AMLD Conference Summary

F. Giordano, B. Salvant


AMLDEPFL
Welcome to
#AMLD2020
Foyer Clouyd
13:30-17:00
Cross-lingual Natural
Language
Processing
Other Workshops
+ Coatcheck & WC
EPFL
VAUD
swisscom
January 25-29
2020
@ EPFL
Lausanne
Switzerland
FRANCESCO
GIORDANO
CERN


Machine Learning = Learning from Data

Workshop Days

Saturday 25th and Sunday 26th

Several workshop on different ML topics.

My choices:

Fraud detection with unsupervised learning

Adding interpretability to ML models

Tensorflow 2.0 basics

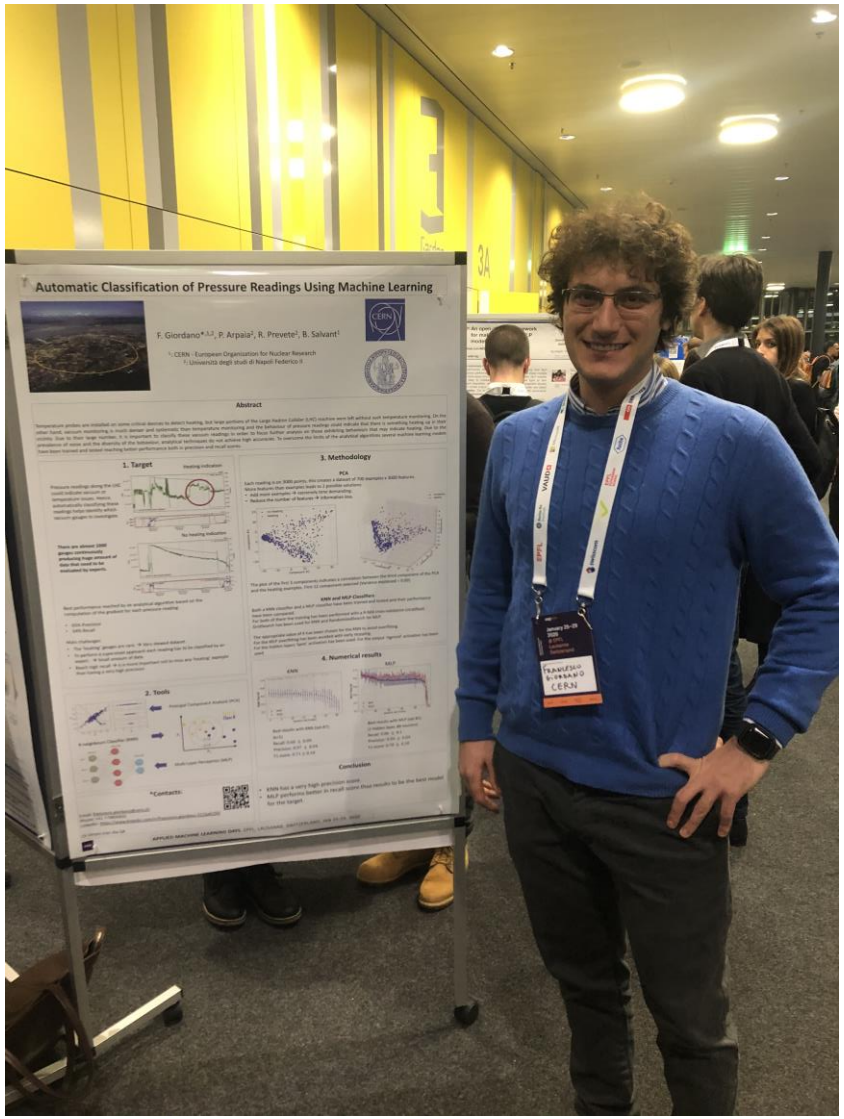
Conference Days

Monday 27th , Tuesday 28th and Wednesday 29th

Different talk in many ML fields:

- AI and Physics
- AI and Pharma
- AI and Industry
- AI and Aviation
- AI and Nutrition
- And many others

Poster Session (during the conference)



