SVKM'S NMIMS Deemed-to-be UNIVERSITY

RESEARCH TREATISE

GROUP NO-8

A Study on Analysing Failure Rate of Storage Devices using Weibull Log Logistic Grey Forecasting Model.

Team Members-

NAME	ROLL NO.	SAP ID
Ritesh Patil	A047	86062400006
Darshana Parle	AO43	86062400032
Sejal Mahadik	A033	86062400036
Reetik Prajapati	A059	86062400005
Nishad Sangale	A062	86062400072

Mentors- Prof. Vaibhav Vasundekar

OVERVIEW

- Introduction
- Objectives
- Dataset
- Methodology
- Results
- Conclusions
- Limitations
- Future Scope

INTRODUCTION

- In today's data-centric world, hard disk drives are crucial for storage systems, and unexpected failures can lead to costly disruptions.

- Our project tested the accuracy and adaptability of the Weibull Log-Logistic Grey Forecasting Model, which combines Weibull and Log-Logistic distributions with grey modeling to capture complex failure patterns even in data with uncertainties.

- Results showed that the model provides reliable predictions with lower error rates than traditional methods, helping data centers anticipate failures, allocate resources effectively, and reduce downtime.

LITERATURE REVIEW

Research Paper Title	Source	Year	Citation
1.New Weibull Log-Logistic grey forecasting model for a hard disk drive failures	Applied Mathematical Modelling (ELSEVIER)	2023	Rongxing Chen , Xinping Xiao,New Weibull Log-Logistic grey forecasting model for a hard diskdrive failures, https://www.sciencedirect.com/science/article/abs/pii/S0307904X24001756
2.Hard disk drive failure prediction using SMART attribute	Materials Today: Proceedings (ELSEVIER)	2021	V. Tomer, V. Sharma, S. Gupta, et al., Hard disk drive failure prediction using SMART attribute, Mater. Today: Proc. 46 (2021) 11258–11262, https://doi.org/10.1016/j.matpr.2021.03.229
3.The whale optimization algorithm	Journal of King Saud University – Computer and Information Sciences (ELSEVIER)	2016	S. Mirjalili, A. Lewis, The whale optimization algorithm, Adv. Eng. Softw. 95 (2016) 51–67, https://doi.org/10.1016/j.advengsoft.2016.01.008

OBJECTIVES

1. Test a forecasting model for accurate HDD failure prediction to enhance data storage reliability by capturing trends and reducing volatility.
2. Implement a grey model with non-linear features to reduce incorrect predictions and unnecessary HDD replacements related costs.
3. Validate the model on real-world HDD data and compare its performance with the Multiple Linear Regression (MLR) model.

