SDC 19

DELL EMC

September 23-26, 2019
Santa Clara, CA

New Perspective on Machine Learning Predictions Under Uncertainty

Rahul Vishwakarma, Jayanth Reddy

Agenda

Understanding Trustworthiness of Prediction

Quantifying Uncertainty

Application in risk-sensitive system

Why should we care?

Classical ML Approach

- Classification (Binary)
  - Input: $(x_1, y_1), (x_2, y_2), (x_n, y_n) \ldots (x_{n+1}, y_{n+1} = ?)$
  - Task: Predict *label* of (y_{n+1}) new data point
  - Prediction model : $f(x_i) \rightarrow (y_i)$

- Model performance
  - $f(x_i) \rightarrow (y_i)$ accuracy is 98% for $(x_1, y_1), (x_2, y_2), (x_n, y_n)$

Can we trust the prediction accuracy of (y_{n+1}) is also 98%
Are we really confident about the specific prediction

Motive

- Prediction models only output bare predictions but not the confidence in those predictions

- Obtain a metrics which explains:
  - Confidence of prediction for new label
  - Informativeness of each new data points

How do we quantify prediction uncertainty

How do we know when to trust our results?

Sources of Uncertainty in ML

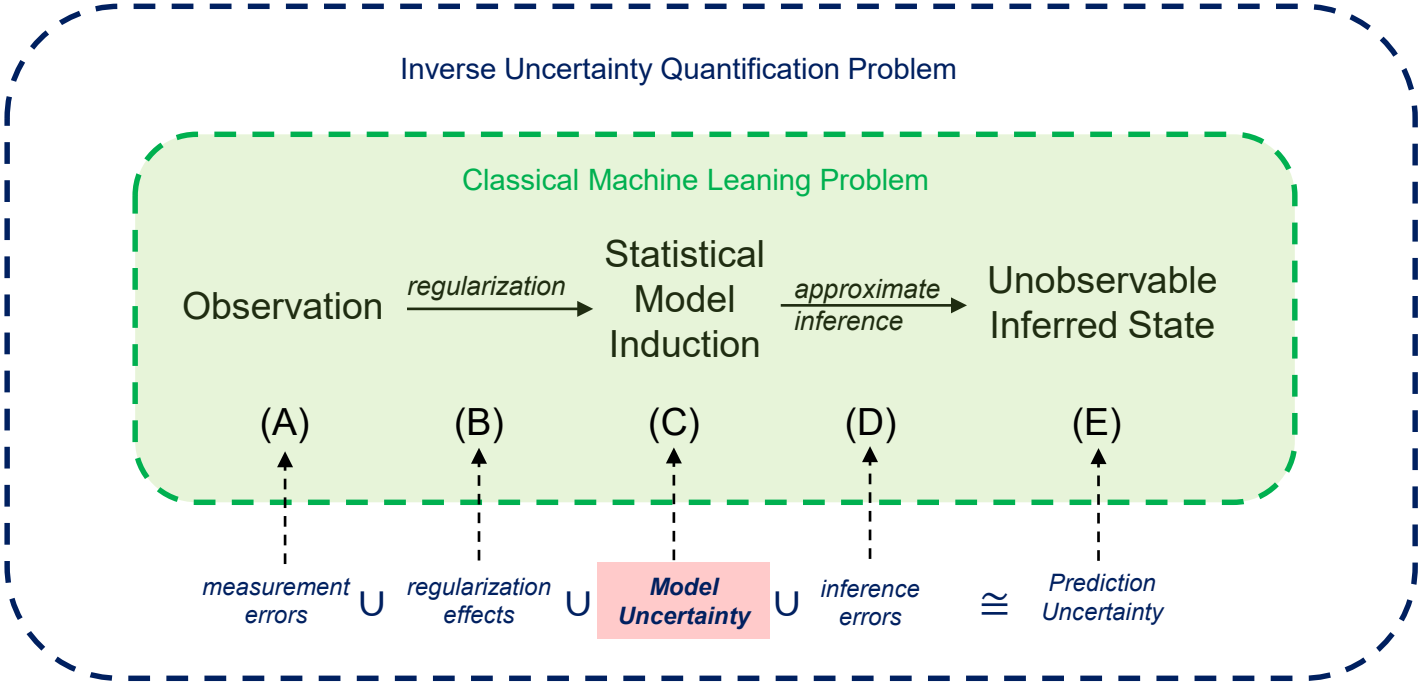


Figure 1: Sources of uncertainty in machine learning

