SmartPred: Unsupervised Hard Disk Failure Detection

Philipp Rombach and Janis Keuper

Institute for Machine Learning and Analytics (IMLA),
Offenburg University, Germany
{rombach,keuper}@imla.ai

Abstract. Due to the rapidly increasing storage consumption worldwide, as well as the expectation of continuous availability of information, the complexity of administration in today's data centers is growing permanently. Integrated techniques for monitoring hard disks can increase the reliability of storage systems. However, these techniques often lack intelligent data analysis to perform predictive maintenance. To solve this problem, machine learning algorithms can be used to detect potential failures in advance and prevent them. In this paper, an unsupervised model for predicting hard disk failures based on Isolation Forest is proposed. Consequently, a method is presented that can deal with the highly imbalanced datasets, as the experiment on the Backblaze benchmark dataset demonstrates.

Keywords: Unsupervised learning · Hard disk drives · Anomaly detection

1 Introduction

The prediction of the reliability of hard disk drives was born out of the need to protect stored information on hard disks from data loss. For this purpose, several hard drive manufacturers have developed a technology that monitors and analyzes the current state of a hard drive, also known as Self-Monitoring, Analysis and Reporting Technology (SMART). SMART uses sensors to collect information about the state of magnetic hard disks and Solid State Drives (SSDs). From the collected sensor values, SMART creates an overview of the hard disk's health and can indicate a current failure. However, no intelligence analysis is performed by combining several sensor values. In particular, a predictive analysis of the sensor values would make it possible to detect potential errors or failures in advance. Such predictive maintenance would make it possible to increase the reliability of storage systems by replacing hard disks before they fail.

The recorded performance values of SMART cover 62 attributes. Each attribute has assigned a threshold value based on the experience of the manufacturer. If an attribute exceeds its threshold value, the drive is marked as faulty [15]. Since it is possible to return the drive to the manufacturer for a warranty replacement if the threshold is exceeded, it is reasonable to assume that manufacturers carefully reduce the false alarm rates of their predictions. This assumption is backed by the evaluation in [6], which showed that the current SMART algorithms implemented in the drives have failure detection rates ranging from only 3 to 10%.

Hard disk failures can be divided into two different categories. On the one hand, there are predictable failures that can be detected before a hard disk fails. On the other hand, there are unpredictable failures. In this case, the sudden death of the hard disk occurs [20]. Unpredictable failures occur quickly and suddenly and mean that no prediction can be made at the time of failure. The SMART attributes remain constant during this type of failure and therefore show no variance in the recorded values. These failures cannot be detected using the values logged with SMART. Predictable failures are caused by the worsening of at least one of the SMART attributes overtime before the hard disk fails. By monitoring these SMART attributes, predictive failure analysis is possible, and it can be determined whether the hard disk needs to be replaced. In [11], the percentage of predictable hard disk failures is given as 60%.

There are already several research papers that deal with the failure detection of hard disks [3, 4, 7, 19, 14, 13, 5, 12, 16, 18], they evaluate the results using a binary classification and thus predict whether a hard disk will fail. However, differences in the experimental setups make it challenging to compare the performance of the models created. This includes the choice of dataset and the choice of metrics used to evaluate the model.

The objective of this paper is to perform a predictive analysis based on the SMART values and to divide the results into two categories. These categories reflect the failure probability of a hard disk over a specified period. The analysis is carried out using an anomaly detection algorithm since faulty hard disks represent only a tiny minority and can be referred to as anomalies.

In summary, the major contributions in this paper are the following:

- Transfer of anomaly detection techniques to the failure prediction in hard disk drives.
- By choosing the Isolation Forest algorithm, a model is presented to handle the highly imbalanced dataset without preprocessing. Therefore no downsampling or upsampling of the instances is necessary.
- The unsupervised method can be performed on a few samples without the need to generate a training data set. Thus, there is no significant delay between data acquisition and prediction.
- The detected anomalies are divided into two categories based on their failure probability. Depending on the classified category, a replacement process for the hard disk is proposed.

The structure of this work is divided into four basic sections. Section 2 deals with the research work done so far and its results. In section 3, the model created to predict hard disk failures is presented. The results achieved are described and evaluated in section 4. Furthermore, possible limitations of the dataset are discussed. The work is rounded off by section 5 with a conclusion.

