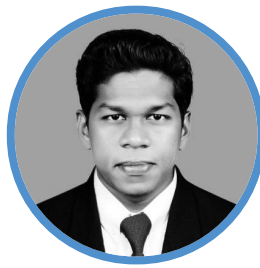


# Emmanuel Raphael

## DATA DEVELOPER

Data Researcher and former Software Developer offering an insatiable intellectual curiosity and passion for turning raw data into products, actionable insights, and meaningful stories. Inspired by open data initiatives, artificial intelligence and data science.



raphelemmanuel@gmail.com ✉

+1-416-807-6971 📞

Toronto, Canada 📍

www.emmanuelraphel.com 🌐

www.github.com/raphelemmanuel 🐙

www.linkedin.com/in/raphelemmanuel in

www.twitter.com/rafa\_emmanuel 🐦

## PROFESSIONAL EXPERIENCE

### Data Research Assistant Intern ThinkData Works Inc. 🔗

05/2017 – 08/2017

Toronto, Canada

ThinkData is an open data startup that allows developers, businesses, and citizens to access high value data in standardized formats through the data management platform - Namara.io.

#### Achievements/Tasks

- Orchestrated systematic researches to discover and import data from various portals and provided better insights to businesses by incorporating relevant open data-sets with their internal data through Namara.io.
- Developed scripts to scrape raw data on-the-fly from about 100 data portals using Ruby and removed coarse data from data-sets of sizes about 5 GB using data wrangling tools.
- Build lightweight Docker containers to develop, test, deploy and ship the Ruby scripts securely and portably, thus to automate on-board deployments for the clients.

Contact: Bryan Smith – [linkedin.com/in/bryan-smith-68282122](https://www.linkedin.com/in/bryan-smith-68282122)

### Software Developer UST Global 🔗

12/2014 – 12/2015

Kochi, India

#### Achievements/Tasks

- Designed and developed applications using best software implementation practices for the full stack development of a Fortune 50 Healthcare client's Shopping Portal project.
- Developed a MEAN-Stack prototype of the legacy application in a team, as proof of concept for a lightweight architecture for the portal for its AWS migration, which received client's recognition and proposal for the project.
- Recommended design and security improvements after a security breach and for the existing web application accessed by 800 million users, earning client's appreciation.

Contact: Premchand Chandrasekharan Nair – +91-9895-173-813

## EDUCATION

### Post Graduate Certificate In Cloud Computing for Big Data LAMBTON COLLEGE 🔗

01/2016 – 08/2017

Toronto, Canada

### Bachelors of Technology in Electrical and Electronics Engineering MAHATMA GANDHI UNIVERSITY 🔗

07/2010 – 04/2014

Kottayam, India

## TECHNICAL EXPERTISE

Machine Learning (Clustering, Classification)

Docker

Web Scraping

Java

Spring MVC Framework

Ruby (Mechanize, Nokogiri)

JSON | XML

Python (NumPy, scikit-learn, nltk)

Node.js

RESTful Web-Services

MEAN Stack

Heroku

MongoDB | mLab

Oracle Database 10g

API.AI

Amazon S3

Google Cloud Storage

AngularJS

GitHub

Bitbucket

Tableau

OpenRefine

Javascript | jQuery

JIRA

Selenium Web-driver

## PROJECT EXPERIENCE

### Unsupervised Learning for SMS Spam Detection (02/2017 – 03/2017)

- Developed machine learning algorithm for clustering an unlabeled data set of 5574 SMS messages into 'spam' and 'ham' using k-means and Python machine learning packages (scikit-learn and NumPy)
- Implemented the feature selection algorithm employing a bypass filter to remove low frequency words from the pre-processed bag-of-words and used 'tfidf' document representation which improved the accuracy of the clustering results from .85 to .94.
- Tuned the k-means settings with different random state values which further boosted the accuracy to .9659 and achieved 2nd rank of 22 participants in the competition hosted in kaggle.

### Multiclass Classification for Sentimental Analysis of Amazon Book Reviews (03/2017 – 04/2017)

- Implemented a multi-class classification algorithm using Python to classify sentiment of a data set of book reviews.
- Employed 'scikit-learn' to use 80% of the pre-processed data for training and 20% for test to avoid overfitting of data and used term-frequency document representation to extract the top-features, which boosted the accuracy by 40%.
- Modeled a 'SVM SVC' classifier using the training data to predict the labels for the test data and obtained F1 score of .76.

## ACHIEVEMENTS

### GRADitude Award (04/2017)

Received GRADitude Award in recognition for outstanding and distinguished service to Lambton College in Toronto.

### Best Project of the Year in Small Enterprises Category (09/2015)

Team member of the 'Shopping Portal Project for Health Insurance Plans' (UST Global) which won Project Management Institute of India's 'Best Project of the Year-2015' at PMI India Awards.