## Fundamentals in Data Science: Data Wrangling, Normalization, Preprocessing of Physiological Signals

## Joseph Picone, Professor, Department of Electrical and Computer Engineering, Temple University

## Iyad Obeid, Professor, Department of Electrical and Computer Engineering, Temple University

## Abstract:

### Data wrangling is defined as the process of mapping data from an unstructured format to another format that enables automated processing. State of the art deep learning systems require vast amounts of annotated data to achieve high performance, and hence, this is often referred to as a Big Data problem. Many decision support systems in healthcare can be successfully automated if such big data resources existed. Therefore, automated data wrangling is crucial to the application of deep learning to healthcare. In this talk, we will discuss data wrangling challenges for physiological signals commonly found in healthcare, such as electroencephalography (EEG) signals. For signal and image data to be useful in the development of machine learning systems, identification and localization of events in time and/or space plays an important role. Normalization of data with respect to annotation standards, recording environments, equipment manufacturers and even standards for clinical practice, must be accomplished for technology to be clinically relevant. We will specifically discuss our experiences in the development of a large clinical corpus of EEG data, the annotation of key events for which there is low inter-rater agreement (such as seizures), and the development of technology that can mitigate the variability found in such clinical data resources. In a companion talk to be given on December 2, data wrangling of unstructured text, such as that found in electronic medical records, will be discussed.

## About The Speaker:

Joseph Picone is currently a professor in the Department of Electrical and Computer Engineering at Temple University. He received his Ph.D. in Electrical Engineering in 1983 from the Illinois Institute of Technology. His primary research interests are machine learning approaches to acoustic modeling in speech recognition. Recently, he has been focusing on the commercialization of technology to automatically interpret EEGs. He has spent significant portions of his career in academia, research and the government, giving him a very balanced perspective on management of R&D. Dr. Picone is a Senior Member of the IEEE and has been active in several professional societies related to human language technology. He has authored numerous papers on the subject and holds several patents in this field. See *www.isip.piconepress.com* and *www.nedcdata.org* to learn more about his research and teaching.

Joseph Picone is a co-PI on an NIH BD2K grant titled “Automatic discovery and processing of EEG cohorts from clinical records” which is a collaboration between Temple University and the University of Texas at Dallas.