Uncertainty Estimation

- Algorithmic Randomness
  - Algorithmic information theory provides universal measures of confidence but these are, unfortunately, **non-computable**
  - *Obtain practicable **approximations** to universal measures of confidence*

Conformal Prediction

Learning Framework

- Generalizable for any machine leaning algorithm

- Framework
  - Algorithmic randomness[1]
    - *problem of assigning confidences to predictions is closely connected to the problem of defining random sequences*
  - Hypothesis testing

[1]Algorithmic Learning in a Random World by Vovk, Gammerman and Shafer. Springer, 2005.

Assumption

Exchangeability

$$P\{(z_1, z_2, \dots) \in Z^{\infty} : (z_1, \dots, z_n) \in E\} = P\{(z_1, z_2, \dots) \in Z^{\infty} : (z_{\pi(1)}, \dots, z_{\pi(n)}) \in E\}$$

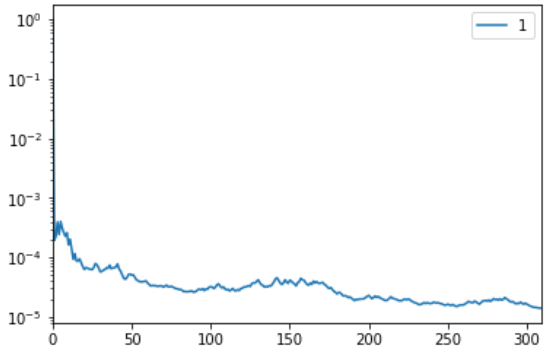


Figure 2: Exchangeable data

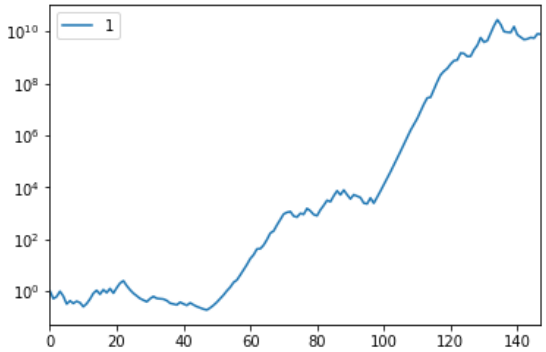


Figure 3: Non Exchangeable data

Conformal Prediction

$$\alpha_i^y = \frac{\sum_{j=1}^{k} D_{ij}^{y}}{\sum_{j=1}^{k} D_{ij}^{-y}}$$

$$p(\alpha_{n+1}^{y_p}) = \frac{\#\{i: \alpha_i^{y_p} \geq \alpha_{n+1}^{y_p}\}}{n+1}$$

