

Predicting Disk Replacement towards Reliable Data Centers

Mirela Botezatu
IBM Research
Zurich Switzerland
bot@zurich.ibm.com

Jasmina Bogojeska
IBM Research
Zurich Switzerland
jbo@zurich.ibm.com

Ioana Giurgiu
IBM Research
Zurich Switzerland
igi@zurich.ibm.com

Dorothea Wiesmann
IBM Research
Zurich Switzerland
dor@zurich.ibm.com

ABSTRACT

Disks are among the most frequently failing components in today's IT environments. Despite a set of defense mechanisms such as RAID, the availability and reliability of the system are still often impacted severely.

In this paper, we present a highly accurate SMART-based analysis pipeline that can correctly predict the necessity of a disk replacement even 10-15 days in advance. Our method has been built and evaluated on more than 30000 disks from two major manufacturers, monitored over 17 months. Our approach employs statistical techniques to automatically detect which SMART parameters correlate with disk replacement and uses them to predict the replacement of a disk with even 98% accuracy.

Keywords

Disk replacement; time series; classification; changepoint

1. INTRODUCTION

Data center downtime costs have increased significantly in the past years from \$5,600/minute in 2010 to \$8,851/minute in 2016 according to a study conducted on 63 data center organizations in the U.S [1]. IT equipment failure is a significant contributor to such downtimes. Disks are among the most frequently failing components in today's IT environments. It appears that field behavior of disks is fairly different than the one described in the datasheet specifications [18]. Factors such as temperature, duty cycles or workloads may significantly affect both the reliability and the performance of hard drives. Reliability issues are by far the most severe and manifest themselves as disk failures leading to replacements.

Disk failures can be either predictable or unpredictable. On the one hand, unpredictable failures, ranging from elec-

tronic components becoming defective to sudden crashes due to improper handling, cannot be foreseen by monitoring. On the other hand, predictable failures mainly result from slow processes such as wear-and-tear that typically progress over months or years. The latter ones make it possible for predictive failure analysis.

In this paper, we introduce a novel data mining approach able to automatically predict disk replacements based on historic disk replacement data from an expert-maintained disk environment and hence minimize the effects of component failure as shown in Figure 1.

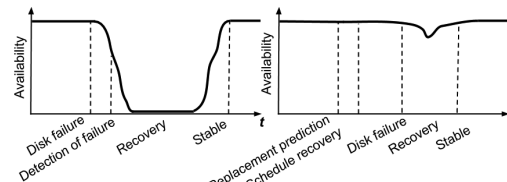


Figure 1: Availability: without proactive replacement (left) vs. with proactive replacement(right)

SMART monitoring (disk sensors' data) can be used to determine when disk failures become more likely. Some manufacturers even use them to deploy drives with embedded predictive models. However, these models are proprietary and often times, simple, threshold-based normalizations, that are designed to avoid false alarms and therefore have a very weak predictive power [18, 17, 12].

In this paper, we focus on the automatic forecasting of predictable disk replacements using SMART attributes. For this purpose, we use data collected from a large population of disks (>30000) monitored over 17 months. A drive is labeled as failed when it stopped working, it is non-responsive to commands, the RAID system reports that the drive cannot be written or read, or it shows evidence of failing soon [2]. Therefore, the model goes beyond the expert knowledge used in proactive replacements and is able to detect failures that this knowledge can not capture (see Section 3.6).

The goals of our analysis are two-fold: (1) to provide the set of SMART attributes that are informative for disk replacements; (2) to use these attributes to build a statistical model that automatically predicts impending replacements with high accuracy (81-98%). Such a model not only automates the disk replacement decision, but also allows ad-

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

KDD '16, August 13 - 17, 2016, San Francisco, CA, USA

© 2016 Copyright held by the owner/author(s). Publication rights licensed to ACM. ISBN 978-1-4503-4232-2/16/08...\$15.00

DOI: <http://dx.doi.org/10.1145/2939672.2939699>

ministrators to proactively replace disks at risk, days in advance.

To achieve these goals, we employ an approach that comprises four steps. First, we use changepoint detection in time series to identify the SMART attributes indicative of impending replacements. Second, we transform the event sequence into a set of examples [20] by encoding multiple events as individual points such that we achieve a compact, yet informative representation of the time series of each disk. Next, we build a predictive classification model that is able to discriminate between healthy and failure-impending drives, by using these data points as inputs. Finally, we propose a transfer learning approach to enable replacement decision prediction on data from novel disk models.

There are several challenges one may encounter when performing the aforementioned steps. In the following, we list a few of them. Since SMART indicators are manufacturer-specific, their encoding and normalization varies widely across manufacturers. This hinders the possibility to fit one predictive model for all different disk manufacturers. A separate model needs to be trained for each individual disk manufacturer. Further, due to the lack of standards when implementing SMART attributes, one needs to discover the ones that are indicative of failures. Finally, the disk data are highly unbalanced (only about 2% of disks are replaced), which makes the task of fitting high quality models very challenging.

Therefore, we build and evaluate our approach for disks from two individual manufacturers. Following our efforts to choose the right SMART indicators and tuning the predictive model, results show up to 98% accuracy in identifying both disks that are about to be replaced and those that are healthy, when using only a small set of SMART parameters.

The remainder of this paper is organized as follows. In Section 2, we describe the predictive pipeline, whereas in Section 3 we present experimental results. We discuss deployment in Section 4 and finally review the state-of-the-art in Section 5 and conclude in Section 6.

2. PREDICTING DISK REPLACEMENT

Given the longitudinal measurements of the SMART attributes for a large set of disks, from a specific disk model of interest and information on their replacements, we develop a fully automated approach for solving the disk replacement prediction problem. Our method is summarized in Algorithm 1 and consists of four consecutive steps: (1) selection of relevant SMART attributes, (2) compact time-series representation, (3) balancing of the healthy and unhealthy disk classes via informed downsampling, and (4) classification model for disk replacements. In the following, we present the details of each step.

2.1 Selection of relevant SMART attributes

The main goal of this step is to automatically discover the set of SMART attributes that are indicative of impending disk replacements. This will reveal the most informative predictors with respect to the disks at risk to the domain experts. As SMART attribute data are gathered over time, we address this feature selection problem through changepoint detection in time series. More specifically, when a SMART attribute is informative of disk replacement, we expect a significant shift in its values at some time point before the

Algorithm 1 Disk replacement prediction algorithm

Input: A time series collection of SMART attributes along with the disk replacement information for a given target disk type.

1. Find the subset of SMART attributes indicative of disk replacements by identifying significant changepoints in their corresponding time series;
2. Compute a highly-informative compact representation for the time series corresponding to each relevant attribute from Step 1 via exponential smoothing;
3. Perform informative downsampling via K-means clustering to address the high class imbalance in the disk replacement datasets;
4. Use the training dataset from Step 3 to fit a classification model that predicts disk replacements.

