Prof. Dr.-Ing. Grit Behrens

EU PVSEC 2016

DATA MINING METHODS FOR FAILURE CLASSIFICATION ON PV-MODULES UNDER FIELD-CONDITIONS

G. Behrens, A. Dercho, H. Quakernack, T. Wächter, S. Hempelmann, I. Kruse

Munich, 23th of June 2016

Solar Computing Lab

SunSniffer®

FH Bielefeld
University of
Applied Sciences

Data-Mining in PV

- ***Data Mining*** ist the computational proces of discovering patterns in large data sets

- ***Big data*** sets can come from
  - monitoring PV-plants - best way module based data, ✔
  - picturing and video streaming in PV-fields – IR, EL,
  - environment data – climate, irradiation, ... ✔
  - Module specification data – data sheet

- ***Goals by data Mining in PV-fields:***
  - Failure classification ✔
  - Failure prediction
  - Predictions for optimal component selection

4 Steps of Data Mining-Process

1. Data Collection

Module based SunSniffer®-Data from 3 PV-Fields

Weather API from Internet

2. Preprocessing

Data-Filtering
- Cleaning
- Interpolation

Normalization

3. Analysis

Correlation

Visualisation

Interview with Experts

4. Classification

Neural Network

Learning Experiments

Interpretation

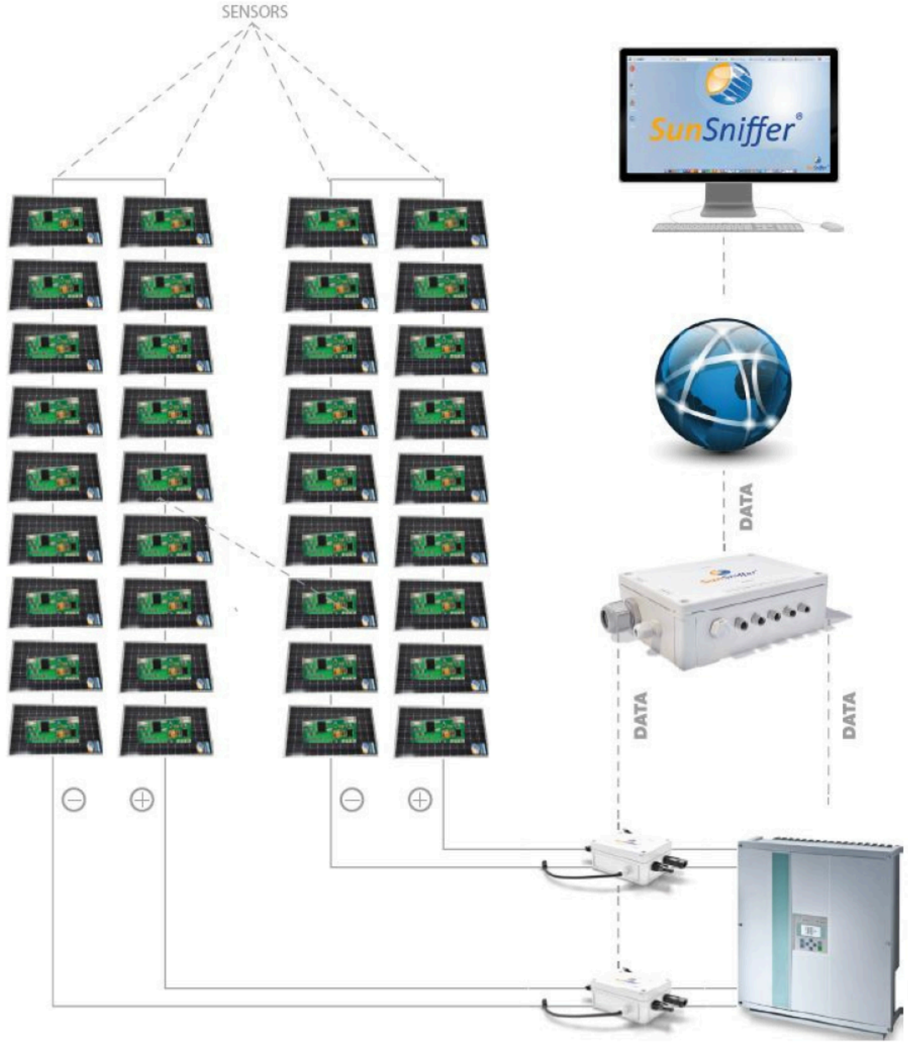
[1] Runkler, „Data Mining - Methoden und Algorithmen intelligenter Datenanalyse“

