Bioscience High School

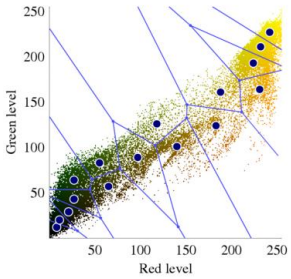
RET Project: Solar Fault Detection and Classification Learning July 2020

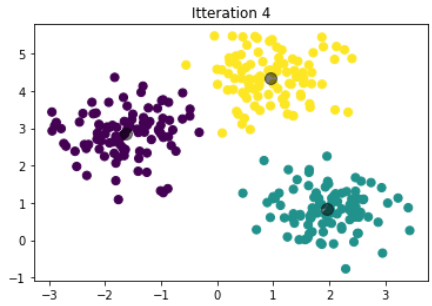
Milton Johnson

12th Grade Physics & Math
Bioscience High Schol, Phoenix, Az
Mentors: Kristen Jaskie, Dr. Andreas Spanias

NSF Award 1953745

https://sensip.engineering.asu.edu/ret/





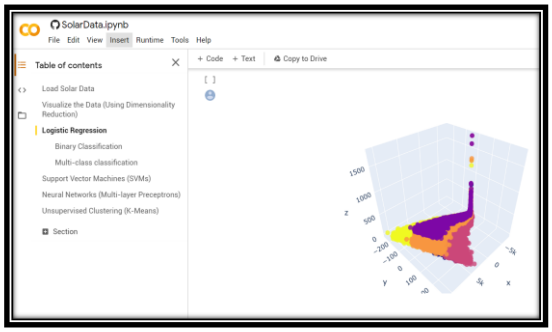
RET Research and Training

https://sensip.engineering.asu.edu/ret/

RET Schedule and Training

Hands On Technical Training

- Canvas: repository for notebooks, lessons, sample work,videos, documentation, resources
- Google Colab: hands on Python training and exploration
- Python Data Science Handbook
- Explored: Kmeans, Linear Regression, Logistic Regression, SVM

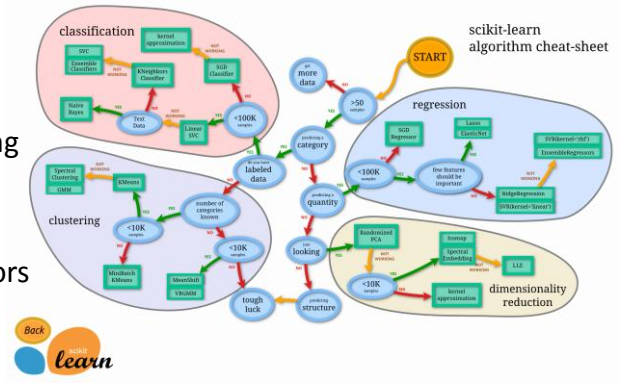


RET Schedule and Training

Technical Exposition

Speakers

- Kristen Jaskie- ML algorithms
- Andreas Spanias- ML and optimizing solar topography
- J-DSP signal processing simulation
- Jennifer Blain Christen- ML & sensors
- Gary Mastin- R&D Management
- Mike Stanley- Tiny ML (MCU's)
- Social Implications of ML



RET Schedule and Training

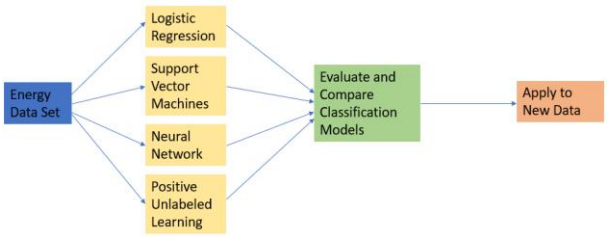
Research Materials

- Canvas- access to resources & videos
- Colab- coding canvas
- SciKitLearn- Python ML library
- ASU Solar Data- classified
- Cyprus Data- commercial

RET Lab Experience Research Summary

Research Objectives

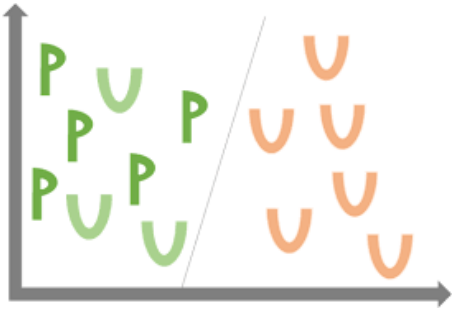
- Create classification models using labeled data for solar arrays that are affected by soiling, weather, ground leakage and short circuits. We then plan to evaluate these models and use our assessments to create new methods that can be used for unlabeled solar array data sets.