Output: Predictive model for disk replacement using a small set of SMART attributes.

actual replacement, i.e., at the changepoint. Moreover, this shift should be permanent and unrecoverable to be indicative of a disk replacement. In the following we provide a more formal description of the approach for detecting the permanent changepoints for SMART attributes.

Let $S_i = (s_1, s_2, \dots, s_p)$ denote the time series for a target SMART attribute comprising p measurements ordered by their timestamps, where s_p is the most recent one when the disk replacement has occurred. If there exists a timestamp $t < p$ when a significant change in the values of the attribute S_i occurs (e.g., the values start increasing), then we consider S_i a potential attribute relevant for the disk replacement. We determine the time point t that indicates a significant change using the approach described in [7]. Briefly, $t = \operatorname{argmax}_\tau ML(\tau)$ provided that $ML(t)$ is significantly larger than $\log p(s_{1:p} | \hat{\theta})$, where:

$$ML(\tau) = \log(p(s_{1:\tau} | \hat{\theta}_1)) + \log(p(s_{\tau+1:p} | \hat{\theta}_2)). \quad (1)$$

Next, we verify whether the change is permanent by checking whether the difference between the time series of the potential SMART attribute and the corresponding time series of the same attribute in the absence of the observed change at time point t is significant. We do this as follows. First, let the time series $\Gamma_t = (s_t, \dots, s_p)$ denote the subsequent values recorded for the potential SMART indicator S_i starting from the timestamp t to the time of the replacement p . Then, we generate a synthetic time series for the same indicator denoted $\Psi = (\tilde{s}_{t+1}, \dots, \tilde{s}_p)$ that has no significant change at time point t . More specifically, we compute the posterior distribution of Ψ , $p(\tilde{s}_{(t+1):p} | s_{1:(t)}, x_{1:p})$ given the value of the series in the pre-change period $s_{1:t}$ along with the values of the control time series $x_{1:p}$ using a Bayesian structural time-series model. The control time series is a sample of the values of the target SMART attribute collected for a healthy disk. Finally, the target SMART attribute is indicative of a disk replacement if the probability distributions of the actual time series measured after the detected change point and the synthetic one generated based on the values of a healthy disk are significantly different. We assess the difference via hypothesis testing. Formally, let Γ_k and

Ψ be samples generated from unknown distributions P and Q , respectively.

The hypothesis to test is the following:

$$\begin{cases} H_0 : P = Q \\ H_1 : P \neq Q \end{cases} \quad (2)$$

Then we check whether we can reject the null hypothesis H_0 that the two probability distributions P and Q are equal with high confidence.

2.2 Compact time series representation

As the previous step results in the set of relevant SMART indicators for the disk replacement problem, the goal of this step is to provide a compact, but highly informative representation the time series of each indicator that can readily be employed in the predictive model.

There are several observations that hint the necessity of a compact representation for the time series data: (i) Each daily observation on its own is not enough - we need to consider a longer time frame - this is because the single day record is not stable due to the recovery mechanisms embedded in the disk. (ii) Also, if we considered as observations for the failed class only the entries from the last day of the life of the disk then the model will not be able to predict replacement in advance as it can only recognize the instances when drive fails not before.

Therefore, we use a window to split the raw data set into segments. We aggregate each of the relevant time series to a single value using exponential smoothing over a specific time window. This way, we assign the highest weights to the most recent observations and exponentially decreasing weights to the remaining observations as they get older. Intuitively we expect that the observations closer to the time point of the disk replacement are more informative compared to the older ones. Formally:

$$S_t = \alpha \cdot Y_t + (1 - \alpha) \cdot S_{t-1} \quad (3)$$

In the equation above, the smoothed value at time t , S_t is computed recursively based on the observation at time t and the smoothed value at time $t - 1$. When fixing the width of the window to a value k , S_t becomes the weighted average of a certain number of the past observations up to Y_{t-k} . A smaller value of k causes a weaker smoothing effect which enables higher sensitivity to new changes in the data. The parameter α controls the speed at which the older observations are dampened. A large α is used for assigning lower weights to observations from the more distant past.

For each relevant SMART attribute, the width of the time window used in the smoothing process is chosen as *the median of the distribution of the time stamps of their corresponding significant change* computed as described in Section 2.1.

2.3 Class balancing via informative downsampling

The data to be used in the predictive model is highly imbalanced, as only a small percentage of all disks are replaced over time. Since classification algorithms are typically optimized to maximize the overall accuracy, when trained using imbalanced datasets they exhibit poor predictive performance. To address this issue, we balance the training dataset for our predictive model by using a representative

subset of the data for the dense class – in our case the healthy disks. This representative subset is chosen such that it comprises the most informative samples with low or no redundancy. We achieve this by clustering the observations pertaining to the healthy disk set into k clusters using the K-means clustering algorithm [15]. Next, for each cluster, we select the data points closest to the respective cluster centroid as representatives for the healthy disk class. Finally, we generate a balanced training dataset by choosing k close to the number of samples available for the replaced disks.

2.4 Classification for disk replacements

In the final step of our approach we fit a model that utilizes the training dataset generated in the previous step and provides high quality disk replacement predictions for new, unseen data. Formally, let $D = \{(\mathbf{x}_i, y_i)\}_{i=1}^n$ denote the training dataset, where $\mathbf{x}_i \in X$ is a multivariate temporal observation aggregating information between time points t_{i-k} and t_i for the set of relevant SMART attributes, and y is a binary response variable ($y \in \{0, 1\}$). We want to learn a function $h: X \rightarrow \{0, 1\}$ that minimizes the loss $\ell(h(\mathbf{x}); y)$ that quantifies the prediction quality. Intuitively the goal is to train a model that correctly predicts whether a disk needs replacement ($y = 1$) or not ($y = 0$).

We tackle this problem using the regularized greedy forests (RGF) [14] approach that is a powerful, non-linear classification method. We show that, for this task, it delivers better quality predictions compared to other tree ensemble based methods such as gradient boosted decision trees (GBDT) [21] or random forests [6] and also outperforms other classification methods such as SVM [8], or logistic regression [9].

The RGF algorithm is a variation of GBDT in which the structure search and the optimization are decoupled. More specifically, the main differences are given as follows:

- RGF introduces an explicit regularization term that takes advantage of individual tree structures.

$$\hat{h} = \operatorname{argmin}_{h \in H} [\ell(h(\mathbf{x}); y) + R(h)] \quad (4)$$

- RGF employs a *fully-corrective* greedy algorithm which iteratively modifies the weights of all the leaf nodes (decision rules) currently obtained while new rules are added into the forest by greedy search. Here, an explicit regularization is also included to avoid overfitting and very large models.
- RGF utilizes the concept of structured sparsity to perform greedy search directly over the forest nodes based on the forest structure.

