Log Anomaly Detection and Resolution

—

Srajan Dube, Miles Woollacott

Summer 2022 Intern Project

THE TEAM

SRAJAN DUBE

Data Science Intern

MSCS in Machine Learning at GT

Orlando, FL

Manager: Karunakaran Karuppiah

MILES WOOLLACOTT

Data Science Intern

BS Stats/BACS, University of Virginia

Los Angeles, CA

Manager: Jenny Shafer

MENTORS

Simao Liu

Xiaotong Liu

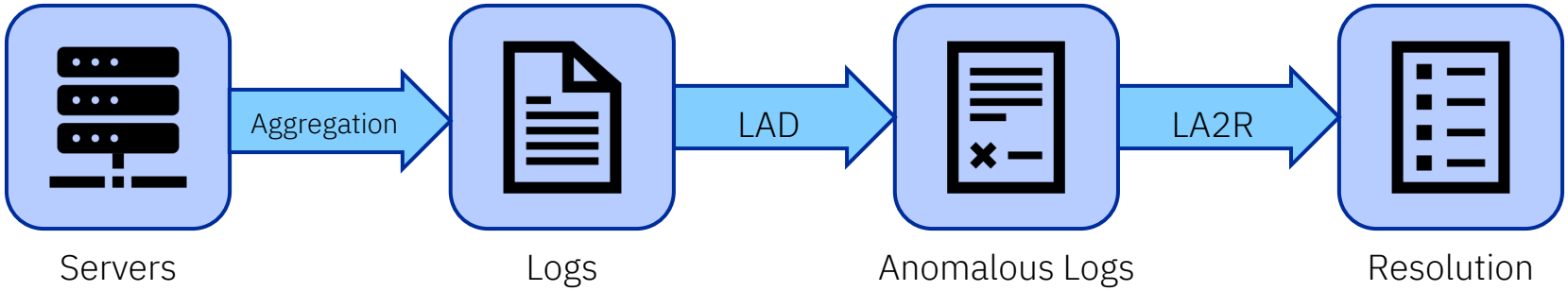
Meenakshi Madugula

Ruchi Mahindru

Andy Tu

Contents

How do we DETECT anomalous logs, and how do we advise the user to RESOLVE them?


Aggregation
LAD
LA2R
Servers
Logs
Anomalous Logs
Resolution

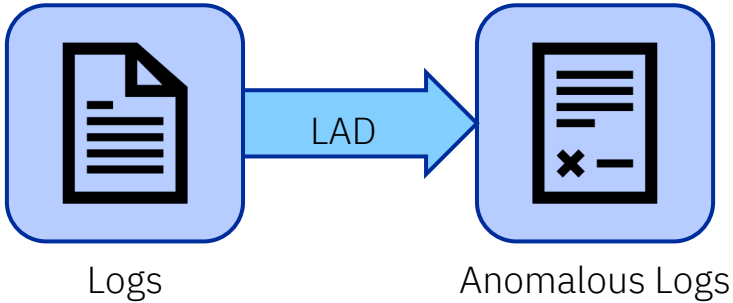
LOG ANOMALY DETECTION WITH OOB

Log Anomaly Detection (LAD) groups logs into windows based on timestamp (inference logs)

GroundTruth identifies which windows in a separate set of windowed logs (reference logs) are anomalous

LAD uses reference to then identify which windows in the inference logs are anomalous

OOB is an efficient method to detect logs



DB2 AGGREGATION

Gathered important information from the raw logs, converted it to a json format using Python

Major fields: “ibm_messageId”, “loglevel”, and “_ts”

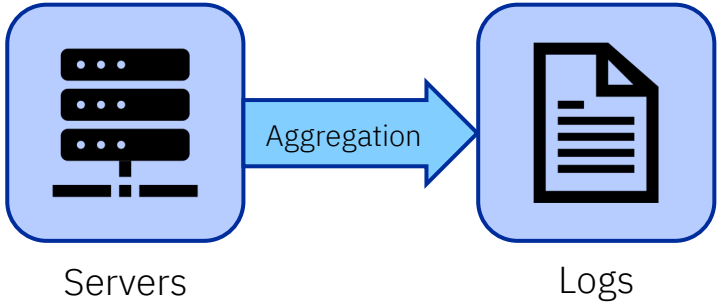
“ibm_messageID” and “loglevel” are used to distinguish anomalous messages

