

|| 산업 빅데이터 분석 실제 - 프로젝트 최종 발표 ||

대규모 그리드 컴퓨팅 미들웨어의 에러 로그 탐지 시스템

통합과정 4학기 유현진

2024.12.09(Mon)

1. 프로젝트 문제 정의

❖ 대규모 시스템에서의 방대한 로그 데이터 중, 머신러닝을 통해 이상 탐지 시스템을 개발하는 것

- 목표 : 정상 / 에러 로그를 구분
- 입력 데이터 : 로그 텍스트 → 미들웨어 구동하면서 출력될 로그가 소스코드에 명시되어 있음
(<https://gitlab.cern.ch/jalien/jalien>)
- 출력 결과 : 1(정상) / 0(에러)

정적 로그

```
logger.log(Level.WARNING, "Error encountered in the JobWrapper process");
```

```
logger.log(Level.WARNING, "Error encountered while running job: ", ex);
```

```
logger.log(Level.SEVERE, "Cannot clean up the job dir even using system commands");
```

```
logger.log(Level.INFO, "Error. Workdir for job could not be created");
```

동적 로그

```
logger.log(Level.SEVERE, "IsolCmd command" + isolCmd);
```

```
logger.log(Level.SEVERE, "Exception running " + launchCommand + " : ", ioe);
```

```
logger.log(Level.WARNING, "Exception parsing a line of output from pgrep: " + readArg, nfe);
```

```
logger.log(Level.INFO, "Exception running " + cmd + " : " + ioe.getMessage());
```

2. 학습 데이터 준비

❖ 1) 데이터 수집

- 실제 운영중인 시스템의 로그 데이터 수집 (운영 시스템 = 글로벌 그리드 시스템)
[KISTI 운영(한국), CERN 사용(스위스)]

시스템의 로그 위치

```
[gsdc24a01@alice-kisti-hpc alien-logs]$ ll
total 33920
drwxr-x---. 2 gsdc24a01 in0190 20480 Dec 6 00:37 2024-12-02
drwxr-x---. 2 gsdc24a01 in0190 61440 Dec 7 00:37 2024-12-03
drwxr-x---. 2 gsdc24a01 in0190 151552 Dec 8 00:37 2024-12-04
drwxr-x---. 2 gsdc24a01 in0190 155648 Dec 9 00:37 2024-12-05
drwxr-x---. 2 gsdc24a01 in0190 106496 Dec 9 10:37 2024-12-06
drwxr-x---. 2 gsdc24a01 in0190 57344 Dec 8 01:37 2024-12-07
drwxr-x---. 2 gsdc24a01 in0190 126976 Dec 9 01:37 2024-12-08
drwxr-x---. 2 gsdc24a01 in0190 69632 Dec 9 10:39 2024-12-09
-rw-r-----. 1 gsdc24a01 in0190 20767 Nov 8 16:21 access_log
-rw-r-----. 1 gsdc24a01 in0190 652 Nov 8 16:21 error_log
```

```
[gsdc24a01@alice-kisti-hpc 2024-12-02]$ ll
total 668
-rw-r-----. 1 gsdc24a01 in0190 3264 Dec 2 00:20 jobagent_40060_000217.err.gz
-rw-r-----. 1 gsdc24a01 in0190 2714 Dec 2 00:20 jobagent_40060_000217.out.gz
-rw-r-----. 1 gsdc24a01 in0190 3062 Dec 2 00:18 jobagent_40060_000217.sh.gz
-rw-r-----. 1 gsdc24a01 in0190 3321 Dec 2 00:24 jobagent_40060_000218.err.gz
-rw-r-----. 1 gsdc24a01 in0190 2714 Dec 2 00:24 jobagent_40060_000218.out.gz
-rw-r-----. 1 gsdc24a01 in0190 3061 Dec 2 00:22 jobagent_40060_000218.sh.gz
-rw-r-----. 1 gsdc24a01 in0190 3292 Dec 2 01:56 jobagent_40060_000219.err.gz
-rw-r-----. 1 gsdc24a01 in0190 2714 Dec 2 01:56 jobagent_40060_000219.out.gz
-rw-r-----. 1 gsdc24a01 in0190 3061 Dec 2 01:54 jobagent_40060_000219.sh.gz
```

로그 파일

```
[gsdc24a01@alice-kisti-hpc 2024-12-02]$ zcat jobagent_40060_000217.err.gz | head
JAlieN/1.9.5-1(10):ERROR:105: Unable to locate a modulefile for 'BASE/1.0'
(10):ERROR:105: Unable to locate a modulefile for 'JDK/12.0.1-5'
(10):ERROR:105: Unable to locate a modulefile for 'XRootD/v5.7.0-16'
JAlieN/1.9.5-1(12):ERROR:102: Tcl command execution failed: set JALIEN_ROOT $::env(BASEDI
R)/JAlieN/$version
```

```
XRootD/v5.7.0-16(11):ERROR:105: Unable to locate a modulefile for 'BASE/1.0'
XRootD/v5.7.0-16(12):ERROR:105: Unable to locate a modulefile for 'libffi/v3.2.1-alice1-2'
XRootD/v5.7.0-16(13):ERROR:105: Unable to locate a modulefile for 'zlib/v1.2.8-12'
XRootD/v5.7.0-16(14):ERROR:105: Unable to locate a modulefile for 'FreeType/v2.10.1-15'
hdf5(15):ERROR:105: Unable to locate a modulefile for 'hdf5/1.10.9-4'
```

```
[gsdc24a01@alice-kisti-hpc 2024-12-02]$ zcat jobagent_40060_000217.err.gz | tail
INFO: Connecting to JCentral on alice-jcentral.cern.ch:8098
Dec 02, 2024 12:20:23 AM alien.api.DispatchSSLClient getInstance
INFO: Connecting with client cert: C=ch,O=AliEn2,CN=JobAgent
Dec 02, 2024 12:20:24 AM alien.api.DispatchSSLClient connectTo
INFO: Printing peer's information:
Dec 02, 2024 12:20:24 AM alien.api.DispatchSSLClient connectTo
INFO: Peer's Certificate Information:
INFO
Dec 02, 2024 12:20:25 AM alien.site.JobRunner run
INFO: JobRunner going to exit from lack of jobs
[gsdc24a01@alice-kisti-hpc 2024-12-02]$
```

2. 학습 데이터 준비

❖ 2) 데이터 전처리

- 수집한 로그 데이터에 CSV 데이터로 변환 & 라벨링 진행 (1 = 정상 / 0 = 에러)

로그 파일

```
Nov 01, 2024 10:23:10 AM alien.api.DispatchSSLClient checkIdleConnection
INFO: Closing idle socket
Nov 01, 2024 10:23:21 AM alien.site.JobAgent computeTimeLeft
OFF: Still have 76725 seconds to live (1714526601850-1714516326732=10275)
Nov 01, 2024 10:23:21 AM alien.site.JobAgent computeTimeLeft
OFF: Certificate timeleft is 162435
Nov 01, 2024 10:23:23 AM alien.site.JobAgent computeTimeLeft
INFO: Still have 76723 seconds to live (1714526603919-1714516326732=10277)
Nov 01, 2024 10:23:23 AM alien.site.JobAgent computeTimeLeft
INFO: Certificate timeleft is 162433
Nov 01, 2024 10:23:23 AM alien.site.JobAgent recomputeDiskSpace
INFO: Recomputing disk space of /scratch/gsd23a01. Starting with a free space of
7437841515282432
Nov 01, 2024 10:23:23 AM alien.site.JobAgent recomputeDiskSpace
INFO: WorkdirSize=10702722, maxSize=5242880000, recomputedDisk=7437836283105154
Nov 01, 2024 10:23:23 AM alien.site.JobRunner run
INFO: No new thread
Nov 01, 2024 10:23:41 AM alien.site.JobAgent computeTimeLeft
OFF: Still have 76705 seconds to live (1714526621982-1714516326732=10295)
Nov 01, 2024 10:23:41 AM alien.site.JobAgent computeTimeLeft
OFF: Certificate timeleft is 162415
JobID 3068977211: Nov 01, 2024 03:24:15 alien.api.DispatchSSLClient getInstance
INFO: Connecting to JCentral on alice-jcentral.cern.ch:8098
JobID 3068977211: Nov 01, 2024 03:24:15 alien.api.DispatchSSLClient getInstance
INFO: Connecting with client cert:
C=ch,O=AliEn2,CN=Jobs,CN=alitrain,OU=alitrain,OU=queueid\=3068977211/resubmission\=0
JobID 3068977211: Nov 01, 2024 03:24:16 alien.api.DispatchSSLClient connectTo
INFO: Printing peer's information:
JobID 3068977211: Nov 01, 2024 03:24:16 alien.api.DispatchSSLClient connectTo
INFO: Peer's Certificate Information:
INFO
JobID 3068977211: Nov 01, 2024 03:24:18 alien.api.DispatchSSLClient dispatchRequest
SEVERE: Exception connecting at attempt no. 0
java.io.EOFException
    at java.base/java.io.ObjectInputStream
$BlockDataInputStream.peekByte(ObjectInputStream.java:3214)
    at java.base/java.io.ObjectInputStream.readObject0(ObjectInputStream.java:1684)
    at java.base/java.io.ObjectInputStream.readObject(ObjectInputStream.java:509)
    at java.base/java.io.ObjectInputStream.readObject(ObjectInputStream.java:467)
    at alien.api.DispatchSSLClient.dispatchRequest(DispatchSSLClient.java:544)
```