Automatic Classification of Pressure Readings Using Machine Learning

F. Giordano*[1,2], P. Arpaia[2], R. Prevete[2], B. Salvant[1]

[1]: CERN - European Organization for Nuclear Research
[2]: Università degli studi di Napoli Federico II

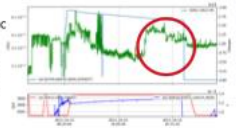
Abstract

Temperature probes are installed on some critical devices to detect heating, but large portions of the Large Hadron Collider (LHC) machine were left without such temperature monitoring. On the other hand, vacuum monitoring is much denser and systematic than temperature monitoring and the behaviour of pressure readings could indicate that there is something heating up in their vicinity. Due to their large number, it is important to classify these vacuum readings in order to focus further analysis on those exhibiting behaviours that may indicate heating. Due to the prevalence of noise and the diversity of the behaviour, analytical techniques do not achieve high accuracies. To overcome the limits of the analytical algorithms several machine learning models have been trained and tested reaching better performance both in precision and recall scores.

1. Target

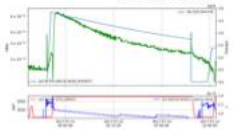
Heating indication

Pressure readings along the LHC could indicate vacuum or temperature issues. Hence, automatically classifying these readings helps identify which vacuum gauges to investigate.



No heating indication

There are almost 1000 gauges continuously producing huge amount of data that need to be evaluated by experts.



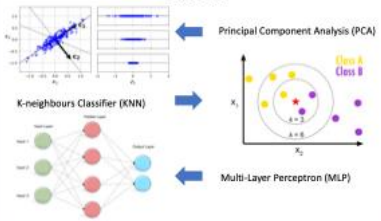
Best performance reached by an analytical algorithm based on the computation of the gradient for each pressure reading:

- 65% Precision
- 54% Recall

Main challenges:

- The 'heating' gauges are rare. → Very skewed dataset.
- To perform a supervised approach each reading has to be classified by an expert. → Small amount of data.
- Reach high recall → It is more important not to miss any 'heating' example than having a very high precision.

2. Tools

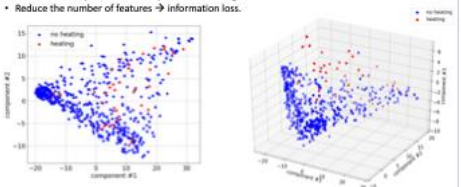


3. Methodology

PCA

Each reading is on 3000 points, this creates a dataset of 700 examples x 3000 features.
More features than examples leads to 2 possible solutions:

- Add more examples → extremely time demanding.
- Reduce the number of features → information loss.



The plot of the first 3 components indicates a correlation between the third component of the PCA and the heating examples. First 12 component selected (Variance explained = 0.99).

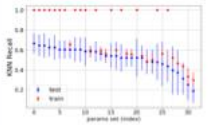
KNN and MLP Classifiers

Both a KNN classifier and a MLP classifier have been trained and tested and their performance have been compared.
For both of them the training has been performed with a 4-fold cross-validation (stratified).
GridSearch has been used for KNN and RandomizedSearch for MLP.

The appropriate value of K has been chosen for the KNN to avoid overfitting.
For the MLP overfitting has been avoided with early stopping.
For the hidden layers 'tanh' activation has been used. For the output 'sigmoid' activation has been used.

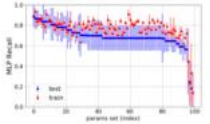
4. Numerical results

KNN



Best results with KNN (set #7):
(k=5)
Recall: 0.60 ± 0.09
Precision: 0.97 ± 0.04
F1-score: 0.71 ± 0.10

MLP



Best results with MLP (set #1):
(1 hidden layer, 88 neurons)
Recall: 0.86 ± 0.1
Precision: 0.58 ± 0.04
F1-score: 0.70 ± 0.10

Conclusion

- KNN has a very high precision score.
- MLP performs better in recall score thus results to be the best model for the target.

