Justin **Charbonneau**

Montréal, QC

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

Industry Experience

Data Scientist

Montréal, Québec

Computer Research Institute of Montréal (CRIM) *Dec. 2019 - Present*

- Model user representations using language models for millions of cardholders and predict probabilities of transacting over time for the client .
- Integrate with their Machine Learning component of their system, by using BERT to extract useful representations from the content of the emails for the client.
- Ability to capture business requirements and transform to actionable project.
- Design different deep learning architectures and select the most appropriate modelling techniques for the task and provide a knowledge transfer to various clients.
- Systematic study of deep neural networks and classical predictive methods on tabular data.
- Systematic study of deep neural networks and classical predictive methods on Word Net 2006.

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 - Dec. 2019*

- 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 - Dec. 2019*

- 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	Proficient and familiar with many machine learning libraries and developing prototyping solutions. (Scikit-learn, Pandas/NumPy, PyTorch, FastAI, Kedro, MLFlow, Dask)
Experiments	Ran jobs on the SLURM cluster and OpenStack virtual machine with GPU (Linux)
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	PyTorch, Keras; Tensorboard; fine-tuning; Convolutional Neural Networks, Recurrent Neural Networks
Time Series	Familiar with classical time series forecasting algorithms and interest in deep learning for time series
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 PostgreSQL) and exposure to ETL tools
Data Visualization	Produced meaningful graphs, used widgets appropriately to showcase notebook demos; some experience using Tableau
Version Control	Followed best-practices using Git (Bit bucket) 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

Automatic representation creation using definitions *May 2020 - Oct. 2020*

- Determine what are the optimal techniques to create dynamic semantic representations that could solve the issue of rare or unknown words in language models.
- Created an experimental setup that would serve as the foundation for research on dynamic semantic representations. Created two tasks, super sense and hypernyms classification using synset definitions from the WordNet 2006 lexical database.
- Explored various methods such as FastText, ULMFiT from FastAI and multiple transformer models using the Simple Transformers library.

Stacked Machine Learning Model for Stock Market Prediction *Sept. 2019 - Dec. 2019*

- 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 Network (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

Rossman Store Sales Prediction *Feb. 2020 - Apr. 2020*

- 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 Network (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

View more projects on GitHub

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