DATASET DESCRIPTION

	A	B	C	D	E	F	G	H	I	J	K	L
1	date	model	smart_5_normaliz	smart_5_raw	smart_187_normalized	smart_187_raw	smart_188_normalized	smart_188_raw	smart_197_normalized	smart_197_raw	smart_198_normalized	smart_198_raw
2	01/06/24	ST14000NM001G	100	0	100	0	100	0	100	0	100	0
3	01/06/24	ST14000NM001G	100	0	100	0	100	0	100	0	100	0
4	01/06/24	ST14000NM001G	100	0	100	0	100	0	100	0	100	0
5	01/06/24	ST14000NM001G	100	0	100	0	100	0	100	0	100	0
6	01/06/24	ST14000NM001G	100	0	100	0	100	0	100	0	100	0
7	01/06/24	ST14000NM001G	100	0	100	0	100	0	100	0	100	0
8	01/06/24	ST14000NM001G	100	0	100	0	100	0	100	0	100	0
9	01/06/24	ST14000NM001G	100	8	92	8	100	0	100	0	100	0
10	01/06/24	ST14000NM001G	100	0	100	0	100	0	100	0	100	0
11	01/06/24	ST14000NM001G	100	0	100	0	100	0	100	0	100	0
12	01/06/24	ST14000NM001G	100	0	100	0	100	0	100	0	100	0
13	01/06/24	ST14000NM001G	100	0	100	0	100	0	100	0	100	0
14	01/06/24	ST14000NM001G	100	0	100	0	100	0	100	0	100	0
15	01/06/24	ST14000NM001G	100	0	100	0	100	0	100	0	100	0
16	01/06/24	ST14000NM001G	100	0	100	0	100	0	100	0	100	0
17	01/06/24	ST14000NM001G	100	0	100	0	100	0	100	0	100	0
18	01/06/24	ST14000NM001G	100	104	100	0	100	0	100	0	100	0
19	01/06/24	ST14000NM001G	100	0	100	0	100	0	100	0	100	0
20	01/06/24	ST14000NM001G	100	0	100	0	100	0	100	0	100	0
21	01/06/24	ST14000NM001G	100	0	100	0	100	0	100	0	100	0
22	01/06/24	ST14000NM001G	100	0	100	0	100	0	100	0	100	0
23	01/06/24	ST14000NM001G	100	0	100	0	100	0	100	0	100	0
24	01/06/24	ST14000NM001G	100	0	100	0	100	4295032833	100	0	100	0
25	01/06/24	ST14000NM001G	100	0	100	0	100	0	100	0	100	0
26	01/06/24	ST14000NM001G	100	0	100	0	100	0	100	0	100	0
27	01/06/24	ST14000NM001G	100	0	100	0	100	0	100	0	100	0
28	01/06/24	ST14000NM001G	100	0	100	0	100	0	100	0	100	0
29	01/06/24	ST14000NM001G	100	0	100	0	100	0	100	0	100	0
30	01/06/24	ST14000NM001G	100	0	100	0	100	0	100	0	100	0
31	01/06/24	ST14000NM001G	100	0	100	0	100	0	100	0	100	0
32	01/06/24	ST14000NM001G	100	0	100	0	100	0	100	0	100	0
33	01/06/24	ST14000NM001G	100	0	100	0	100	0	100	0	100	0
34	01/06/24	ST14000NM001G	100	0	100	0	100	0	100	0	100	0
35	01/06/24	ST14000NM001G	100	0	100	0	100	0	100	0	100	0
36	01/06/24	ST14000NM001G	100	0	100	0	100	0	100	0	100	0
37	01/06/24	ST14000NM001G	100	0	100	0	100	0	100	0	100	0
38	01/06/24	ST14000NM001G	100	0	100	0	100	0	100	0	100	0
39	01/06/24	ST14000NM001G	100	0	100	0	100	0	100	0	100	0
40	01/06/24	ST14000NM001G	100	0	100	0	100	0	100	0	100	0
41	01/06/24	ST14000NM001G	100	0	100	0	100	0	100	0	100	0
42	01/06/24	ST14000NM001G	100	0	100	0	100	0	100	0	100	0
43	01/06/24	ST14000NM001G	100	0	100	0	100	0	100	0	100	0
44	01/06/24	ST14000NM001G	100	0	100	0	100	0	100	0	100	0