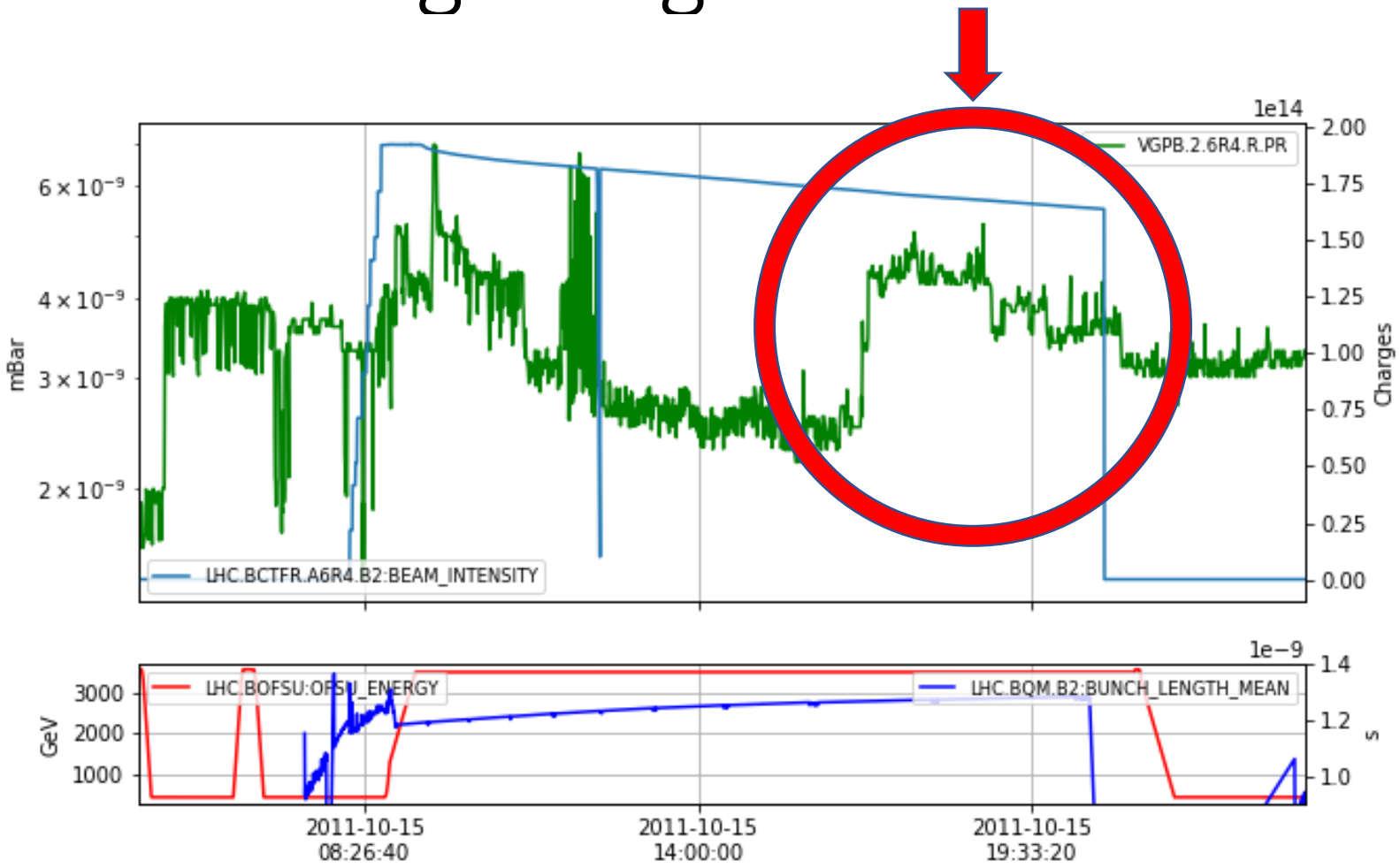
*Contacts:

Email: francesco.giordano@cern.ch
Phone: +41 779800820
LinkedIn: https://www.linkedin.com/in/francesco-giordano-3115a6150/

