Daniel de Souza Severo

More information: `https://dsevero.com`

EDUCATION	**University of Toronto**	
	Electrical & Computer Engineering	
	Master of Applied Science (M.A.Sc.)	Starting Fall 2020
	Undergraduate Exchange Program (1 year)	2013 - 2014
	Federal University of Santa Catarina, Brazil	2010 - 2015
	Bachelor of Science in Electronics Engineering	
	First Class Honours, 99th percentile.	

AWARDS

Virtual Design Challenge Winner   2019
Won 1st place at the VDC hosted by The University of British Columbia with my
paper *Proof of Novelty*. Received a cash prize of $ 3,000.00.
`https://github.com/dsevero/Proof-of-Novelty`

Student Merit Award   2015
Graduated with the highest GPA ever obtained (at the time) for my major.

Student Merit Medal   2015
Elected "Best Student" by the faculty of Electrical & Electronics Engineering at
the Federal University of Santa Catarina.

Science Without Borders Scholarship   2013
Awarded a full scholarship that covered tuition, transportation, necessary materials
and living costs to study 2 academic semesters at the University of Toronto.

**RESEARCH
EXPERIENCE**

Syrian-Lebanese Hospital   2018 - Current

ICD-10 Classification of Brazilian-Portuguese Medical Notes.
Work in Progress

Ward2ICU: A Vital Signs Dataset of Inpatients from the General Ward
We present a proxy dataset of vital signs with class labels indicating patient transitions
from the ward to intensive care units called Ward2ICU. Patient privacy is protected using
a Wasserstein Generative Adversarial Network to implicitly learn an approximation of the
data distribution, allowing us to sample synthetic data. The quality of data generation is
assessed directly on the binary classification task by comparing specificity and sensitivity
of an LSTM classifier on proxy and original datasets. We initialize a discussion of uninten-
tionally disclosing commercial sensitive information and propose a solution for a special
case through class label balancing. **arXiv:1910.00752**

Predicting Patient Health Risk through Financial Data.
Adoption of Electronic Health Records in Brazilian hospitals is slow due to administrative
inertia. Most hospitals collect and organize only financial claims of patient spendings,
making it difficult to apply ML techniques to predict patients with high health risk. We
show that it is possible to use financial risk as a proxy for health risk and apply it to a
Learning to Rank task for medical prevention. **Writing in progress.**

Reducing Readmission Rates with Time Series Data of Patient Vital Signs.
Hospital Readmission Rates are a key indicator of efficiency as it depends on a hospital's
ability to prioritize admitted patients and administer the allocation of intense-care units.
Since patient prioritization must be transparent, explainability in ML applications are an
important factor. Hence, traditional ML is still preferred over Neural Networks despite the
latter having superior performance. We propose a middle-ground approach by using Deep

Feature Synthesis to learn features from patient vital signs as inputs to highly explainable models such as Random Forests. **Writing in progress.**

Linx Impulse   2016 - 2017

Hypothesis Testing for Competitive A/B Experiments with High Data Volume.
In the ecommerce sector, high paying customers usually demand that an A/B experiment take place. This pins competitors against each other in an online testing environment devised to measure the performance of each recommender system at once. Due to the dynamic nature of evaluation metric (e.g. click-through-rate, revenue-per-user) chosen for each test, my research focused on developing a generic statistical hypothesis test that could be used to prove the effect our product had on the website, while handling unstructured data ingestion values averaging 1TB per day. Our results showed that a specific sub-sampling strategy together with Bootstrapping significantly reduced the computational complexity sacrificing very little power. **Publishing not authorized by employer.**

University of Toronto   2014 - 2015

A Report on the Ziggurat Method.
Pseudo-random number generators (PRNG's) are crucial in the context of simulating noise in communication channels. We present a report on an efficient method for generating pseudo-random samples from any decreasing probability distribution called the Ziggurat Method. Specifically, we will show the latest and most efficient version presented by McFarland. In the latter paper, the method shows a speedup of over 3 times compared to traditional algorithms such as Marsaglia's Polar Method. We present a speed comparison in C implemented on an Intel i7-4790 clocked at 3.60 GHz. A proof that the samples from this method are truly Gaussian is also provided. **Source code and writing available at** *https://github.com/dsevero/A-Report-on-the-Ziggurat-Method*

TEACHING EXPERIENCE	**Federal University of Santa Catarina** *Teaching Assistant for Communications Theory (2013), Non-linear Electronic Circuits (2013) and Single-Variable Calculus (2010).*

CERTI Foundation   2010 - 2013
LabVIEW Programming Instructor for Interns

PROFESSIONAL SERVICE
NeurIPS 2019: Conference on Neural Information Processing Systems
Reviewer for the Machine Learning for Health (ML4H) workshop.

PROFESSIONAL EXPERIENCE

3778 Healthcare,	*Machine Learning Engineer & Researcher*	2018 - Now
Linx Impulse,	*Head of Data Science*	2016 - 2018
CERTI Foundation,	*Research Engineer*	2015 - 2016
Wavetech Technology Solutions,	*Embedded Systems Eng. Intern*	2015 - 2015
WEG Industries,	*Electrical Engineering Intern.*	2010 - 2013

OPEN SOURCE CONTRIBUTIONS

Dask: Scalable analytics in Python
https://github.com/dask/dask/pulls?q=author:dsevero

Dask-ML: Scalable Machine Learn with Dask
https://github.com/dask/dask-ml/pulls?q=author:dsevero

Ward2ICU: A Vital Signs Dataset of Inpatients from the General Ward
https://github.com/3778/Ward2ICU