Solar Computing Lab

1. Step: Data Collection with SunSniffer®

FH Bielefeld
University of
Applied Sciences

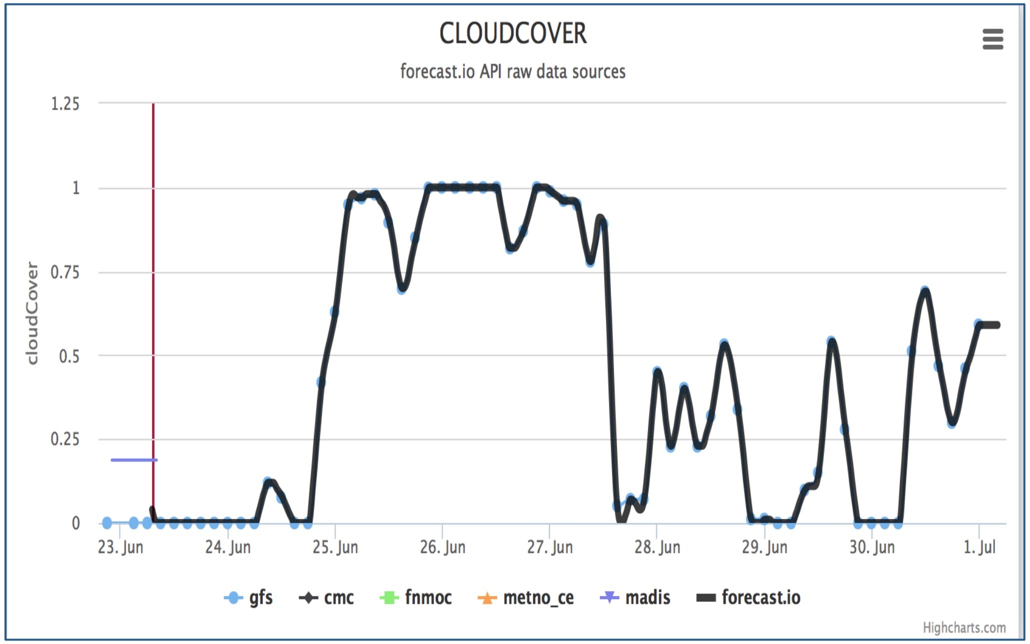
* *SunSniffer®*



- **Sunsniffer®-Technology**
  - Modulbased measuring of **temperature and voltage**
  - Powerline- data communcation
  - Integrated in junction box

[2] Kilper et.al. , "A new generation of PV Monitoring system with high-grade remote diagnostics based on module level monitoring and integrated yield simulation" , PVSEC2015

1. Step : Data Collection by Weather-API



http://www.forecast.io, MUNICH 23.06.2016

- **weather data for example:**
  - CloudeCover [0,1]✔
  - Temperature [ºF]✔
- **Worldwide locations**

Request-Types:

- Current conditions
- Forecasts
- **historical data ✔**

- **1000 requests a day for free**
  >1000 requests : 0.0001$ each

HTTP-Request:

https://api.forecast.io/forecast/APIKEY/LATITUDE,LONGITUDE,TIME

HTTP-Response(JSON):