logs.csv 파일

20241101	Nov 01, 2024 10:23:10 AM alien.api.DispatchSSLClient checkIdleConnection	1
20241101	INFO: Closing idle socket	1
20241101	Nov 01, 2024 10:23:21 AM alien.site.JobAgent computeTimeLeft	1
20241101	INFO: Still have 76725 seconds to live (1714526601850-1714516326732=10275)	1
20241101	Nov 01, 2024 10:23:21 AM alien.site.JobAgent computeTimeLeft	1
20241101	INFO: Certificate timeleft is 162435	1
20241101	Nov 01, 2024 10:23:23 AM alien.site.JobAgent computeTimeLeft	1
20241101	INFO: Still have 76723 seconds to live (1714526603919-1714516326732=10277)	1
20241101	Nov 01, 2024 10:23:23 AM alien.site.JobAgent computeTimeLeft	1
20241101	INFO: Certificate timeleft is 162433	1
20241101	Nov 01, 2024 10:23:23 AM alien.site.JobAgent recomputeDiskSpace	1
20241101	INFO: Recomputing disk space of /scratch/gsd23a01. Starting with a free spa	1
20241101	Nov 01, 2024 10:23:23 AM alien.site.JobAgent recomputeDiskSpace	1
20241101	INFO: WorkdirSize=10702722, maxSize=5242880000, recomputedDisk=7437836	1
20241101	Nov 01, 2024 10:23:23 AM alien.site.JobRunner run	1
20241101	INFO: No new thread	1
20241101	Nov 01, 2024 10:23:41 AM alien.site.JobAgent computeTimeLeft	1
20241101	INFO: Still have 76705 seconds to live (1714526621982-1714516326732=10295)	1
20241101	Nov 01, 2024 10:23:41 AM alien.site.JobAgent computeTimeLeft	1
20241101	INFO: Certificate timeleft is 162415	1
20241101	JobID 3068977211: Nov 01, 2024 03:24:15 alien.api.DispatchSSLClient getInstan	1
20241101	INFO: Connecting to JCentral on alice-jcentral.cern.ch:8098	1
20241101	JobID 3068977211: Nov 01, 2024 03:24:15 alien.api.DispatchSSLClient getInstan	1
20241101	INFO: Connecting with client cert: C=ch,O=AliEn2,CN=Jobs,CN=alitrain,OU=al	1
20241101	JobID 3068977211: Nov 01, 2024 03:24:16 alien.api.DispatchSSLClient connectT	1
20241101	INFO: Printing peer's information:	1
20241101	JobID 3068977211: Nov 01, 2024 03:24:16 alien.api.DispatchSSLClient connectT	1
20241101	INFO: Peer's Certificate Information:	1
20241101	INFO	1
20241101	JobID 3068977211: Nov 01, 2024 03:24:18 alien.api.DispatchSSLClient dispatchR	1
20241101	SEVERE: Exception connecting at attempt no. 0	0
20241101	java.io.EOFException	0
20241101	at java.base/java.io.ObjectInputStream\$BlockDataInputStream.peekByte(Object	0
20241101	at java.base/java.io.ObjectInputStream.readObject0(ObjectInputStream.java:168	0
20241101	at java.base/java.io.ObjectInputStream.readObject(ObjectInputStream.java:509)	0
20241101	at java.base/java.io.ObjectInputStream.readObject(ObjectInputStream.java:467)	0
20241101	at alien.api.DispatchSSLClient.dispatchRequest(DispatchSSLClient.java:544)	0

2. 학습 데이터 준비

❖ 3) 로그 데이터 추가 수집

- " 운영 시스템에서 추출한 로그로는 학습할 에러 로그의 개수가 적음 " (정상 90%, 에러 10%)

- 미들웨어 소스코드에서 샘플 에러 로그를 추가 수집

그리드 컴퓨팅 환경에서 자주 사용하는 소스코드를 중점으로 추출
([jalien/src/main/java/alien/site](#) 디렉터리)

master ▾

jalien / src / main / java / alien / site

batchqueue	Include queued in getNumberActive() for SLURM	2 months ago
containers	Remote logging	5 months ago
packman	Remove Singularity remnants from CVMFS	2 months ago
CgroupUtils.java	Catch IOE directly in method	3 weeks ago
ComputingElement.java	Revert "Provide stracktrace on unknown exceptions"	2 months ago
Functions.java	Remove unused imports	1 year ago
JobAgent.java	Do not allow custom kill timeout above 10m	2 days ago
JobRunner.java	Remove duplicate timestamp	1 week ago
JobWrapper.java	Always include the kill timeout	1 week ago
KillReason.java	Add the JobBroker killing reason	1 month ago
MachineJobFeatures.java	Fix reading back content from files; more robust number parsing	2 years ago
MemoryController.java	Cleanup before tag	3 weeks ago
NUMAExplorer.java	More Eclipse warnings fixed	5 months ago
OutputEntry.java	Stop parsing the options, they are handed by 'cp' now	1 year ago
ParsedOutput.java	Revert parsing	1 year ago
ResourceManager.java	Moved ResourceManager to using primitives instead of objects	5 months ago
SiteMap.java	Avoid duplicate sites in advertising close locations	3 months ago
SiteStatusDTO.java	Remove non-ASCII to prevent possible build errors	6 months ago



```
// TODO: Hack to exclude alihyperloop jobs from nodes without avx support. Remove me soon!
try {
    if (!Files.readString(Paths.get("/proc/cpuinfo")).contains("avx") && (System.getProperty("os.arch").contains(
        final ArrayList<String> noUsers = (ArrayList<String>) siteMap.computeIfAbsent("NoUsers", (k) -> new A
    if (noUsers.contains("alihyperloop"))
        ((ArrayList<String>) currentSiteMap.get("NoUsers")).add("alihyperloop");
    if (noUsers.contains("alitrain"))
        ((ArrayList<String>) currentSiteMap.get("NoUsers")).add("alitrain");
    if (noUsers.contains("aliprod"))
        ((ArrayList<String>) currentSiteMap.get("NoUsers")).add("aliprod");
    logger.log(Level.WARNING, "This host appears to be missing AVX support, and will be blocked from runn
}
} catch (IOException | NullPointerException ex) {
    logger.log(Level.WARNING, "Unable to check for AVX support", ex);
}

containerizer = ContainerizerFactory.getContainerizer();

// Verify environment if there are no containers, before matching
if (containerizer == null && env.getOrDefault("DISABLE_CONTAINER_ENFORCE", "").toLowerCase().contains("true")) {
    logger.log(Level.SEVERE, "This host does not appear to support containers. Please verify that user namespace
    throw new EufException("Job matching aborted due to missing containers");
}

// Verify that apmon is not null
if (apmon == null) {
    logger.log(Level.SEVERE, "Apmon is null. Please verify that connection and DNS settings are correct");
    throw new EufException("Job matching aborted due to null Apmon");
}

// Check for matching overrides
if (env.containsKey("Disk"))
    currentSiteMap.put("Disk", Long.valueOf(Long.parseLong(env.get("Disk").toString())));
if (env.containsKey("Memory"))
    currentSiteMap.put("Memory", Double.valueOf(Double.parseDouble(env.get("Memory").toString())));

logger.log(Level.INFO, "Attempting to match job using the following SiteMap: ");
logger.log(Level.INFO, currentSiteMap.toString());

setStatus(jaStatus.REQUESTING_JOB);

// Broker expects disk space to be expressed in KB (JOBAGENT table content)
final GetMatchJob jobMatch = commander.q.api.getMatchJob(new HashMap<>{currentSiteMap});

matchedJob = jobMatch.getMatchJob();
```