RET Lab Experience Research Summary

Research Background

- Similar work has been done already with applying classification algorithms to fault detection in solar systems. Our goal is to also apply the Positive Unlabeled algorithm to the data set as well.



RET Lab Experience Research Summary

Research Proposal

- ***Proposal Abstract*** **– Solar energy is becoming more affordable and more common in local businesses and residential use. We propose to create classification models using labeled data for solar arrays that are affected by soiling, weather, ground leakage and short circuits. We then plan to evaluate these models and use our assessments to to create new methods that can be used for unlabeled solar array data sets. If successful, these new methods or algorithms can be used to inform customers of deficiencies within their rooftop systems and the probable cause. Our studies will enable improved system power output and better management of the solar array. We will compare several supervised learning algorithms including logistic regression, support vector machines, neural net classifiers. We will also simulate semi-supervised positive unlabeled algorithm to evaluate its effectiveness for PV faults.**

PV Fault Detection and Classification using PU Learning

Milton Johnson, Bioscience High School, Kristen Jaskie and Andreas Spanias, SenSIP Center, School of ECEE, ASU

Abstract **– Solar energy is becoming more affordable and more common in local businesses and residential use. We propose to create classification models using labeled data for solar arrays that are affected by soiling, weather, ground leakage and short circuits. We then plan to evaluate these models and use our assessments to to create new methods that can be used for unlabeled solar array data sets. If successful, these new methods or algorithms can be used to inform customers of deficiencies within their rooftop systems and the probable cause. Our studies will enable improved system power output and better management of the solar array. We will compare several supervised learning algorithms including logistic regression, support vector machines, neural net classifiers. We will also simulate semi-supervised positive unlabeled algorithm to evaluate its effectiveness for PV faults.**

Index **terms:** Machine Learning, Logistic regression, neural nets, PU Learning, Fault detection, Photovoltaics, solar arrays.

PROJECT DESCRIPTION

Remote fault detection is a very relevant problem today. As solar energy systems become more common around the world, the need for remote monitoring in regards to performance and maintenance will increase tremendously. Training machine learning systems to monitor their output is a critical step. This will facilitate utility companies and other customers in getting optimal output from their systems. This RET research will explore machine learning methods (Fig. 1) for fault detection with the emphasis on positive unlabeled learning. The goals of this research include: performing a survey of machine learning and classification techniques [1,2], b) studying specifically solar array faults [3-6,9-11] with the aid of machine learning algorithms, and c) identify faults in PV systems using data from working systems. We begin with a dataset produced using Simulink that emulates an 18KW solar array [6]. The set contains over 21,000 labeled points. The labels include soiled, degraded, shaded, short circuit and standard. We will classify these data using machine learning (ML) algorithms [7,8], including logistic regression, support vector machines and neural net classifiers. Each algorithm can identify the decision boundary between two labels in order to effectively separate both sets. We will use confusion matrices to compare the effectiveness of each learning algorithm. We also intend to evaluate the data using a semi-supervised positive unlabeled learning algorithm [2]. In this scenario only a portion of the dataset will be labeled (Fig. 2) and identified as positive. The rest of the set is then considered unlabeled. The goal is to simulate several real world situations in which fully labeled data sets are not available.

One of the key objectives of our initial study is to produce reliable classification models for fault detection in solar arrays. Using our established models, we can then overlay new unlabeled data onto our existing classifications with the hope of predicting the unlabeled fault type. Anticipated outcomes include documented results and comparisons using the PU learning [13] method. The algorithm will be profiled in terms of performance and complexity and comparative results will be given relative to well established ML methods. This study is part of the SenSIP workforce development programs [12].

Fig. 1. Block Diagram of using ML for Solar Energy Systems.

Fig. 2. Positive Unlabeled Learning Concept [13].