All weather data for given day, hour-by-hour resolution

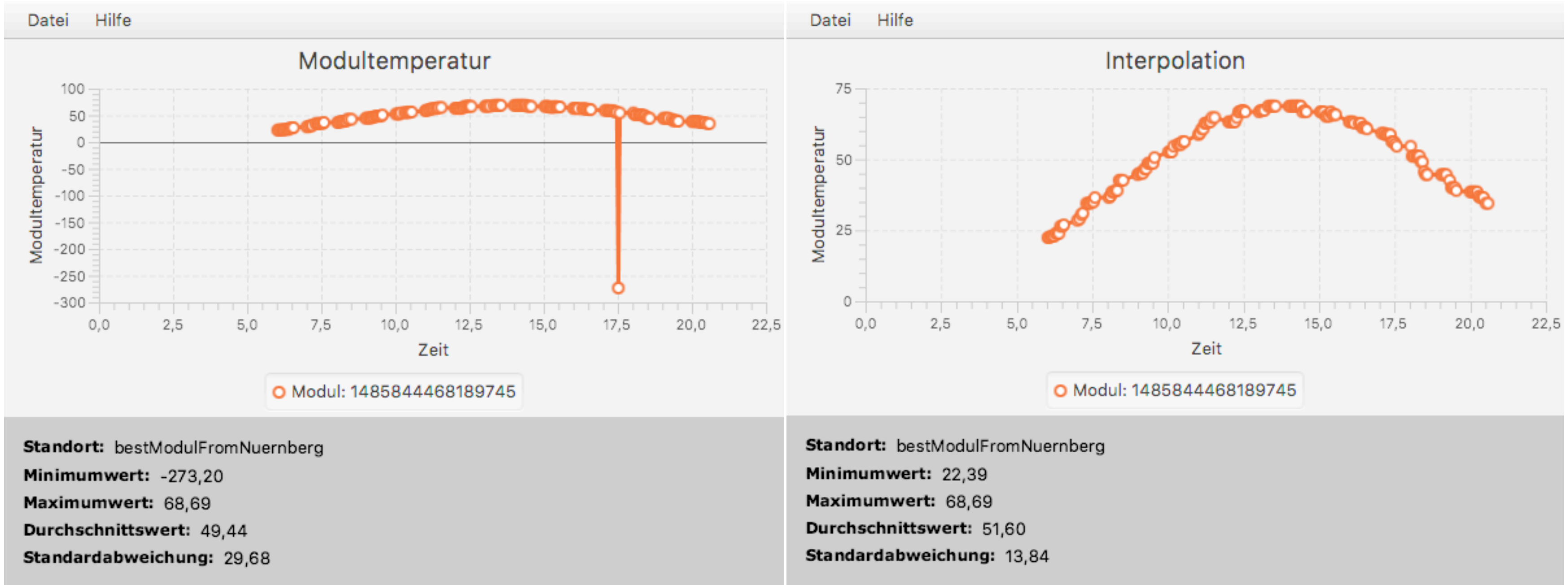
forecast.io

2. Step : Data PreProcessing
Filtering

Example module temperature:

- ***outliers*** correction
- ***Interpolation*** outliers ore missing
  - by next neighbours
  - time equidistant values

$$x_f^{(i)} = \frac{x_{f-1}^{(i)} + x_{f+1}^{(i)}}{2}$$



Modul temperature values before (left) and after interpolation operations (right)

2. Step: Data PreProcessing
Normalisation I/II

- Individual features have very often ***very different rages of values***.
- Numerical classification methods would ignore features with less value ranges

Feature vector in project:

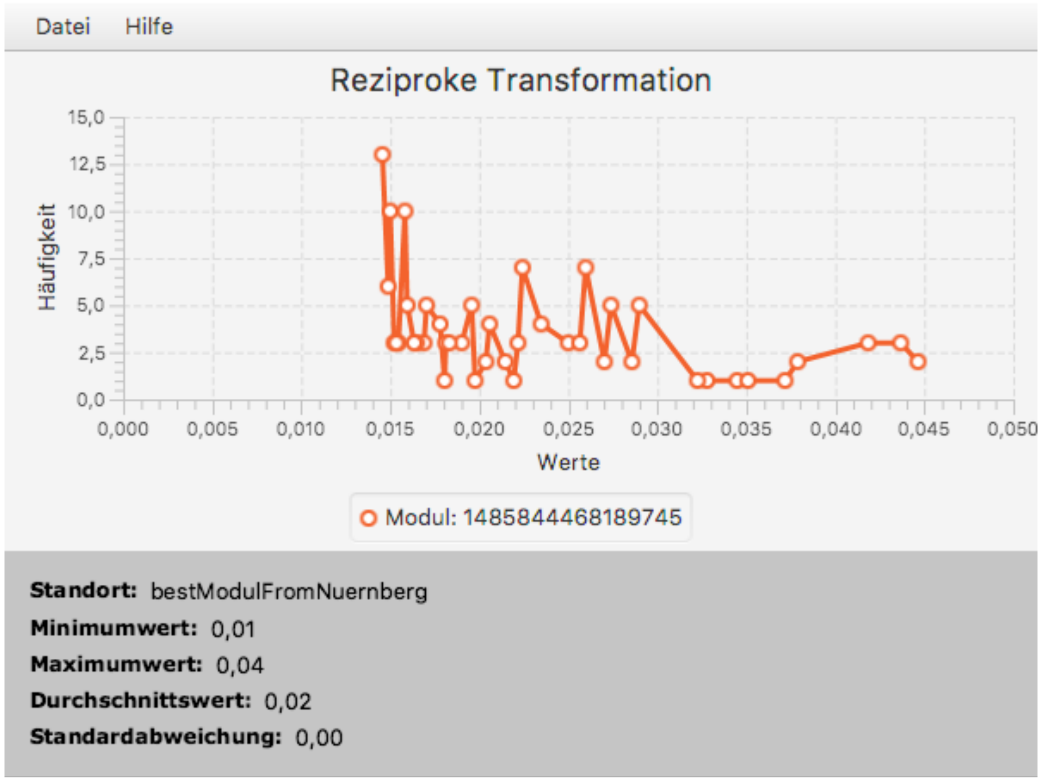
feature	unit	Value range
Module Temperatur	[ºC]	20 - 60
Module Voltage	[V]	0 - 100
String Current	[A]	0 - 20
Weather Temperatur	[ºC]	-20 - 50
Solar efficiency *		~(0 – 10)
Cloud Covering		0 - 1