The general framework of RGF is given in Algorithm 2 which we describe in the following.

F represents a forest, and each node v of F is associated with the pair (b_v, a_v) , where b_v denotes the basis function of node v and a_v the weight assigned to this node. The model of F is given by $h_{F(x)} = \sum_{v \in F} a_v b_v(\mathbf{x})$ with $a_v = 0$ for any internal node v .

In this setting, the regularized loss specified in Eq. 4 is a function of F : $Q(F) = \ell(h_F(\mathbf{x}), y) + R(h_F)$. Further, $S(F)$ represents the set of all structure-changing operations applicable to F (i.e. the split of a node or the addition of a new tree).

Algorithm 2 Regularized Greedy Forest framework

```

 $F \leftarrow \{\}$ 
while stopping criterion not met do
  Fix weights and adjust forest structure  $s$ :
   $\hat{s} \leftarrow \operatorname{argmin}_{s \in S(F)} Q(s(F))$  (the optimum  $s$  that
  minimizes  $Q(F)$  among all the structures that can be
  obtained by applying one structure-changing operation
  to  $F$ ).
  if some criterion is met then
    Fix the structure and change the weights in  $F$  s.t.
    the loss is minimized in  $Q(F)$  (it can be optimized using
    a standard procedure (such as coordinate descent) if the
    regularization penalty is standard e.g.,  $L2$ -loss
  end if
end while
Optimize leaf weights in  $F$  to minimize loss in  $Q(F)$ 
return  $h_F(\mathbf{x})$ 

```

2.5 Transfer learning

As illustrated in Figure 5 in Section 3.5, the data collected from different disk models are different. We observe the fact that different models of a single disk manufacturer have similar SMART reporting but different distributions of the values reported for the SMART attributes. Therefore, utilizing an existing predictive model created on the training data of a specific disk model will not deliver the optimal predictive performance when directly applied on the data collected from a different disk model from the same manufacturer. In data mining, this problem is referred to as *sample selection bias*, *covariate shift* or *dataset shift*. Therefore, we apply a transfer learning approach in order to be able to use a prediction model trained on specific disk model for a new disk model of the same manufacturer.

Note that such an approach is valuable as it transfers the expert knowledge gathered over the years through historic data to a new disk model from a given manufacturer of interest. We tackle the described dataset shift issue we have across disk models of a given manufacturer as follows. We leverage the unlabeled data for the target (new) disk model to conduct a sample selection de-biasing, as described in Algorithm 3. The idea behind the algorithm is to train a classifier that can rank the observations linked to a specific disk model based on their similarity to observations pertaining to the target disk model. Furthermore, this enables to sample the observations from the original disk model (which are already labeled) that are more representative for learning the class labels for the target disk model, i.e. that matches the distribution of the original disk model to the target disk model. Learning a predictive model using a training sample that reflects the distribution of the new disk model results in higher quality predictions.

3. EVALUATION

In the following, we present our experimental setup and the results obtained in each step of our approach.

3.1 Data description and experimental setup

Our analysis is based on the Backblaze dataset¹. The set contains data collected from 50984 hard disks, moni-

¹<https://www.backblaze.com/hard-drive-test-data.html>

Algorithm 3 Transfer learning for different models

Input: $D_{DM_1} = \{x_i, y_i\}_i^n$, the labeled data collected from disk model 1, and $D_{DM_2} = \{x'_i, y'_i\}_i^m$ the unlabeled data from disk model 2.

1. Let $D_{DM_1} = \{x_i, y_i\}_i^n$ be the labeled data collected from disk model 1, and $D_{DM_2} = \{x'_i, y'_i\}_i^m$ be the unlabeled data from disk model 2.
2. Let $D_{aug} = \{x_i, "DM_1"\}_i^n \cup \{x'_i, "DM_2"\}_i^m$
3. Use D_{aug} to learn a function $f : X \rightarrow [0, 1]$, such that $f(x)$ represents the probability of a disk being of type "DM₁" or "DM₂".
4. Sample a subset D_{sub} from D_{DM_1} according to f .
5. Use D_{sub} to learn a function $g : X \rightarrow [0, 1]$ (call the procedure in Algorithm 2) such that $g(x)$ represents the probability of a disk of type DM_2 needing replacement.

Output: Predictive model for disk replacement for disk model 2.

tored over 27 months (April 2013 to June 2015) with daily granularity. The data collected contains the following: (1) timestamp, (2) disk serial number, (3) disk model, (4) disk capacity, (5) failure - '0' if the drive is alive and '1' if the disk has been replaced the following day, and (6) SMART statistics. From the disk models, we extract the manufacturers and we restrict our analysis to Hitachi² and Seagate¹ due to the fact that for the other manufacturers there are only few samples in the dataset, or poor population of the SMART parameters. We also exclude all monitoring data between April 2013 and January 2014, as more than 70% of the SMART parameters are not collected. Thus the dataset we consider is gathered over 17 months.

First, we build and evaluate the predictive model described in this paper for Seagate ST4000DM000 (SgtA) and Hitachi HDS722020ALA330 (HitA). Then, we evaluate the transfer learning approach on Seagate ST31500541AS (SgtB) and Hitachi HDS5C3030ALA630 (HitB), respectively. Further details on the data are presented in Table 1.

	Original		Post-aggregation	
	H	R	H	R
SgtA	247524	543	17769	457
SgtB	30859	375	2188	227
HitA	75618	150	4616	115
HitB	74040	80	4662	73

Table 1: Healthy (H) vs. replaced (R) disks in the raw dataset and after data cleaning and aggregation for Hitachi and Seagate.

3.2 Selection of relevant SMART attributes

First, note that for each SMART indicator there are two values recorded – the raw value, and the normalized value. The raw value often represents counts or a physical unit (e.g., degrees Celcius or milliseconds). The normalized values are a very specific mapping of the raw values such that, typically, higher values indicate healthy disks with some exceptions (e.g., the *temperature* attribute for Seagate models). A detailed breakdown of the SMART parameters is found in [4].

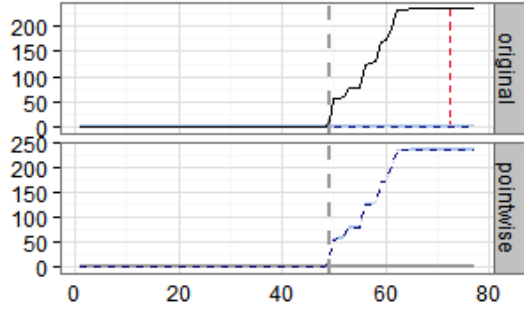


Figure 2: Differences between the forecasted and the observed values for SMART_187_raw.