$$\Gamma_{1-\varepsilon} = \{y_i : P^{yi} > \varepsilon, y_i \in Y\}$$

$$Confidence: 1 - p_{max}^{2nd}$$
$$Credibility: p_{max}$$

Method

Define nonconformity score for *k-NN* classifier

Method

Calculate p-value for each of the hypothesis

Method

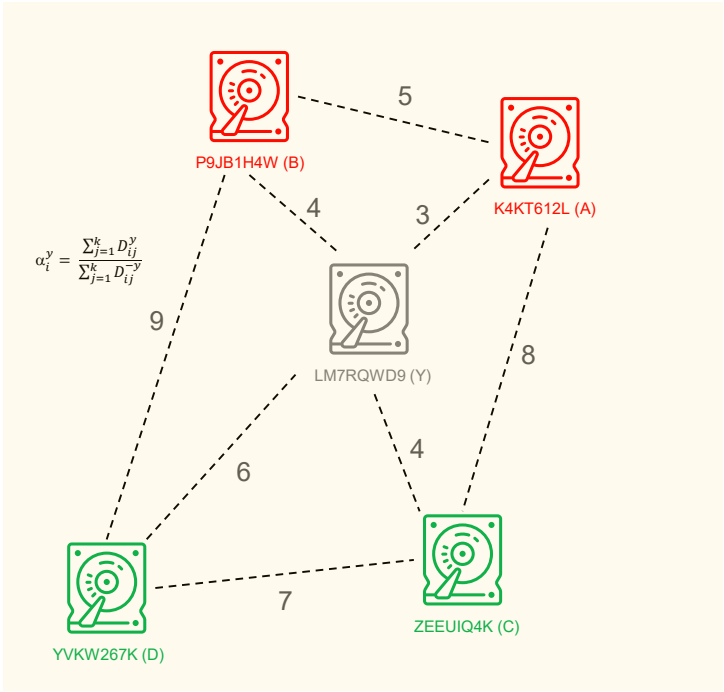
Prediction set whose p-value satisfies the confidence level

Method

Explainable reliability of each new prediction

Problem statement

Predict the label (Y) of new data point

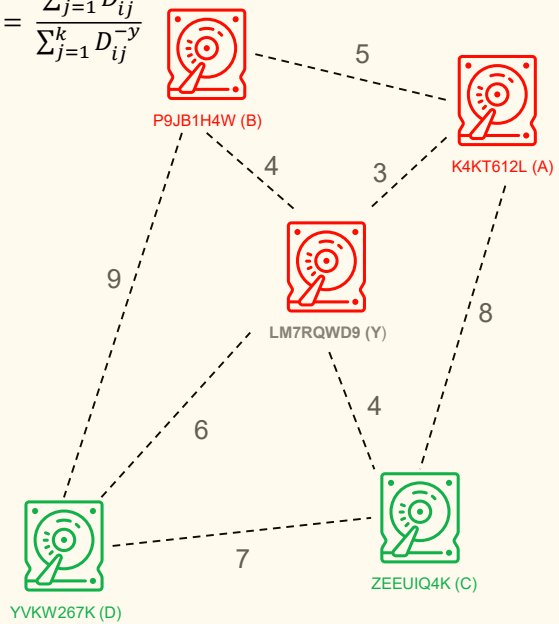


- Given
  - Disk Drives dataset
- Predict
  - Label of new Disk
  - Assign Confidence and Credibility
- Model
  - k-Nearest Neighbor (k =1)

Nonconformity score

Assuming the predicted label is a **Failed** disk: Calculate non conformity score