2 Related Work

This section provides an overview of research work focusing on predicting disk failure. All studies of investigation take the SMART attributes into account. The metrics used to evaluate the results are Failure Detection Rate (FDR), False Alarm Rate (FAR), and the failure rate of hard disks in the dataset. Recall, or in the context of hard disk failure detection, also called FDR, shows how complete the results are. FDR is a good metric for unbalanced datasets because it only refers to the anomalies. In equation 1 the calculation of FDR is shown.

$$FDR = \frac{TP}{TP + FN} \tag{1}$$

False Positive Rate (FPR) is referred to as the FAR value when detecting hard disk failure. FAR is the ratio of correctly detected normal instances to false positive anomalies. In anomaly detection, the goal is to keep the FAR value as low as possible. In equation 2 the calculation of FAR is shown.

$$FAR = \frac{FP}{FP + TN} \tag{2}$$

Hamerly and Elkan [3] use two Bayesian approaches, Naive Bayes Expectation Maximization (NBEM), and naive Bayesian classifier, to create semi-supervised models. Their data is provided by Quantum Inc. and includes 1927 good drives and nine failed drives. They achieve failure detection rates of 35% to 40% for NBEM and 55% for the naive Bayes classifier at approximately 1% FAR. Hughes et al. [4] use the Wilcoxon rank-sum test to create predictive models. Since they observed that most of the SMART attributes are distributed more non-parametrically, their model has tested 3780 drives, with a failure rate of 0.9%. They achieve a detection rate of 60% with a false alarm rate of 0.5%. In their later research [7], they used several methods, including rank-sum testing, Support Vector Machine (SVM), and unsupervised clustering. The dataset was substantially smaller, with a population of 369 drives and a failure rate of 51%. SVM achieved the best results with a FDR of 50.6% and a FAR of 0%. 25 SMART attributes were used to create the SVM model. By additionally using the change rates of the SMART attributes, Zhu et al. in [19] were able to improve the SVM model to achieve a FDR of 80% at 0.3% FAR.

Wang et al. [14] proposed a strategy to predict drive anomalies based on Mahalanobis distance. They used the same dataset as in [4, 7] and showed that the method with prioritized attributes selected by the Failure Modes, Mechanisms and Effects Analysis (FMMEA) performed better than the method with all attributes. In their subsequent study [13], minimum Redundance Maximum Relevance (mRMR) was used to remove redundant information from the attribute selected by the FMEA. Using these critical parameters, they built a baseline Mahalanobis space. This model could detect 68% of the faulty disks with 0% FAR.

The research work of Zhu et al. in [19] not only included the improvement of the SVM model, they furthermore created a second model with Back Propagation Artificial Neural Network (BP ANN). BP ANN aimed to increase the failure detection rate significantly while keeping the FAR low. The models were created on a real dataset with 23395 drives, and this dataset has a failure rate of 1.9%, which is significantly lower than the dataset used in [4, 7]. The collected SMART attributes covered eight weeks and were provided from the Baidu data center. The BP ANN failure detection rate was 95%, and the FAR was reported as reasonably low. In another paper by Zhu et al. [5], the Classification and Regression Trees (CART) algorithm is used to create a model for predicting hard disk failures. Compared to the BP ANN, the advantages of a CART model are improved prediction results and better stability and interpretability. On the Baidu dataset, they achieve a FDR of 95% with a FAR of 0.1%.

The proposed model of Xu et al. [17] used a Recurrent Neural Network (RNN) to predict hard disk failures and to assess the health of hard disks. The SMART attributes divided by their timestamp are used as input data. The results from the model are divided into six levels that reflect the health of a hard disk. The smaller the level, the higher the risk of hard disk failure. Level 6 means that there is no limitation on the hard disk, and it is functioning reliably. If the hard disk is assigned to level 1, it means that the hard disk will fail in less than 72 hours. Several RNN-based models were created, focusing on maximizing the FDR and minimizing the FAR. As a result, the predicted results were a FDR of 87% with a very low FAR of 0.004% and a FDR of 97.7% and a FAR of 0.59%.