수집 기준

- 각각의 소스코드를 직접 확인
- logger.log 라인 중,
 - Level.SEVERE
 - Level.WARNING
 - Level.INFO 중 에러 메시지에 해당

2. 학습 데이터 준비

❖ 4) 추가 수집 데이터 전처리

최종적으로 정상 60%, 에러 40%

- 동적 로그가 다수 존재하여, 임의의 데이터 삽입하는 과정 진행
 - <> = 5 ~ 8자리 숫자(num) / [] = 임의의 파일 경로(path) / [[]] = 10 ~ 30자리의 문자열(str)
- 'INFO 레벨 + 에러 로그' 인 경우를 잘 학습시키기 위해, 해당 로그의 비율을 높임 (실제로 사람이 자주 놓치는 케이스)

```
"SEVERE: Could not fetch machine's idle CPU"
"WARNINg: JobRunner main loop caught an exception"
"SEVERE: Could not get the node's memory to set up the oversubscription pool"
"WARNINg: Error creating top cgroup:"
"WARNINg: Exception moving <> to cgroup <>"
"WARNINg: Exception moving <> to cgroup <>"
"WARNINg: Exception moving <> to cgroup <>"
"WARNINg: Exception setting low memory <> to cgroup <>"
"WARNINg: Exception setting low memory <> to cgroup <>"
"WARNINg: Exception setting max memory <> to pid <>"
"WARNINg: Exception killing process inferred from []"
"WARNINg: Exception cleaning cgroups of pid <>"
"WARNINg: Exception getting the CPU cores in cgroup []"
"WARNINg: Exception assigning CPU cores [[]] to cgroup []"
"WARNINg: Exception setting CPU quota <> to cgroup []"
"SEVERE: Problem in construction of ComputingElement: "
"SEVERE: Unable to fetch config for \"host_host\""
"WARNINg: Could not create a pidfile in: /scratch/gsd24a01/ALICE/alien-log/CE.pid"
"SEVERE: Exception reloading the identity: [[]]"
"WARNINg: Error syncing with LDAP: [[]]"
"SEVERE: Exception checking the identity: [[]]"
"SEVERE: Can't send parameter to ML (getNumberFreeSlots): [[]]"
"INFO: Error creating Agent Startup file: /scratch/gsd24a01/ALICE/alien-log/2024-11-02/jobagent-<->.sh"
"INFO: Agent Startup file not found: /scratch/gsd24a01/ALICE/alien-log/2024-11-02/jobagent-<->.sh"
"INFO: Getting JobAgent TokenCertificate failed: [[]]"
"WARNINg: Exception writing token to"
"WARNINg: Getting JobAgent TokenCertificate failed"
"SEVERE: Cannot find class for type: "
"SEVERE: Cannot find class for ceConfig:"
"SEVERE: Cannot instantiate queue class for type: PBS.java"
"SEVERE: Could not obtain ALIEn jar path:"
"WARNINg: Problem with getting logger:"
"WARNINg: Problem with the monitoring objects IO Exception: "
"WARNINg: Problem with the monitoring objects ApMon Exception: "
"WARNINg: Unknown exception: "
"WARNINg: Unable to check for AVX support"
"WARNINg: Another exception matching the job"
"ERROR: Unable to handle job"
"SEVERE: Could not generate JobWrapper launch command "
"SEVERE: Exception running [[]] :"
```

추가 로그 데이터

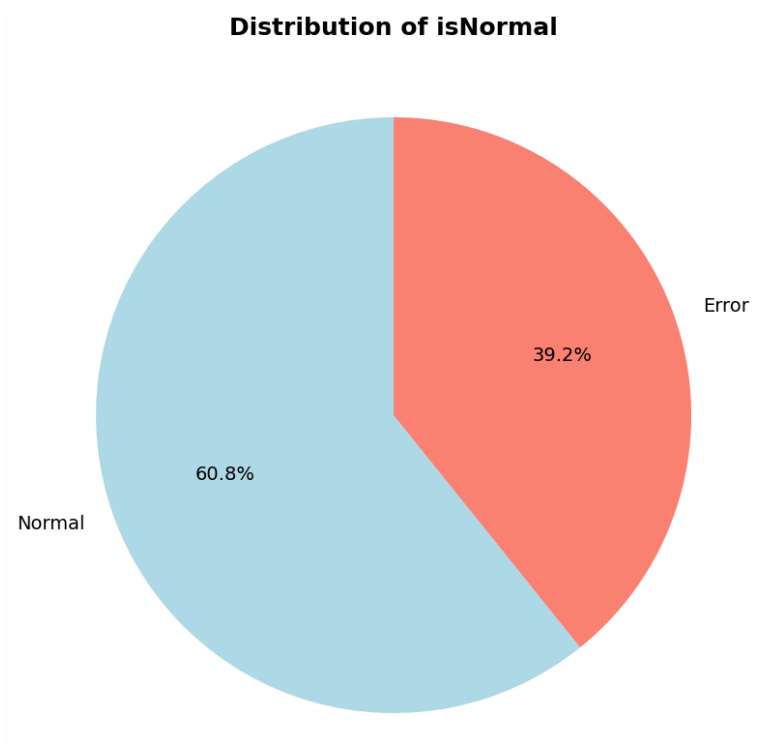
```
20241104 "WARNING: Unknown exception: " 0
20241104 "INFO: Did not get any OOM error from dmesg" 0
20241104 "WARNING: Data list format not understood: [/abs/path/to/random/file_pcqwc]" 0
20241104 "INFO: Unable to kill the JobWrapper" 0
20241104 "SEVERE: Error. Could not receive data from JobAgent: " 0
20241104 "INFO: .alienValidation.trace does not exist." 0
20241104 "INFO: There is not enough space left: 81633530MB" 0
20241104 "INFO: Can not set NUMA architecture aware cpulsoation" 0
20241104 "WARNING: Invalid containerizer: " 0
20241104 "INFO: Could not parse memory limit from proc file" 0
20241104 "WARNING: Exception saving content to file" 0
20241104 "INFO: Error reading /abs/path/to/random/file_pmphr" 0
20241104 "WARNING: Problem with the monitoring objects IO Exception: " 0
20241104 "INFO: There is not enough space left: 70359905MB" 0
20241104 "INFO: Exception running [/abs/path/to/random/file_vvyhz] : [/abs/path/to/random/file_dpqwe]" 0
20241104 "INFO: Error reading /abs/path/to/random/file_arcmu" 0
20241104 "INFO: Error reading /abs/path/to/random/file_xrtco" 0
20241104 "WARNING: An error occurred while registering [/abs/path/to/random/file_pldvt]. Bad connection?" 0
20241104 "WARNING: Can't upload output file [/abs/path/to/random/file_tiqgk], does not exist or has zero size." 0
20241104 "INFO: Did not change pinning of cgroup " 0
20241104 "INFO: Did not change pinning of cgroup " 0
20241104 "INFO: Format error on getting NUMA range" 0
20241104 "SEVERE: Failed to get Site sonar probe list for node node95343443 in [/abs/path/to/random/file_qdqbdl]" 0
20241104 "INFO: Failed getting or inserting host in getNumberFreeSlots" 0
20241104 "INFO: No probes returned from Site Sonar. Skipping ..." 0
20241104 "INFO: Failed getting or inserting host in getNumberFreeSlots" 0
20241104 "INFO: There is not enough space left: 30027389MB" 0
20241104 "WARNING: Exception creating temporary file" 0
20241104 "WARNING: Problem with getting logger:" 0
20241104 "WARNING: Could not create process list: " 0
20241104 "WARNING: Could not access [/abs/path/to/random/file_weihrl]" 0
20241104 "WARNING: Error deleting the job workdir, using system commands instead" 0
20241104 "SEVERE: IO Error in calling the url [/abs/path/to/random/file_lwjja] for node node5006437 in [/abs/path/to/" 0
20241104 "SEVERE: Error installing [/abs/path/to/random/file_xpzen]" 0
20241104 "INFO: Unable to kill the JobWrapper" 0
```