$$\alpha_i^y = \frac{\sum_{j=1}^{k} D_{ij}^{y}}{\sum_{j=1}^{k} D_{ij}^{-y}}$$



Let Y = 

$$\alpha_A = \frac{d_{|A,Y|}}{d_{|A,C|}} = \frac{3}{8}$$

$$\alpha_B = \frac{d_{|B,Y|}}{d_{|B,D|}} = \frac{4}{9}$$

$$\alpha_C = \frac{d_{|C,D|}}{d_{|C,Y|}} = \frac{7}{4}$$

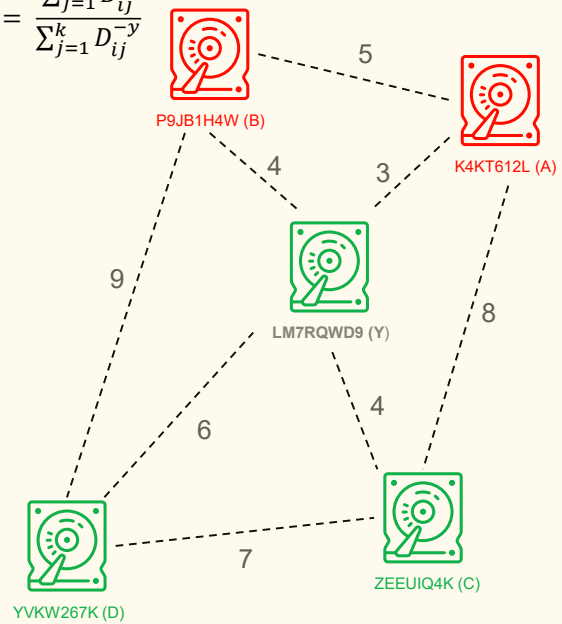
$$\alpha_D = \frac{d_{|D,C|}}{d_{|D,Y|}} = \frac{7}{6}$$

$$\alpha_Y = \frac{d_{|Y,A|}}{d_{|Y,C|}} = \frac{3}{4}$$

Nonconformity score

Assuming the predicted label is a **Normal** disk: Calculate non conformity score

$$\alpha_i^y = \frac{\sum_{j=1}^{k} D_{ij}^{y}}{\sum_{j=1}^{k} D_{ij}^{-y}}$$



Let Y = 

$$\alpha_A = \frac{d_{|A,B|}}{d_{|A,Y|}} = \frac{5}{3}$$

$$\alpha_B = \frac{d_{|B,A|}}{d_{|B,Y|}} = \frac{5}{4}$$

$$\alpha_C = \frac{d_{|C,Y|}}{d_{|C,A|}} = \frac{4}{8}$$

$$\alpha_D = \frac{d_{|D,Y|}}{d_{|D,B|}} = \frac{6}{9}$$

$$\alpha_Y = \frac{d_{|Y,C|}}{d_{|Y,A|}} = \frac{4}{3}$$

p-value and Prediction

Let Y = 	Let Y = 
$\alpha_A = \frac{3}{8} = 0.37$	$\alpha_A = \frac{5}{3} = 1.66$
$\alpha_B = \frac{4}{9} = 0.44$	$\alpha_B = \frac{5}{4} = 1.25$
$\alpha_C = \frac{7}{4} = 1.75$	$\alpha_C = \frac{4}{8} = 0.5$
$\alpha_D = \frac{7}{6} = 1.16$	$\alpha_D = \frac{6}{9} = 0.66$
$\alpha_Y = \frac{3}{4} = 0.75$	$\alpha_Y = \frac{4}{3} = 1.33$
p  $= \frac{3}{5} = 0.6$	p  $= \frac{2}{5} = 0.4$

$$p(\alpha_{n+1}^{y_p}) = \frac{\#\{i: \alpha_i^{y_p} \geq \alpha_{n+1}^{y_p}\}}{n+1}$$

Higher non-conformity : Lower p-value

$$Credibility = p_{max}$$
$$Confidence = 1 - p_{max}^{2nd}$$

 = 

LM7RQWD9 (Y) LM7RQWD9 (Y)

$Confidence = 0.6$

$Credibility = 0.6$

Variants of Conformal Prediction

- Inductive Conformal Prediction
  - Computationally efficient
- Aggregated Conformal Prediction
- Cross Conformal Prediction