Shen et al. [12] and Xiao et al. [16] both use the random forest as the underlying technology in their research. They also use datasets provided by Backblaze [1] to train and test their models. In [12], the Random Forest (RF) based model is improved by additional voting. For this purpose, a sliding window is created that always trains samples from a hard disk of 30 days, and if the number of bad samples exceeds a limit, this hard drive is marked as faulty. The model achieves a FDR of 95% with a FAR of 0.4%. In [16], the focus is on training the model to handle streaming data and process it on-the-fly. This is also called Online Random Forest (ORF) [9]. By using ORF, performance is increased, and less memory is used. However, the ORF model takes up to six months to converge to offline random forest models' performance. The FDR on ORF is 98%, with a FAR of 0.7%.

Anomaly detection is performed by Zhang et al. in [18] using Isolation Forest. The dataset is again from Backblaze in the second quarter of 2018 and is trained on several training datasets where the number of failed disks varies from 2% to 10%. The FDR was not specified in the research work, but the accuracy is described with up to 95% and an average FAR of 5%. Also, the better performance in terms of training duration compared to the Random Forest is highlighted.

The presented research papers are sorted chronologically, starting with the oldest research [3] from the year 2001 up to the newest paper [18] from the year 2019. Most of the research has been focused on the development of supervised models, resulting in improved predictive performance over the years as measured by the FDR and FAR values. However, it must be taken into account that the used supervised models provide their results in the form of a binary classification, which cannot reflect the deterioration of a hard disk in reality.

3 The Proposed Method

This section describes the process of creating the Isolation Forest models to predict hard drive failures, intending to predict a failure probability for all hard drives, group the drives based on this probability, and identifying the suspect drives by their serial number. As a first step, exploratory data analysis is performed, and relevant features are selected prior to creating the model and optimize the hyperparameters.

3.1 Dataset

The dataset used to predict disk failure is derived from Backblaze [1]. Backblaze is a cloud storage provider that enables users to store their backups online. The company provides data storage for both private and business purposes. All hard drives used by Backblaze are monitored, and the SMART attributes are logged daily. The collected SMART data is provided in datasets that are publicly available and can be used freely.

Since 2013, Backblaze provides its SMART datasets for each quarter. These are compressed and contain one file for each day, which is stored as structured data. The number of features has been continuously increased since 2013 and reached 129 features in 2019, of which the first five features are reserved for identifying the hard disk.

The *date* feature contains the day on which the values were recorded. The *serial_number* feature is used to identify the disk in the dataset. All hard disks of the same type are combined in the feature *model*. The storage capacity of the hard disk is specified in the feature *capacity_bytes* in bytes. The *failure* feature is an integer in the value range $[0, 1]$ and contains the value 0 if the disk is healthy and the value 1 if this is the last recorded sample of the disk before the failure.

The remaining 124 features are used for the SMART attributes, divided into *smart_x_raw* and *smart_x_normalized*. The features for the raw values contain the real values recorded; these are stored as floating-point numbers. In contrast, for the normalized values, a vendor-specific function has been executed based on the raw values to store them in a specific value range.

To highlight the deterioration of a hard disk, the list of features is extended. For this purpose an additional feature *smart_x_raw_diff* is created for each feature of *smart_x_raw*. These features contain the difference value of a disk to the previous day. The hard disks must be grouped by their serial numbers and sorted by date to calculate the difference value. Afterward, for each feature of *smart_x_raw*, the calculation of the difference value to the previous value is performed. Since no difference can be calculated for the first entry in the dataset, this value is set to 0.

The Backblaze dataset consists of all in all 136568 hard disks with 40737546 samples for the year 2019 and is, therefore, one of the largest SMART datasets publicly available. The failure rate and wear and tear of hard disks vary according to the model and manufacturing process [10, 8]. In order to minimize the effects of the different models, this work concentrates on one specific model. This limitation is negligible as the same methodology can be conducted for other hard disk models in the same manner. The number of hard disks and the number of failed disks were used as selection criteria for the model. Thus, the model *ST12000NM0007* from Seagate is taken into account and, for simplification, is referred to as *ST1* in the following. The model *ST1* contains a total of 38256 hard disks, of which 1155 hard disks fail over a period of one year. Over this period, 12721076 samples were recorded.

3.2 Feature Selection

To determine the relevant features for prediction, the correlation coefficient between all SMART Attributes is calculated concerning the feature *failure*. The correlation is a bivariate analysis that measures the strength of the association between two features and the direction of the relationship.