Or simply scan the QR

APPLIED MACHINE LEARNING DAYS. EPFL, LAUSANNE, SWITZERLAND, JAN 25-29, 2020

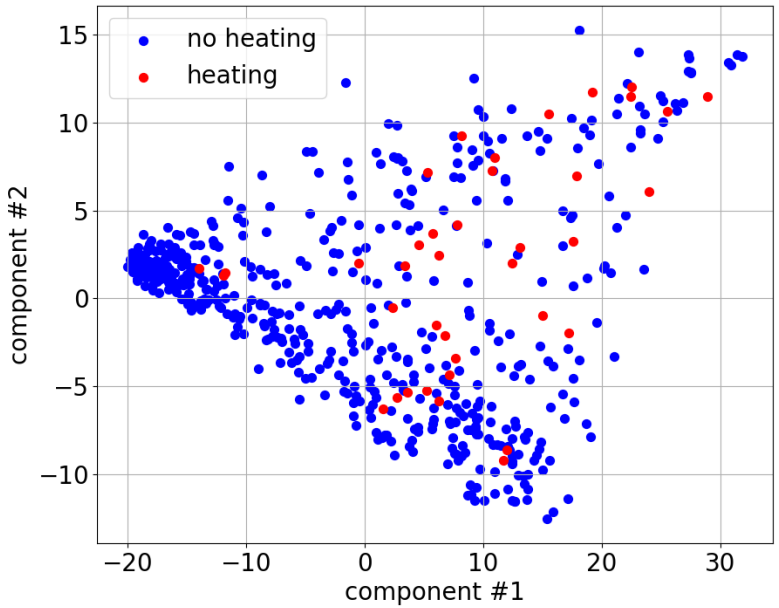
Target: identify possible heating from pressure reading using machine learning

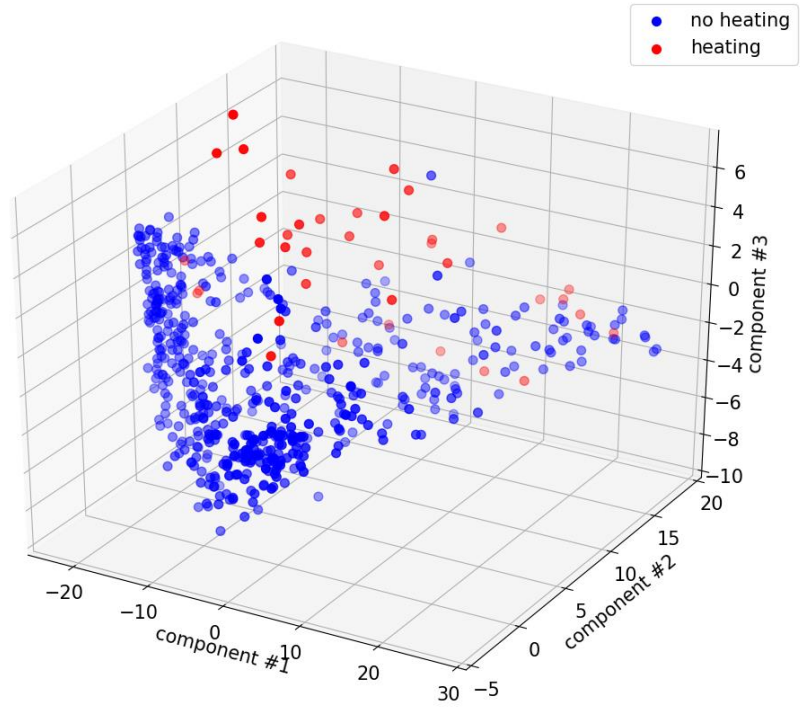


Brief recap: PCA

Each reading is on 3000 points, this creates a dataset of 700 examples x 3000 features.
More features than examples leads to 2 possible solutions:

- Add more examples → extremely time demanding.
- Reduce the number of features → information loss.





The plot of the first 3 components indicates a correlation between the third component of the PCA and the heating examples.
First 12 component selected (Variance explained = 0.99).

