

# Alger B. Remirata

DATA SCIENTIST · RESEARCH SCIENTIST  
MS MATHEMATICS

EXTENSIVE BACKGROUND IN HEALTHCARE/BIO TECHNOLOGY, EXPERIENCE IN PAYMENTS AND RISK ANALYTICS

08-56 Blk 502 Bedok North St 3 Singapore 460502

☎ (+65) 87099187 | ✉ abremirata21@gmail.com | 🌐 aremirata | 📧 alger-remirata | 📞 alger-remirata-066011107

## Highlights

### HIGHLIGHTS

Singapore(5 years), PH(5 years)

**DATA SCIENCE/RESEARCH** (HEALTHCARE, PAYMENTS/ECOMMERCE)

May 2013 - Present

- More than nine years of experience in data science starting from ideation, research, development, deployment and maintenance
- Vast experience with machine learning (supervised, unsupervised, semi-supervised)
- Experience in deep learning (convolutional neural networks, recurrent neural networks, transformer models)
- Comprehensive experience with various databases (MSSQL/Postgres, MongoDB, Neo4j, DynamoDB/Athena, Big Query, Elasticsearch)
- Significant experience with big data analytics such as Apache Spark (PySpark, Scala), Hive, Hadoop
- Substantial experience in Data Visualization, Natural Language Processing, Speech Analytics, Signal Processing, Computer Vision
- 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, videos, etc.)

## Data Science Work Experiences

### NEUROGLEE THERAPEUTICS PTE LTD

Singapore

**ARTIFICIAL INTELLIGENCE SCIENTIST** [FULL TIME PERMANENT] (HEALTHCARE, DIGITAL THERAPEUTICS)

November 2021 - October 2023

#### • **Proof of Concept Projects:**

- Collaborate with neuroscientists, data engineers and cross-functional teams to design and build recommendation system framework for personalized care delivery. The framework developed has been handed over to engineering team for application integration.
- Design and construct 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 and natural language.
- Design and create topic modeling 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.
- Collaborate with neuroscientist to identify, extract, process and analyze digital biomarkers from application interaction data, wearable, audio and text data coming from four main modules: games, learning, wellness and reminiscence.

#### • **Production-Based Project:**

- Plan, design and execute data and predictive analytics pipeline/framework and scope of work for the company's main product. Responsible for the creation of reporting, data analytics and predictive analytics dashboards that has been utilized by multiple departments including leadership, clinical operations, content, product, and technology teams.
- Design, create and manage data collection processes and pipelines, ensuring data accuracy and integrity for the patient application. Experience with open-source customer data platforms such as rudderstack.
- Experience in using Tableau, Bokeh, Holoviews and iPython widgets to create interactive visualization for reporting, data analytics and predictive analytics dashboard.

### JEWEL PAYMENTECH PTE LTD

Singapore

**SENIOR RESEARCH SCIENTIST** [FULL TIME PERMANENT] (FINANCIAL RISK COMPLIANCE TECHNOLOGY COMPANY)

November 2018 - November 2021

- Lead the data science team on its priority projects, objectives and key results. Work with the products team, engineering team, support team, professional services and devops for planning and decision making of the company's existing products. Conduct training and mentoring to junior research scientists and interns.

#### • **Production-Based Project:**

- Improve real-time fraud detection system by introducing new plugins based on trends identified by fraud analysts. The new plugins created decrease false negative rates obtained from chargeback resulting to less than 0.01 false negative rates across all clients.
- Enhance fraud detection models using autoencoder and one-class adversarial networks.
- Experience working with API supporting real-time fraud detection systems including logging, configuration setting and AI engine management. Strong experience in working with No-SQL databases demonstrated by successfully transitioning mongodb-based cloud application platform to elasticsearch. This technique allowed current cloud application to process concurrently 46 transactions per second from previously 23 transactions per second.
- Experience in working with Graph databases. Extensive experience with Neo4j and graph-based models including creating queries, database setup, data migration and query optimization.

#### • **Proof of Concept Projects:**

- Build a recommendation engine to identify corresponding merchant categories based on their website contents. The recommendation engine achieves over 0.8 average F1-score on more than 60 categories.
- Develop interactive visualization using bokeh applications for fraud monitoring.

