

6 курс М05-313в, кафедра “Облачные технологии” МТС

Физтех-школа прикладной математики и информатики МФТИ





Содержание

- › Актуальность
- › Введение
- › Цель и задачи
- › Существующие решения
- › Байесовская нейросеть/CatBoost
- › Предлагаемое решение
- › Сравнение результатов
- › Итоги и дальнейшие планы



Актуальность

- › Типичный уровень отказов накопителей **~1%** в год;
На масштабе облака МТС это десятки и сотни сбоев ежегодно;
Например, у Backblaze ~5 выходов из строя в день при 260k дисков;
- › Отказ накопителя провоцирует
 - внеплановые затраты на замену
 - нагрузку на L2 поддержку
 - деградацию производительности кластера.
- › Непредсказуемые сбои усложняют планирование емкости и угрожают SLA.
Прогнозирование отказов критично для снижения расходов и выполнения SLA.





Введение

- › **Время до отказа (TTF):**

случайная величина – период работы накопителя от ввода в эксплуатацию до сбоя.

- › **Функция выживаемости $S(t)$:**

вероятность, что накопитель проработает дольше времени t ($S(0)=1$, убывает к 0).

- › **Плотность отказов $f(t)$:**

распределение вероятностей времени отказа; характеризует относительную частоту сбоев в момент t .

- › **Интенсивность отказов $\lambda(t)$:**

мгновенный риск отказа в момент t , условно при дожитии до t (отражает скорость “старения” устройства).



Введение

- Функция выживаемости

$$S(t) = \mathbb{P}(T > t),$$

- Множительная оценка Каплана-Мейера

$$\hat{S}(t) = \prod_{i: t_i \leq t} \left(1 - \frac{d_i}{n_i}\right),$$

- Распределение Вейбулла (наработка до отказа)

$$f_X(x) = \begin{cases} \frac{k}{\lambda} \left(\frac{x}{\lambda}\right)^{k-1} e^{-\left(\frac{x}{\lambda}\right)^k}, & x \geq 0 \\ 0, & x < 0 \end{cases}.$$



Введение

$$\text{CVaR}\alpha = \mathbb{E}[L \mid L \geq \text{VaR}\alpha]$$

где:

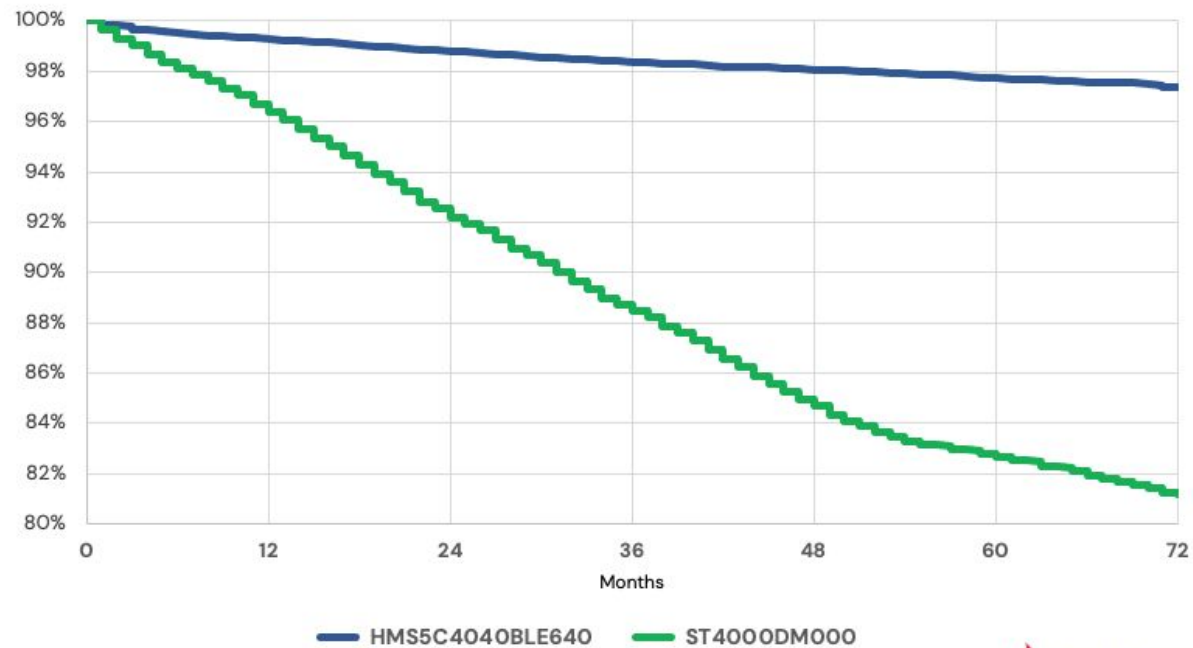
- $\text{CVaR}\alpha$ — условная стоимость под риском на уровне доверия α ;
- $\mathbb{E}[L \mid L \geq \text{VaR}\alpha]$ — математическое ожидание потерь L при условии, что они превышают значение $\text{VaR}\alpha$;
- $\text{VaR}\alpha$ — стоимость под риском на уровне доверия α , то есть пороговое значение потерь, которое не будет превышено с вероятностью α .