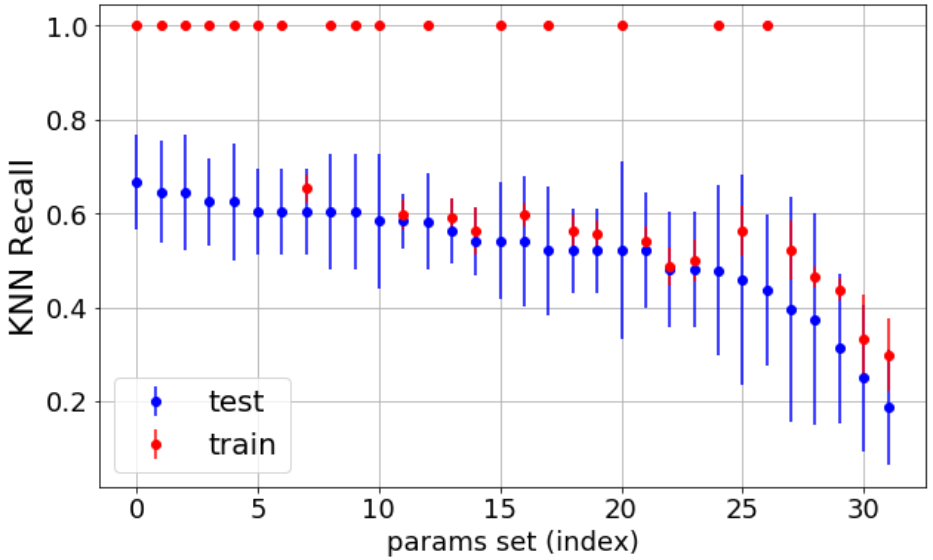
Brief recap: KNN and MLP Classifiers

- Both a KNN classifier and a MLP classifier have been trained and tested and their performance have been compared.
- For both of them the training has been performed with a 4-fold cross-validation (stratified).
- GridSearch has been used for KNN and RandomizedSearch for MLP.

- The appropriate value of K has been chosen for the KNN to avoid overfitting.
- For the MLP overfitting has been avoided with early stopping.
- For the hidden layers ‘tanh’ activation has been used. For the output ‘sigmoid’ activation has been used.

Brief recap: Results

KNN



Best results with KNN (set #7):

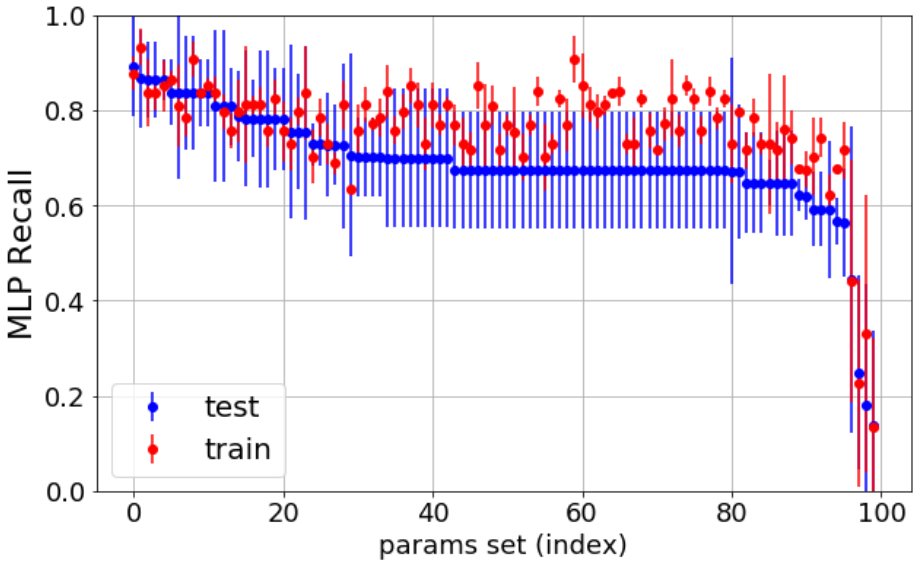
(k=5)

Recall: 0.60 ± 0.09

Precision: 0.97 ± 0.04

F1-score: 0.71 ± 0.10

MLP



Best results with MLP (set #1):

(1 hidden layer, 88 neurons)

Recall: 0.86 ± 0.1

Precision: 0.58 ± 0.04

F1-score: 0.70 ± 0.10

Conclusions

- KNN has a very high precision score.
- MLP performs better in recall score thus results to be the best model for the target.

Frequently asked questions:

- Can we arrange a visit at CERN?
- Have you ever seen the tunnel?
- … and of course a lot of useful suggestion to improve the classification of pressure readings

THANKS!