Justin **Charbonneau**

Montréal, QC

(819) 319-3622 | justin.m.charbonneau@gmail.com | justincharbonneau.github.io | in/charbonneaujustin/

Industry Experience

Data Scientist (CO-OP)

Ottawa, Ontario

Ario Platform — *Jan. 2019 - Aug. 2019*

- Created an R package for collecting Google Trends data and aggregating results across time and location
- Performed data collection and data cleaning while ensuring the data is representative of the distribution that is expected for specific projects
- Developed an image processing and computer vision pipeline to extract information from image files using Python and open-source libraries such as OpenCV for its image processing capabilities; applied and compared several OCR engines (AWS, Google and Microsoft)
- Prepared and presented a Jupyter Notebook to the product development team to showcase the image processing pipeline

Programmer

Ottawa, Ontario

Fisheries and Oceans Canada — *Jan. 2018 - Dec. 2018*

- Improved the accuracy of the information stored in the database by developing a PHP web application that identified discrepancies and allowed users to validate the information and track the validation progress
- Exploited APIs to update the information in the configuration management database
- Displayed summary data using Google Chart API

Junior System Analyst

Ottawa, Ontario

Fisheries and Oceans Canada — *Apr. 2017 - Dec. 2017*

- Integrated parameterized SQL views in Cognos Query Studio to allow clients to query user specific subsets of the data
- Produced multiple ad-hoc reports using SQL queries for providing detailed server and database metrics to management
- Wrote a PowerShell script to send customized emails according to configurable RDBMS stored information

Academic Experience

Teacher Assistant | Introduction à l'Intelligence Artificielle

Ottawa, Ontario

University of Ottawa — *Sep. 2019 - PRESENT*

- Prepared Python Jupyter Notebooks utilizing libraries (Scikit-Learn, NetworkX, Matplotlib) for fourth year students
- Provided help to students on topics such as searching algorithms (Best-First, A*, Greedy) and machine learning algorithms

Research Assistant | Knowledge Discovery and Data Mining (KDD) Lab

Ottawa, Ontario

University of Ottawa — *Sep. 2019 - PRESENT*

- Researched and applied deep learning techniques (Word Embeddings, Convolutional Neural Networks, Recurrent Neural Networks) using Keras and acquired all experimental data for other researchers
- Supervised master students working on the research project (helped with Python, strategizing their research topics and taught how to manage their source control); led meetings and demos

Skills

Python	Great working experience with Python and familiarity with many libraries (Scikit-learn, Pandas/NumPy, NLTK, Gensim)
R	Intermediate level using R and familiarity using popular libraries (Tidyverse, Shiny, FlexDashboard, RStan)
Machine Learning	Classification (Naive Bayes, K-Nearest Neighbor, Support Vector Machines); clustering; recommender systems
Deep Learning	Keras framework in Python; Tensorboard; fine-tuning; Convolutional Neural Networks, Recurrent Neural Networks
Time Series	Regression Analysis, Auto Regressive Integrated Moving Average (ARIMA), Prophet
Statistics	Good applied statistics skills, hypothesis testing, regression analysis, and Bayesian inference using STAN
SQL	Experience with several SQL dialects (MS SQL, SQL Server and Oracle) and exposure to MongoDB and ETL tools
Data Visualization	Good working knowledge of producing graphs using ggplot2 in R and matplotlib in Python; experience using Tableau
Version Control	Familiarity using Git and some experience using Subversion
Languages	French (Fluent), English (Fluent)

Education

Master in Electronic Business Technologies

University of Ottawa, Ontario

Research project: Deep Learning Models for Stock Market Prediction using market data and news articles — *Jan. 2018 - Exp. Dec. 2019*

B.com in Management Information Systems and Analytics (MISA)

University of Ottawa, Ontario

Received Merit Scholarship for maintaining an above average of 85 percent — *Jan. 2014 - Dec. 2017*

Projects and Certifications

Stacked Machine Learning Model for Stock Market Prediction (In Development) *Sept. 2019 - Present*

- Used Natural Language Processing algorithms for producing word embeddings using the gensim package in Python. Used those weights in a Keras embedding layer and attached a 1D Convolutional Neural Netowork (CNN) for learning word group filters to classify text
- Used the Keras implementation of a Bidirectional Long Short-Term Memory (BiLSTM) on technical indicators and stock time series
- Combined both output to engineer trading signals for a financial investment platform
- Evaluated results using receiver operating characteristic (ROC) curve, precision and recall

Smile Detection using ensemble of Convolutional Neural Networks (In Development) *Oct. 2019*

- Trained a convolutinal neural network on 35 thousand images from the FER2013 dataset and used transfer learning for training a smile detection CNN using 200 images
- Built an ensemble of Convolutional Neural Networks with Keras for classifying neutral and smiling face images
- Evaluated results using stratified K-Fold from scikit-learn using the F1-measure due to small size of the dataset
- Evaluated against a baseline approach that extracted features from applying gabor filters (OpenCV) and using Support Vector Machines (SVM)

Cannabis Recommender System (Course Project) *July. 2019*

- Wrote a Python script using Beautiful Soup for scraping and parsing social media information (Leafly reviews)
- Applied Natrual Language Processing (NLP) algorithms (word2vec, TF-IDF, n-grams) to transform textual data (reviews) to build a content-based recommendation system in Python
- Used the cosine similarity to provide recommendations
- Converted project into R and built an R Flex Dashboard to display the results

News Text Classification and Clustering (Personal Project) *June. 2019*

- Used Python's scikit-learn and keras (feed forward neural network) to classify news articles
- Visualized univariate and multivariate clusters of data using (EM and K-Means)
- Evaluated the outcomes by calculating Kappa and Silhouette

IMDB Movie Rating Prediction (Coursera Project) *Apr. 2018*

- Built a parsimonious step-wise regression model to predict the Internet Movie Database (IMDb) movie ratings which was part of the Coursera course
- Extended my project by integrating my model into a Shiny application

2019 **Bayesian Statistics: From Concept to Data Analysis**, Coursera - Certificate received

2018 **Linear Regression and Modelling**, Coursera - Certificate received

Links

Blog **Introduction to Bayesian Modelling with Stan and R**, Aug. 2019

Blog **Doing Data Science as a Co-op**, Aug. 2019

Portfolio **Personal Website and Portfolio** , oct. 2019