* Solar efficiency = (sunset – sunrise)* (1- cloud covering)

2. Step: Data PreProcessing
Normalisation II/II

For example:

- ***reziprocal transformation*** of values of module temperatures
- Gives values between [-1,1]



$$f : R \setminus \{0\} \rightarrow R \setminus \{0\}$$
$$f(x) = f^{-1}(x) = 1/x$$

- Modul temperature values after filtering **[20, 60]**
- After reziprocal calculations **[-0.1, 0.04]**

3. Step : Data Analysis

Feature vector optimisation

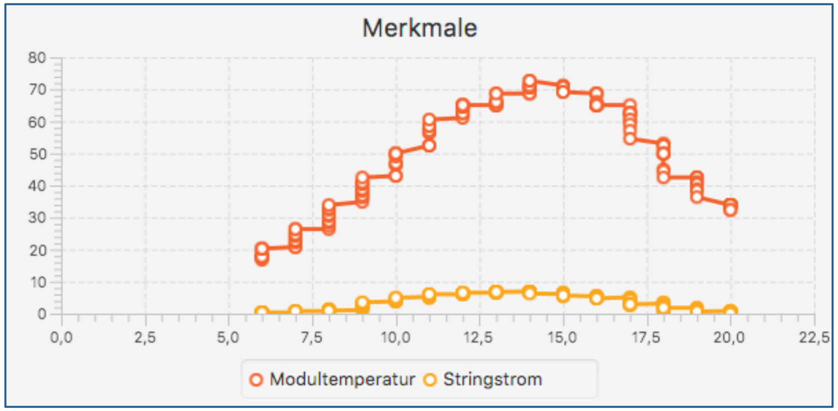
- Find out ***dependencies*** between components and ***most significant*** features
- Vizualisation
- Expert intervies

Correlation matrix indices for relations between feature i and j :

$$c_{ij} = \frac{1}{n-1} \sum_{k=1}^{n} \left(x_k^{(i)} - \overline{x}^{(i)} \right) \left(x_k^{(j)} - \overline{x}^{(j)} \right), \quad mit \quad i, j = 1, .., p$$

