

# VALENTINO CONSTANTINOU

## PERSONAL INFORMATION

*email* [vc@valentino.io](mailto:vc@valentino.io)  
*phone* +1 (865) 387-6121  
*github* [github.com/vc1492a](https://github.com/vc1492a)

## EDUCATION

2015-2016 Northwestern University

*Master of Science  
in Analytics*

Exploratory and visual analysis of medical data from past Chicago Marathons, providing critical information to Chicago Event Management for use in medical staffing. Experimental design and analysis of consumer activity monitor data in order to understand and evaluate potential use cases toward mass running events.

Contributing team member of a nine month practicum project with GE Transportation, involving both unsupervised and supervised machine learning methods, in addition to visualization in Tableau. Designed and evaluated predictive models using random forests, gradient boosted trees, neural networks, among others.

Ten week capstone design project with HERE Maps. Developed visualization tools and machine learning models for determining the spacial extent and criticality of roadway congestion from construction activity for the German Bundesautobahn 5.

2010-2014 University of Tennessee

*Bachelor of Science  
in Economics*

Studies abroad at the University of Melbourne and the London School of Economics. Alan Wendy Wilson, CIE, David Smith Chambers, and HOPE Scholarship recipient.

## EXPERIENCE

2016-present NASA Jet Propulsion Laboratory

*Principal  
Investigator*

**Sep. 2018 - present:** Responsible for overall technical strategy for multi-year fault detection and alarm management effort that utilized numerous machine and deep learning approaches, from Principal Components Analysis (PCA) to various time-series deep learning architectures. Responsible for managing and mentoring co-investigators and interns, as well as project budget and communication with external and internal stakeholders.

*Data Scientist II*

**Mar. 2018 - present:** Research and development of anomaly detection systems for the Soil Moisture Active Passive (SMAP) spacecraft using machine and deep learning approaches for time-series applications such as LSTMs. Developed numerous full-stack prototype internal product or services on top of Natural Language Processing (NLP) and classification approaches, such as classification-based recommendation systems and tools to assess water resource variable (e.g. drought management, water reservoir height) priorities across water resource agencies. Responsible for architecture design, model training, API development, and front-end interfaces of numerous product or service prototypes.

*Data Scientist*

**Feb. 2017 - Feb. 2018:** Lead the development effort from early concept to near-production classification-based recommendation system for the Problem Reporting System (PRS). Developed small scale prototype of a tool that utilized graph theory to assess employees' relationship strength's with external

companies and organizations and released the Python open-source software PyNomaly.

*Data Scientist  
Intern*

**Jun. 2016 - Aug. 2016:** Contributed to the development an application that aids facilities planners and other organizations throughout the facility by overlaying internal data on interactive d3 visualizations of building floor plans.

2013-2014

### Alcoa Materials Management

*Intern*

**Feb. 2013 - Apr. 2014:** Performed market analysis of primary aluminum market as a member of the strategic marketing team. Developed a predictive model for US Canadian extruded products that aided planning of 2014 strategy.

## PUBLICATIONS

2019

### An Overview of Chicago Marathon Prehospital Care Demographics, Patient Care Operations and Injury Patterns

*Prehospital and  
Disaster Medicine.  
Cambridge  
University Press.*

**Authors:** Chan J., Constantinou V., Fokas J., Van Deusen Phillips S., Chiampas G. **Abstract:** During large scale events, disaster planning and preparedness between event organizers, EMS, local, state and federal agencies is critical to ensure participant and public safety. The Bank of America Chicago Marathon is one of the largest marathons in the world and with the improvement of technology, the use historical patient and event data during various environmental conditions can allow organizers and public safety officials a way to plan based on injury patterns and patient demands for care by predicting the placement and timing of needed medical support and resources. This study took a unique approach of reviewing digital runner data retrospectively over a five year period to establish patterns of medical demand geographically, temporally, and by the presenting diagnoses. Most medical complaints were musculoskeletal in nature, however there were life threatening conditions such as hyperthermia and cardiac incidents that highlight the need for detailed planning, coordination and multi-disciplined coordination and communication to ensure a safe and secure event. As technology has evolved, the use, analysis and implementation of historical digital data with various environmental conditions can provide organizers and public safety officials a map to plan on injury patterns, patient demands by predicting the placement and timing of needed medical support, personnel and resources.