The “_ts” field is a conversion of the “ibm_datetime” field to milliepochs, used to generate time windows

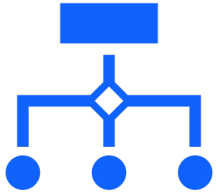
The aggregated logs are sorted by “_ts”

Included “function”, “ibm_serverName”, “host”, “module”, “ibm_datetime”, “recordId”, and “type”

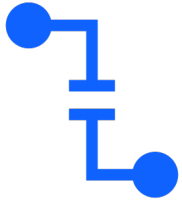
We retained all info from “DATA,” “CHANGE,” “BLU,” “START,” and “FUNCTION” fields



DB2/MQ TESTING



We ran our aggregated DB2 and MQ log files through the pipeline to generate windowed logs and then ran multiple tests to understand the effects of reference/inference logs.



We also slightly adjusted the pipeline to accommodate the differences between WebSphere and DB2/MQ logs.

DB2 RESULTS

```
Normal + abnormal combined :  True
svtdbm7
78 abnormal window
2013 normal window
78 TP
2013 TN
0 FP
0 FN
app accuracy: 1.0
app precision: 1.0
app recall: 1.0
app normal accuracy 1.0
app abnormal accuracy 1.0
78 total TP
2013 total TN
0 total FP
0 total FN
total accuracy: 1.0
total precision: 1.0
total recall: 1.0
F1 score: 1.0
Normal accuracy 1.0
Abnormal accuracy 1.0
```

Dataset 1 full dataset test

```
Normal + abnormal combined :  True
regress1
892 normal window
0 TP
706 TN
186 FP
0 FN
app accuracy: 0.7914798206278026
app precision: 0.0
app recall: None
app normal accuracy 0.7914798206278026
app abnormal accuracy None
0 total TP
706 total TN
186 total FP
0 total FN
total accuracy: 0.7914798206278026
total precision: 0.0
total recall: None
F1 score: None
Normal accuracy 0.7914798206278026
Abnormal accuracy None
```

Dataset 2 full dataset test with no GroundTruth

```
Normal + abnormal combined :  True
regress1
168 abnormal window
726 normal window
168 TP
706 TN
20 FP
0 FN
app accuracy: 0.9776286353467561
app precision: 0.8936170212765957
app recall: 1.0
app normal accuracy 0.9724517906336089
app abnormal accuracy 1.0
168 total TP
706 total TN
20 total FP
0 total FN
total accuracy: 0.9776286353467561
total precision: 0.8936170212765957
total recall: 1.0
F1 score: 0.9438202247191011
Normal accuracy 0.9724517906336089
Abnormal accuracy 1.0
```

Dataset 2 full dataset test with GroundTruth

LOG ANOMALY TO RESOLUTION

Goal: Explore anomalous message codes to provide a resolution

Takes anomalous message codes as input

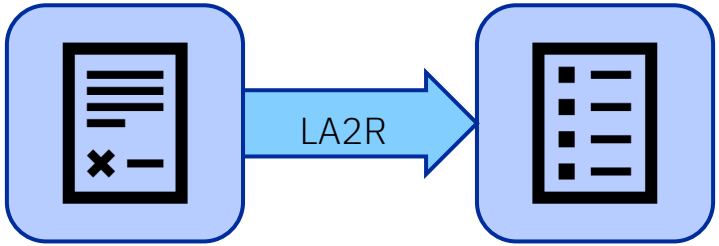
Queries message codes to gather ticket data