** n – number of features vectors, p – number of components of feature vector*

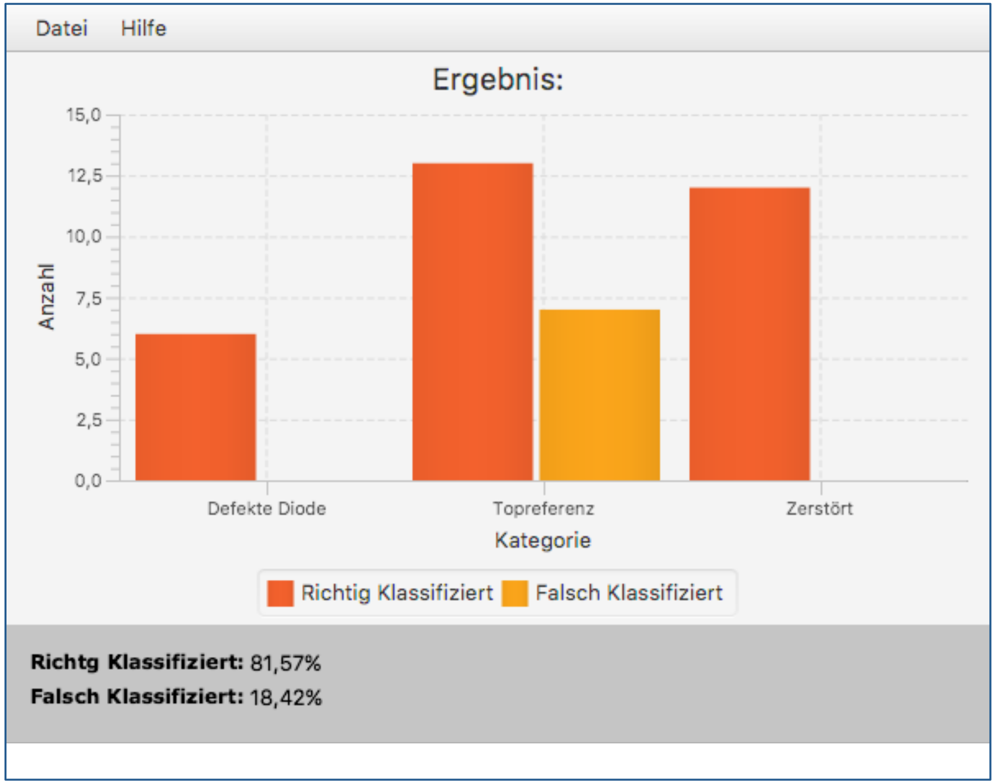
Feature 1	Feature 2	Correlation
Modul Temp.	Temperatur	82,02
Modul Temp.	String Current	43,98
Modul Temp.	Module Vol.	13,42
String Current	Temperatur	7,26
Module Voltage	Temperatur	2,34
...	...	...



- *yellow – Module Temperature,*
- *orange – String Currency*

4. Step : Classification by Neural Network Classifier I/II

- ***WEKA*** – **W**aikato **E**nvironment for **K**nowledge **A**nalysis **[3]**
  - Open Source Library for Java
  - Preprocessing, normalisation, classification, visualiszation,..



Results for training on 3 classes, all data:

- Defect diode
- No failure
- Mechanically destroyed

Orange :right classified 81,57%
Yellow : bad classified 18,42%

- false positives:
- some modules with failure classified in the „no failure" result class

WEKA
The University of Waikato

[3] I.H. Witten, et. al. "Data Mining: Practical Machine Learning Tools and Techniques"

4. Step : Classification by Neural Network Classifier II/II

FH Bielefeld
University of
Applied Sciences

Results for training on 4 classes, 3 PV-Fields monitored by Sunsniffer
Only sunny days:

- Defect bypass diode
- No failure
- Mechanically destroyed
- Hailstorme damage

- Shadowing can recognized with 99% accuracy (SVM better than NN) **[4]**

Right classified	Epochs	Momen-tum	Learning Rate	Training Data Ratio
100%	584	0,1	0,1	98%
96%	**69**	**0,2**	**0,5**	**66%**

[4] G. Behrens, et. al. "Machine Learning Methods for partial shading detection in monitoring data on PV-sytems" PVSEC2014

Results

- ***Data Mining Algorithms are useful*** on large data sets from ***module based*** monitoring data taken with ***low-cost SunSniffer®-Sensor***.
- Weather data ***taken from low cost - Internet API*** are useful
- Our Software application is ***easy to adopt to any monitoring system with possibility of failure labeling***.
- ***High recognition rates*** give up **to 96% accuracy** in failure detection.
- Recognition ***of different failure types is possible partly based based on additional information from weather sensors***
- ***Different module types*** in the 3 PV-plants were not a problem.
- ***The recognition rate results are better than better are the training data sets***:
  - Quantity: than more labeled failure types exist (up to 100%)
  - Quality: than more optimized is the training data set (excellent conditions, significant features)


Solar
Computing
Lab
Thank You for Your Attention!