To select the relevant features for the Isolation Forest, features that show a positive correlation with the feature *failure* are filtered. The table 1 shows all features with a positive correlation.

Table 1. Descriptive statistics of *ST1*

	5_raw	**187_raw**	**197_raw**	**5_diff**	**187_diff**	**197_diff**
mean	27.85	1.1	0.14	0.65	0.03	0.01
std	697.15	201.08	15.52	117.97	40.52	9.09
min	0.0	0.0	0.0	-63368.0	-13.0	-280.0
25%	0.0	0.0	0.0	0.0	0.0	0.0
50%	0.0	0.0	0.0	0.0	0.0	0.0
75%	0.0	0.0	0.0	0.0	0.0	0.0
max	65528.0	65535.0	30960.0	57056.0	65535.0	30944.0

The measurements representing the selected raw features are described below:

Reallocated Sectors Count (SMART 5): is the total number of defective sectors that have been detected and reallocated.

Reported Uncorrectable Errors (SMART 187): is the number of errors that could not be recovered with Error Correcting Codes (ECCs).

Current Pending Sector Count (SMART 197): is the number of sectors waiting to be reassigned to the spare area due to uncorrectable errors in reading and writing a sector.

3.3 Preprocessing

To determine whether a hard disk will fail, the values in each feature must have a variance greater than 0 concerning this hard disk. If the disk features have a constant value of 0, they are among the 40% of disks that fail without significant SMART attributes [11] and are therefore considered false negative by the model. This is due to the fact that the values do not differ from healthy disks and, therefore, cannot be detected as anomalies. For this reason, the failed disks without variance are removed from the dataset.

An important step to avoid incorrect predictions is to inspect the dataset for possible inconsistent data and to correct them. This process is also called data cleaning and is divided into four steps:

1. Check the dataset for possible duplicates, if duplicates exist, remove these samples from the dataset
2. Search in the dataset for values without content. This can result from incorrect data collection and processing. If there are missing values in the dataset, delete these samples from the dataset.
3. Check for each disk whether the values are complete. This means that for a working disk, the last sample must match the last day of the dataset, and for a failed disk, the last sample in the *failure* feature must contain a value of 1. If these conditions are not met, all samples from that disk will be removed from the dataset.
4. For each failed hard disk, check if it has a variance greater than 0 in its features. If not, the hard disk will be removed from the dataset.

After data cleaning, the dataset *ST1* contains 37768 disks with 12614746 samples. Thus 1.28% of the data is removed by the cleanup process. With the failing disks, 287 disks fail without the required variance in features, leaving 868 disks marked as failed in the dataset, which is 75.2%.

3.4 Setup

In order to be able to group the failure probabilities for all hard disks, two Isolation Forest models are created, which differ in their parameters. The metrics used to evaluate the Isolation Forest models are also different. For the first Isolation Forest Model *IF_FDR*, the failure detection rate is used as the relevant metric to detect as many faulty disks as possible. In the second Isolation Forest Model *IF_FAR*, the false alarm rate and precision are the metrics used to minimize the number of false positive instances.

The implementation of Isolation Forest in scikit-learn provides several hyperparameters that have a decisive influence on the prediction results. The most essential hyperparameter is the *contamination* value. This value is critical for mapping between anomalies and normal instances, as it defines the relationship between abnormal and normal instances for the dataset.

In figure 1 a grid search is performed for the parameter *contamination* and evaluated with the metrics FDR and FAR. To achieve a FDR of 100%, the value

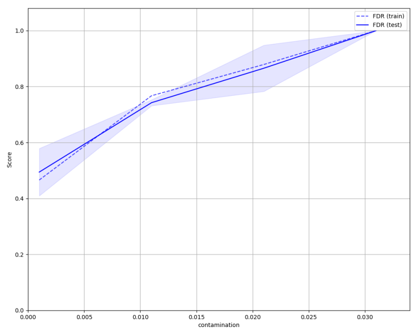

Fig. 1. GridSearch CV for contamination FDR

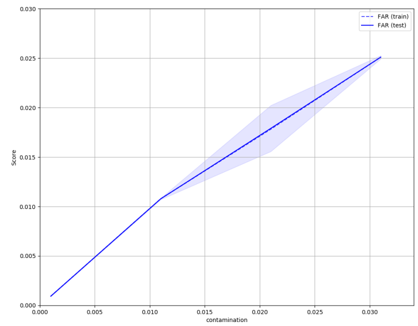

Fig. 2. GridSearch CV for contamination FAR

for *contamination* must be 0.031, which is 3.1% of the complete dataset. However, the FAR value must also be considered, which is 3%. Since the FAR value is calculated from the ratio between the true negative and false positive instances, the number of false positive instances is too high. A detailed representation of the FAR value is illustrated in figure 2. To achieve a tradeoff between FDR and FAR, the value for *contamination* is defined as 0.01 for the model *IF_FDR*.