최종 logs.csv

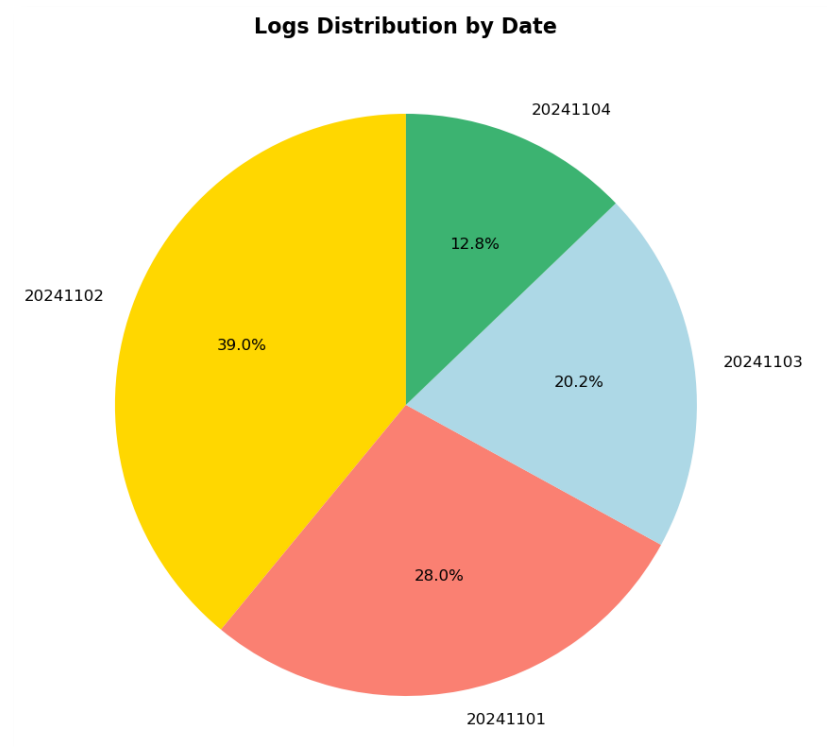
2. 학습 데이터 준비

```
RangeIndex: 27459 entries, 0 to 27458
Data columns (total 3 columns):
#   Column    Non-Null Count  Dtype
---  -
0   Date      27459 non-null  int64
1   Msg       27459 non-null  object
2   isError   27459 non-null  int64
dtypes: int64(2), object(1)
```

❖ 2) 데이터 분석 (에러 로그 비율, 날짜별 로그 비율)



총 행 수: 27459
'1' 값 비율: 60.78% (16689/27459)
'0' 값 비율: 39.22% (10770/27459)

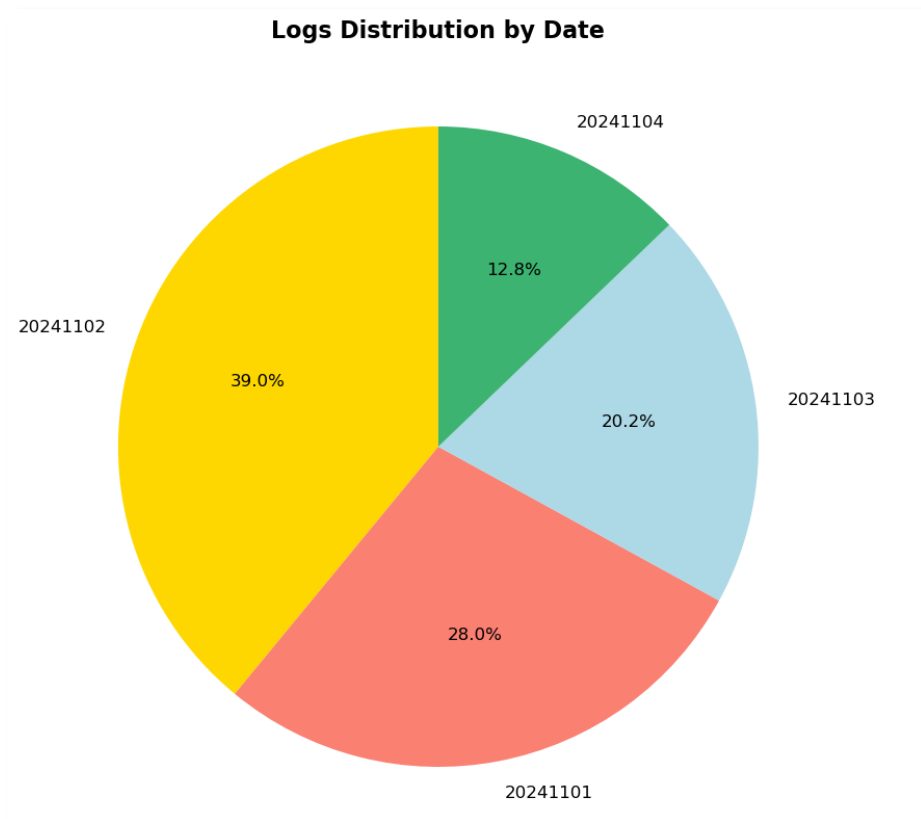


총 행 수: 27459
Date 값 비율:
'20241101': 27.98% (7684/27459)
'20241102': 39.04% (10719/27459)
'20241103': 20.16% (5535/27459)
'20241104': 12.82% (3521/27459)

2. 학습 데이터 준비

```
RangeIndex: 27459 entries, 0 to 27458
Data columns (total 3 columns):
#   Column      Non-Null Count  Dtype
---  -
0    Date        27459 non-null  int64
1    Msg         27459 non-null  object
2    isError     27459 non-null  int64
dtypes: int64(2), object(1)
```

❖ 2) 데이터 분석 (에러 로그 타입)



Log Counts and Proportions (isNormal = 0):
 WARNING: 29.07% (3131/10770)
 SEVERE: 8.53% (919/10770)
 INFO: 48.97% (5274/10770)
 OTHER: 13.43% (1446/10770)

WARNING

20241103	"WARNING: Found exception while processing cgroups exploration "	0
20241103	"WARNING: Site Sonar probe [/abs/path/to/random/file_jpicq] didn't	0
20241103	"WARNING: JobInfo or DiskInfo monitor null"	0
20241103	"WARNING: Exception assigning CPU cores [/abs/path/to/random/file	0

SEVERE

20241103	"SEVERE: Could not fetch machine's idle CPU"	0
20241103	"SEVERE: Monitor has detected an error: [/abs/path/to/random/file_u	0

INFO

20241103	"INFO: There is not enough space left: 73276140MB"	0
20241103	"INFO: Did not change pinning of cgroup "	0
20241103	"INFO: Switching to core check using shell instead"	0
20241103	"INFO: Can not set NUMA architecture aware cpusolation"	0
20241103	"INFO: Can not set NUMA architecture aware cpusolation"	0