Processes/extracts/ aggregates ticket data

Explores and gathers data from urls and checks for url aliveness

Normalizes url data and collapses urls based on document number

Combines ticket and url data into final output



Anomalous Logs

Resolution

Objective: Scalable Creation of OOB Resolution Recommendation Knowledge Base

Improve Code **Stability**

Code **generalizability**

Configurability to handle multiple products

End to End Automation requiring **zero to minimal human touch**

Flexibility in pipeline execution

CODE REFACTORING

Initial Refactoring

- Remove unnecessary code

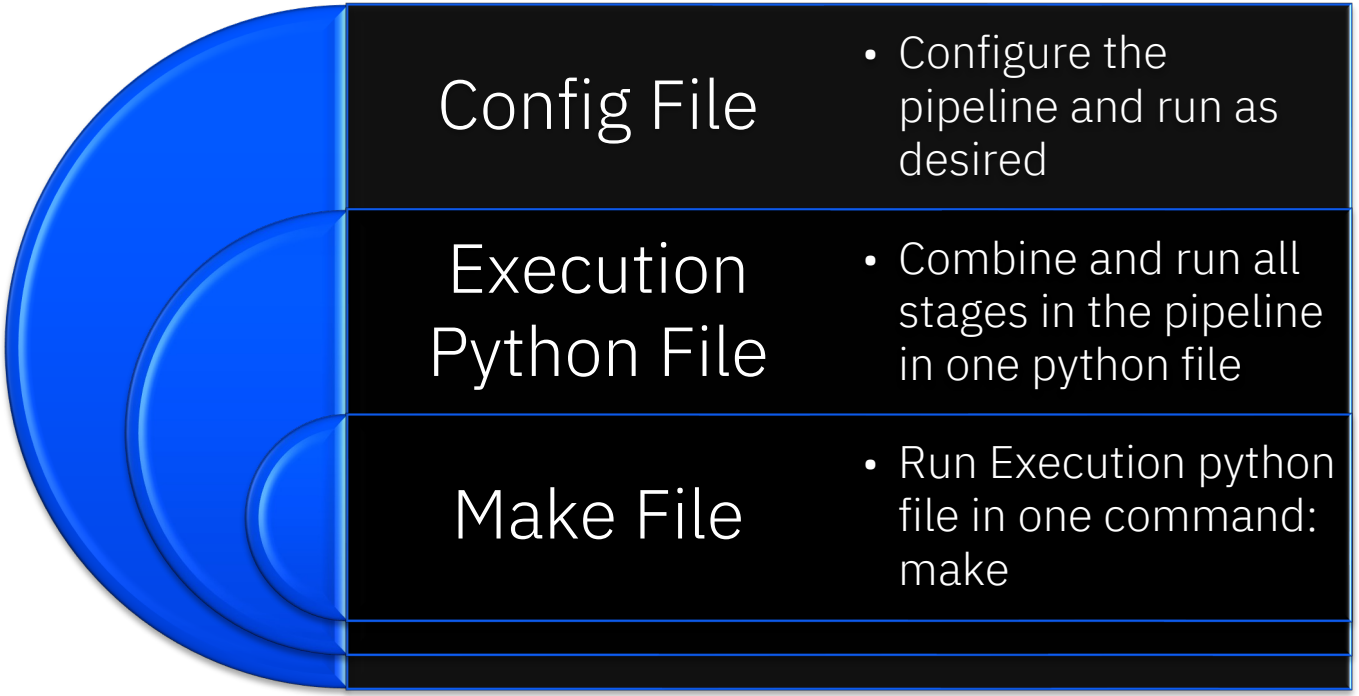
Further Refactoring

- Simplify code and Increase efficiency, generalizability, stability, & usability

Other Refactoring

- Refactor normalize and finalize to use 1 .csv file as input/output rather than 4 .pkl files
- Remove Title Aggregator and Refactor Normalize and Finalize to accommodate for the change

CUSTOMIZABILITY



PREP ADDITIONS

- Converts .dita, .html files into a .csv file with processed information
- Supports file hierarchies of .zip files and folders

CRAWL ADDITIONS

Save state

- Allow Crawl (and Normalize) to be stopped and continued at any point
- Saves partially completed runs

Incremental update

- Allows users to append new results to existing run
- Run Crawl by message type

PROCESS ADDITIONS

- Ticket data can timeout, logs all tickets that timed out
- Gathered more information from ticket data
- Exception handling for deviant ticket formats

NORMALIZE ADDITIONS

Fix normalize so data is actually extracted

Gather additional data from Urls

Filter Url text and title by language

Filter out Landing pages and pdfs

Normalize rows in output

Refactor Normalize

Was this topic helpful?