For the model *IF_FAR*, the main focus is on the false alarm rate and precision, so that the model only predicts hard disk failures if they have a high probability of failure. For the model *IF_FAR* the value 0.0002 is set as suitable for *contamination*.

Once the parameters for the models are set, the models can be created, and the prediction for the dataset can be made. This process is described below:

1. The cleaned-up dataset *ST1* is loaded, it contains the date of the sample and the serial number as indices, as well as the six selected features. In addition, the dataset *STF* with the feature *failure* is loaded for later evaluation of the prediction.
2. The period for which the predictions are to be made is specified.
   (a) The dataset *ST1* is reduced to the samples of one day.
   (b) The Isolation Forest model is created with the parameters, and the prediction for the samples is performed.
   (c) The anomaly score is attached to a list together with the serial number and date.
3. Steps a), b), and c) are repeated for each day in the specified period.
4. The anomaly score and the *contamination* parameter are used to classify each sample in the list as an anomaly or normal instance.
5. The predicted values are compared with the true values from *STF*. If the failure date of a disk is within one week after prediction, it is considered true positive.
6. Based on the analysis of the data, the confusion matrix is built, and the values for FDR, FAR, and precision are calculated.

The different values of *contamination* and the calculated value of precision in both models can be used to define the failure probability of the hard disk. The failure probability is stored in a structured text file together with the prediction date and the serial number.

4 Experimental Results

In this section, the results of the model are presented and evaluated with different metrics. Furthermore, a comparison with other common models is carried out. Finally, gained insights and possible limitations of the dataset are discussed.

Figure 3 shows the results for the Isolation Forest model *IF_FDR*. In this model, the focus is on detecting the largest possible number of faulty disks with the lowest possible FAR. The FDR is 84.54%, with a FAR of 0.0073%. Also, the precision is 44.21%, reflecting the probability of failure of the hard drive over seven days.

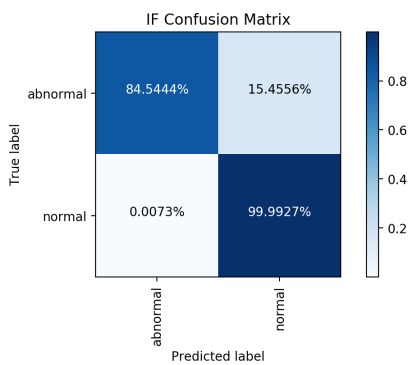

Fig. 3. FDR confusion matrix

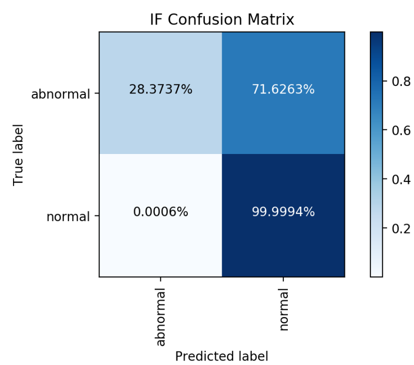

Fig. 4. Precision confusion matrix

In contrast to the *IF_FDR* model, the *IF_FAR* model was developed with the focus on keeping the number of false positive instances as low as possible. The results are shown in figure 4. All instances detected as faulty have a probability of failure of 77.85%, which reflects the precision. Due to the high value of the failure probability, the FAR is also significantly lower, with a value of 0.0006%. The FDR is 28.37% and thus far below the value of the *IF_FDR* model.