OTHER

보통 JAVA
Exception
Message에 해당

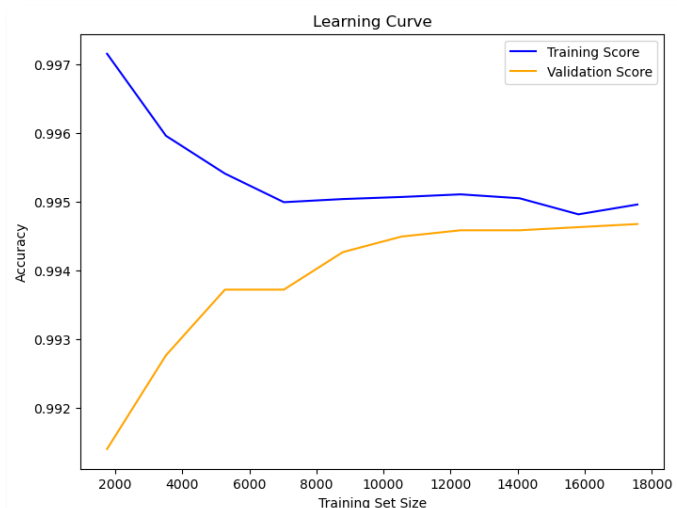
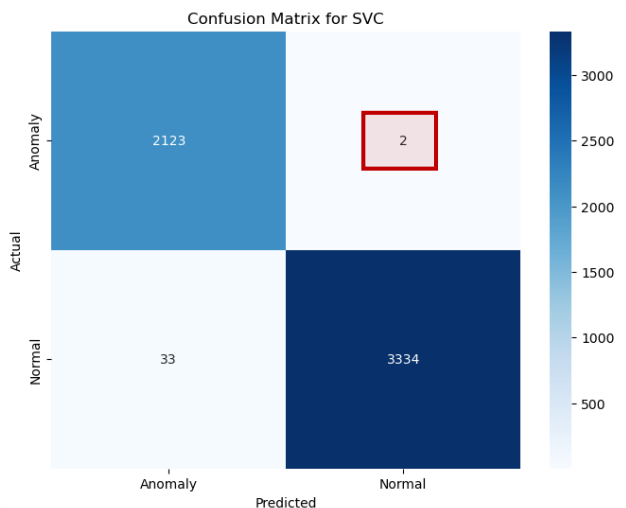
20241101	java.lang.NullPointerException: Cannot invoke "java.lang.Long.lc	0
20241101	at alien.site.JobRunner.lambda\$static\$0(JobRunner.java:39)	0
20241101	at alien.monitoring.Monitor.callMonitoringObjects(Monitor.java:	0
20241101	at alien.monitoring.Monitor.run(Monitor.java:340)	0
20241101	at java.base/java.util.concurrent.Executors\$RunnableAdapter.call	0
20241101	at java.base/java.util.concurrent.FutureTask.runAndReset(FutureT	0
20241101	at java.base/java.util.concurrent.ScheduledThreadPoolExecutor\$	0
20241101	at java.base/java.util.concurrent.ThreadPoolExecutor.runWorker(0
20241101	at java.base/java.util.concurrent.ThreadPoolExecutor\$Worker.run	0
20241101	at java.base/java.lang.Thread.run(Thread.java:833)	0

3. 모델 선택 및 학습

❖ 지도 학습 / 비지도 학습 방법을 모두 사용하여, 해당 로그에 적합한 학습 방법 선택

- 1) 지도 학습 (라벨이 있는 경우)
- SVM** : 데이터를 특정 카테고리로 분류하는데 유용

Classification Report:				
	precision	recall	f1-score	support
0	0.98	1.00	0.99	2125
1	1.00	0.99	0.99	3367
accuracy			0.99	5492
macro avg	0.99	0.99	0.99	5492
weighted avg	0.99	0.99	0.99	5492



```
import pandas as pd
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.svm import SVC
from sklearn.model_selection import train_test_split
from sklearn.metrics import classification_report, confusion_matrix
import matplotlib.pyplot as plt
import seaborn as sns

# 1. 데이터 로드
file_path = 'data/logs.csv' # CSV 파일 경로
data = pd.read_csv(file_path)

# 주요 열 확인
if 'Msg' not in data.columns or 'isNormal' not in data.columns:
    raise ValueError("CSV 파일에 'Msg' 또는 'isNormal' 열이 없습니다.")

# 2. 텍스트 데이터 및 레이블 준비
texts = data['Msg'] # 텍스트 데이터
labels = data['isNormal'] # 레이블 (1: 정상, 0: 비정상)

# 3. TF-IDF 벡터화
vectorizer = TfidfVectorizer(max_features=5000)
X = vectorizer.fit_transform(texts) # 텍스트 데이터 벡터화

# 4. 데이터 분할 (훈련/테스트)
X_train, X_test, y_train, y_test = train_test_split(X, labels, test_size=0.2, random_state=42)

# 출력 확인
print(f"TF-IDF 행렬 크기: {X.shape}")
print(f"훈련 데이터 크기: {X_train.shape}")
print(f"테스트 데이터 크기: {X_test.shape}")

# 5. SVC 모델 학습
model = SVC(kernel='rbf', gamma='scale', C=1.0, probability=True)
model.fit(X_train, y_train)

# 6. 모델 평가
y_pred = model.predict(X_test)

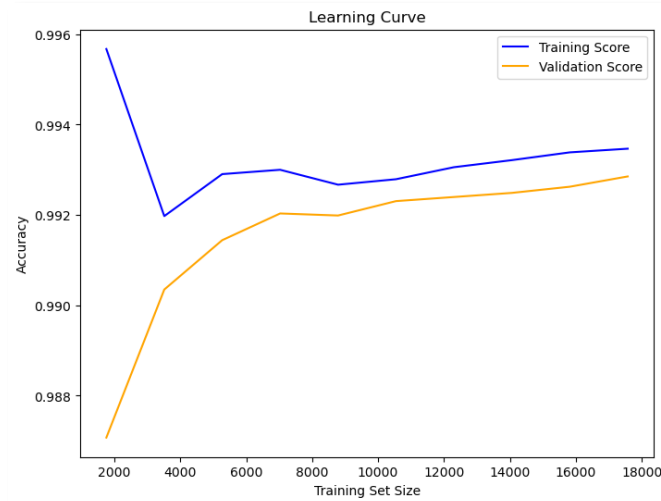
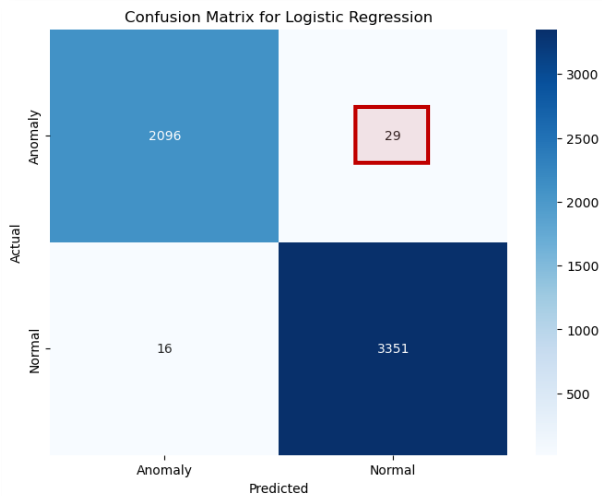
# 7. 성능 보고
print("Classification Report:")
print(classification_report(y_test, y_pred))
```

3. 모델 선택 및 학습

❖ 지도 학습 / 비지도 학습 방법을 모두 사용하여, 해당 로그에 적합한 학습 방법 선택

- 1) 지도 학습 (라벨이 있는 경우)
 - Logistic Regression** : 다양한 데이터 요소를 기반으로 값을 예측하는 데 유용

	precision	recall	f1-score	support
0	0.99	0.99	0.99	2125
1	0.99	1.00	0.99	3367
accuracy			0.99	5492
macro avg	0.99	0.99	0.99	5492
weighted avg	0.99	0.99	0.99	5492



```
import pandas as pd
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import classification_report

# CSV 파일 경로
file_path = 'data/logs.csv'

# 데이터 로드
data = pd.read_csv(file_path)

# 주요 열 확인 (Msg와 isError 열을 사용한다고 가정)
if 'Msg' not in data.columns or 'isNormal' not in data.columns:
    raise ValueError("CSV 파일에 'Msg' 또는 'isNormal' 열이 없습니다.")

# 1. 텍스트 데이터와 타겟 데이터 준비
texts = data['Msg'] # 텍스트 데이터
labels = data['isNormal'] # 타겟 데이터

# 2. TF-IDF 벡터화
vectorizer = TfidfVectorizer(max_features=5000) # 최대 5000개의 피처로 제한
X = vectorizer.fit_transform(texts) # TF-IDF 벡터화

# 3. 데이터 분할 (훈련 세트와 테스트 세트)
X_train, X_test, y_train, y_test = train_test_split(X, labels, test_size=0.2, random_state=42)

# 출력 확인
print(f"TF-IDF 행렬 크기: {X.shape}")
print(f"훈련 데이터 크기: {X_train.shape}")
print(f"테스트 데이터 크기: {X_test.shape}")

# 4. 모델 학습
model = LogisticRegression()
model.fit(X_train, y_train)

# 5. 테스트 데이터 평가
y_pred = model.predict(X_test)
print(classification_report(y_test, y_pred))
```