Applications

Applications

- Hardware failure prediction
  - Disk drive
- Anomaly detection
  - Time Series

- Feature selection
- Prediction quality assessment

Disk Drive Failure Prediction

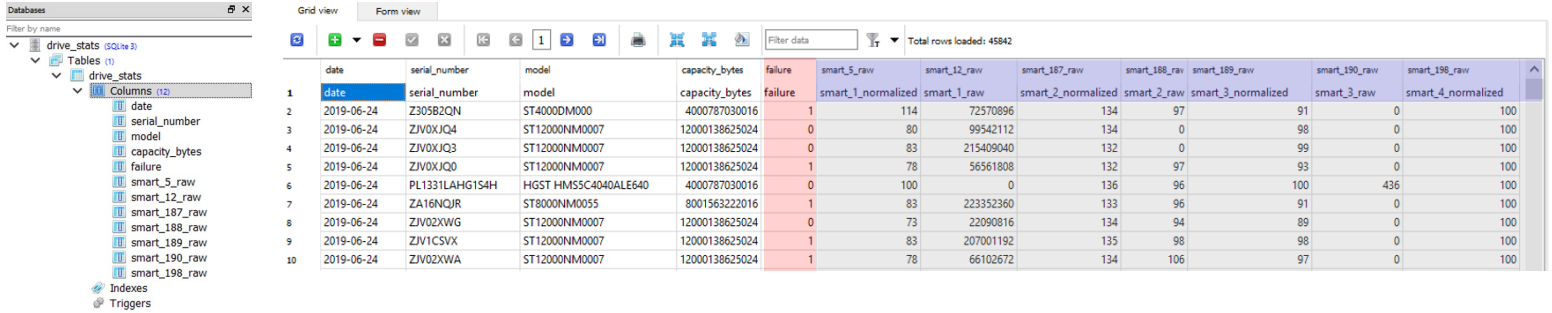
- Dataset: Black Blaze Q2_2019
- Classification using Conformal Framework
  - *K*- Nearest Neighbors

  Want to try out Conformal Classification package?

  https://pypi.org/project/ConfClr/ (pip install ConfClr)
- Interpretation of results
  - Confidence
  - Credibility
- Application
  - Ability to Rank the Labels

https://f001.backblazeb2.com/file/Backblaze-Hard-Drive-Data/data_Q2_2019.zip

Dataset



	date	serial_number	model	capacity_bytes	failure	smart_5_raw	smart_12_raw	smart_187_raw	smart_188_raw	smart_189_raw	smart_190_raw	smart_198_raw
1	date	serial_number	model	capacity_bytes	failure	smart_1_normalized	smart_1_raw	smart_2_normalized	smart_2_raw	smart_3_normalized	smart_3_raw	smart_4_normalized
2	2019-06-24	Z305B2QN	ST4000DM000	4000787030016	1	114	72570896	134	97	91	0	100
3	2019-06-24	ZJV0XJQ4	ST12000NM0007	12000138625024	0	80	99542112	134	0	98	0	100
4	2019-06-24	ZJV0XJQ3	ST12000NM0007	12000138625024	0	83	215409040	132	0	99	0	100
5	2019-06-24	ZJV0XJQ0	ST12000NM0007	12000138625024	1	78	56561808	132	97	93	0	100
6	2019-06-24	PL1331LAHG1S4H	HGST HMS5C4040ALE640	4000787030016	0	100	0	136	96	100	436	100
7	2019-06-24	ZA16NQJR	ST8000NM0055	8001563222016	1	83	223352360	133	96	91	0	100
8	2019-06-24	ZJV02XWG	ST12000NM0007	12000138625024	0	73	22090816	134	94	89	0	100
9	2019-06-24	ZJV1CSVX	ST12000NM0007	12000138625024	1	83	207001192	135	98	98	0	100
10	2019-06-24	ZJV02XWA	ST12000NM0007	12000138625024	1	78	66102672	134	106	97	0	100

