Michael Armanious

https://michaelarman.github.io
michael.arman7@gmail.com | +1 647 997 3306 | linkedin.com/in/michael-armanious

SKILLS

PROGRAMMING

Python and iPython Notebook
R and R Studio
SQL and CQL
SAS
HTML and Markdown

MACHINE LEARNING

Regression
Decision Trees
Support Vector Machines
Random Forests & XGBoost
Cluster Analysis
Neural Networks/Deep Learning:
Image Classification
Text Classification
Audio Classification

TOOLS & SOFTWARE

SQL (PostgreSQL, MySQL)
IBM Netezza
SAS & Neo4j
Git and Github
Jupyter, Google-Colab
Anaconda + packages
ggplot2, matplotlib, seaborn
WEKA

EDUCATION

MSC STATISTICS

Carleton University
Sept 2019 | Ottawa, ON Master's Project: Men's U-Sports Basketball Analysis

BSC, MATH & STATS

Carleton University
May 2018 | Ottawa, ON

CERTIFICATES

Deep Learning, Tensorflow, Computer Vision, NLP
Coursera, DataCamp, Udacity

LINKS

Github:// michaelarman
LinkedIn:// michael-armanious

EXPERIENCE

THOMSON REUTERS | Data Scientist Intern

January 2020 – Present | Toronto, ON

- Tools: Python: numpy, pandas, sklearn, tensorflow, pytorch, nltk, spacy, Git, AWS Sagemaker, Azure Data Studio, SQL, Jupyter, Anaconda
- Worked with business and Product team to translate data science tasks into clear business value functions
- Created workflows in an AGILE environment and managed tasks on weekly timelines to present to a business unit
- Classified Legal Matters by creating a Deep Learning BERT model with narratives as inputs and achieved an 80% accuracy for 231 classes
- Performed data extraction cleaning and PII redaction using Spacy NER
- Aggregated, cleaned, obfuscated data, and explored text data using NLP libraries & techniques, ranked text using gloVe embeddings, and used Exploratory Data Analysis methods for all types of data
- Explored different clustering methods and compared algorithms to find the best suited clustering

CANADA REVENUE AGENCY | Data Scientist

May 2017 – December 2019 | Ottawa, ON

- Tools: SAS, Anaconda, R, Python; numpy, pandas, sci-kit learn, tensorflow, keras, pytorch, Neo4J, Netezza SQL, Excel, HTML
- Collected and interpreted large datasets of taxpayer information and generated Relational Graph Databases using Neo4j and SAS to classify and predict compliance.
- Created classification and prediction models on large imbalanced datasets using various machine learning algorithms such as Logistic Regression, Random Forests, XGBoost, Neural Networks, etc. to classify and predict which taxpayers should be audited.

CARLETON UNIVERSITY | Research & Teaching Assistant

Sept 2018 – Sept 2019 | Ottawa, ON

- Data Mining - STAT 4601/STAT 5703: Tasked with assisting students with Data Mining projects using R Studio and taught data mining techniques including data cleaning, data preparation, data analysis and write-ups.
- Implemented several data mining algorithms in datasets such as Ontario University Athletics (OUA) basketball using the Synergy Sports Technology to.create reports to the coach of the Carleton Basketball team
- Scraped and cleaned raw data from the OUA website and Synergy database to use for analysis. Incorporated classification and clustering algorithms in order to gain insights on Basketball data.

RESEARCH & PROJECTS

- Tools: R , R Markdown, RShiny, SQL, Python; BeautifulSoup, scrapy, selenium, requests, scikit-learn, pandas, numpy, matplotlib, tensorflow, pytorch, bokeh, librosa, OpenCV
- Master's Project: Data Mining on OUA Basketball
- Project: Audio Classification of Accents using Deep Learning
- Project: Steganalysis
- Project: Heart Disease Analysis

Relevant Coursework

Categorical Data Analysis, Bayesian Analysis, High Dimensional Analysis, Advanced Data Mining, Time Series Forecasting, Multivariate Analysis, Survey Sampling, Combinatorial Optimization, Regression Analysis, Stochastic Processes & Applications, Analysis of Variance & Experimental Design