References

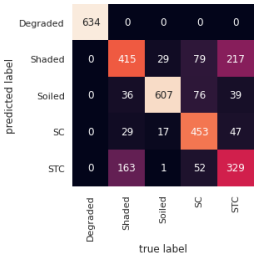
[1] Zhao, Ye, et al. "Graph-based semi-supervised learning for fault detection and classification in solar photovoltaic arrays." *IEEE Transactions on Power Electronics* 30.5 (2014): 2848-2858.
[2] Mehta, Sachin, et al. "Deepsolareye: Power loss prediction and weakly supervised soiling localization via fully convolutional networks for solar panels." 2018 *IEEE WACV*. IEEE, 2018.
[3] Gerossifa, Elyas, et al. "An enhanced machine learning based approach for failures detection and diagnosis of PV systems." *Energy conversion and management* 151 (2017): 496-513.
[4] S. Rao, A. Spanias, and C. Tepedelenlioglu, "Solar array fault detection using neural networks." *2019 ICPS. IEEE*, Taipei, 2019.
[5] Rao, Sunil, et al. "An 18 kW solar array research facility for fault detection experiments." *2016 19th MELECON, IEEE*, Limassol, 2016.
[6] Harrou, Fouzi, et al. "An unsupervised monitoring procedure for detecting anomalies in photovoltaic systems using a one-class Support Vector Machine." *Solar Energy* 179 , pp. 48-58, 2019.
[7] U. Shanthamallu, A. Spanias, C. Tepedelenlioglu, M. Stanley, "A Brief Survey of Machine Learning Methods and their Sensor and IoT Applications," *Proc. 8th IISA (IISA 2017)*, Larnaca, Aug. 2017.
[8] K. Jaskie and A. Spanias, "Positive and Unlabeled Learning Algorithms and Applications: A Survey," *Proc. IEEE IISA 2019*, Patras, July 2019.
[9] H. Braun, S. T. Buddha, V. Krishnan, A. Spanias, C. Tepedelenlioglu, T. Takehara, S. Takada, T. Yeider, and M. Banavar, Signal Processing for Solar Array Monitoring, Fault Detection, and Optimization, Synthesis Lect. Power Electronics, J. Hudgins, Ed. Morgan & Claypool, vol. 3, Sep. 2012.
[10] A. Spanias, "Solar Energy Management as an Internet of Things (IoT) Application," *Proceedings 8th IEEE IISA 2017*, Larnaca, August 2017.
[11] S. Rao, S. Katoch, V. Narayanaswamy, G. Muralgia, C. Tepedelenlioglu, A. Spanias, P. Turaga, R. Ayyanar, and D.Srinivasan, "Machine Learning for Solar Array Monitoring, Optimization, and Control,"Synth. Lect. Power Electronics, Morgan & Claypool, Ed. J. Hudgins, 91 pages, Aug. 2020.
[12] A. Spanias, Machine Learning Workforce Development Programs on Health and COVID-19, *Proc. IEEE IISA 2020*, Piraeus, July 2020.
[13] K. Jaskie, C. Elkan, A.Spanias, A Modified Logistic Regression For Positive and Unlabeled Learning, *IEEE Asilomar Conf. Sign., Syst. & Comp.*, Monterrey, 2019.

RET Lab Experience Research Summary

Research Conclusions

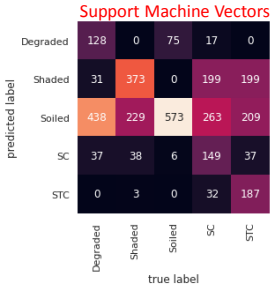
No conclusions yet… So far Logistic Regression has provided best results

Logistic Regression



Validation accuracy = 0.762333229
Test accuracy = 0.76233322991

Support Machine Vectors

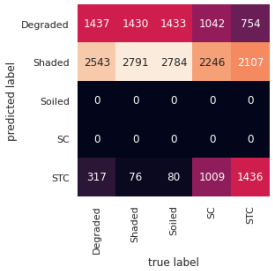


Validation accuracy = 0.4185541421
Test accuracy = 0.4374806081

Neural Network

Test accuracy =0.679801427

KMeans



Test accuracy = 0.2636

Next STEPS in Research

- Continue to examine models and fine-tune hyperparameters
- Attempt PU Learning
- Results will be presented in October at industry meeting
- Continue to train models in hopes of applying them to Cyprus solar data
- Continue to attend meetings with Cyprus team

RET Instructional Lesson Implementation

Lesson Objectives

- Create and represent a procedure to find the mathematical center (centroid) of a cluster
- Create and represent a procedure to mathematically place two centroids in a cluster
- Describe the steps followed in a KMeans clustering algorithm
- Modify existing Python code to create a new cluster of data with new centroids
- Apply KMeans algorithm to a new situation (cellular towers) and make informed recommendations
- Share recommendations in a report

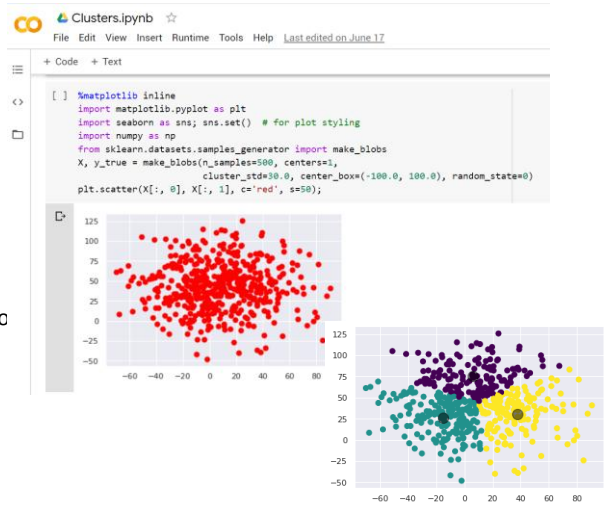
RET Instructional Lesson Implementation

Lesson Description

Introduction to KMeans Algorithm:

Context: New restaurant locations

- Step 1- Manually place single centroid in 'single' cluster. Then design a procedure to find mathematical center. Diagram the procedure.
- Step 2- Create procedure to place two centroids. Create a diagram and spreahsheet model to calculate points
- Step 3- Use Python Kmeans Algorithm in Colab to place multiple centroids. Use cost function to find elbow
- Step 4- Apply Colab Kmeans algorithm to new problem (cellular towers)
- Create report with recommendations of number of sites and locations



Questions & Feedback

- In teaching using Kmeans as an intro to Machine Learning, what do you think are the 'big ideas' that students need to comprehend? Technical or General

Self Assessment

Benefits of Program:

- Great overview of Machine Learning
- Flexible with my schedule
- Independent learning portions
- Addressing actual research topic (ML & Solar)
- Learned to use Colab (great for classroom)
- Gained more experience with Python & Arduinos
- Good classroom lesson ideas

Challenging part:

- Writing abstract and doing literature search without deep understanding of research topic

References

[1] Zhao, Ye, et al. "Graph-based semi-supervised learning for fault detection and classification in solar photovoltaic arrays." *IEEE Transactions on Power Electronics* 30.5 (2014): 2848-2858.

[2] Mehta, Sachin, et al. "Deepsolareye: Power loss prediction and weakly supervised soiling localization via fully convolutional networks for solar panels." 2018 *IEEE WACV.* IEEE, 2018.

[3] Garoudja, Elyes, et al. "An enhanced machine learning based approach for failures detection and diagnosis of PV systems." *Energy conversion and management* 151 (2017): 496-513.

[4] S. Rao, A. Spanias, and C. Tepedelenlioglu. "Solar array fault detection using neural networks." *2019 ICPS. IEEE*, Taipei, 2019.

[5] Rao, Sunil, et al. "An 18 kW solar array research facility for fault detection experiments." *2016 18th MELECON, IEEE*, Limassol, 2016.

[6] Harrou, Fouzi, et al. "An unsupervised monitoring procedure for detecting anomalies in photovoltaic systems using a one-class Support Vector Machine." *Solar Energy* 179 , pp. 48-58, 2019.

[7] U. Shanthamallu, A. Spanias, C. Tepedelenlioglu, M. Stanley, "A Brief Survey of Machine Learning Methods and their Sensor and IoT Applications," *Proc. 8th IISA (IISA 2017),* Larnaca, Aug. 2017.

[8] K. Jaskie and A. Spanias, "Positive and Unlabeled Learning Algorithms and Applications: A Survey," *Proc. IEEE IISA 2019,* Patras, July 2019.

[9] H. Braun, S. T. Buddha, V. Krishnan, A. Spanias, C. Tepedelenlioglu, T. Takehara, S. Takada, T. Yeider, and M. Banavar, Signal Processing for Solar Array Monitoring, Fault Detection, and Optimization, Synthesis Lect. Power Electronics, J. Hudgins, Ed. Morgan & Claypool, vol. 3, Sep. 2012.

[10] A. Spanias, "Solar Energy Management as an Internet of Things (IoT) Application," *Proceedings 8th IEEE IISA 2017*, Larnaca, August 2017.

[11] S. Rao, S. Katoch, V. Narayanaswamy, G. Muniraju, C. Tepedelenlioglu, A. Spanias, P. Turaga, R. Ayyanar, and D.Srinivasan, "Machine Learning for Solar Array Monitoring, Optimization, and Control,"Synth. Lect. Power Electronics, Morgan & Claypool, Ed. J. Hudgins, 91 pages, , Aug. 2020.

[12] A. Spanias, Machine Learning Workforce Development Programs on Health and COVID-19*, Proc. IEEE IISA 2020*, Piraeus, July 2020.

[13] K. Jaskie, C. Elkan, A.Spanias, A Modified Logistic Regression For Positive and Unlabeled Learning, *IEEE Asilomar Conf. Sign., Syst. & Comp.*, Monterrey, 2019.

Thank You

This has been a great experience. I am really excited to share Machine Learning with my students and fellow teachers.

- Andreas Spanias
- Kristen Jaskie
- Jean Larson
- Jennifer Blain Christen
- Ruby Sayed