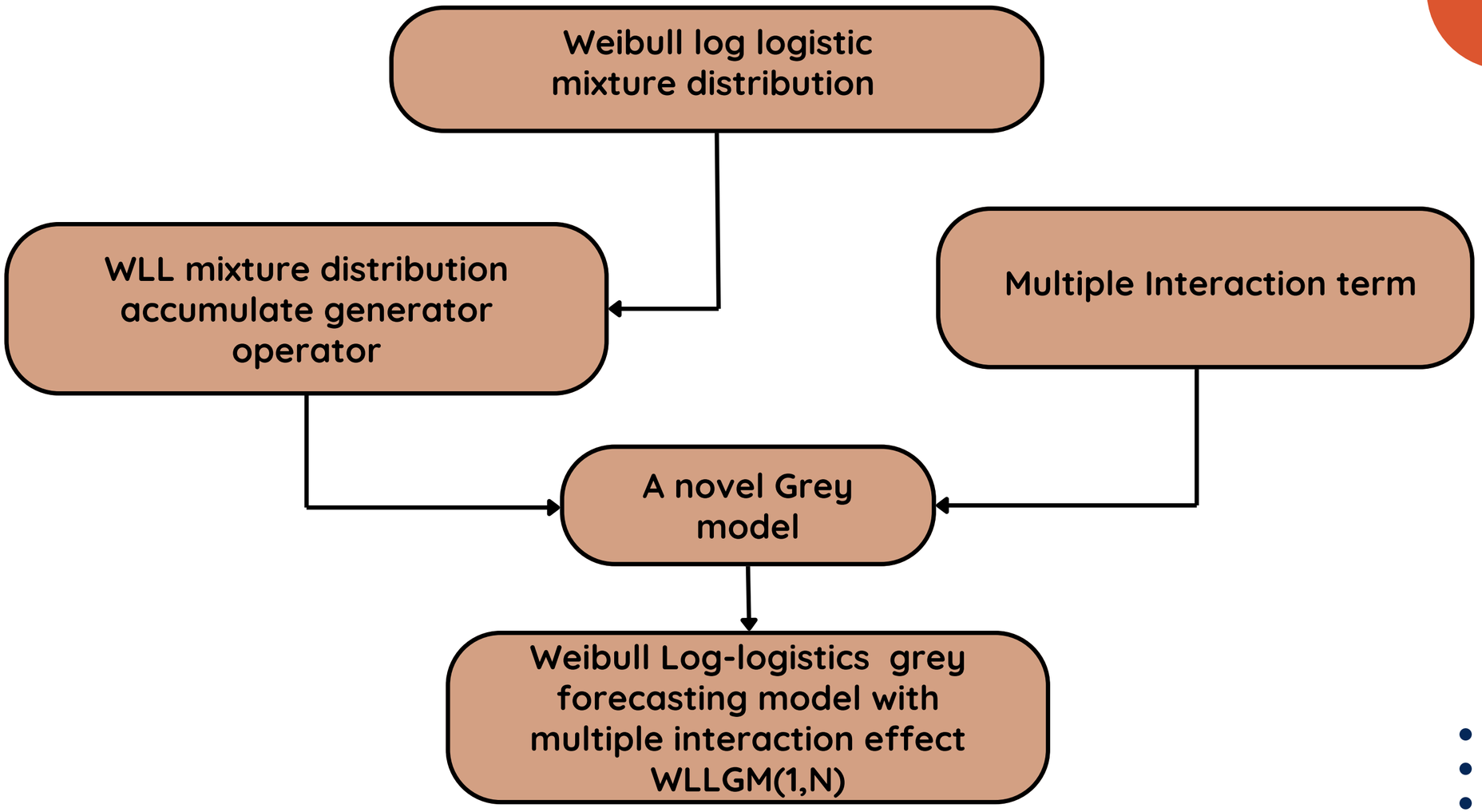
SOURCE- BACKBLAZE.COM

Attribute	Name
SMART5	Reallocated Sectors Count
SMART187	Reported UncorrecTable Errors
SMART188	Command Timeout
SMART197	Current Pending Sector Count
SMART198	UncorrecTable Sector Count

SMART - Self Monitoring And ReportingTechnology

Models-
ST14000NM001G
ST4000DM000
ST14000NM0138

WLLGM ESTABLISHMENT



Weibull distribution

Weibull distribution is one of the most used distributions in reliability and survival analysis analysis

$$f_1(t,\psi_1) = (\beta_1 / \alpha_1)(t/\alpha_1)^{\beta_1 - 1} e^{-(t/\alpha_1)^{\beta_1}},\ t > 0,\ \alpha_1 > 0,\ \beta_1 > 0,$$

scale parameter = α1
shape parameter = β1

Log-Logistic distribution

The log-logistic distribution has large-scale applications in analyzing time-to-event data

$$f_2(t,\psi_2) = \frac{(\beta_2/\alpha_2)(t/\alpha_2)^{\beta_2 - 1}}{\left[1 + (t/\alpha_2)^{\beta_2}\right]^2},\ t > 0,\ \alpha_2 > 0,\ \beta_2 > 0,$$

scale parameter = α2
shape parameter = β2

Mixture distribution

A mixture distribution is a probability distribution created by combining two or more component distributions, each representing a different subpopulation within the data.