Введение

Kaplan-Meier Life Expectancy Curves for Select Backblaze 4TB Drives

Source: Backblaze Drive Stats data from 4/2013 through 3/2022 inclusive, as of 31 March 2022



MFG	MODEL	DRIVES IN OPERATION	LIFETIME DRIVE FAILURES	LIFETIME DRIVE DAYS	LIFETIME AFR
HGST	HMS5C4040BLE640	12,728	343	30,025,871	0.40%
Seagate	ST4000DM000	18,495	4,581	68,104,520	2.45%



Введение

- У разных производителей некоторые из метрик могут отличаться по смыслу
- Не все метрики влияют на вероятность отказа диска

Table 1

Descriptions of relevant SMART metrics.

SMART	Attribute name	Description
3	Spin-up time	Average time (in milliseconds) of spindle spin up from zero RPM to fully operational.
5	Reallocated sectors count	Count of bad sectors that have been found and reallocated. A hard drive which has had a reallocation is very likely to fail in the immediate months.
7	Seek error rate	Rate of seek errors of the magnetic heads, due to partial failure in the mechanical positioning system.
187	Reported uncorrectable errors	Count of errors that could not be recovered using hardware ECC (Error-Correcting Code), a type of memory used to correct data corruption errors.
188	Command timeout	Count of aborted operations due to hard drive timeout.
190	Temperature difference	Difference between current hard drive temperature and optimal temperature of 100°C.
197	Current pending sectors count	Count of bad sectors that have been found and waiting to be reallocated, because of unrecoverable read errors.
198	Offline uncorrectable sectors count	Total count of uncorrectable errors when reading/writing a sector, indicating defects of the disk surface or problems in the mechanical subsystem.



Цель и задачи

- › BNN/CatBoost модель прогноза отказов накопителей оптимизирует средний доход облака МТС на ~10% снижает риск (CVaR) на ~20% по сравнению с текущим подходом.
- › Сбор и подготовка данных (SMART атрибуты, Backblaze + HWLinux)
обучение моделей (baseline vs. BNN/CatBoost)
симуляция финансовых последствий (доход/риск)



Существующие решения

- › Точечный прогноз (MTBF, AFR):
 - + простота, минимальные требования к данным
 - не учитывает неопределенность (риск)
- › Weibull MLE:
 - + простота (распределение исходов)
 - не учитывает сложные зависимости, предполагает постоянный уровень риска отказа.
- › Регрессии (Cox Proportional Hazard):
 - + интерпретируемость
 - линейность, частотный подход
- › Байесовские модели(BNN/Catboost):
 - + учёт неопределенности
 - высокая вычислительная сложность, требуется больше ресурсов



multiple instance Naive Bayes

Algorithm 1 mi-NB Train (for SMART failure prediction)