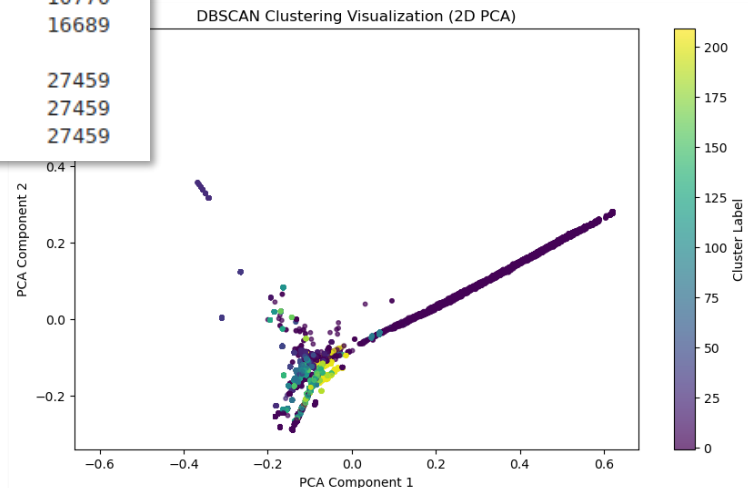
3. 모델 선택 및 학습

❖ 지도 학습 / 비지도 학습 방법을 모두 사용하여, 해당 로그에 적합한 학습 방법 선택

- 2) 비지도 학습 (라벨이 없는 경우)
 - DBSCAN : 라벨이 없는 데이터를 유사성 또는 차이점에 따라 그룹화하는 기술

Adjusted Rand Index (ARI): 0.181
Homogeneity: 0.891
Completeness: 0.166
V-Measure: 0.280

Classification Report:				
	precision	recall	f1-score	support
0	0.99	0.95	0.97	10770
1	0.97	0.99	0.98	16689
accuracy			0.98	27459
macro avg	0.98	0.97	0.98	27459
weighted avg	0.98	0.98	0.98	27459



```
import pandas as pd
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.cluster import DBSCAN
from sklearn.metrics import classification_report, adjusted_rand_score

# 1. 데이터 로드
file_path = 'data/logs.csv' # CSV 파일 경로
data = pd.read_csv(file_path)

# 주요 열 확인
if 'Msg' not in data.columns or 'isNormal' not in data.columns:
    raise ValueError("CSV 파일에 'Msg' 또는 'isNormal' 열이 없습니다.")

# 2. 텍스트 데이터 및 레이블 준비
texts = data['Msg'] # 텍스트 데이터
labels = data['isNormal'] # 레이블 (1: 정상, 0: 비정상)

# 3. TF-IDF 벡터화
vectorizer = TfidfVectorizer(max_features=5000)
X = vectorizer.fit_transform(texts).toarray() # 벡터화 및 배열 변환

# 4. DBSCAN 모델 학습
dbscan = DBSCAN(eps=0.5, min_samples=10, metric='cosine')
clusters = dbscan.fit_predict(X)

# 5. 결과 평가
print("클러스터 레이블 분포:")
print(pd.Series(clusters).value_counts())

# DBSCAN은 비지도 학습이므로 레이블이 없지만, 기존 레이블과 비교 가능
if labels is not None:
    print("\nAdjusted Rand Index (ARI):", adjusted_rand_score(labels, clusters))
```

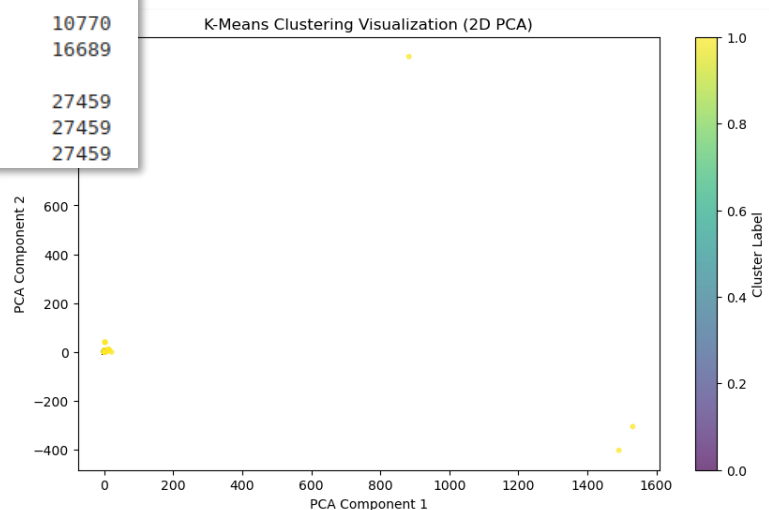
3. 모델 선택 및 학습

❖ 지도 학습 / 비지도 학습 방법을 모두 사용하여, 해당 로그에 적합한 학습 방법 선택

- 2) 비지도 학습 (라벨이 없는 경우)
 - K-Means** : 라벨이 없는 데이터를 유사성 또는 차이점에 따라 그룹화하는 기술

Adjusted Rand Index (ARI): 0.118
Homogeneity: 0.264
Completeness: 0.296
V-Measure: 0.279

Classification Report:				
	precision	recall	f1-score	support
0	0.55	1.00	0.71	10770
1	1.00	0.47	0.64	16689
accuracy			0.68	27459
macro avg	0.77	0.73	0.67	27459
weighted avg	0.82	0.68	0.67	27459



```
import pandas as pd
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.cluster import KMeans
from sklearn.decomposition import PCA

# 1. 데이터 로드
file_path = 'data/logs.csv' # CSV 파일 경로
data = pd.read_csv(file_path)

# 주요 열 확인
if 'Msg' not in data.columns:
    raise ValueError("CSV 파일에 'Msg' 열이 없습니다.")

# 2. 텍스트 데이터 준비
texts = data['Msg'] # 텍스트 데이터

# 3. TF-IDF 벡터화
vectorizer = TfidfVectorizer(max_features=5000)
X = vectorizer.fit_transform(texts).toarray() # 벡터화된 데이터

# 4. K-Means 모델 학습
n_clusters = 2 # 클러스터 개수 설정
kmeans = KMeans(n_clusters=n_clusters, random_state=42, n_init=10)
kmeans.fit(X)
clusters = kmeans.labels_ # 클러스터 레이블

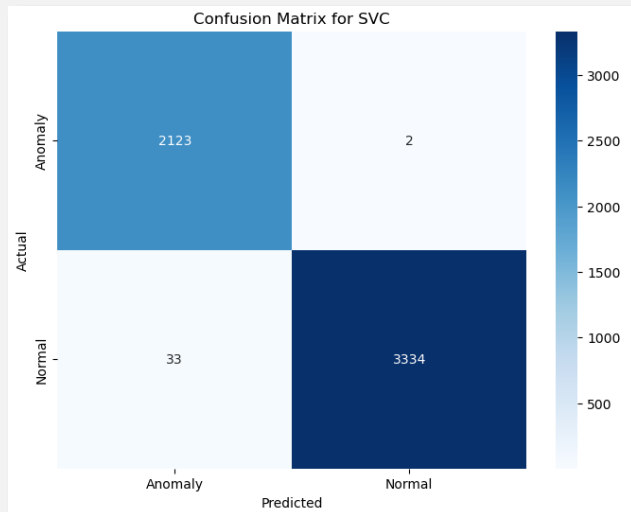
# 5. 실루엣 점수 평가
sil_score = silhouette_score(X, clusters)
print(f"Silhouette Score: {sil_score:.3f}")
```

3. 모델 선택 및 학습

❖ 학습 결과 – 지도 학습 (SVM vs. Logistic Regression)