4.1 Comparative Analysis

Table 2 shows a comparison between different anomaly detection algorithms. On the dataset *ST1*, the predictions were additionally performed for the algorithms One-class SVM and Local Outlier Factor (LOF). One-class SVM and LOF were selected because they are among the most widely used algorithms in anomaly detection and provide good results for imbalanced datasets [2]. The FDR is higher for both algorithms than for the Isolation Forest, but the FAR is also markedly higher, which harms precision. Additionally, the computation time, calculated on the prediction of one day, is a multiple of the Isolation Forest. Precision is an essential factor for the evaluation of the results, as the probability of failure in the replacement process serves as the underlying principle. Therefore, the Isolation Forest, out of these three algorithms investigated, is best suited for prediction.

Table 2. Comparison of Isolation Forest, One-class SVM and LOF

	FDR	**FAR**	**Precision**	**Computation Time**
IF_FDR	84.54%	0.0073%	44.21%	1.10 s ± 10 ms
IF_FAR	28.37%	0.0006%	77.85%	1.10 s ± 10 ms
One-class SVM	95.39%	0.2925%	2.19%	15.9 s ± 780 ms
LOF	96.19%	1.1916%	0.55%	16.0 s ± 671 ms

By combining the two models, a categorization of the results can be carried out. For this purpose, the predictions are made, and the anomalies are saved in the form of a serial number and the date. Besides that, each instance is provided with an additional label. For all instances from the model *IF_FDR* the hard disk receives the label *Warning* and *Failure* for the instances from the *IF_FAR* Model. The *Failure* label is weighted higher and replaces an existing label on an instance.

The replacement program of the marked hard disks depends on the number of data storage systems as well as the Redundant Array of Independent Disks (RAID) technology used, since the reliability of a RAID array changes depending on the RAID level.

The goal is to distribute the hard disks with the label *Warning* over several RAID arrays in such a way that in each array, a maximum of one marked hard disk operates. This means that there are no additional costs for the provision of a new hard disk, and the reliability of the array is not endangered. If it is not possible to replace the marked hard disks in the array, they must be replaced with a new hard disk. The marked hard disks from model *IF_FAR* with the label *Failure* are to be regarded as critical due to their high failure probability and should, therefore, be replaced with a new hard disk.

4.2 Limitations

It was discovered that the features *smart_5_raw* and *smart_187_raw* are in a value range of $[0, 65535]$, which is exactly 2 bytes in size. Thus it can be concluded that a buffer overflow occurs with these two features, and then the values are reset to 0. A buffer overflow can lead to a degradation of the prediction, but was taken into account in the created models and compensated by the additional features *smart_5_raw_diff* and, *smart_187_raw_diff*.

The logging interval of one day is not optimal because there are disks whose values deteriorate significantly within one day and fail the same day. Hourly logging would make it easier to identify these disks, but it would also increase the dataset's storage requirements considerably. For 10% of the disks detected as faulty, only one sample indicates the possible failure. Hourly logging could increase the number of samples that indicate a failure.

5 Conclusion

In this work, it has been shown that the prediction of faulty disks by techniques of anomaly detection achieves good results, particularly for models that were created based on the Isolation Forest. The decision for the algorithm was based on the highly imbalanced dataset, the comprehensible behavior, and low linear time complexity of the models. Many machine learning methods cannot handle imbalanced datasets without preprocessing as they tend to overestimate the majority class. The Isolation Forest does not require preprocessing for an imbalanced dataset, so there is no need to downsample or upsample the instances. Furthermore, the training phase is omitted with the unsupervised approach, which significantly reduces the time between data acquisition and prediction.

Two models were created with a different focus. The model *IF_FDR* concentrates on a high failure detection rate and a low false alarm rate. This model achieved a FDR of 84.54% with a FAR of 0.0073%. Furthermore, the failure probability of the predicted failing disks is 44.21%. The second model *IF_FAR* focused on providing a prediction to find the faulty disks that were very likely to fail within the next seven days. A failure probability of 77.85% was achieved for the hard disks marked as faulty. The model *IF_FAR* determines which hard disks require urgent action by infrastructure administration.

Several aspects can improve the accuracy of the models. For applications in data centers, it would make sense to include the location of the hard disks and the corresponding array in the dataset, so that the decision whether a hard disk should be replaced does not have to be performed manually, and can be calculated by the model. It was also shown that the prediction is limited to the SMART attributes alone since 25% to 40% of all hard disks fail without any variance in the data. To improve the prediction, a combination of system log files and SMART attributes could be used to perform a hard disk assessment in the future.