As presented in Section 2, we find SMART indicators relevant for disk replacements via changepoint analysis.

In Figure 2, we illustrate the evolution of the time series for the parameter SMART_187_raw (reported uncorrectable errors) over 80 days for SgtA disk. Note that after 50 days of usage the disk starts to accumulate uncorrectable errors, up to the point where a replacement is necessary. Since there is a significant difference between the time series observed on days 1 to 50 and the one observed on days 50 to 80, our algorithm detects a changepoint 30 days before the disk has been replaced.

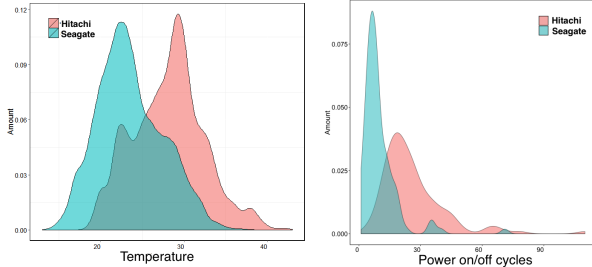


Figure 3: Distribution of the temperature and of the power on off cycles across the replaced disks for Hitachi and Seagate.

We perform the changepoint analysis for both Seagate and Hitachi disks and present the results in Table 2. For each of the considered SMART parameters we report the percentage of drives for which a correlation with disk is observed.

For Seagate, 63% of the replaced drives correlate with an increase in SMART_193_raw (the load cycle count), and between 19 and 26% of them also correlate with SMART_7_raw (seek error count), SMART_1_normalized (read error rate), SMART_240_raw (transfer error rate), SMART_197_raw (nr. of pending sectors), SMART_198_raw (uncorrectable sector count), SMART_187_raw (number of uncorrectable errors), as well as SMART_5_raw (reallocated sector count).

For Hitachi, only some of these SMART parameters are indicative of drive replacements. Among the top correlated indicators for Hitachi (30-47%), we note SMART_196_raw (reallocation event count), SMART_194_normalized (internal temperature), SMART_5_raw and SMART_197_raw.

We also note that it's mostly the raw values of SMART indicators that correlate with impending replacements. This is expected, since, the normalized values are computed based on generous thresholds, where a replacement can also occur before the normalized value changes at all.

The changepoint analysis also shows that some SMART indicators correlate stronger with the replacements of the

	SgtA		HitA	
	Ratio	Inp.	Ratio	Inp.
SMART_1_norm	23%	✓	28%	✓
SMART_1_raw	2%	✓	15%	✓
SMART_3_norm	—	×	13%	✓
SMART_3_raw	—	×	15%	✓
SMART_5_norm	2%	✓	22%	✓
SMART_5_raw	19%	✓	31%	✓
SMART_7_norm	14%	✓	—	×
SMART_7_raw	26%	✓	—	×
SMART_183_norm	0.5%	×	—	×
SMART_183_raw	0.5%	×	—	×
SMART_184_norm	1%	✓	—	×
SMART_184_raw	1%	✓	—	×
SMART_187_norm	21%	✓	—	×
SMART_187_raw	21%	✓	—	×
SMART_188_norm	0%	×	—	×
SMART_188_raw	10%	✓	—	×
SMART_189_norm	1%	✓	—	×
SMART_189_raw	1%	✓	—	×
SMART_190_norm	2%	✓	—	×
SMART_190_raw	2%	✓	—	×
SMART_193_norm	10%	✓	—	×
SMART_193_raw	63%	✓	—	×
SMART_194_norm	2%	✓	31%	✓
SMART_194_raw	2%	✓	2%	✓
SMART_196_norm	—	×	20%	✓
SMART_196_raw	—	×	26%	✓
SMART_197_norm	5%	✓	4%	✓
SMART_197_raw	27%	✓	22%	✓
SMART_198_norm	6%	✓	—	×
SMART_198_raw	27%	✓	—	×
SMART_199_norm	0%	×	—	×
SMART_199_raw	0.5%	×	—	×
SMART_240_norm	0.5%	×	—	×
SMART_240_raw	21%	✓	—	×
SMART_241_norm	0%	—	—	×
SMART_241_raw	15%	✓	—	×
SMART_242_norm	0%	×	—	×
SMART_242_raw	19%	✓	—	×

Table 2: SMART correlation frequencies for SgtA and HitA. A ✓ indicates the predictor is included in the classification task.

Seagate model than with those of the Hitachi model, and vice versa. We discuss these differences in the context of SMART_194 (the disk internal temperature). Relative to temperature, 31% of the Hitachi replaced disks correlate, compared to only 2% for Seagate. We attribute this to the overall higher temperatures that characterize the Hitachi disks, as shown through the comparative plots in Fig. 3. Although the distributions are similar, there is a clear shift towards higher temperature for Hitachi, by 5 to 10 degrees Celsius.

3.3 Compact time series representations

Fig. 4 shows the distribution of the number of days before replacement when the changepoint was observed for six SMART indicators (read error rate, the number of reallocated sectors, the number of pending sectors, the reported uncorrectable errors, the seek error count and the transfer error rate). Note how the median values are different from one predictor to another. We use these median values to select the length of the window of the time series when creating the compact representation.

We notice that on the one hand, for the number of real-

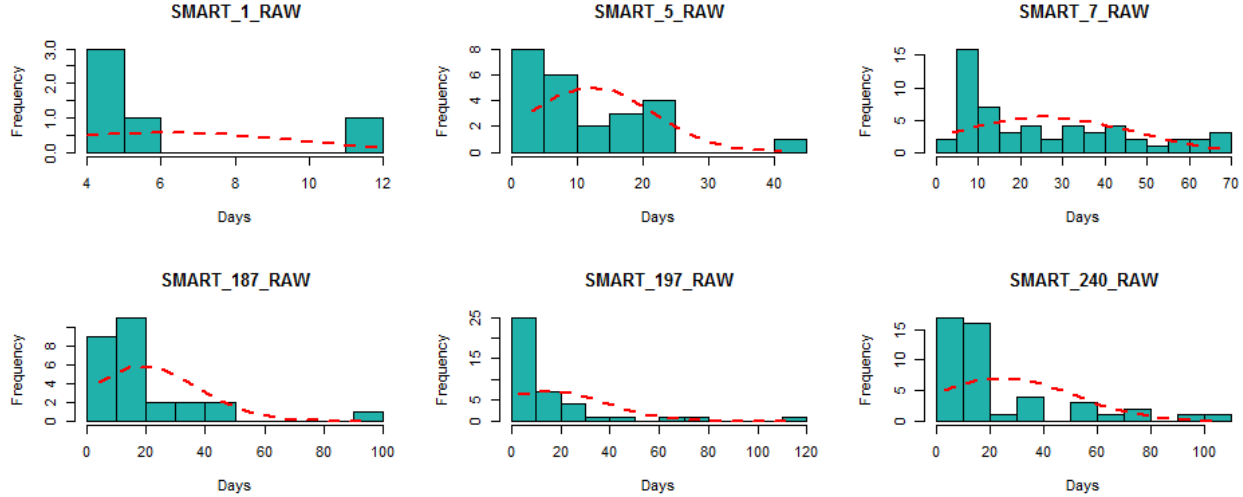


Figure 4: Distribution of the number of days before replacement when the changepoint was observed.