Yes No

Document Information

More support for:
IBM MQ

Component:
Product Documentation

Software version:
All Version(s)

Operating system(s):
AIX, HP-UX, IBM i, Linux, Solaris, Windows, z/OS

Document number:
708247

Modified date:
03 June 2022

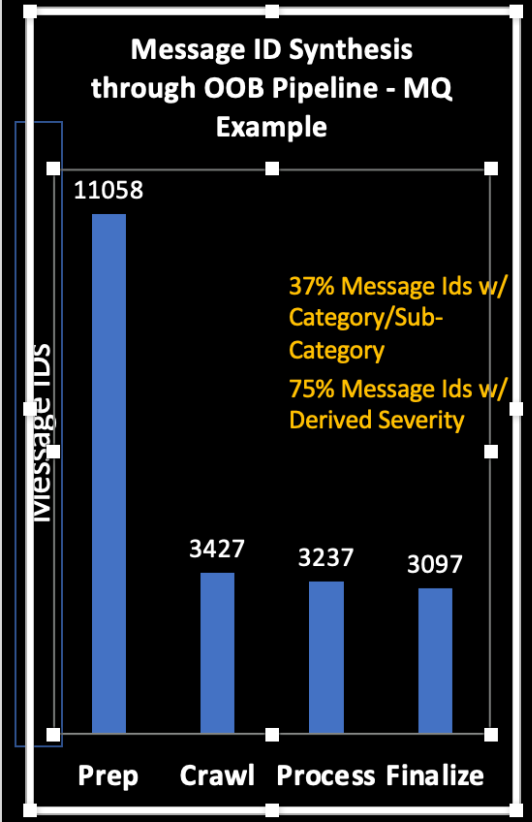
FINALIZE ADDITIONS

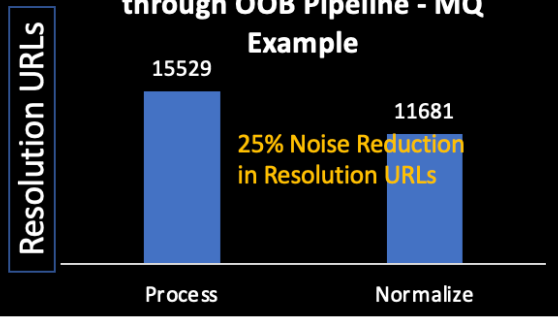
Added additional parameters

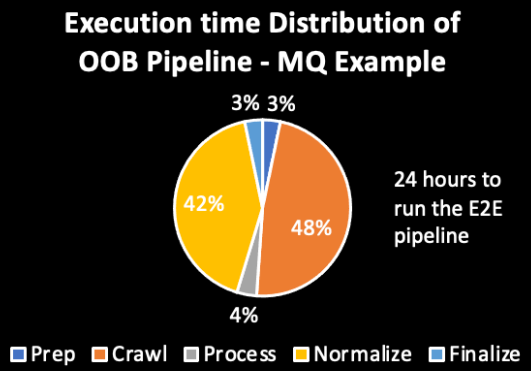
Refactored to use a single csv instead of multiple pkl files