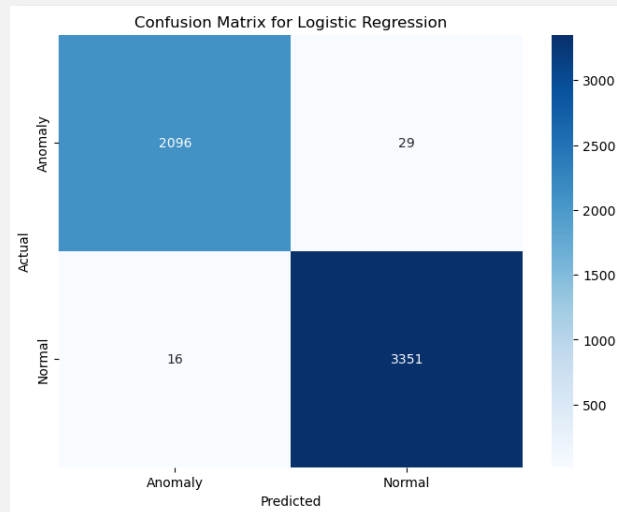
SVM

	precision	recall	f1-score	support
0	0.98	1.00	0.99	2125
1	1.00	0.99	0.99	3367
accuracy			0.99	5492
macro avg	0.99	0.99	0.99	5492
weighted avg	0.99	0.99	0.99	5492



Logistic Regression

	precision	recall	f1-score	support
0	0.99	0.99	0.99	2125
1	0.99	1.00	0.99	3367
accuracy			0.99	5492
macro avg	0.99	0.99	0.99	5492
weighted avg	0.99	0.99	0.99	5492



실제 *Anomaly* 로그를 놓칠 확률이
LR 모델보다 *SVM* 모델이 적음

[Confusion Matrix]

SVC's FP = 2 / LR's FP = 29

- Precision :
0 / 1으로 예측할 때 얼마나 정확한지를 평가
- Recall :
실제 데이터를 얼마나 잘 찾아냈는지를 평가
- F1-Score :
Precision과 Recall 사이의 균형을 평가

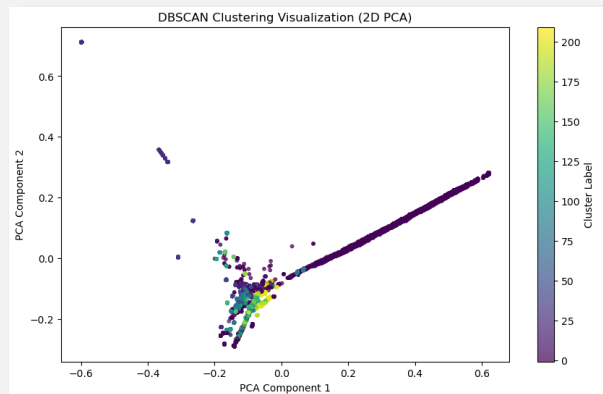
3. 모델 선택 및 학습

❖ 학습 결과 - 비지도 학습 (DBSCAN vs. K-Means)

DBSCAN

Adjusted Rand Index (ARI): 0.181
Homogeneity: 0.891
Completeness: 0.166
V-Measure: 0.280

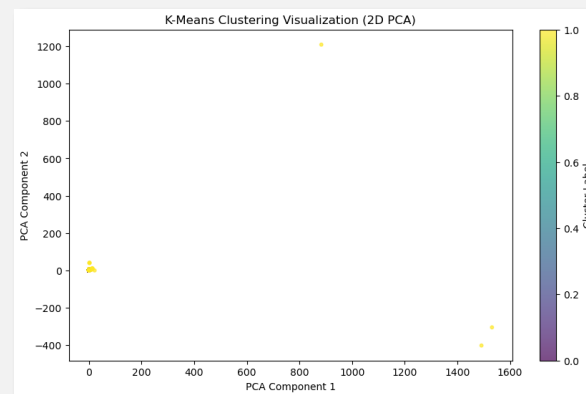
Classification Report:				
	precision	recall	f1-score	support
0	0.99	0.95	0.97	10770
1	0.97	0.99	0.98	16689
accuracy			0.98	27459
macro avg	0.98	0.97	0.98	27459
weighted avg	0.98	0.98	0.98	27459



K-Means

Adjusted Rand Index (ARI): 0.118
Homogeneity: 0.264
Completeness: 0.296
V-Measure: 0.279

Classification Report:				
	precision	recall	f1-score	support
0	0.55	1.00	0.71	10770
1	1.00	0.47	0.64	16689
accuracy			0.68	27459
macro avg	0.77	0.73	0.67	27459
weighted avg	0.82	0.68	0.67	27459



전체적인 지표로는

DBSCAN이 더 나은 성능을 보임

- ARI :
기존 레이블과 클러스터링 결과 간의 유사도
- Homogeneity :
각 클러스터가 단일 클래스의 데이터만 포함하는지
- Completeness :
같은 클래스의 모든 데이터가 동일한 클러스터에 포함되는지
- V-Measure :
Homogeneity와 Completeness의 조화 평균

4. 학습 모델 활용



❖ ‘학습한 데이터 + 새로운 데이터’를 통해 에러 데이터 탐지가 제대로 이루어지는지 확인

- 학습데이터 : 정적 + 동적 로그를 대상으로 테스트

⇒ 동적 에러 로그를 잘 탐지하는지 확인하는 것이 목표

```
INFO: [ NTPDate ] Initial NTP sync finished. Status:
Nov 01, 2024 8:23:11 AM alien.api.DispatchSSLClient checkIdleConnection
WARNING: Exception filling the values for resource_status / null
WARNING: Exception moving 87184252 to cgroup 1923
SEVERE: Problem in construction of ComputingElement:
SEVERE: Exception checking the identity: /home01/gsd24a01/test/
INFO: Exception while waiting for release of CPU pinning lock
INFO: Error creating Agent Startup file: /scratch/gsd24a01/ALICE/alien-log/2024-11-02/jobagent-4231-00039.sh
```

- 새로운 데이터 : 학습하지 않은 정상 + 에러 로그를 대상으로 테스트

⇒ 에러 로그에서 보이는 패턴을 잘 학습했는지 확인하는 것이 목표 (not, null, cannot, exception, unable ...)

```
INFO: We have jobs that might be optimised (TTL)
INFO: Printing client information:
INFO: Running JCentral with host cert:
WARNING: HTTP port is defined but not as a number for
WARNING: Exception accepting a client
WARNING: Cannot setSoTimeout for the initial communication
SEVERE: No TTL optimizations could be applied. Errors while parsing the entries.
SEVERE: Host certificate is not valid any more, please renew it and restart the service
INFO: Queue is not open! Check queueinfo
INFO: Did not shutdown gracefully
INFO: Can't instantiate JobOptimizer!
INFO: Can not instantiate JobOptimizer!
```

4. 학습 모델 활용

❖ 에러 로그 탐지 시스템 구조

학습된 모델 / 모델별 TF-IDF

- dbscan_model.pkl
- dbscan_tfidf_vectorizer.pkl
- kmeans_model.pkl
- kmeans_tfidf_vectorizer.pkl
- lr_model.pkl
- lr_tfidf_vectorizer.pkl
- svc_model.pkl
- svc_tfidf_vectorizer.pkl

탐지할 로그 데이터

```
INFO: [ NTPDate ] Initial NTP sync finished. Status: 0
Nov 01, 2024 8:23:11 AM alien.api.DispatchSSLClient checkIdleConnection
WARNING: Exception filling the values for resource_status / null
WARNING: Exception moving 87184252 to cgroup 1923
SEVERE: Problem in construction of ComputingElement:
SEVERE: Exception checking the identity: /home01/gsd24a01/test/
INFO: Exception while waiting for release of CPU pinning lock
INFO: Error creating Agent Startup file: /scratch/gsd24a01/ALICE/alien-
log/2024-11-02/jobagent-4231-00039.sh
INFO: We have jobs that might be optimised (TTL)
INFO: Printing client information:
INFO: Running JCentral with host cert:
WARNING: HTTP port is defined but not as a number for
WARNING: Exception accepting a client
WARNING: Cannot setSoTimeout for the initial communication
SEVERE: No TTL optimizations could be applied. Errors while parsing the entries.
SEVERE: Host certificate is not valid any more, please renew it and restart the service
INFO: Queue is not open! Check queueinfo
INFO: Did not shutdown gracefully
INFO: Can't instantiate JobOptimizer!
INFO: Can not instantiate JobOptimizer!
```