located sectors and of pending sectors (SMART_5_raw and SMART_197_raw), considering the last 12 and 10 days, respectively, before the disk replacement in the predictive model is sufficient. This is because an increase in either of these two parameters indicates that a remapping operation is necessary (i.e., data from the defective sector is transferred to a spare area). As shown in [17], a drive which has had any re-allocated or pending sectors at all is significantly more likely to fail in the nearest future. An even stronger indicator for replacement is the read error rate (SMART_1_raw), which represents the rate of hardware read errors while reading data from the disk surface. It indicates a critical problem with the read/write heads (e.g., head resonance or contamination, broken head, etc.) and in most cases an imminent failure. From our dataset, we find that for this predictor, looking back only 4 days in the past is sufficient for the predictive model. Similarly, a drive with uncorrectable errors as indicated by SMART_187_raw (i.e., cannot be recovered using hardware ECC) may need to be replaced about 15 days after the event.

On the other hand, for the seek and transfer error rates (SMART_7_raw and SMART_240_raw), the algorithm considers the past 25 days in the aggregation process. Both are indicative of malfunctions of the magnetic heads, but hinder performance primarily and lead to failures only in a second phase. For instance, seek errors hint at the drive overshooting or undershooting the correct track when it moves the heads. This implies it will need to perform another seek to acquire the track before it can read or write data.

3.4 Classification for disk replacements

Since only 2.5 to 3% of the disks for both SgtA and HitA models are replaced, the classifier will be biased towards the healthy drives. Thus, we downsample the healthy class to an amount that is close to the size of the replaced class. We choose to downsample to 1000 for SgtA and to 500 for HitA. These values are chosen based on the error estimate of the Regularized Greedy Forest (RGF) classifier. Consequently, we run K-means with 100 and 50 clusters as inputs and subsequently for each cluster we select the top 10 data points closest to the centroid of each cluster.

The SMART attributes used to build the predictive model correspond to the rows that have non-null entries and values higher than 1% in Table 2. In essence, for the Seagate model we use 26 SMART predictors and for the Hitachi model only 12. This discrepancy in the amount of predictors we feed to the Seagate model versus the Hitachi one will be reflected in the difference in performance of the classifiers.

To evaluate the classifier’s performance, we measure precision, recall and F-score as defined below, for both replaced and healthy classes. Precision is used to measure the ability of the classifier to correctly identify disks at risk. Recall measures the classifier’s sensitivity, i.e. the ability of the classifier to capture all replaced disks. A higher recall is equivalent to minimizing the number of false negatives (i.e., the number of disks labeled as healthy when they were actually replaced). The F-score is the combined score between precision and recall, or the weighted harmonic mean.

$$P = \frac{tp}{tp + fp} \quad R = \frac{tp}{tp + fn} \quad F\text{-score} = \frac{2PR}{P + R}$$

We perform a systematic comparison with different classifiers. In order to assess the goodness of each classifier we run the following experiment. We generate 100 random splits of the dataset into training (80%) and test (20%), and for each such split, we train the model on the training set and evaluate it on the test set and compare the performance of RGF with that of other classifiers such as Random Forests (RF), Gradient Boosted Decision Trees (GBDT), Support Vector Machines (SVM), Logistic Regression (LR) and decision trees (DT).

For a fair comparison, we have performed parameter tuning (grid search on parameter space to maximize accuracy) for all the parametric classifiers. For RGF we have obtained the best performance when using the L2 regularizer. There were two L2 regularization parameters to be tuned: one for weight optimization – which was set to 1 and the other for tree learning which was set to 0.005. The model size in terms of the number of leaf nodes in the forest was set to 10000 leaves. The results are given in Table 3.

In case of the replaced disks, the model exhibits better prediction quality for Seagate, where we have 4x more data

		RGF		GBDT		RF		SVM		LR		DT	
		SgtA	HitA	SgtA	HitA	SgtA	HitA	SgtA	HitA	SgtA	HitA	SgtA	HitA
<i>Replaced</i>	P	0.98	0.84	0.97	0.82	0.93	0.82	0.93	0.72	0.73	0.72	0.89	0.74
	R	0.98	0.79	0.96	0.78	0.94	0.76	0.95	0.65	0.81	0.59	0.87	0.61
	F	0.98	0.81	0.96	0.80	0.94	0.79	0.94	0.68	0.77	0.65	0.88	0.67
	Sd	0.01	0.02	0.01	0.04	0.05	0.08	0.02	0.05	0.07	0.1	0.04	0.03
<i>Healthy</i>	P	0.99	0.93	0.98	0.92	0.97	0.92	0.97	0.87	0.89	0.85	0.94	0.86
	R	0.98	0.95	0.98	0.94	0.96	0.93	0.96	0.90	0.85	0.90	0.95	0.91
	F	0.98	0.94	0.98	0.93	0.97	0.92	0.96	0.88	0.87	0.87	0.94	0.88
	Sd	0.01	0.02	0.02	0.03	0.04	0.05	0.02	0.04	0.08	0.05	0.02	0.02

Table 3: Precision, Recall, F-score, Deviation of different classifiers - median on 100 runs , each of which using randomly-drawn training and test data points

points and 2x more non-null SMART indicators, with 98% accuracy and 1-2% error over 100 runs. The precision, recall and F-score for Hitachi are lower by 14-19% and the error is higher – 2%, due to a smaller number of drives in the set and 60% less predictors.

For the healthy class, the model achieves similar performance for Seagate as on the replaced class, with 99% precision and 98% recall and F-score. In the case of Hitachi, the model achieves better performance in discriminating the healthy drives as compared to the faulty ones, by 15% on average. We attribute this boost in accuracy to the fact that healthy disks are easier to identify due to the lower variability in the values of the SMART parameters recorded for them.

3.5 Transfer learning

Figure 5 illustrates the covariate shift for various relevant predictors between different disk models from the same manufacturer. This demonstrates that if we want to reuse the data from one model to build a predictive model for another one we need to employ appropriate transfer learning.

