

Siddharth Arora

 [linkedin.com/in/sidd-arora](https://www.linkedin.com/in/sidd-arora)  github.com/sarora  sarora.github.io
 +1 (438) 342 8552  sidd.arora.05@gmail.com
 Canada

An experienced data scientist and software developer, with expertise in developing robust, scalable data pipelines, and using machine learning techniques to extract key insights from large complex datasets.

I excel at succinctly communicating these insights to broader, non-technical audiences.

TECHNICAL SKILLS

Programming Languages	Python (numpy, pandas, scikit-learn, keras, tensorflow, open-cv, spacy) R (tidyverse, ggplot, forecast, Shiny)
Machine Learning	Classification, regression, clustering, feature engineering, hyper-parameter tuning
DataBases	PostgreSQL, MS SQL, MySQL
Statistical Methods	Hypothesis testing, confidence intervals, survival analysis, principal component analysis and dimensionality reduction, Bayesian inference
Cloud Services	AWS (EC2, S3, EMR)
DevOps	Travis CI, Docker, Docker-Compose, Kubernetes
Workflow Tools	Agile, Jira, Confluence, BitBucket, Github, Team Foundation Server, Trello, RStudio, Visual Studio, PyCharm, Lucid Chart

EXPERIENCE

Nov 2018 Present	Data Scientist Stradigi AI, Montreal, Canada Stradigi AI provides state of the art AI platform that empowers users to build and execute solutions to complex business challenges faster and in a secure way. In a team of 5: <ul style="list-style-type: none">➤ Designed and developed data pre-processing and transformation libraries for data cleaning and manipulation using Python for Stradigi's AI platform.➤ Built Machine Learning (ML) pipelines using classical and deep learning algorithms from data ingest through to solutions for specific use cases like sentiment analysis, document classification, name-entity recognition, image classification, object detection, and image segmentation.➤ Devised and developed a benchmarking framework to analyze and optimize Stradigi's Machine Learning pipelines against competitors to achieve state of the art performance.➤ Worked closely with Researchers and AI Developers in an Agile setting to ensure production quality software. <div>Python Pandas Numpy Tensor Flow Docker BitBucket Scikit AWS</div>
Apr 2018 Jun 2018	Data Science Intern (UBC Capstone Project) Unbounce, Vancouver, Canada Unbounce provides an online platform that enables marketers or digital agencies to create, publish, and test landing pages, pop-ups or sticky bars to improve conversion rates. In a team of 4: <ul style="list-style-type: none">➤ Applied survival analysis modeling techniques using R to determine underlying behaviors that impact customer retention at Unbounce.➤ Developed and automated the data pipeline using python to extract and transform data into usable components for modeling.➤ Utilized easily decipherable visualizations to report findings to a non-technical audience. <div>Python Pandas R Survival Analysis Trello Github</div>

Siddharth Arora

 [linkedin.com/in/sidd-arora](https://www.linkedin.com/in/sidd-arora)  github.com/sarora  sarora.github.io
 +1 (438) 342 8552  sidd.arora.05@gmail.com
 Canada






May 2016
Dec 2016

Data Scientist (Contract) | [Brewbound.co.nz](https://brewbound.co.nz), Auckland, New Zealand

Brewbound is an e-commerce startup providing fast online liquor delivery service in New Zealand.

In a team of 2:

- Cleaned and wrangled data on a regular basis to generate meaningful datasets from marketing data.
- Identified trends, patterns, and relationships in marketing data to provide answers to business questions, particularly regarding the effectiveness of various advertising channels.
- Communicated findings to key decision-makers to help facilitate the decision making process.

Sep 2013
Feb 2016

Software Developer | [IMS Health](https://ims.health), Auckland, New Zealand

IMS Health (now IQVIA) is a leading global company providing clients in the healthcare industry with comprehensive solutions to measure and improve their performance.

- Developed and updated functionalities for IMS Health's Nexus Mobile Intelligence - a CRM solution, that helps drive customer engagement for life sciences industry using C#, JavaScript, XML.
- Performed maintenance and bug fixing on the existing code base to ensure production quality software.
- Participated in daily stand-ups and bi-weekly sprint meetings in an Agile setup and developed interpersonal skills.

EDUCATION

Aug 2017 - Jun 2018 **Master of Data Science**
University of British Columbia, Vancouver, Canada

Aug 2009 - Aug 2013 **Bachelor of Engineering with First Class Honors - Software Engineering**
University of Waikato, New Zealand








Personal Projects

Ongoing

[Twitter Sentiment Analysis](#) | Personal Project.

A project to analyze tweets and classify them as either positive, negative or neutral.

- Gathered open-source datasets and performed cleaning, pre-processing, tokenization, and vectorization to construct features for training and testing step.
- Utilized classical and deep learning algorithms to train and test on created features, to classify tweets as positive, negative or neutral.
- Queried the Twitter API via Tweepy - a python library, to get the most recent tweets and apply created models to classify sentiment.

For more projects and blog posts please visit sarora.github.io