Weibull Log-Logistic distribution

$$f(t,\psi) = pf_1(t,\psi_1) + (1-p)f_2(t,\psi_2),\ 0 < p < 1,$$

p and 1-p are mixing proportion

METHODOLOGY

1. Data Preprocessing

- Data Cleaning
- SMART Attribute Selection

2. Model Formulation

- Weibull
- Log-Logistic
- Weibull Log-Logistic Mixture Distribution
- Accumulation Generation Operator

4. Multiple Interaction Effects Modeling

- Defining Interaction Terms

3. Parameter Estimation

- Linear & Nonlinear Parameters
- Structural Parameters- Whale Optimization Algorithm (WOA)

5. Model Evaluation

- MAPE
- RMSE
- MAE
- STD

RESULTS

Model - ST14000NM001G			
Month	Actual	WLLGM	MLR
Fitting			
February	12	12.7933	16.2408
March	13	12.5304	15.2672
April	16	18.3272	14.0256
May	19	20.3961	21.0256
Predicted			
June	13	12.7079	14.0256

Experiment results of WLLGM and MLR model for hard disk drive failure.

Evaluation Metrics of models for fitting and prediction.

Model - ST14000NM001G		
Criterion	WLLGM	MLR
Fitting		
MAPE(%)	0.2002	1.2036
RMSE	0.1890	0.3624
STD%	0.0580	0.1254
MAE	0.1027	0.2145
Prediction		
MAPE(%)	0.6077	4.1824
RMSE	0.5737	0.3214
STD%	0.3664	0.2823
MAE	0.557	0.7523