2018

### PyNomaly: Anomaly detection using Local Outlier Probabilities (LoOP)

*Journal of Open  
Source Software,  
3(30), 845.*

**Authors:** Constantinou V. **Abstract:** PyNomaly is a Python 3 implementation of LoOP (Local Outlier Probabilities) (Kriegel, Krger, Schubert, Zimek, 2009). LoOP is a local density based outlier detection method by Kriegel, Krger, Schubert, and Zimek which provides outlier scores in the range of [0,1] that are directly interpretable as the probability of a sample being an outlier. PyNomaly also implements a modified approach to LoOP (Hamlet, Straub, Russell, Kerlin, 2017), which may be used for applications involving streaming data or where rapid calculations may be necessary.

2018

### Detecting Spacecraft Anomalies Using LSTMs and Nonparametric Dynamic Thresholding

*Proceedings of the  
24th ACM  
SIGKDD  
International  
Conference on  
Knowledge  
Discovery and  
Data Mining*

**Authors:** Hundman K., Constantinou V., Colwell I., Laporte C., Soderstrom T. **Abstract:** As spacecraft send back increasing amounts of telemetry data, improved anomaly detection systems are needed to lessen the monitoring burden placed on operations engineers and reduce operational risk. Current spacecraft monitoring systems only target a subset of anomaly types and often require costly expert knowledge to develop and maintain due to challenges involving scale and complexity. We demonstrate the effectiveness of Long Short-Term Memory (LSTMs) networks, a type of Recurrent Neural Network (RNN), in overcoming these issues using expert-labeled telemetry anomaly data

from the Soil Moisture Active Passive (SMAP) satellite and the Mars Science Laboratory (MSL) rover, Curiosity. We also propose a complementary unsupervised and nonparametric anomaly thresholding approach developed during a pilot implementation of an anomaly detection system for SMAP, and offer false positive mitigation strategies along with other key improvements and lessons learned during development.

2017                      **Applicability of Consumer Activity Monitor Data in Marathon Events: An Exploratory Study**

*Journal of Medical  
Engineering  
Technology*

**Authors:** Constantinou V., Felber A., Chan J. **Abstract:** Emerging opportunities to measure individual and population-level health data with activity monitors during recreational running activities may set the stage for new research possibilities in mass participation running events and marathon medicine. This study explores the applicability of consumer activity monitor data in a preliminary study for future marathon health research with a cohort of 12 (n1412) participants completing a 3.379km walking or running course. This study explored the feasibility of collecting pace and distance data from Fitbit brand consumer activity monitors, from access to user data to reporting of data characteristics and data analysis. We show that a large percentage of participant data can be successfully retrieved from Fitbit consumer activity monitor devices for analysis in marathon health research, and that identifying variations in pace across participants is a practical possibility. We note a mean absolute percent- age error of 13% over the true distance of 3.379 km, a higher error than that reported by other studies. We also observe a Pearson correlation coefficient between participant variation in pace and absolute distance error of 0.61. This study provides preliminary evidence to support the applicability of consumer activity monitor data in marathon health research.

TALKS

2019                      **Telemanom: an Extensible Framework for Time-series Anomaly Detection**

*Oak Ridge  
National  
Laboratory*

**Authors:** Constantinou V., Hundman H., Francis C., Laporte C., Colwell I., Mattmann C., Soderstrom T. **Abstract:** The NASA Jet Propulsion Laboratory (JPL) is a leader in the robotic study and exploration of the solar system. As spacecraft send back increasing amounts of telemetry data, improved anomaly detection systems are needed to lessen the monitoring burden placed on operations engineers and reduce operational risk. The majority of spacecraft monitoring systems only target a subset of anomaly types and often require costly expert knowledge to develop and maintain due to challenges involving scale and complexity. In this NSET Geoinformatics Engineering Group Seminar, we demonstrate the effectiveness of Long Short-Term Memory (LSTMs) networks, a type of deep learning model called a Recurrent Neural Network (RNN), in overcoming these issues using expert-labeled telemetry anomaly data from two JPL spacecraft - the Soil Moisture Active Passive (SMAP) satellite and the Mars Science Laboratory (MSL) rover, Curiosity. A complementary unsupervised and nonparametric anomaly thresholding approach developed during a pilot implementation of an anomaly detection system for these missions is detailed, along with a false positive mitigation strategy. We also highlight current work in the generalization of a time-series telemetry anomaly detection framework and challenges to overcome in applying telemetry anomaly detection for missions such the Mars Science Laboratory (MSL).