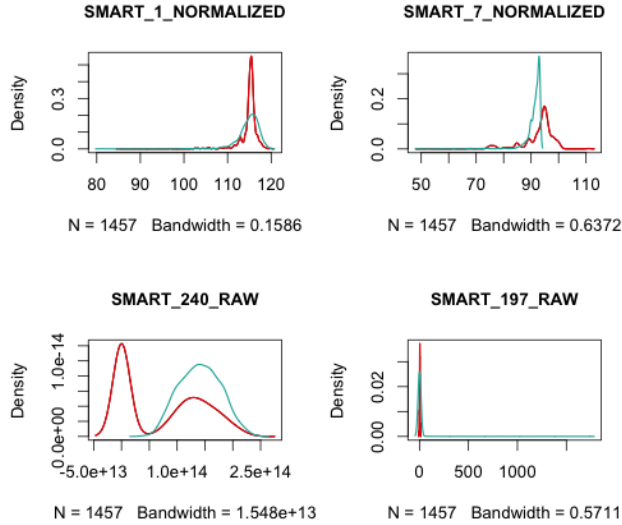


Figure 5: Covariate shift for the two Seagate models

To illustrate the usefulness of our transfer learning approach we compare the models trained and evaluated on SgtA and HitA with the models built with transfer learning and tested on SgtB and HitB, respectively. The results are given in table Table 4.

The gain in predictive performance achieved from using transfer learning when building the new disk models (SgtB

		SgtB		HitB	
		Base	Tr. Learn.	Base	Tr. Learn.
<i>Replaced</i>	P	0.65	0.90	0.53	0.76
	R	0.52	0.82	0.84	0.78
	F	0.58	0.86	0.65	0.77
<i>Healthy</i>	P	0.89	0.96	0.92	0.83
	R	0.93	0.98	0.73	0.82
	F	0.91	0.97	0.81	0.83

Table 4: Precision, recall and F-score to illustrate the importance of transfer learning

and HitB, respectively) compared to directly evaluating the base model (trained on SgtA and HitA, respectively) on the new disk models is shown in Table 4. We obtain 50% increase in the accuracy of the model with transfer learning compared to the accuracy of the base for model SgtA. Also for Hitachi, transfer learning boosts the accuracy of the model by 20%.

3.6 Comparison with human designed replacement policies

A drive is labeled as failed when it stopped working, it is non-responsive to commands, the RAID system reports that the drive cannot be written or read, or it shows evidence of failing soon [2]. Currently, datacenter administrators at Backblaze only focus on a very small set of SMART indicators (5, 187, 188, 197, 198) [2]. However, we illustrate that if one were to do proactive replacement using only this small subset of indicators, the number of disks one could correctly identify drops by almost 50%.

In order to mimic a set of such replacement rules we train a decision tree on the aforementioned subset of SMART indicators. We report the results in Table 5.

	DT on the reduced subset		
		SgtA	HitA
<i>Replaced</i>	Precision	0.95	0.66
	Recall	0.53	0.44
	F-score	0.68	0.51
	Sd	0.06	0.15
<i>Healthy</i>	Precision	0.70	0.84
	Recall	0.98	0.96
	F-score	0.81	0.92
	Sd	0.02	0.12

Table 5: Simple decision tree with (insufficient but commonly used) subset of SMART indicators

Note the differences in recall for SgtA and HitA for our model (98% and 81% respectively) compared to a simple rules based model (53% and 44% respectively) – see Table 5 vs. Table 3. Our solution employs powerful learning methods, leverages a larger set of relevant SMART attributes and hence has numerous advantages: captures a higher amount

of the failure patterns of disks (high recall), it has low false alarm rate, early detection of disks that need to be replaced, and enables transferring the knowledge acquired by expert data center administrators on specific disk models to new disk models from the same manufacturer.

3.7 Early vs. late replacement detection

While one would prefer to use as much as possible from the lifespan of a disk, being able to detect an impending failure early on allows administrators to plan properly for replacements. Therefore, we evaluate how many of the replaced disks our model correctly captures based on snapshots of the SMART indicators taken 1, 3, 10 and 30 days prior to the actual replacement. We expect this amount to be higher when using the snapshots closer to the failure event, since SMART attributes would become more indicative of the impending replacement, thus making the model more accurate. Figure 6

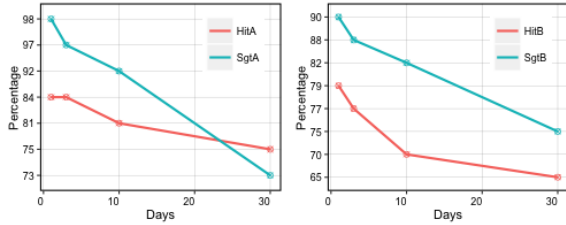


Figure 6: Percentage of disks correctly predicted as replaced on snapshots taken 1,3,10 and 30 days before the actual replacement event.

shows the results obtained for both Seagate and Hitachi. On the one hand, for SgtA, the model correctly identifies 97% of the faulty disks 3 days prior, 92% of them 10 days in advance and, then, sees a more significant decrease to 73% for up to 30 days in the past. On the other hand, for HitA, the decrease in percentage is less dramatic, as 84% of replaced disks are predicted 3 days before the event and 75% of them 30 days prior. The ratio late vs. early detection is similar for the model built with transfer learning for SgtB and HitB. Compared to the current predictive model implemented in SMART, which mostly warns about a disk failure in the last minute, our model has a major advantage. For both Seagate and Hitachi, an administrator can identify 73 to 75% of the disks to replace a month in advance, which provides her/him with the possibility of planning the replacement in advance, while still using the drives for another 25-30 days.

3.8 SMART indicator rules

Once we have the balanced training dataset comprising the informative SMART attributes in a compact form (this is obtained by running steps 1 through 3 in Algorithm 1) we can use it to fit a decision tree and extract a set of rules that can be used to predict disk replacements. We present rules that are of the form of the underlying learner, more specifically:

$$Rule(\mathbf{x}) = \prod_j \mathcal{I}(\mathbf{x}[i_j] \leq t_j) \prod_k \mathcal{I}(\mathbf{x}[i_k] > t_k) \quad (5)$$

where $\{(i_j, t_j), (i_k, t_k)\}$ represent a set of (smart attribute index, threshold) pairs and $\mathcal{I}(z) = 1$ if z is true and 0 otherwise. The rules provide a detailed insight into the information on the relation among the relevant SMART attributes and the disk replacements available in our training dataset.

