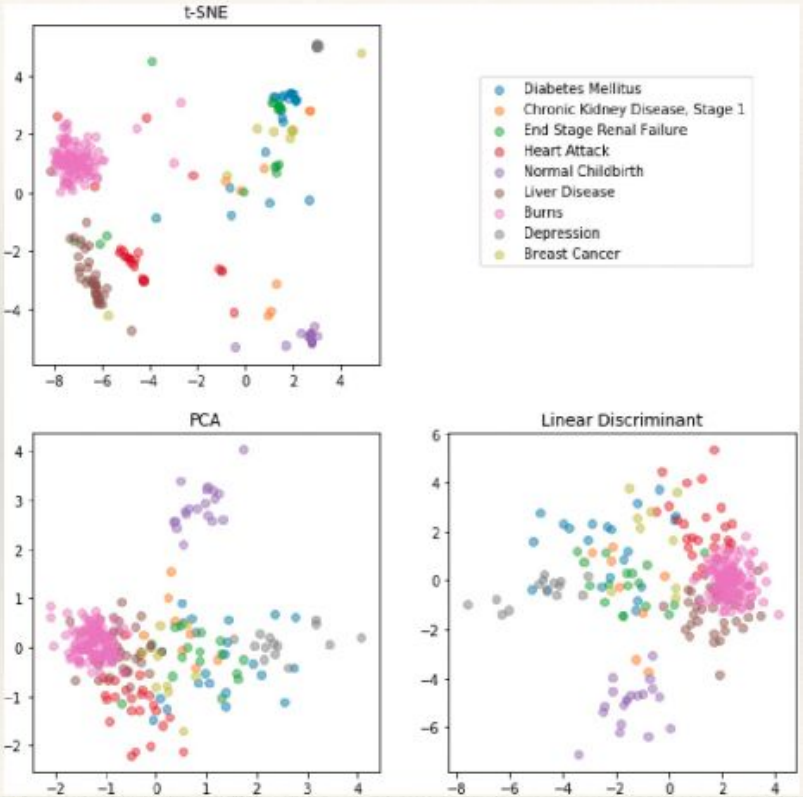
May 2013 – October 2018

Natural Language Processing Application to Healthcare

Problem Statement: Construct tools for document query and disease comorbidities discovery using word embedding trained from insurance claims data.

Solution: The use of word embedding to discover discover disease comorbidities is based on the paper “Code2Vec: Embedding and Clustering Medical Diagnosis Data”

For document query, we create a word embedding model trained on all articles of PubMed in 2017. Apache Spark has been used to preprocess the data.



code2vec: Learning Distributed Representations of Code

SAVVYSHERPA INC – (RESEARCH-BASED PROJECTS)

May 2013 – October 2018

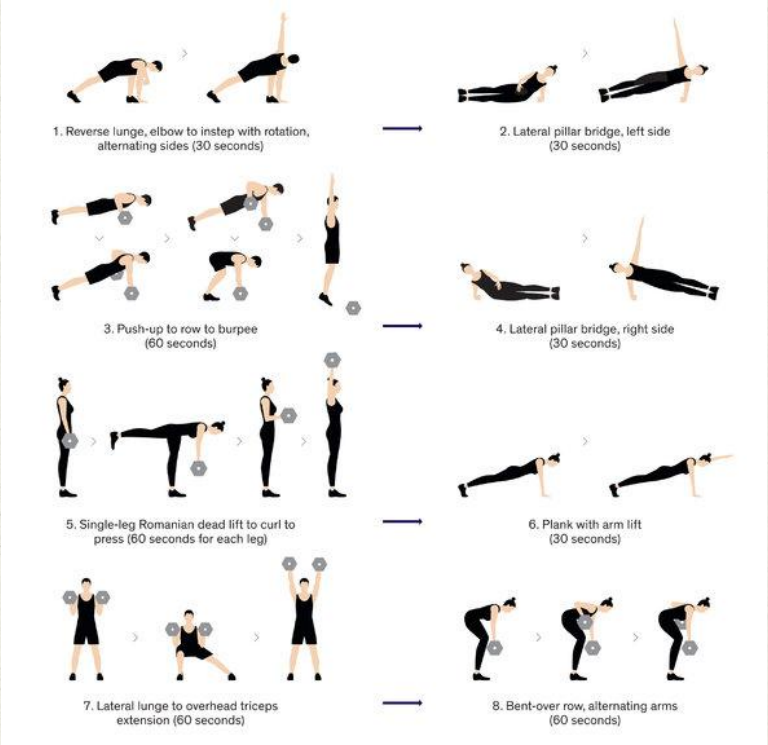
Application of Digital Signal Processing Healthcare and Motion Programs

Problem Statement: Can we classify activity within a 7 minute workout?

Solution:

1. Use convolutional neural network to tag activities within a 7 minute workout video. This is a very hard problem wherein we tried to classify each frame (images) that consists the video. The approach has been dropped.

2. Use LSTM to classify activities in an accelerometer data used within the 7 minute workout. We get better results for this approach getting more than 0.8 overall accuracy.



7 minute workout activities

Other Responsibilities and Skill Sets

OTHER RESPONSIBILITIES AND SKILL SETS

- Experience in data engineering as one of the employees heavily involved in transitioning mongodb-based cloud application platform to elasticsearch.
- Point person in maintaining neo4j databases which is used primarily by our Fraud Detection AI Engine
- Experience in SQL and nosql: Mongodb, elastic search, Hadoop and Hive, Microsoft SQL (T-SQL), Postgres
- Experience in creating interactive visualisation using Bokeh and Jupyter Widgets
- Experience in Big Data such as Apache Spark and deep learning using Keras, tensor flow
- Experience in model interpretability using approaches like LIME (Local Interpretable Model-Agnostic Explanations), ICE (Individual Conditioning Explanation) and others
- Participant of Summer 2017 Interdisciplinary Mentoring Program in Analysis, Computation and Theory held in Brigham Young University in Utah. Sent by the company for this training on scientific computing, data science, machine learning and optimisation
- Lead the data science team on its priority projects, objectives and key results. Involve with manager’s decision planning and resource prioritisation representing data science team. Conduct mentoring to junior research scientist.

THANKS!

DO YOU HAVE ANY QUESTIONS?

abremirata21@gmail.com
+65 8709 9187
https://www.linkedin.com/in/alger-remirata-066011107/