ML PAPER MAIN DRAFT

INTRODUCTION:

Perovskite Solar Cells are one of the major futuristic alternative of the traditional Si based solar cell. Yet considering the sensitivity and over all stability of the Perovskite materials, we are a little far than the traditional 1st generation solar cells. Following last few years of research in this domain, we can practically notice the scientific progress both in efficiency and stability of PSCs. The development can be more persistent if the results of the collective research could be more organized, accessible and be used in a more efficient manner. There comes the advantage of the machine learning that enables us to develop and optimize our scientific insights by utilizing the exiting data. Using different ML algorithms the development of predictive models has been tested and modified into deployment for research purposes. This can give us advantage to minimize the experimentation redundancy and to examine the effect of compositional materials and their amount on the resulting overall efficiency of PSC. The process doesn’t only develops the ML models for research purposes but also demonstrates the application of synthetic data generation algorithms to avoid the imbalance from of data by over sampling of the minority data classes. We have taken more than 100 parameters involved in the process of the both synthesizing and assembling of the PSCs and processed into ML algorithms after data cleaning and feature engineering. Considering over-fitting and hyper-parameter optimization, the training process is done for numerous iteration to achieve the average accuracy of 85%. (The accuracy can be achieved as high as 90% with proper and delayed hyper-parameter optimization.)

PROCESS SEGMENTATION:

Section 1: Imp : DISCUSSION

Section 2: EXPERIMENTATION AND APPLICATION

1. EXPERIMENT FUNCTIONS
2. UNDERSTANDING THE PROCESS OF PSC DEVELOPEMENT
3. EXAMINING THE FACTORES FROM THE DEVELOPMENT PROCESS
4. EXTRACTING MEANINGFUL SET OF DATA
5. PREPARING DATA ACCORDING TO NUMERICAL OR CATEGORICAL SEGREGATION.
6. ML FUNCTIONS
7. DATA COLLECTION
8. DATA CLEANING
9. SYNTHETIC DATA GENERATION
10. LABEL ENCODING
11. SCALING
12. DEPLOYING INPUT DATA PIPE LINE
13. DISTRIBUTED TRAINING OF K-FOLD CROSS-VALIDATION
14. DEVELOPING DEPLOYING PIPE LINE
15. CREATING USER INTERFACE FOR THE DEPLOYMENT USING GRADIO

Section 3: ADVANTAGES AND POSSIBLITIES