2018                      **Henosis - a generalizable, cloud-native Python form recommender framework for data scientists**

*Open Data Science  
Conference - East*

**Authors:** Constantinou V., Colwell I. **Abstract:** In October 2017, the Office of Safety and Mission Success (5X) desired to improve processes for reporting in the Problem Reporting System (PRS). PRS is an internal tool that allows engineers to submit Problem Failure Reports (PFRs) and Incident Surprise, Anomaly reports (ISAs), which document pre-launch test failures and post-launch operational anomalies experienced by spacecraft. These reports not

only serve as a record of past problems but also of past solutions to the problems described. Despite their value, the reports contained within the PRS are costly to fill out and submit. With dozens of textual, categorical, and other inputs in the forms, the PFRs and ISAs draw valuable time away from mission staff to the annotation of internal forms time better spent with spacecraft operations and mission work. A solution was needed that would reduce the time needed to file reports in PRS while ensuring ease of use for users already familiar with the current PRS system, such as a recommendation system for form fields. What we needed from a data science and IT operations perspective was a straightforward process to deploy a simple recommendation system for use in enterprise applications containing categorical form inputs (like dropdown menus). Henosis is a cloud-native, lightweight Python-based recommender framework that facilitates providing recommendations to users of applications. Henosis brings together model training and testing, storage and deployment, and querying under a single framework. The framework provides data scientists with a straight-forward and generalizable environment in which to train, test, store, and deploy categorical machine learning models for making form field recommendations, while also providing software engineers and web developers with a REST API that can be easily queried for recommendations and integrated across different enterprise applications.

2017 Pre-Hospital Care at a Mass Endurance Event:  
The Chicago Marathon Experience

*Congress on  
Disaster and  
Emergency  
Medicine*

**Authors:** Goyal A., Constantinou V., Fokas J., Van Duesen Phillips S., Chan J., Chiampas G. **Abstract:** The objective is to investigate the impact of pre-hospital care on patient outcomes and public health systems for the Chicago Marathon from 2012 to 2015. The Bank of America Chicago Marathon is a 26.2 mile race with approximately 40,000 runners annually. One of six World Marathon Majors, it serves as a model for mass event field medicine. The goals of this medical infrastructure are to decrease time to medical attention and transport, triage effectively, decrease hospital burden, and optimize public safety. This study examines impacts of medical structures at the Chicago Marathon, highlighting patient care volumes and trends. A comprehensive retrospective study was conducted of the Chicago Marathon during 2012-2015. Pre-hospital medical care involved 21 medical aid stations on course, and two major medical tents in the finish area, including general care and intensive care units. Each facility, staffed with medical personnel, conducted its own triaging. Additionally, each site was equipped with an EMS Superior ambulance for transferring patients to a higher level of care if needed. Medical data was collected via paper and a digital Medical Patient Tracking System designed by Nika Tec. During 2012 to 2015, in total 4,963 people encountered pre-hospital care at the marathon. Of these, 175 people (3.53%) were transferred to a hospital for further care (Figure 1). One-hundred-thirty (2.62%) people were treated in the on-site ICU facilities, suggesting that the triage system imperatively connects patients to appropriate care without over-burdening particular personnel and resources. **Conclusion** The majority of individuals seeking care were successfully triaged and treated by the personnel and resources on site, addressing the needs of an average of over 1,200 individuals per event. The small fraction transferred to hospitals minimized the public health burden, while identifying and mobilizing those who required transfer. This event can be likened to mass events both planned and unexpected, and it demonstrates strategic pre-hospital medical preparedness.

POSTERS

2019 Enabling Earth Science Research Through Use of  
Aura MLS Data with Outside Science Data Products

*Aura Science Team  
Meeting (accepted,  
to appear)*

**Authors:** Knosp B., Constantinou V., Livesey N. **Abstract:** The Aura Microwave Limb Sounder (MLS) instrument has produced nearly 15 years worth of atmospheric science data. This data is publicly available through the Goddard Earth Science Data and Information Services Center (GES-DISC). While on its own it can reveal the workings of multiple physical processes that

occur in the Earth's atmosphere, MLS data is often used with data from other instruments to answer complex research questions. Studies using multi-instrument data can also help assess and validate atmospheric models and many data products are typically used together. Over the last decade, the 22 MLS Level 2 data products have been used in over 900 publications. To mine these publications for information on trends and associations that might normally be overlooked, we have developed a new publication exploration tool.