Fixed file formatting

E2E Automated Pipeline for Scalable OOB Resolution Bootstrapping



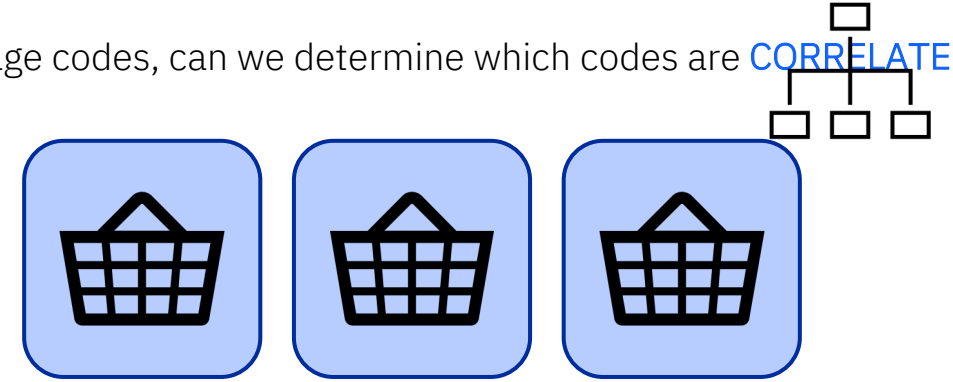




- SDK for Augmented Knowledge Base Creation for OOB Models for IBM Incumbency
  - Configurable pipeline
  - Tested on MQ, and DB2/System Z (partial data)
  - Allow easier Content Updates
  - Proposed Wireframes for User-friendly UI for KB Creation
- Exploiting Additional Metadata:
  - *OS supported*, *Versions Supported*, *Date Modified* (Product Documentation)
  - *Up Votes*, *Down Votes* (Asset Reuse Manager)
  - *Severity*, *Category*, *Sub-Category* (Support Tickets)
- Improved Recommendation Re-ranking
  - 3.4 – max(*#URL*) across tickets + *Content Type* (Technote, Product Documentation etc.)
  - 3.5 - added *Date Modified*, max(*Up Votes*)
- Enhanced Significant Alert Prioritization
  - 3.4 - max(*#Message Code*) across the anomalous window
  - 3.5 – weighted score (derived *Severity* per *Message Code* + *Up Votes* + *Content Source* +

MESSAGE CODE CORRELATIONS

Given a list of lists of message codes, can we determine which codes are CORRELATED with each other?

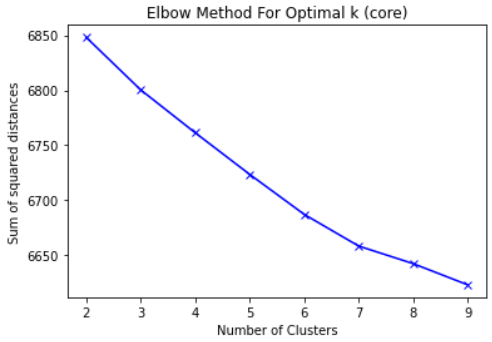


APRIORI ALGORITHM

WebSphere Dataset Example

Index	lhs	rhs	support	confidence	coverage	lift	count
1	{DSRA0304E}	{DSRA0302E}	0.004662	0.868852	0.005366	147.4261	106
2	{DSRA0302E}	{DSRA0304E}	0.004662	0.791045	0.005893	147.4261	106

MESSAGE ID CLUSTERING



	title	cluster
5007	0803 9:39 Queue manager becomes unresponsive i...	1
4629	Queue manager ended unexpectly.	1
4628	Queue manager not starting	1
4166	Probe ID HL206037 in Multi Instance Queue Manager	1
6211	RDQM HA Queue Manager not responding	1
4199	Multi-instance queue manager became unresponsi...	1
5210	IBM MQ queue manager error	1
3141	Problems to downgrade a Queue manager	1
6092	Queue manager has lot of hung processes due to...	1
4627	Queue manager status is not available.	1

Create clusters of message IDs based on ticket titles

Find problem centric / Generic Message IDs to use for prioritization

Used TF-IDF Vectorization along with Kmeans Clustering to create the clusters

NORMALIZE OUTPUT SIMILARITY MATRICES

Take Title and Text columns from Normalize output

Clean them up by removing punctuation, stop words (“the”, “and”, “as”, etc.), and making them lowercase