RESULTS

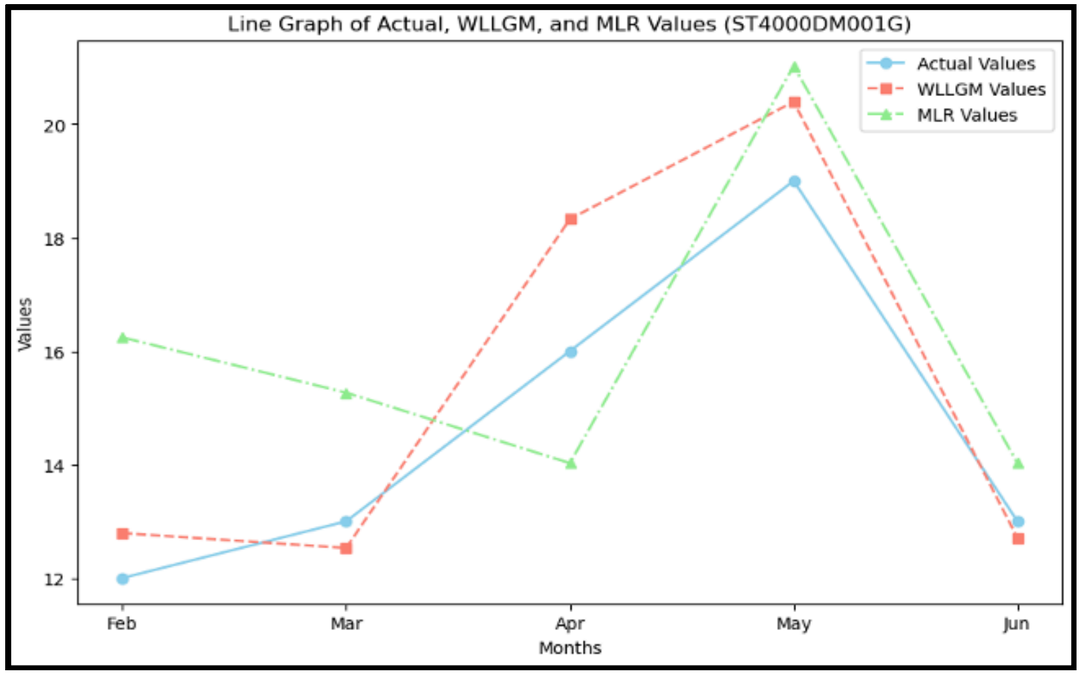
The results of Parameters of WLLGM

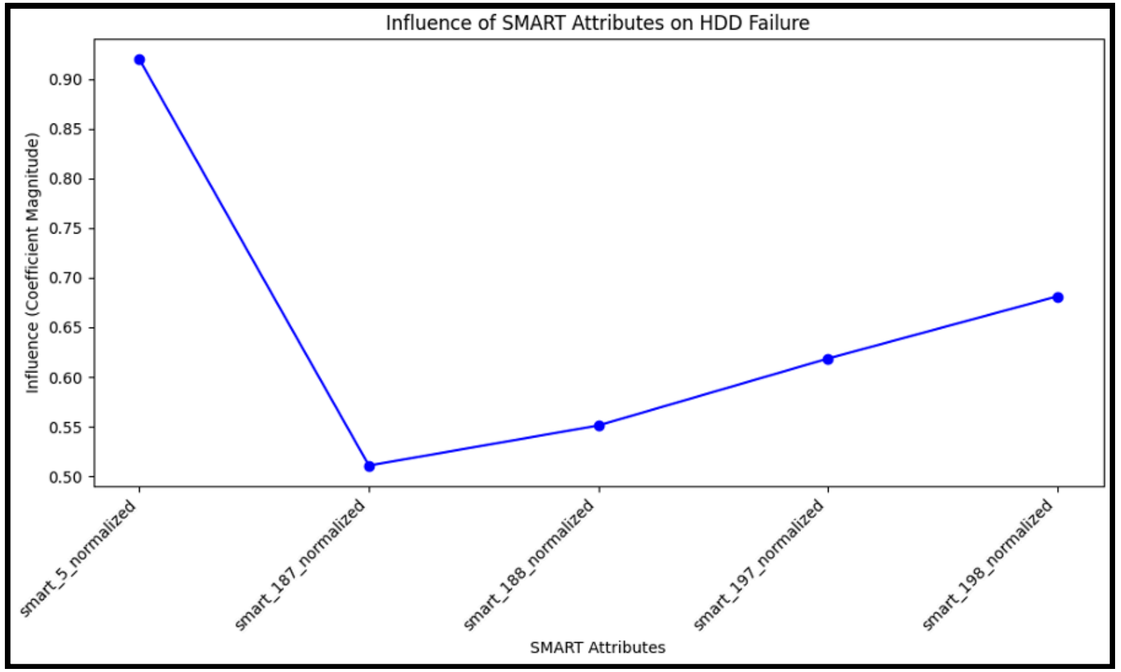
Parameter	ST4000DM000	ST14000NM0138	ST14000NM001G
v	1.3953904	1.90212157	0.11779333
u	1.20138130	0.86652997	0.64586261
a	-0.00348812	-0.00049662	0.00600237
b2	4.04635767e-01	4.55979758e-01	2.84539493e-01
b3	1.23023154e-04	5.02074157e-06	1.27294058e-05
b4	-1.30001196e-01	3.43576572e-06	-1.30049882e-01
b5	-2.25882886e-06	-3.39541682e-06	2.18662073e-04
b6	-2.25882888e-06	-3.39541674e-06	2.18662073e-04
b7	0.50247033	0.49206755	0.49895195

v = Power Coefficient
a = Development Coefficient
bi = Influencing Coefficients

RESULTS

The visual representation





CONCLUSIONS

01

The Weibull Log-Logistic grey forecasting model with multiple interaction effects accurately predicts hard disk drive failure.

02

Comparing 5 attributes we can colclude that SMART 5 (Reallocated Sector Count) attribute influence the most on the failure rate of HDD

03

WLLGM is more adapt at handling hard disk drive failure data comapared to the MLR model

04

The model effectively reduces false positive rate, this leads to fewer unnecessary HDD replacements, saving resources and costs

LIMITATIONS

01 Missing other important failure factors

02 Real time prediction challenges

03 Dependence on smart data

04 Limited to HDD

FUTURE SCOPE

FUTURE SCOPE

1. Extend the model to predict failures in other storage technologies like SSDs and cloud storage, where failure patterns differ.

2. Integrate the Weibull Log-Logistic Grey model with machine learning for hybrid models that enhance prediction accuracy across datasets.

3. Apply the model to other mechanical systems, such as turbines or vehicles, to broaden its use beyond HDD reliability.

Thank You !!!