Disk Drive Dataset

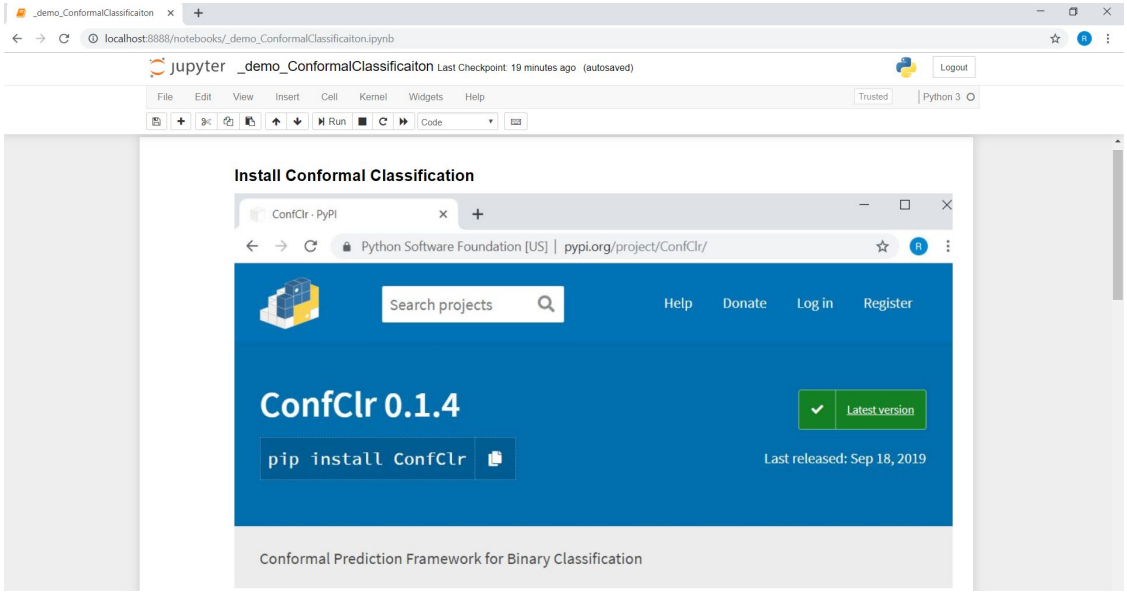
In [2]:
```
import pandas as pd
data = pd.read_csv('HDD.csv')
data.head()
```

Out[2]:

	0.419355	7	3.2	4.7	4.7.1	1.4	1.4.1	NORMAL
0	0.419355	6.4	3.2	4.5	4.5	1.5	1.5	NORMAL
1	0.225806	6.9	3.1	4.9	4.9	1.5	1.5	NORMAL
2	0.290323	5.5	2.3	4.0	4.0	1.3	1.3	NORMAL
3	0.290323	6.5	2.8	4.6	4.6	1.5	1.5	NORMAL
4	0.451613	5.7	2.8	4.5	4.5	1.3	1.3	NORMAL

https://f001.backblazeb2.com/file/Backblaze-Hard-Drive-Data/data_Q2_2019.zip

Package ConfClr




```
In [1]: !pip install ConfClr
Collecting ConfClr
  Using cached https://files.pythonhosted.org/packages/ee/83/6948bdb787bc048020410e50a10ae1ca34ce90c9f55da81fe31b84b44ab3/ConfClr-0.1.4-py3-none-any.whl
Installing collected packages: ConfClr
Successfully installed ConfClr-0.1.4
```

Prediction with Explainable Reliability

Compute p-values for each label and calculate Confidence and Credibility

In [4]:
```
import ConfClr
CC.conformalScore(x,y)
```

Out[4]:

	p-FAILED	p-NORMAL	Status	Credibility	Confidence
0	0.360656	0.147541	Failed	0.360656	0.639344
1	0.213115	0.147541	Failed	0.213115	0.786885
2	0.180328	0.508197	Normal	0.508197	0.819672
3	0.229508	0.147541	Failed	0.229508	0.770492
4	0.163934	0.131148	Failed	0.163934	0.836066
5	0.180328	0.147541	Failed	0.180328	0.819672
6	0.147541	0.721311	Normal	0.721311	0.852459
7	0.163934	0.081967	Failed	0.163934	0.836066
8	0.655738	0.147541	Failed	0.655738	0.344262
9	0.180328	0.721311	Normal	0.721311	0.819672
10	0.163934	0.737705	Normal	0.737705	0.836066
11	0.065574	0.655738	Normal	0.655738	0.934426
12	0.163934	0.852459	Normal	0.852459	0.836066
13	0.622951	0.147541	Failed	0.622951	0.377049
14	0.721311	0.147541	Failed	0.721311	0.278689
15	0.770492	0.147541	Failed	0.770492	0.229508
16	0.163934	0.147541	Failed	0.163934	0.836066
17	0.163934	0.803279	Normal	0.803279	0.836066
18	0.852459	0.147541	Failed	0.852459	0.147541
19	0.409836	0.147541	Failed	0.409836	0.590164

