APPLIED DATA SCIENCE UG Minor

DSCI 351: Exploratory Data Science for Energy & Manufacturing

Spring 2015 Tuesday, Thursday 11:30 am to 12:45 pm Olin 303

Prof. Roger H. French, rxf131@case.edu, 3 Credits

This counts as a 3rd level class in the **Applied Data Science UG Minor**

For more information see http://datascience.case.edu/minor

Prerequisites: ENGR131 or equivalent, STAT312 or equivalent

Textbook: **Doing Data Science**, C. O'Neil, R. Schutt, O'Reilly.

R Cookbook, P. Teetor, O'Reilly.

Course description: Data Sources, Assembly & Exploratory Data Analytics

In this course, we will learn data science and analysis approaches applicable to energy and manufacturing technologies, to identify statistically significance relationships and better model and predict the behavior of these systems. We will assembly and explore real-world datasets, perform clustering and pair plot analyses to investigate correlations, and logistic regression will be employed to develop associated predictive models. Results will be interpreted, visualized and discussed.

We will introduce the basic elements of data science and analytics using R Project for Statistical Computing. R Analytics will be applied to the case of energy systems (such as PV power plant degradation, and building energy efficiency) over time. And it will be applied to manufacturing systems to understand the principles of statistical process control and identify critical factors of variability and uniformity.

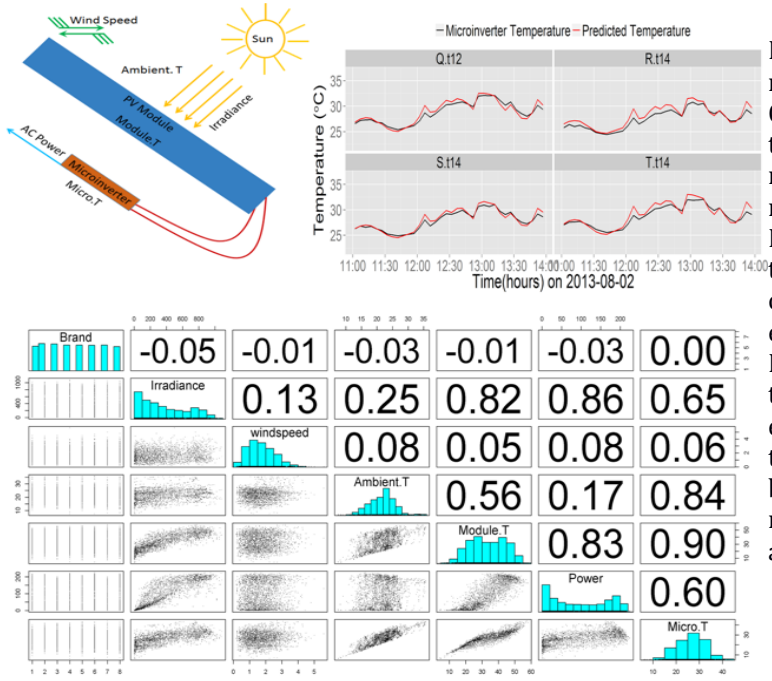

Figure. a) Schematic diagram of PV module and microinverter setup. b) Comparison of actual microinverter temperature and predicted microinverter temperature for the microinverters connected to 4 different PV module brands during noon time on a typical cloudy day. c) Pairs plot and correlation coefficient between different environmental and application stressors. Irradiance,wind speed and ambient temperature (Ambient.T) are the enviromental stressors. PV module temperature (Module.T), PV module brand(Brand), AC power (Power) and microinverter temperature (Micro.T) are application stressors.

DSCI351: Exploratory Data Science for Energy and Manufacturing

Roger French, Spring 2015
3 credit course, Undergraduate
Instructor permission required
Format: Seminar

Course Description:
Data Sources, Data Assembly and Exploratory Data Analytics

In this course, we will learn data science and analysis approaches applicable to energy and manufacturing technologies, to identify statistically significance relationships and better model and predict the behavior of these systems. We will assembly and explore real-world datasets, perform clustering and pair plot analyses to investigate correlations, and logistic regression will be employed to develop associated predictive models. Results will be interpreted, visualized and discussed.

We will introduce the basic elements of data science and analytics using R Project for Statistical Computing. R is an open-source software project with broad abilities to access machine-readable open-data resources, data cleaning and munging functions, and a rich selection of statistical packages, used for data analytics, model development and prediction. This will include an introduction to R data types, reading and writing data, looping, plotting and regular expressions, so that one can start performing variable transformations for linear fitting and developing structural equation models, while exploring for statistically significant relationships.

R Analytics will be applied to the case of energy systems (such as PV power plant degradation, and building energy efficiency) over time, by analyzing system responses, combined with results of experiments to identify fundamental principles that are statistically significant in the observed system performance. And it will be applied to manufacturing systems to understand the principles of statistical process control and identify critical factors of variability and uniformity.

Learning Outcomes:
Familiarity with R Statistics, scripting, functions, packages, automated data analysis.
Familiarity with exploratory data analysis, statistical model building
Applications of domain knowledge and statistical analytics to identify important predictors and develop initial predictive models
Data set characteristics will include:
Variety of types of information, including both structured and unstructured data,
Volume: Data from human sources (vendors, suppliers, distributors, customers, etc.) and sensor networks of the energy system of factory, both small and large data volumes.
Velocity: Energy system and manufacturing supply chains changes will be included.