References

1. Backblaze: Hard drive stats (2013-2019), https://www.backblaze.com/b2/hard-drive-test-data.html (visited on 04/16/2020)
2. Goldstein, M.B.: Anomaly detection in large datasets. Verlag Dr. Hut (2014)
3. Hamerly, G., Elkan, C., et al.: Bayesian approaches to failure prediction for disk drives. In: ICML. vol. 1, pp. 202–209 (2001)
4. Hughes, G.F., Murray, J.F., Kreutz-Delgado, K., Elkan, C.: Improved disk-drive failure warnings. IEEE transactions on reliability **51**(3), 350–357 (2002)

5. Li, J., Ji, X., Jia, Y., Zhu, B., Wang, G., Li, Z., Liu, X.: Hard drive failure prediction using classification and regression trees. In: 2014 44th Annual IEEE/IFIP International Conference on Dependable Systems and Networks. pp. 383–394. IEEE (2014)
6. Murray, J.F., Hughes, G.F., Kreutz-Delgado, K.: Hard drive failure prediction using non-parametric statistical methods. In: Proceedings of ICANN/ICONIP (2003)
7. Murray, J.F., Hughes, G.F., Kreutz-Delgado, K.: Machine learning methods for predicting failures in hard drives: A multiple-instance application. Journal of Machine Learning Research **6**(May), 783–816 (2005)
8. Pinheiro, E., Weber, W.D., Barroso, L.A.: Failure trends in a large disk drive population. In: 5th USENIX Conference on File and Storage Technologies (FAST 07). USENIX Association, San Jose, CA (Feb 2007)
9. Saffari, A., Leistner, C., Santner, J., Godec, M., Bischof, H.: On-line random forests. In: 2009 ieee 12th international conference on computer vision workshops, iccv workshops. pp. 1393–1400. IEEE (2009)
10. Schroeder, B., Gibson, G.A.: Understanding disk failure rates: What does an mttf of 1,000,000 hours mean to you? ACM Transactions on Storage (TOS) **3**(3), 8–es (2007)
11. Seagate: Get s.m.a.r.t. for reliability. technical report. Seagate Technology Paper (1999)
12. Shen, J., Wan, J., Lim, S.J., Yu, L.: Random-forest-based failure prediction for hard disk drives. International Journal of Distributed Sensor Networks **14**(11) (2018)
13. Wang, Y., Ma, E.W., Chow, T.W., Tsui, K.L.: A two-step parametric method for failure prediction in hard disk drives. IEEE Transactions on industrial informatics **10**(1), 419–430 (2013)
14. Wang, Y., Miao, Q., Ma, E.W., Tsui, K.L., Pecht, M.G.: Online anomaly detection for hard disk drives based on mahalanobis distance. IEEE Transactions on Reliability **62**(1), 136–145 (2013)
15. Wang, Y., Miao, Q., Pecht, M.: Health monitoring of hard disk drive based on mahalanobis distance. In: 2011 Prognostics and System Health Managment Confernece. pp. 1–8. IEEE (2011)
16. Xiao, J., Xiong, Z., Wu, S., Yi, Y., Jin, H., Hu, K.: Disk failure prediction in data centers via online learning. In: Proceedings of the 47th International Conference on Parallel Processing. pp. 1–10 (2018)
17. Xu, C., Wang, G., Liu, X., Guo, D., Liu, T.Y.: Health status assessment and failure prediction for hard drives with recurrent neural networks. IEEE Transactions on Computers **65**(11), 3502–3508 (2016)
18. Zhang, T., Wang, E., Zhang, D.: Predicting failures in hard drivers based on isolation forest algorithm using sliding window. Journal of Physics: Conference Series **1187**(4) (2019)
19. Zhu, B., Wang, G., Liu, X., Hu, D., Lin, S., Ma, J.: Proactive drive failure prediction for large scale storage systems. In: 2013 IEEE 29th symposium on mass storage systems and technologies (MSST). pp. 1–5. IEEE (2013)
20. Züfle, M., Krupitzer, C., Erhard, F., Grohmann, J., Kounev, S.: To fail or not to fail: Predicting hard disk drive failure time windows. In: International Conference on Measurement, Modelling and Evaluation of Computing Systems. pp. 19–36. Springer (2020)