- Confidence level of 95% means that in 95% the predicted class is also a true class
- $Credibility = p_{max}$
- $Confidence = 1 - p_{max}^{2nd}$

Interpretation

Plot Confidence and Credibility of each new Label

Expected: High Confidence and Credibility not too low

In [5]:
```
import ConfClr
CC.plotConfCred(x,y)
```

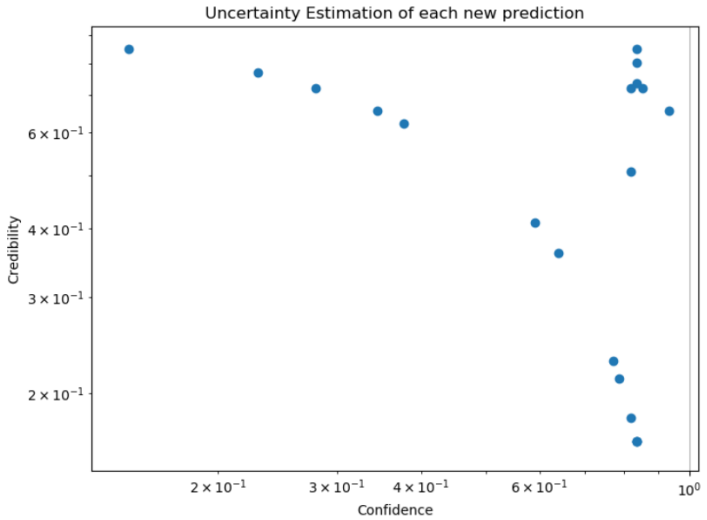


Figure 4: Interpreting quality of prediction

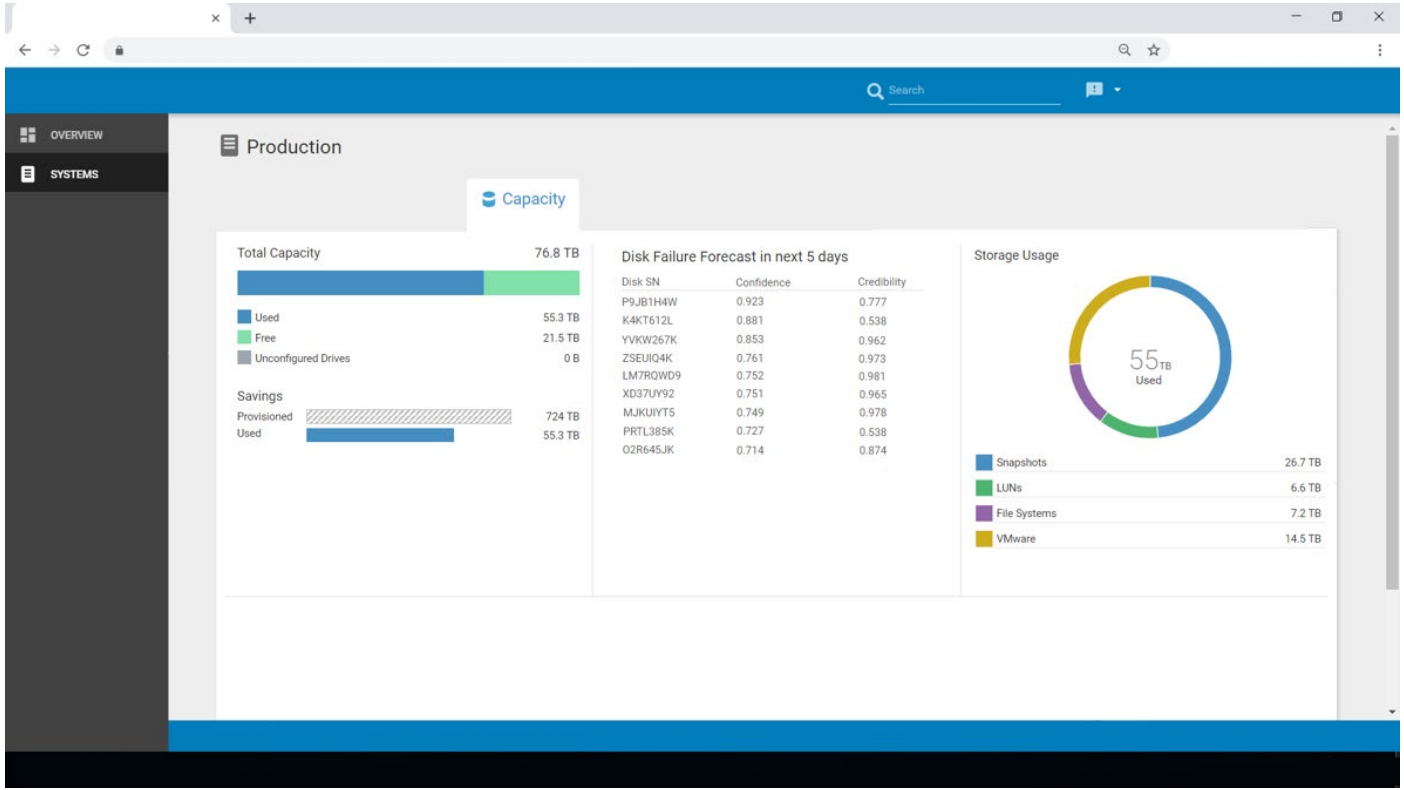
Framework Application

Disk Serial Number	Status	Confidence
Z305B2QN	FAILED	0.950820
ZA16NQJR	FAILED	0.918003
ZJV1CSVX	FAILED	0.868852
ZA18CEBF	FAILED	0.737705
ZJV02XWA	FAILED	0.672131
ZJV0XJQ0	FAILED	0.426230
ZJV02XWV	FAILED	0.114754
PL2331LAG9TEEJ	FAILED	0.098361

Figure 4: Ranking the failed drive (Confidence)

- Priority based insightful action
- Efficient Inventory management

Sample Interface - Application



Conclusion and Discussion

Challenges

- Designing nonconformity score
- Application in Time Series

Summary

- Standard prediction models are based on bare predictions
  - Classification: discrete classes
  - Regression: real-valued point prediction
- Model with high accuracy does not guarantee reliable forecasting

- Conformal prediction can be used with any machine learning
- Conformal Framework
  - Confidence of each single prediction
  - Credibility tells about the quality of data point

Why should we care?

It matters how much we can rely upon a given prediction in risk-sensitive applications.

Horizon 2020
European Union Funding
for Research & Innovation

- Drug design

LVE

- Insurance
- Investment

AstraZeneca

- Development of new drugs
- Precision medicine & Genomics

Seven scientific experts in Italy were convicted of manslaughter and sentenced to six years in prison for failing to give warning before the April 2009 earthquake that killed 309 people.

Diacu F. Is failure to predict a crime? October 2012. New York Times; October 2012. https://www.nytimes.com/2012/10/27/opinion/a-failed-earthquake-prediction-a-crime.html

References

- V. Vovk, A. Gammerman, and G. Shafer, Algorithmic learning in a random world. Springer, 2005.
- G. Shafer and V. Vovk, “A tutorial on conformal prediction,” The Journal of Machine Learning Research, vol. 9, pp.
- Balasubramanian, V., Ho, S. and Vovk, V. (2014). Conformal Prediction for Reliable Machine Learning.

"*As far as the laws of mathematics refer to reality, they are not certain; and as far as they are certain, they do not refer to reality*." – Albert Einstein