Vectorize the text and title

Use cosine similarity to generate similarity matrices (text, title, and text + title)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB
1	1	1	1	0.042346	0	0	0	0	0	0.048094	0	0	0	0	0.098573	0	0	0	0.047727	0	0	0	0	0	0	0	0	0
2	1	1	1	0.042346	0	0	0	0	0	0.048094	0	0	0	0	0.098573	0	0	0	0.047727	0	0	0	0	0	0	0	0	0
3	1	1	1	0.042346	0	0	0	0	0	0.048094	0	0	0	0	0.098573	0	0	0	0.047727	0	0	0	0	0	0	0	0	0
4	.042346	0.042346	0.042346	1	0.175526	0.208389	0.20684	0.177522	0.177522	0.045484	0.080379	0.292615	0.017288	0.036573	0.052177	0.106232	0.262685	0.025665	0.045137	0.156758	0.312536	0.252458	0.225993	0.119872	0.0344	0.031563	0.089616	0.253891
5	0	0	0	0.175526	1	0.949249	0.929602	0.985523	0.985523	0.183947	0.235772	0.320571	0.115378	0.214796	0.208009	0.317538	0.246923	0.112849	0.182545	0.111339	0.237343	0.095046	0.120391	0.049232	0.203349	0.04213	0.069049	0.197413
6	0	0	0	0.208389	0.949249	1	0.98494	0.963183	0.963183	0.170688	0.228926	0.352407	0.089622	0.19908	0.194756	0.305389	0.294098	0.117399	0.169387	0.137928	0.300452	0.089235	0.132657	0.050193	0.197444	0.055224	0.082152	0.245991
7	0	0	0	0.20684	0.929602	0.98494	1	0.942267	0.942267	0.153091	0.210921	0.336307	0.044811	0.1659	0.171313	0.284883	0.287704	0.127182	0.151924	0.137928	0.298878	0.084659	0.124776	0.052583	0.197444	0.051133	0.082152	0.245991
8	0	0	0	0.177522	0.985523	0.963183	0.942267	1	1	0.186039	0.241133	0.324217	0.11669	0.217239	0.210375	0.32115	0.249732	0.114132	0.184621	0.112605	0.240042	0.096127	0.121761	0.049792	0.205661	0.042609	0.069834	0.199658
9	0	0	0	0.177522	0.985523	0.963183	0.942267	1	1	0.186039	0.241133	0.324217	0.11669	0.217239	0.210375	0.32115	0.249732	0.114132	0.184621	0.112605	0.240042	0.096127	0.121761	0.049792	0.205661	0.042609	0.069834	0.199658
10	.048094	0.048094	0.048094	0.045484	0.183947	0.170688	0.153091	0.186039	0.186039	1	0.268233	0.16068	0.196343	0.33749	0.957633	0.253498	0.022411	0.000857	0.967129	0.057167	0.018609	0.057479	0.078267	0.002095	0	0.019716	0.033099	0.025196
11	0	0	0	0.080379	0.235772	0.228926	0.210921	0.241133	0.241133	0.268233	1	0.284142	0.344407	0.55101	0.323393	0.46624	0.057328	0.046368	0.266189	0.045364	0.089668	0.136777	0.164881	0.113282	0.044873	0.09432	0.079833	0.038876
12	0	0	0	0.292615	0.320571	0.352407	0.336307	0.324217	0.324217	0.16068	0.284142	1	0.179641	0.283973	0.199203	0.411023	0.402966	0.010459	0.159455	0.32296	0.560543	0.107353	0.188383	0.052167	0.017022	0.134835	0.217876	0.462476
13	0	0	0	0.017288	0.115378	0.089622	0.044811	0.11669	0.11669	0.196343	0.344407	0.179641	1	0.37022	0.261574	0.245145	0.035669	0	0.194846	0	0.017552	0.051061	0.087932	0	0	0.045644	0	0
14	0	0	0	0.036573	0.214796	0.19908	0.1659	0.217239	0.217239	0.33749	0.55101	0.283973	0.37022	1	0.398002	0.493549	0.030184	0.003464	0.330795	0.073696	0.067766	0.108021	0.179822	0.018336	0.003759	0.045866	0.08916	0.052788
15	.098573	0.098573	0.098573	0.052177	0.208009	0.194756	0.171313	0.210375	0.210375	0.957633	0.323393	0.199203	0.261574	0.398002	1	0.310424	0.033014	0.001757	0.939357	0.053563	0.022603	0.072603	0.095541	0.004293	0	0.027552	0.034768	0.037296