Input: $\mathbf{x}, \mathcal{Y}, FA_{\text{desired}}$ (desired false alarm rate)

Initialize:

Good drives: For drives with $\mathcal{Y}_i = 0$ initialize $y_j = 0$ for $j = 1 \dots N_i$

Failed drives: For drives with $\mathcal{Y}_i = 1$ initialize $y_j = 0$ for $j = 1 \dots N_i - 1$, and $y_{N_i} = 1$

Learn NB model

$\hat{y}_j = \arg \max_{c \in \{0,1\}} f_c(\mathbf{x}_j)$ Classify each pattern using the NB model

Find FA and DET rate

while $FA < FA_{\text{target}}$ **do**

for all Misclassified failed drives, $\hat{y}_j = 0 \forall j = 1 \dots N_i$ **do**

$j^* = \arg \max_{j \in \{1 \dots N_i | y_j = 0\}} f_1(\mathbf{x}_j)$ Find pattern closest to decision surface with label $y_j = 0$

$y_{j^*} \leftarrow 1$ Reclassify the pattern as failed

 Update NB model

end for

$\hat{y}_j = \arg \max_{c \in \{0,1\}} f_c(\mathbf{x}_j)$ Reclassify each pattern using the NB model

 Find FA and DET rate

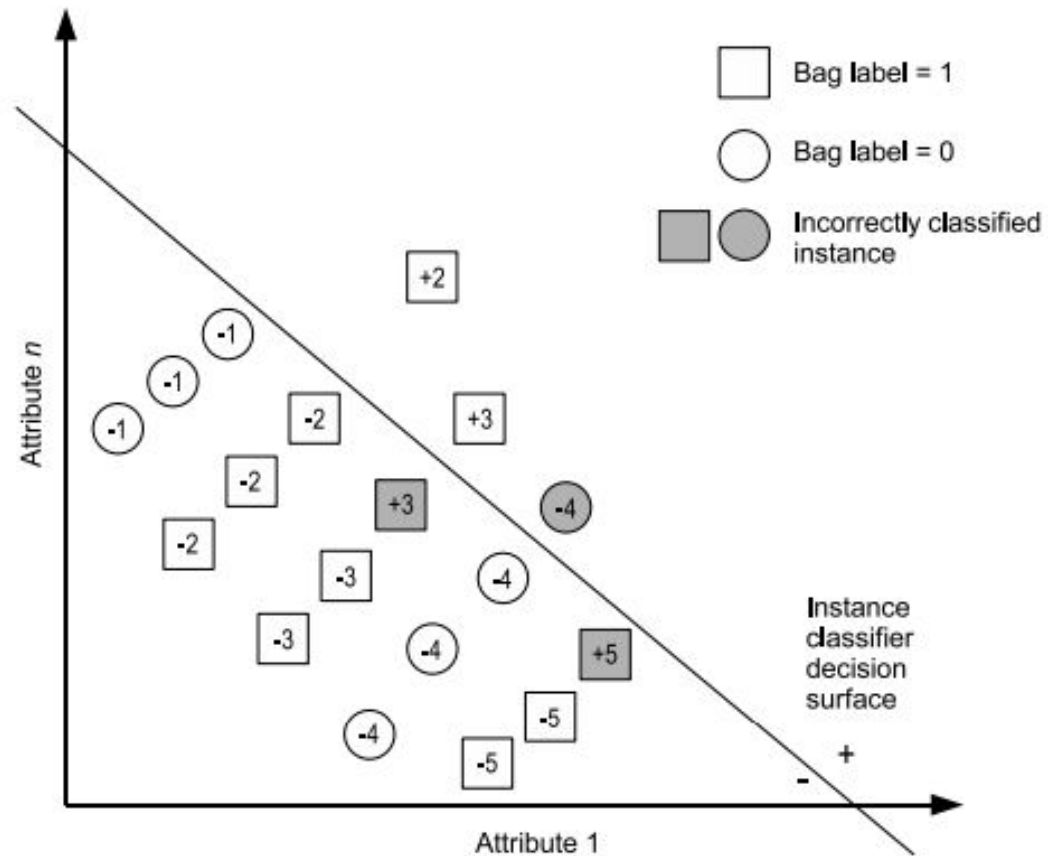
end while

Return: NB model



multiple instance Naive Bayes

METHODS FOR PREDICTING FAILURES IN HARD DRIVES





multiple instance Naive Bayes

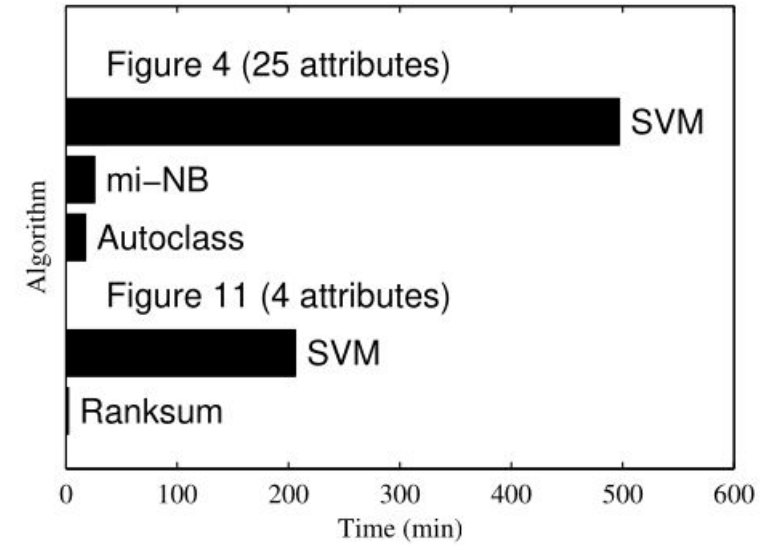
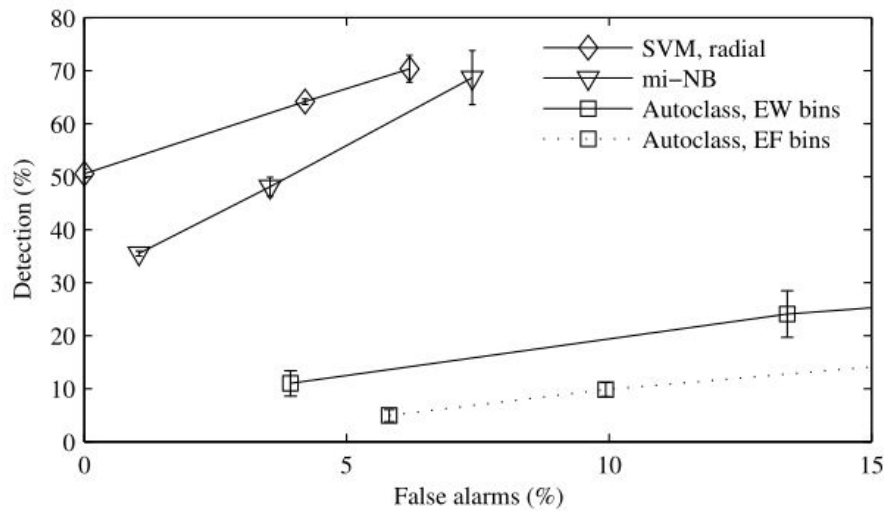


Figure 4: Failure prediction performance of SVM, mi-NB and Autoclass using 25 attributes (one sample per pattern) measured per drive. For mi-NB, the results shown are for equal-width binning. Autoclass is tested using both equal-width (EW) and equal-frequency (EF) binning (results with 5 bins shown). Error bars are ± 1 standard error in this and all subsequent figures.



