LIEBY CARDOSO

Toronto, ON M4P 2W3 | (416) 871-4548

liebycardoso@gmail.com

Website: liebycardoso.github.io   **Github**: github.com/liebycardoso
Linkedin: linkedin.com/in/liebycardoso   **Kaggle**: kaggle.com/liebycardoso

SUMMARY

Data scientist with a strong programming background experienced at storytelling, flask API, machine learning algorithms, and cloud functions. Over 4 years of experience working with credit risk analysis and shopper behaviour data. A Data-driven, detail-oriented professional focused on delivering valuable and meaningful solutions.

SKILLS

Coding: Python, JavaScript, R, HTML, CSS, Delphi, SQL, Visual Studio, MVC

Data: Oracle, SQL Server, MySQL, Bigquery, JSON, CSV, pickle, google Analytics, Kubernetes yaml

Data visualization: d3.js, Microstrategy, Klipfolio, Tableau, Crystal Reports, Matplottlib, Seaborn, SSRS, Excel

Machine Learning: Pandas, Numpy, XGBoost, linear and logistic regression, Basket analysis (MLxtend - Apriori), NLTK, Sklearn, Feature engineering, GridSearchCV, Pipeline, Supervised and Unsupervised algorithms, Clustering, times series analysis

API: Web app framework Python Flask, Visual Studio Flask Web App

Cloud Computing: Google Cloud, AWS S3, GPC Storage, GPC App engine, Serverless GPC functions, Pub/Sub, Docker container, Kubernetes, GPC SDK, Hyper-v, Virtual Box, Cloudera

Statistics: A/B testing, Confidence Intervals, Dimensionality Reduction, Sanity Checks

Other: Unittest, Git, Subversion, Bitbucket, Agil, Waterfall

Soft Skills: Positive, Curious, fast learner with the ability to multi-task in a fast-paced work environment, team Player, willing to take on more responsibilities as needed and work to complete the team deadlines

PROFESSIONAL EXPERIENCE

Data Scientist   **Total Credit Recovery**   **Jan 2020-Present**

• Designed and Developed a machine learning algorithm capable of predicting the likelihood of debtor delinquency. Improved the model accuracy by 18% applying, in an imbalanced dataset, ML techniques as SMOTE
• Extract and analyze raw data from various databases and legacy system to drive agent performance improvement and support IT operations
• Designed and implemented 12 business overview metrics; also conduct weekly meetings with executives and stakeholders to monitor results achieved
• Created and maintain +20 data ingestion routine to import, clean and transform data consumed by analytics applications
• Reduced by 15% the agent idle time by developing metrics to control the time spent between calls and making this data available on an online dashboard accessed by all agents and managers

Technologies: SQL Server, SSRS, Python, Flask RESTful API, SQL, XGBoost, Sklearn

Data Scientist   **Mediar IDXP Analytics**   **2017-2018**

• Improved the searching timing process by 20% by redesign the data warehouse nested data model
• Implemented algorithms and cloud functions to process, cleaning and correlate shopper behavior data, brand promotion plans and video of shoppers' journeys – Basket analysis
• Developed a Flask API that queries the Bigquery tables based on endpoint parameters, formatted the result and delivery it to Klipfolio Dashboard
• Reduced by 70% the cost with a third-party contract by implementing a cloud function solution that standardizes brands name and updates the manufacturer database

Technologies: Python, Flask RESTful API, SQL, GPC Functions, Bigquery, NLTK, MLxtend, Apriori

Quality Assurance/Programmer   CPS   2012-2017

• Extraction, analysis, and distribution of financial, strategic and statistical information using PL / SQL, Crystal Reports, MicroStrategy, Excel and Oracle
• Developed several functionalities for the calculation of mortgages financial routines as monthly installment and interest rates
• Created more than 15 interface files for communication with banks (Citibank, HSBC, Santander) and financial institutions following current legislation

Technologies: Python, Crystal Reports, MicroStrategy, Excel, Oracle, SOX, ISO 9001, ISO 27000, ISO 20000, ITIL, MPSBR

EDUCATION

Certificate in Cloud Computing   MAY/2020
Lambton College, Toronto, ON

Udacity - **Nanodegree Data Science / Analyst**   2017
Estácio de Sá - **MBA – Project Manager PMI**   2014
Newton Paiva - **Information System, B.Sc**.   2009

PROJECTS

Spot a Twitter Bot
github.com/liebycardoso/twitterstreamkubernetes
• Used Pub/Sub technology deployed on a Kubernetes machine to ingest and support continuous data streaming from the Twitter API.
• A Classification algorithm determines whether a user i is a bot or a human, based on their account profile
• Application performance and final result analysis available to the end-user through a data studio dashboard.

Technologies used: Docker containers, Kubernetes, Cloud Functions, Pub/Sub, Bigquery, Bigquery ML, Data Studio, Continuous Delivery (Cloud Build + container registry), Sentiment Analysis

Refugee journey
 liebycardoso.github.io/refugee/
• Developed an interactive graphics website using D3.js to demonstrate the size of the refugee population present in each country for each year from 1951 to 2016
Technologies used: D3.js, JavaScript, HTML and CSS

Enron - Fraud detection
github.com/liebycardoso/ML_Enron
• Predicted which Enron employees were involved in fraud
• Evaluated Machine Learning Algorithms
• Parameters tuning
Technologies used: Python, Sklearn, GridSearchCV, Pipeline, SelectKBest, PCA and Cross-validation

ACTIVITIES

Tanzania Development Trust / United Nations - Humanitarian OpenStreetMap Team Volunteer   2017 – Present