16	0	0	0	0.106232	0.317538	0.305389	0.284883	0.32115	0.32115	0.253498	0.46624	0.411023	0.245145	0.493549	0.310424	1	0.229677	0.072074	0.253476	0.210052	0.185306	0.116271	0.151371	0.092401	0.053427	0.092498	0.240383	0.171274
17	0	0	0	0.262685	0.246923	0.294098	0.287704	0.249732	0.249732	0.022411	0.057328	0.402966	0.035669	0.030184	0.033014	0.229677	1	0.004672	0.02224	0.219577	0.419459	0.044925	0.13382	0.03044	0.010139	0.553537	0.162352	0.342022
18	0	0	0	0.025665	0.112849	0.117399	0.127182	0.114132	0.114132	0.000857	0.046368	0.010459	0	0.003464	0.001757	0.072074	0.004672	1	0.000851	0.011805	0.032955	0.035672	0.008959	0.069869	0.062062	0.025909	0.015642	0.010117
19	.047727	0.047727	0.047727	0.045137	0.182545	0.169387	0.151924	0.184621	0.184621	0.967129	0.266189	0.159455	0.194846	0.330795	0.939357	0.253476	0.02224	0.000851	1	0.056732	0.018468	0.057041	0.07767	0.002079	0	0.019566	0.032847	0.025004
20	0	0	0	0.156758	0.111339	0.137928	0.137928	0.112605	0.112605	0.057167	0.045364	0.32296	0	0.073696	0.053563	0.210052	0.219577	0.011805	0.056732	1	0.480383	0.039645	0.08656	0.045481	0.01478	0.08923	0.59339	0.509302
21	0	0	0	0.312536	0.237343	0.300452	0.298878	0.240042	0.240042	0.018609	0.089668	0.560543	0.017552	0.067766	0.022603	0.185306	0.419459	0.032955	0.018468	0.480383	1	0.060943	0.167713	0.063661	0.032431	0.118568	0.436439	0.77269
22	0	0	0	0.252458	0.095046	0.089235	0.084659	0.096127	0.096127	0.057479	0.136777	0.107353	0.051061	0.108021	0.072603	0.116271	0.044925	0.035672	0.057041	0.039645	0.060943	1	0.624586	0.093506	0.019353	0.052827	0.053799	0.045503
23	0	0	0	0.225993	0.120391	0.132657	0.124776	0.121761	0.121761	0.078267	0.164881	0.188383	0.087932	0.179822	0.095541	0.151371	0.13382	0.008959	0.07767	0.08656	0.167713	0.624586	1	0.040648	0.06249	0.037459	0.088941	0.141049
24	0	0	0	0.119872	0.049232	0.050193	0.052583	0.049792	0.049792	0.002095	0.113282	0.052167	0	0.018336	0.004293	0.092401	0.03044	0.069869	0.002079	0.045481	0.063661	0.093506	0.040648	1	0.079602	0.075471	0.058446	0.038026
25	0	0	0	0.0344	0.203349	0.197444	0.197444	0.205661	0.205661	0	0.044873	0.017022	0	0.003759	0	0.053427	0.010139	0.062062	0	0.01478	0.032431	0.019353	0.06249	0.079602	1	0.02595	0.017971	0.043065
26	0	0	0	0.031563	0.04213	0.055224	0.051133	0.042609	0.042609	0.019716	0.09432	0.134835	0.045644	0.045866	0.027552	0.092498	0.553537	0.025909	0.019566	0.08923	0.118568	0.052827	0.037459	0.075471	0.02595	1	0.105801	0.112264
27	0	0	0	0.089616	0.069049	0.082152	0.082152	0.069834	0.069834	0.033099	0.079833	0.217876	0	0.08916	0.034768	0.240383	0.162352	0.015642	0.032847	0.59339	0.436439	0.053799	0.088941	0.058446	0.017971	0.105801	1	0.508531
28	0	0	0	0.253891	0.197413	0.245991	0.245991	0.199658	0.199658	0.025196	0.038876	0.462476	0	0.052788	0.037296	0.171274	0.342022	0.010117	0.025004	0.509302	0.77269	0.045503	0.141049	0.038026	0.043065	0.112264	0.508531	1