## Research Work Experiences

### SAVVYSHERPA INC (now part of UNITED HEALTHGROUP (UHG) OPTUM LABS)

Cebu City, Philippines

RESEARCHER [FULL TIME PERMANENT] (SPECIALIZATION: HEALTHCARE, MACHINE LEARNING, DEEP LEARNING, BIG DATA) [FULL TIME PERMANENT]

May, 2013 - October 2018 (5.5 years)

- Reports to Head of Algorithms Team based in Draper, Utah, USA. Work with senior and principal scientist based in Utah, Minnesota and Cebu. Train junior researchers under the company's version of BYU's IMPACT program.
- **Research-based Projects:**
  - Create machine learning and deep learning framework to predict future diagnosis and cost of some chronic diseases using historical healthcare claims data.
  - Construct tools for document query and disease comorbidities discovery using word embeddings trained from insurance claims data and medical text data. The trained word embedding has been used by the entire team as embedding layer for LSTM models used for disease progression projects.
  - Conducted research on off label drugs by training medical word embeddings on healthcare claims data.
  - Implement algorithms of matrix and tensor factorization in scala and Apache Spark. The algorithms created are used in an attempt to create scalable analytics tools in big data platforms that can be used for prediction problems and recommender systems, pattern discovery, community detection and anomaly analysis on large set of medical claims data from United Health Group.
  - Construct algorithms for meal detection using digital signal processing from accelerometer and continuous glucose monitor. Interpret results using LIME (Local Interpretable Model-Agnostic Explanations) and ICE (Individual Conditional Expectation). The models has been used for research studies on Type 2 Diabetes.
  - Apply long short term memory (LSTM) networks and convolutional neural networks (CNN) to classify activities based on their accelerometer data. The project is used to support existing motion programs offered as services by the company.
  - Apply factor analysis techniques to discover latent demographic information of prepaid subscribers based on location and duration of calls they have made. The project is part of the client's marketing segmentation strategies.
  - Apply supervised machine learning methods to predict sales and call duration for telemarketing cost optimization. The models has been successfully deployed to telemarketing operations of the client in European countries.
  - Create advanced SQL (MSSQL (T-SQL), Hive, Postgres) queries to process healthcare claims data and aid in answering problems in healthcare such as medical adherence and risk computation. The tables generated has been utilized by other researchers for understanding medical adherence throughout an existing clinical program.

## Education

### UNIVERSITY OF THE PHILIPPINES (LOS BAÑOS)

Off-Campus Offering at UP-Cebu

M.S. IN MATHEMATICS

Aug. 2014 - June 23, 2018

- **Thesis:** *Compressed Separable Nonnegative Matrix Factorization With Random Projections and its Application to Biomedical Image Processing*
- Enter the graduate program after 1 year working in Savvysherpa. Manages to balance studying theoretical mathematics while learning skills in data science for work.

### BRIGHAM YOUNG UNIVERSITY

Provo, Utah

VISITING SHORT-TERM SCHOLAR

June 5, 2017 - July 21, 2017

- Participant of Summer 2017 Interdisciplinary Mentoring Program in Analysis, Computation and Theory
- Coursework covers Inner Product Spaces, Spectral Theory, Probability, Statistics, Constrained and Unconstrained Optimization, Convex Analysis, Linear Programming, Logistic Regression, Bayesian Statistics, Linear SVM and Regularization, Expectation Maximization, Nearest Neighbor and Mixture Models, Neural Networks, Decision Trees, Random Forests, Boosting and Aggregation
- Laboratory problems on scientific computing and data science

### UNIVERSITY OF THE PHILIPPINES (CEBU)

Lahug, Cebu City, Philippines

B.S MATHEMATICS

June 2007 - April 2011

- Department of Science and Technology scholar; Working student and a regular tutor while in the university

## Skills and Technology

- **Core Skills:** Python, Numpy, Scikit-Learn, Pandas, Natural Language Processing, Audio and Digital Signal Processing, Machine Learning, Deep Learning, Statistical Analysis, Data Analytics, Data Visualization
- **Big Data and Functional Programming:** Apache Spark, Scala, Dask, Hadoop
- **Deep Learning Programming:** Keras, TensorFlow, PyTorch, MXNet
- **Visualization:** Tableau, Bokeh, Matplotlib, Holoviews, Panel, Vega
- **Databases:** Elasticsearch, MongoDB, Neo4J, Redis, PostgreSQL, MSSQL, Google Big Query, AWS (Athena, Sagemaker)
- **Others:** Git, Bitbucket, Confluence, Latex, Linux, Unix, SQL, NLTK, Spacy, Haproxy, Sonarqube, Amazon S3, Flask, Kibana, Logstash, SearchGuard, AWS SSM, Supervisor, Apache, Rudderstack, JIRA, Confluence, Rudderstack