Модель предсказания отказа

CatBoost

- › Прогноз распределения: модель возвращает не единичное значение, а распределение времени до отказа
- › Учет неопределенности: важно для управления рисками (меньше ложноположительных и ложноотрицательных срабатываний).
- › Устойчива при малом количестве отказов (редкие события) за счет априорных ограничений
- › Возможность оценивать эпистемическую неопределенность, используя постериорное сэмплирование

```
1 model = CatBoostClassifier(iterations=300,  
2                             depth=6,  
3                             learning_rate=0.1,  
4                             loss_function='Logloss',  
5                             posterior_sampling=True,  
6                             verbose=False)  
7 model.fit(X_feat, y_fail)  
8 res = model.virtual_ensembles_predict(X_feat, virtual_ensembles_count=100, prediction_type="VirtEnsembles")
```



Учет экономических рисков

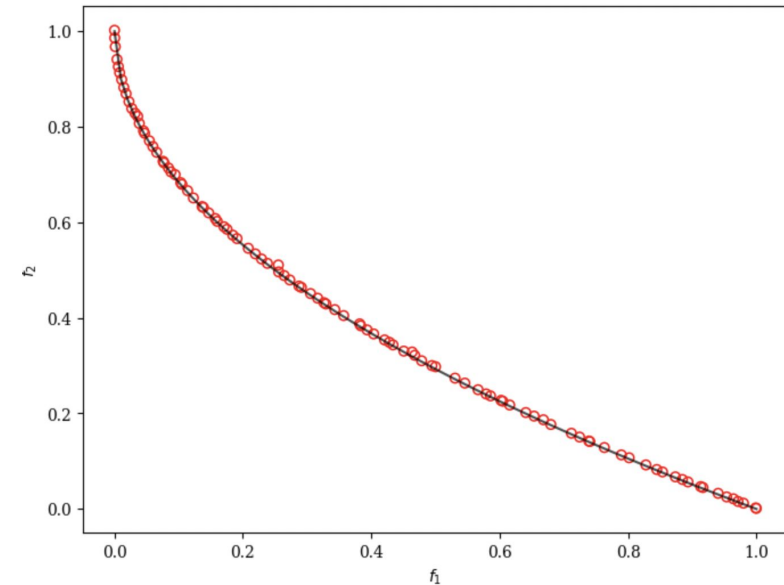
СИМВОЛ	СМЫСЛ
p	вероятность отказа за H (из BNN)
t	порог плановой замены (решение)
$\delta = \mathbf{1}_{[p \geq t]}$	1 — меняем заранее, 0 — ждём
$F \sim \text{Bernoulli}(p)$	фактический отказ, учитывается только если $\delta = 0$
Стоимости	C_{plan} — плановая смена, C_{swap} — аварийная, P_{SLA} — штраф
Доходы	R_{op} — выручка от живого диска за H ; M_{upsell} — маржа от продажи нового
I_{buf}	удельная стоимость хранения «холодного» запасного диска

$$\Pi = -\delta C_{\text{plan}} + \delta M_{\text{upsell}} + (1 - \delta) \left[(1 - F) R_{\text{op}} - F (C_{\text{swap}} + P_{\text{SLA}}) \right] - I_{\text{buf}}$$



Предлагаемое решение

- › Данные → дополнены HWLinux, учитывает данные с российского рынка дисков.
- › Модель отказов (Bayesian NN/CatBoost) → вероятности отказа на разных сроках, функция выживаемости
- › Модель доходов → выбираем политику замены дисков $\Pi(t, b)$.
 - b - количество закупленных дисков
 - t - пороговая вероятность для замены
 - Если $p \geq t \rightarrow$ плановая замена,
 - Если $p < t \rightarrow$ оставляем диск
- › Оптимизация рисков → рунгоо NSGA-II