GroundTruth Definition

VS.

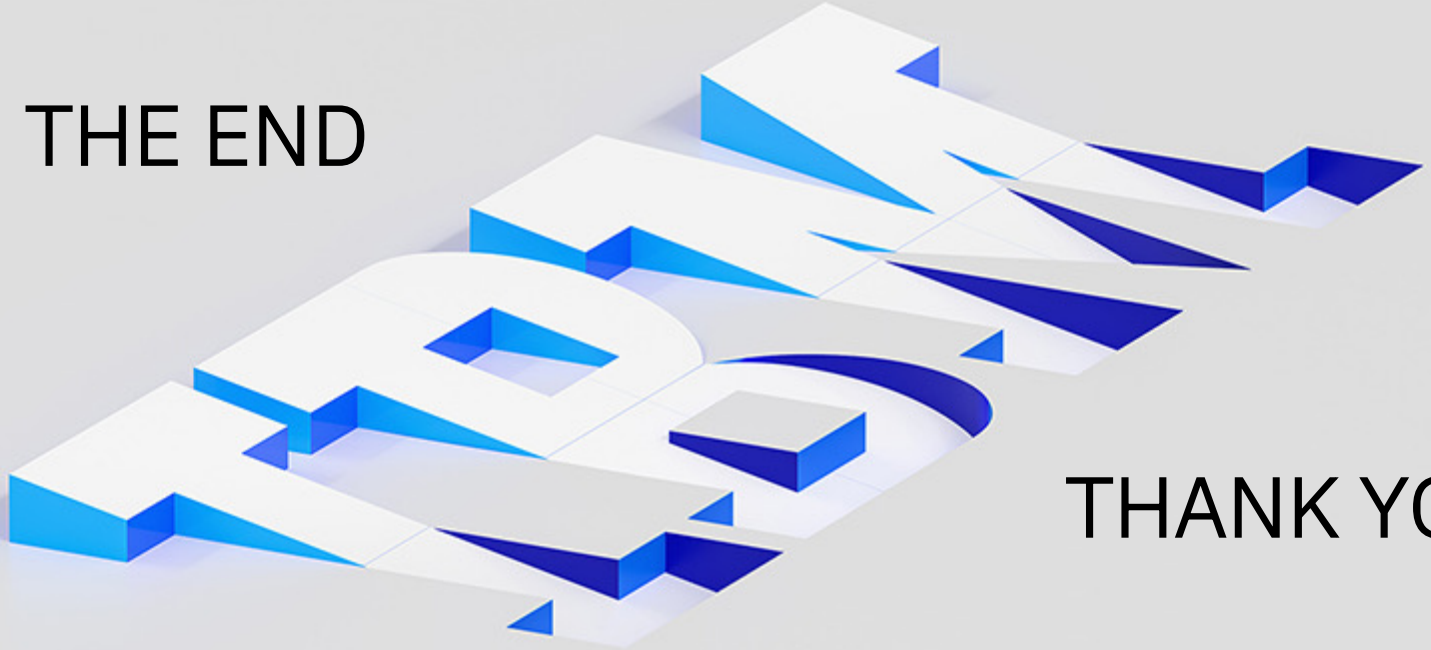
MESSAGE CODE CATEGORY PREDICTION

Could we **PREDICT** an erroneous log, given the current message codes returned?

Determine lagged correlations between message codes

Resolve serious issues before they arise

THE END



THANK YOU!