Examples of such rules for both SgtA and HitA are provided in Table 6, together with the predicted outcome for the disks adhering to these rules and the prediction confidence. Each rule is composed of one or more single SMART parameter conditions. The fewer conditions, the higher the correlation of the corresponding SMART indicators to the healthy or faulty state of the disks. For instance, in the case of Seagate, the second rule states with 100% confidence that if SMART_197_raw is at least 2, that is the disk has at least two pending sectors, it should be replaced. On the other hand, if its value is below 2, the outcome of the prediction can go both ways, depending on other parameters' values. As an example, consider rules 1 and 3, in which the one indicator that changes the prediction output is the normalized read error rate (SMART_1_normalized). As predicted by our decision tree model, if the number of read errors exceeds 800 thousand, then the disk status is unhealthy.

For Hitachi, the majority contain at most three conditions. As seen in lines 5–8, an indicative combination of attributes is (SMART_197_raw, SMART_3_raw). Line 5 shows that if the number of pending sectors is higher than 1 and the average time spent during a spin up operation exceeds 626 milliseconds, the model predicts an impending faulty state of the disk with 100% confidence. However, if the spin up time is lower, it also considers the number of reallocated sectors in its decision and determines that even with less than 17 such sectors, the disk should be replaced (Line 7). A healthy state is predicted with 97% confidence when the disk has no pending and at most 7200 reallocated sectors, a slow spin up time (i.e., higher than 629 milliseconds) and less than 109 read errors.

Comparing Seagate and Hitachi, we make the following remarks. First, the primarily important SMART indicators are somewhat different. The pending sector count and the read error rate seem to be model and even manufacturer agnostic, while the command timeout (SMART_188), the average spin up time and the reallocated sectors count are disk model-specific. Second, we note a very large difference in the number of read errors that determine a faulty disk state. For Seagate, this threshold is in hundreds of millions, while for Hitachi they are 6 orders of magnitude lower. We attribute this gap to the fact that this indicator is vendor specific, and therefore a comparison across manufacturers is not feasible.

4. DEPLOYMENT

Our predictive model has been designed to reduce disk failures and allow for more efficient, scheduled maintenance processes in place of the inefficient, reactive repair procedures. Especially for enterprise workloads, where more than 99.9% data availability needs to be guaranteed, current storage systems use incorporated Predictive Failure Analysis (PFA) components to anticipate certain forms of disk failures. Such models are often threshold-based and use only read and write error counts to nominate disks for replacement. As shown in Table 4, thresholds lead to less accurate replacement decisions, therefore integrating our approach enables more precise replacement strategies.

We exemplify how our component could be integrated for rebuilding a RAID 5 array when a disk is signaled as likely to fail, through smart rebuild [3].

An early signal enables the disk to still be available for I/O operations, and thus be kept in the array, rather than

Line	Model	Rule	Outcome	Confidence
1	Seagate	If $SMART_197_raw < 2$ and $SMART_188_raw > 0$ and $SMART_1_normalized \in [0, 117)$	Healthy	100%
2	Seagate	If $SMART_197_raw \geq 2$	Replace	100%
3	Seagate	If $SMART_197_raw < 2$ and $SMART_188_raw > 0$ and $SMART_1_normalized > 117$	Replace	80%
4	Seagate	If $SMART_197_raw < 2$ and $SMART_188_raw = 0$ and $SMART_187_normalized < 100$ and $SMART_240_raw < 14780$ billion	Replace	97%
5	Hitachi	If $SMART_197_raw > 1$ and $SMART_3_raw > 626$	Replace	100%
6	Hitachi	If $SMART_197_raw > 5$ and $SMART_3_raw < 626$ and $SMART_5_raw > 17$	Replace	92%
7	Hitachi	If $SMART_197_raw > 1$ and $SMART_3_raw < 626$ and $SMART_5_raw < 17$	Replace	100%
8	Hitachi	If $SMART_197_raw < 1$ and $SMART_5_raw < 7200$ and $SMART_3_raw > 629$ and $SMART_1_raw \in [0, 109]$	Healthy	97%

Table 6: Examples of rules extracted from a decision tree model trained on the Seagate and Hitachi datasets obtained with Algorithm 1.

being rejected because of a standard rebuild. A spare disk can be either used from the array or brought in if none is available. The signaled drive and the spare are put in a temporary RAID 1 (full mirroring). This allows the duplication of the faulty drive onto the spare, rather than performing full RAID reconstruction which slows down the entire array’s performance. The spare becomes a regular member of the array and the signaled-to-be-faulty disk can be safely removed from the disk, without any risk of data loss. By using such models that can detect failures early in advance and have low ratios of false positives, the array would never have to go through a time consuming $n - 1$ disks stage where it would be exposed to complete RAID failure if an additional drive fails in the meantime. Therefore, the benefits – time saving and increased availability – are substantial.

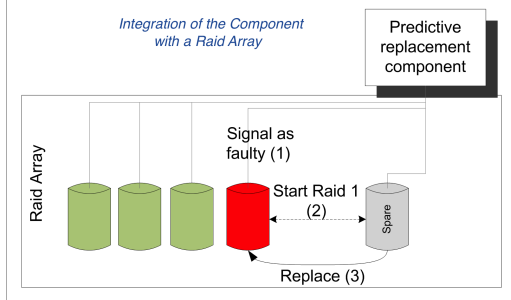


Figure 7: Integration of the predictive replacement component with storage arrays

In Figure 7, we show how our predictive component could be used in interaction with the RAID array and the steps necessary for the automatic rebuild: (i) the predictive component signals a disk with impending failure, (ii) the mirroring process is started on the spare, (iii) the unhealthy disk is replaced by its healthy mirror. In this setting, the process falls back to RAID rebuild only if necessary (e.g., mirroring is not possible).

Our failure model can be deployed in large scale environments (e.g data centers), provided that two conditions are fulfilled. First, the SMART parameters identified as relevant (see Table 2) are continuously measured by the manufacturer. Second in order to learn such model for different manufacturers, even though failed disks represent only a small fraction compared to the healthy ones, in absolute terms, these need to be in the order of hundreds for the model to achieve precision and recall higher than 80%.


5. RELATED WORK

Researchers have performed a couple of large-scale studies on disk failures, the most notable ones pertaining to the authors of [18, 17]. They observed that the field replacement rates of drives are significantly higher than those in the technical datasheets – 2-10 times higher for disks aged less than 5 years and up to 30 times higher for disks between 5 and 8 years. They also demonstrate a significant overestimation of MTTF by the manufacturer. The authors also observed a continuous increase in the replacement rates, starting already in the second year of operation and a high correlation between the first error and a later disk failure. Our analysis also confirms these findings.

This line of research is complemented by works of [5, 11, 13, 16, 19] where the focus is on building a predictive model for the timely discovery of impending disk failures. In [11], the authors employ Bayesian methods to model disk drive failures based on SMART data. First, they solve an anomaly detection problem (i.e., by looking back in the life-span of the drive and establishing if any of the previous observations is an anomaly). They achieve this by applying a mixture model based on naive Bayes clusters trained using expectation-maximization. Second, they train a naive Bayes classifier which predicts that a drive will fail if any of its snapshots are identified as anomalous or as failures. They evaluate the approach on a smaller dataset, consisting of 1936 drives, out of which only 9 were marked as failed, with a detection rate of up to 55% only.