2018                      Pairing NASA Capabilities and Western US  
Water Resource Stakeholders with Data Mining and Machine  
Learning

*American  
Geophysical Union  
(AGU) Fall  
Meeting*

**Authors:** Lee C., Constantinou V., Mishra A. **Abstract:** Partnerships between water resource agencies (stakeholders) and principal investigators (capabilities) that work to improve management of water resources (water supply, water quality, water use) exist largely as the result of ad-hoc domain-expert driven pairings between capabilities and stakeholders. In this paper, we describe and report results for an approach that enables pairing of stakeholders with appropriate capabilities to enhance data-driven decision-making for water management. By mining publication abstracts, combined with domain expertise, natural language processing and machine learning, principal investigators can be recommended for stakeholders based on shared research interests, resource management goals, and other values present in natural language. This work shows encouraging preliminary results for pairing stakeholders and capabilities through the use of a Naive Bayes model, and describes avenues for future work.

2018                      Detection and Interactive Exploration of  
Telemetry Anomalies

*American  
Geophysical Union  
(AGU) Fall  
Meeting*

**Authors:** Constantinou V., Laporte C., Colwell I., Hundman K. **Abstract:** As scientific instruments collect increasing amounts of data, improved systems are needed to analyze, condense, and present this data to engineers and scientists. Spurred by the availability and affordability of GPUs, deep learning methods are proving extremely effective at processing and condensing large volumes of data, yet there has been noticeably less innovation in the presentation of the results of these systems. We've developed one such system that utilizes long short-term memory networks (LSTMs) to predict telemetry data from spacecraft and identify a variety of anomalous behaviors (Hundman et al. 2018). In tandem, we've also developed a visual interface which allows operators to intuitively digest and explore these anomalies. The main component of this interface is an interactive hierarchical heatmap that summarizes anomalous activity using the same hierarchical engineering structure of the instrument. We allow users to navigate the hierarchy down to individual telemetry channels where the raw telemetry values are presented and anomalous regions are highlighted. Then, operators are given the ability to review and flag anomalies as either a true or false positive, which is used to tune the system. Anomalies vary in complexity and scale, and our interface helps users to better understand root causes by correlating anomalies across hundreds of telemetry channels in a seamless and intuitive way. As data rates from scientific instruments continue to increase, these sorts of interfaces will become a necessity for efficient operation. Our system and interface are currently being piloted for the Soil Moisture Active Passive spacecraft and for the Mars Science Laboratory rover, Curiosity, but are generalized to work with any multivariate time series data. These tools have utility in many scientific use cases for example, exploring and correlating unique patterns in climate variables precursing natural disasters.

## PROGRAMMING AND QUERY LANGUAGES

<i>Basic</i>	JULIA, SPARK, JAVA, MONGODB
<i>Intermediate</i>	JAVASCRIPT, R, SQL, L <sup>A</sup> T <sub>E</sub> X, HTML <sub>5</sub> , CSS <sub>3</sub>

*Advanced* PYTHON, ELASTICSEARCH

PROGRAMMING FRAMEWORKS AND TOOLS

*Basic* ANGULAR

*Intermediate* TORCH, D3.JS, REACT, APACHE TIKA, KUBERNETES

*Advanced* SCIKIT-LEARN, SCIPY, NUMPY, Keras, Pandas, SPACY, NLTK, PLOTLY, Matplotlib, PYTEST, Travis-CI, DOCKER

OTHER INFORMATION

*Volunteering*

- 2019 · General Assembly: A day in the life of a data scientist (guest speaker)
- Reviewer for the Journal of Open Source Software (JOSS)
- 2018 · STEAM:CODERS Family Code Night teaching assistant
- 2017 · STEAM:CODERS Introduction to JavaScript teaching assistant
- STEAM:CODERS Arduino Robotics teaching assistant

*Spoken Languages*

- ENGLISH · Native Language
- GERMAN · Intermediate (conversationally fluent)
- GREEK · Basic (simple words and phrases only)

*Interests* Open-Source Code · Traveling · Cooking · History · Art & Design

*Additional experience and education details available upon request.*

June 19, 2019