Report 1

LI, Ruizhe | 1155076990

Github: <https://github.com/rzli6/ML-Storage.git> (private)

Target Paper: Predicting Disk Replacement towards Reliable Data Centers

Goal: Realize the method descripted in the paper.

Dataset: 2017 Q1 – 2018 Q1

Model: ST4000DM000 (SgtA)

|  |  |  |  |
| --- | --- | --- | --- |
| Serial\_number | Percentage | Serial\_number | Percentage |
| smart\_242\_raw | 0.571197 | smart\_193\_normalized | 0.279935 |
| smart\_7\_raw | 0.554207 | smart\_1\_raw | 0.224919 |
| smart\_193\_raw | 0.529126 | smart\_5\_raw | 0.204693 |
| smart\_9\_raw | 0.505663 | smart\_3\_normalized | 0.135113 |
| smart\_240\_raw | 0.501618 | smart\_183\_raw | 0.127832 |
| smart\_190\_normalized | 0.487055 | smart\_183\_normalized | 0.127832 |
| smart\_190\_raw | 0.486246 | smart\_198\_normalized | 0.11246 |
| smart\_194\_raw | 0.486246 | smart\_197\_normalized | 0.11246 |
| smart\_194\_normalized | 0.486246 | smart\_5\_normalized | 0.064725 |
| smart\_9\_normalized | 0.467638 | smart\_192\_raw | 0.045307 |
| smart\_198\_raw | 0.443366 | smart\_189\_raw | 0.021036 |
| smart\_197\_raw | 0.443366 | smart\_189\_normalized | 0.020227 |
| smart\_241\_raw | 0.412621 | smart\_184\_normalized | 0.018608 |
| smart\_7\_normalized | 0.384304 | smart\_184\_raw | 0.018608 |
| smart\_187\_normalized | 0.337379 | smart\_188\_raw | 0.0089 |
| smart\_187\_raw | 0.337379 | smart\_199\_raw | 0.004854 |
| smart\_4\_raw | 0.308252 | smart\_4\_normalized | 0.000809 |
| smart\_12\_raw | 0.307443 | smart\_192\_normalized | 0.000809 |
| smart\_1\_normalized | 0.291262 | smart\_241\_normalized | 0 |
| smart\_3\_raw | 0 | smart\_10\_raw | 0 |
| smart\_191\_normalized | 0 | smart\_191\_raw | 0 |
| smart\_240\_normalized | 0 | smart\_188\_normalized | 0 |
| smart\_199\_normalized | 0 | smart\_242\_normalized | 0 |
| smart\_10\_normalized | 0 | smart\_12\_normalized | 0 |

Features Selected (percentage > 0.01)

'smart\_242\_raw', 'smart\_7\_raw', 'smart\_193\_raw', 'smart\_9\_raw',

'smart\_240\_raw', 'smart\_190\_normalized', 'smart\_190\_raw',

'smart\_194\_raw', 'smart\_194\_normalized', 'smart\_9\_normalized',

'smart\_198\_raw', 'smart\_197\_raw', 'smart\_241\_raw', 'smart\_7\_normalized',

'smart\_187\_normalized', 'smart\_187\_raw', 'smart\_4\_raw', 'smart\_12\_raw',

'smart\_1\_normalized', 'smart\_193\_normalized', 'smart\_1\_raw',

'smart\_5\_raw', 'smart\_3\_normalized', 'smart\_183\_raw',

'smart\_183\_normalized', 'smart\_198\_normalized', 'smart\_197\_normalized',

'smart\_5\_normalized', 'smart\_192\_raw', 'smart\_189\_raw',

'smart\_189\_normalized', 'smart\_184\_normalized', 'smart\_184\_raw'

Features not in paper:

3: Spin-Up Time (NA)

4: Start/Stop Count (not in)

9: Power-On Hours (not in)

12: Power Cycle Count (not in)

183: SATA Downshift Error Count or Runtime Bad Block (0.5%)

192: Power-off Retract Count, Emergency Retract Cycle Count (not in)

Precision, Recall, F-score, Deviation of diﬀerent classiﬁers

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | RGF | GBDT | RF | SVM | LR | DT |
| F1 | 0.988 | 0.990 | 0.986 | 0.990 | 0.897 | 0.991 |
| Recall | 0.986 | 0.986 | 0.992 | 1.0 | 0.870 | 0.989 |
| Precision | 0.990 | 0.993 | 0.980 | 0.980 | 0.948 | 0.993 |

Apply RGF model to *ST8000DM002* **without** transfer learning:

F1: 0.022

Recall: 0.991

Precision: 0.011

Conclusion: It nearly predicts all the test cases to be failed, which is far from the truth (only 114 truly failed in test cases, but in prediction 9999 failed). Hence, without transfer learning, the trained-model is not applicable to another disk model, even if they are from the same producer.