The authors of [13] explore the capabilities of statistical tests such as the multivariate rank sum test to improve failure warning accuracy and lower false alarms. Their dataset is also fairly small with only 3744 drives (out of which 36 failures), coming from two different models and with each set containing at most 3 months of reliability design test data. The highest accuracies achieved were modest (40%-60%) at 5% AFR. A different model for predicting failures is proposed in [16], comprising of an algorithm based on the multiple-instance learning framework and the naive Bayes classifier. The dataset used was again very small - only data from 369 drives.

There are several key differences between the aforementioned studies and ours. First, the number of disks we consider is significantly larger, with over 23000 drives. Second, our approach focuses on selecting the SMART indicators that correlate with disk replacements and proposes stable representations of the time series data for each disk as input

to the predictive model. Last, but not least, some studies are based on monitoring data from drives used in accelerated life tests, whereas we rely only on field  collected when the disks were in actual use. The problem with data collected during testing in uniform controlled environments is that although it can be insightful in understanding the role of certain environmental factors, it has been shown to be not informative enough with respect to actual failure rates observed in the field [10].

Finally, we note that some manufacturers deploy the disks with embedded failure predictive models. However, these models are based on simple methods, such as threshold-based normalizations which according to field observations these models are built such that they avoid false alarms at the expense of a weak predictive power [18, 17, 12].

6. CONCLUSIONS

In this paper, we present a machine learning-based pipeline for predicting disk replacements, built and evaluated on real data from a large disk population from two different manufacturers. We demonstrate the ability of our model built using SMART data to predict disk replacements with high accuracy. A changepoint based feature selection and a compact representation of the time series data for the SMART indicators plugged into a RGF classifier achieves up to 98% accuracy in predicting replacements, 10-15 days in advance. As expected, such models are sensitive to the number of SMART attributes they learn from and the size of the training data. Given that in our original dataset there were considerably less indicators with non-null values for Seagate, we were able to build a model with 24 attributes for Seagate contrary to 12 only for Hitachi. This together with the dataset size explains the 17% difference in accuracy for the two disk models – 98% and 81%, respectively. We also demonstrate how transfer learning can be used to reuse the information available in the labeled dataset for a disk model from a specific manufacturer to build a high quality predictive model for a new disk model from the same manufacturer with no available labeled data.

We believe that such high quality models have many practical benefits. First of all, they can be easily applied to any disk model or manufacturer as long as SMART data is collected. Second, they provide an automatic tool for the disk replacement problem that can be a valuable asset enabling the administrators to identify faulty disks in due time. Last but not least, the predictive models mitigate the reliability issues of storage service providers by allowing administrators to backup the data and plan the actual replacement in advance. Also, note that all these benefits are achievable based only on data that is automatically collected from the disk and no extra effort is necessary.

¹Seagate is a trademark of Seagate Technology LLC.

²Hitachi is a registered trademark of Hitachi, Ltd., and/or its affiliates in the United States and other countries.

7. REFERENCES

- [1] Data center downtime costs. <http://www.emerson.com/en-us/News/Pages/Net-Power-Study-Data-Center.aspx>.
- [2] Hard drive smart stats. <https://www.backblaze.com/blog/hard-drive-smart-stats/>.
- [3] IBM system storage DS8000 architecture and implementation. <http://www.redbooks.ibm.com/redbooks/pdfs/sg248886.pdf>.
- [4] S.M.A.R.T. <https://en.wikipedia.org/wiki/S.M.A.R.T.>
- [5] V. Agarwal, C. Bhattacharyya, T. Niranjana, and S. Susarla. Discovering rules from disk events for predicting hard drive failures. ICMLA '09, pages 782–786, Dec 2009.
- [6] L. Breiman. Random forests. *Machine Learning*, 45(1):5–32.
- [7] K. H. Brodersen, F. Gallusser, J. Koehler, N. Remy, and S. L. Scott. Inferring causal impact using bayesian structural time-series models. *Annals of Applied Statistics*, 9:247–274, 2015.
- [8] C. J. C. Burges. A tutorial on support vector machines for pattern recognition. *Data Min. Knowl. Discov.*, 2(2):121–167, June 1998.
- [9] D. R. Cox. The regression analysis of binary sequences (with discussion). *J Roy Stat Soc B*, 20:215–242, 1958.
- [10] J. Elerath and S. Shah. Server class disk drives: how reliable are they? In *Reliability and Maintainability, 2004 Annual Symposium - RAMS*, pages 151–156.
- [11] G. Hamerly and C. Elkan. Bayesian approaches to failure prediction for disk drives. ICML '01, pages 202–209, San Francisco, CA, USA, 2001. Morgan Kaufmann Publishers Inc.
- [12] How does S.M.A.R.T. function of hard disks work? www.hdsentinel.com/smart/index.php.
- [13] G. F. Hughes, J. F. Murray, K. Kreutz-Delgado, and C. Elkan. Improved disk-drive failure warnings. *IEEE Transactions on Reliability*, 51(3):350–357, 2002.
- [14] R. Johnson and T. Zhang. Learning nonlinear functions using regularized greedy forest. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(5):942–954, May 2014.
- [15] J. MacQueen. Some methods for classification and analysis of multivariate observations. In *Berkeley Symposium on Mathematical Statistics and Probability*, pages 281–297, 1967.
- [16] J. F. Murray, G. F. Hughes, and D. Schuurmans. Machine learning methods for predicting failures in hard drives: A multiple-instance application. *Journal of Machine Learning research*, 6:816, 2005.
- [17] E. Pinheiro, W.-D. Weber, and L. A. Barroso. Failure trends in a large disk drive population. FAST 2007, Berkeley, CA, USA, 2007. USENIX Association.
- [18] B. Schroeder and G. A. Gibson. Disk failures in the real world: What does an mttf of 1,000,000 hours mean to you? FAST 2007, Berkeley, CA, USA, 2007. USENIX Association.
- [19] Y. Tan and X. Gu. On predictability of system anomalies in real world. In *IEEE MASCOTS, 2010*, pages 133–140, Aug 2010.
- [20] G. M. Weiss and H. Hirsh. Learning to predict rare events in event sequences. In *KDD 1998*, pages 359–363. AAAI Press, 1998.
- [21] J. Ye, J.-H. Chow, J. Chen, and Z. Zheng. Stochastic gradient boosted distributed decision trees. CIKM 2009, pages 2061–2064. ACM, 2009.