фронт Паретто из рунгоо



Сравнение результатов

Датасет ~12 GB

	date	serial_number	model	capacity_bytes	failure	datacenter	cluster_id	vault_id	pod_id	pod_slot_num
0	2024-12-22	S2ZYJ9GGB01000	ST500LM012 HN	500107862016	0	sac0	0	1028	0	NaN
1	2024-12-22	VKH7A84X	HGST HUH728080ALE600	8001563222016	0	sac0	0	1028	0	22.0
2	2024-12-22	ZA106RPJ	ST8000DM002	8001563222016	0	sac0	0	1028	0	28.0
3	2024-12-22	ZA106YX1	ST8000DM002	8001563222016	0	sac0	0	1028	0	23.0
4	2024-12-22	ZA106Z5M	ST8000DM002	8001563222016	0	sac0	0	1028	0	39.0

- › Baseline (Weibull MLE) [<https://arxiv.org/pdf/1812.07142>]
ROC-AUC: 0.72
- › BNN
ROC-AUC = 0.88
- › CatBoost
ROC-AUC = 0.93
- › CatBoost + Pymoo optimization
CVaR = -13% (по сравнению со случайным набором)



Итоги и дальнейшие планы

- › Достигнут прирост среднего дохода ~10% и снижение рисков SLA по сравнению с baseline.

Дальнейшие планы

- › Оформить результаты в тексте диплома
- › Улучшить качество решения за счет подбора гиперпараметров
- › Внедрить в перспективе production-ready решение для облака МТС, автоматическое дообучение
- › Планировать расписание закупок для поддержания фиксированного числа живых дисков



Список литературы

- › <https://www.backblaze.com/cloud-storage/resources/hard-drive-test-data>
- › <https://github.com/linuxhw/SMART/tree/master>
- › S. Guo, H. Li, Y. Dong, J. Li, Y. Song and J. Xie, "A Reliability Evaluation Model for Smart Meters Based on Survival Analysis Theory," 2023 4th International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE), Hangzhou, China, 2023, pp. 205-208, doi: 10.1109/ICBAIE59714.2023.10281232.
- › Ping Wang, Yan Li, and Chandan K. Reddy. 2019. Machine Learning for Survival Analysis: A Survey. ACM Comput. Surv. 51, 6, Article 110 (November 2019), 36 pages. <https://doi.org/10.1145/3214306>
- › Jahanbani Fard, Mahtab & Wang, Ping & Chawla, Sanjay & Reddy, Chandan. (2016). A Bayesian Perspective on Early Stage Event Prediction in Longitudinal Data. IEEE Transactions on Knowledge and Data Engineering. PP. 10.1109/TKDE.2016.2608347.
- › Maxime Amram, Jack Dunn, Jeremy J. Toledano, Ying Daisy Zhuo, Interpretable predictive maintenance for hard drives, Machine Learning with Applications, Volume 5, 2021, ISSN 2666-8270, <https://doi.org/10.1016/j.mlwa.2021.100042>.
- › Jayanta Basak and Randy H. Katz, Significance of Disk Failure Prediction in Datacenters, 2017, <https://arxiv.org/abs/1707.01952>
- › Murray, Joseph & Hughes, Gordon & Kreutz-Delgado, Ken. (2005). Machine Learning Methods for Predicting Failures in Hard Drives: A Multiple-Instance Application.. Journal of Machine Learning Research. 6. 783-816.
- › I. C. Chaves, M. R. P. de Paula, L. G. M. Leite, L. P. Queiroz, J. P. P. Gomes and J. C. Machado, "BaNHFaP: A Bayesian Network Based Failure Prediction Approach for Hard Disk Drives," 2016 5th Brazilian Conference on Intelligent Systems (BRACIS), Recife, Brazil, 2016, pp. 427-432, doi: 10.1109/BRACIS.2016.083.