탐지 시스템

```
import joblib # SVC 모델 로드

# 1. SVC 모델 및 TF-IDF 벡터화 로드
model_path = 'trained-model/svc_model.pkl' # SVM 모델 경로
vectorizer_path = 'trained-model/svc_tfidf_vectorizer.pkl' # TF-IDF 벡터화 모델 경로

svc_model = joblib.load(model_path) # SVM 모델 로드
tfidf_vectorizer = joblib.load(vectorizer_path) # TF-IDF 벡터화 모델 로드

# 2. 에러 로그 판단 함수
def predict_log(log_message):
    """
    주어진 로그 메시지가 에러 로그인지 판단.
    :param log_message: str, 로그 메시지
    :return: str, "Error Log" 또는 "Normal Log"
    """
    if not log_message.strip():
        return "Empty log message."

    # TF-IDF 변환
    transformed_message = tfidf_vectorizer.transform([log_message])

    # SVM 예측
    prediction = svc_model.predict(transformed_message)

    # 결과 반환
    return "Error Log" if prediction[0] == 0 else "Normal Log"

# 3. 입력 파일에서 로그 판단
def process_input_file(input_file):
    try:
        with open(input_file, 'r') as infile:
            lines = infile.readlines()
    except FileNotFoundError:
        print(f"Error: File '{input_file}' not found.")
        return

    results = []
    for idx, line in enumerate(lines, start=1):
        log_message = line.strip()
        if log_message: # 빈 줄 무시
            prediction = predict_log(log_message)
            results.append((idx, log_message, prediction))

    # 결과 출력
    for idx, log_message, prediction in results:
        print(f"[Line {idx}] {log_message} \n >> {prediction}")

# 4. 실행
if __name__ == "__main__":
    input_file = 'data/logs-input.txt' # 입력 파일 경로

    print("[Log Detection with SVC]")
    process_input_file(input_file)
```

Prediction
(SVM, Logistic
Regression)

"Normal Log"

Clustering
(DBSCAN,
K-Means)

"Error Log"

4. 학습 모델 활용



SVM

```
[Log Detection with SVC]
[Line 1] INFO: [ NTPDate ] Initial NTP sync finished. Status: 0
>> Normal Log
[Line 2] Nov 01, 2024 8:23:11 AM alien.api.DispatchSSLClient checkIdleConnection
>> Normal Log
[Line 3] WARNING: Exception filling the values for resource_status / null
>> Error Log
[Line 4] WARNING: Exception moving 87184252 to cgroup 1923
>> Error Log
[Line 5] SEVERE: Problem in construction of ComputingElement:
>> Error Log
[Line 6] SEVERE: Exception checking the identity: /home01/gsd24a01/test/
>> Error Log
[Line 7] INFO: Exception while waiting for release of CPU pinning lock
>> Error Log
[Line 8] INFO: Error creating Agent Startup file: /scratch/gsd24a01/ALICE/alien-log/2024-11-02/jobagen
t-4231-00039.sh
>> Error Log
[Line 9] INFO: We have jobs that might be optimised (TTL)
>> Normal Log
[Line 10] INFO: Printing client information:
>> Normal Log
[Line 11] INFO: Running JCentral with host cert:
>> Normal Log
[Line 12] WARNING: HTTP port is defined but not as a number for
>> Error Log
[Line 13] WARNING: Exception accepting a client
>> Error Log
[Line 14] WARNING: Cannot setSoTimeout for the initial communication
>> Error Log
[Line 15] SEVERE: No TTL optimizations could be applied. Errors while parsing the entries.
>> Error Log
[Line 16] SEVERE: Host certificate is not valid any more, please renew it and restart the service
>> Error Log
[Line 17] INFO: Queue is not open! Check queueinfo
>> Error Log
[Line 18] INFO: Did not shutdown gracefully
>> Error Log
[Line 19] INFO: Can't instantiate JobOptimizer!
>> Normal Log
[Line 20] INFO: Can not instantiate JobOptimizer!
>> Error Log
```

Can't 가 포함된
로그를 학습하지
않아 발생한 문제로
보임

Logistic Regression

```
[Log Detection with Logistic Regression]
[Line 1] INFO: [ NTPDate ] Initial NTP sync finished. Status: 0
>> Normal Log
[Line 2] Nov 01, 2024 8:23:11 AM alien.api.DispatchSSLClient checkIdleConnection
>> Normal Log
[Line 3] WARNING: Exception filling the values for resource_status / null
>> Error Log
[Line 4] WARNING: Exception moving 87184252 to cgroup 1923
>> Error Log
[Line 5] SEVERE: Problem in construction of ComputingElement:
>> Error Log
[Line 6] SEVERE: Exception checking the identity: /home01/gsd24a01/test/
>> Error Log
[Line 7] INFO: Exception while waiting for release of CPU pinning lock
>> Error Log
[Line 8] INFO: Error creating Agent Startup file: /scratch/gsd24a01/ALICE/alien-log/2024-11-02/jobagen
t-4231-00039.sh
>> Error Log
[Line 9] INFO: We have jobs that might be optimised (TTL)
>> Normal Log
[Line 10] INFO: Printing client information:
>> Normal Log
[Line 11] INFO: Running JCentral with host cert:
>> Normal Log
[Line 12] WARNING: HTTP port is defined but not as a number for
>> Error Log
[Line 13] WARNING: Exception accepting a client
>> Error Log
[Line 14] WARNING: Cannot setSoTimeout for the initial communication
>> Error Log
[Line 15] SEVERE: No TTL optimizations could be applied. Errors while parsing the entries.
>> Error Log
[Line 16] SEVERE: Host certificate is not valid any more, please renew it and restart the service
>> Error Log
[Line 17] INFO: Queue is not open! Check queueinfo
>> Error Log
[Line 18] INFO: Did not shutdown gracefully
>> Error Log
[Line 19] INFO: Can't instantiate JobOptimizer!
>> Normal Log
[Line 20] INFO: Can not instantiate JobOptimizer!
>> Error Log
```

4. 학습 모델 활용



DBSCAN

정상 로그도 에러 로그로 판단
(모델의 학습 상태가 나쁨)

[Log Detection with DBSCAN]

```
[Line 1] INFO: [ NTPDate ] Initial NTP sync finished. Status: 0
>> Error Log
[Line 2] Nov 01, 2024 8:23:11 AM alien.api.DispatchSSLClient checkIdleConnection
>> Error Log
[Line 3] WARNING: Exception filling the values for resource_status / null
>> Error Log
[Line 4] WARNING: Exception moving 87184252 to cgroup 1923
>> Error Log
[Line 5] SEVERE: Problem in construction of ComputingElement:
>> Error Log
[Line 6] SEVERE: Exception checking the identity: /home01/gsd24a01/test/
>> Error Log
[Line 7] INFO: Exception while waiting for release of CPU pinning lock
>> Error Log
[Line 8] INFO: Error creating Agent Startup file: /scratch/gsd24a01/ALICE/alien-log/2024-11-02/jobagen
t-4231-00039.sh
>> Error Log
[Line 9] INFO: We have jobs that might be optimised (TTL)
>> Error Log
[Line 10] INFO: Printing client information:
>> Error Log
[Line 11] INFO: Running JCentral with host cert:
>> Error Log
[Line 12] WARNING: HTTP port is defined but not as a number for
>> Error Log
[Line 13] WARNING: Exception accepting a client
>> Error Log
[Line 14] WARNING: Cannot setSoTimeout for the initial communication
>> Error Log
[Line 15] SEVERE: No TTL optimizations could be applied. Errors while parsing the entries.
>> Error Log
[Line 16] SEVERE: Host certificate is not valid any more, please renew it and restart the service
>> Error Log
[Line 17] INFO: Queue is not open! Check queueinfo
>> Error Log
[Line 18] INFO: Did not shutdown gracefully
>> Error Log
[Line 19] INFO: Can't instantiate JobOptimizer!
>> Error Log
[Line 20] INFO: Can not instantiate JobOptimizer!
>> Error Log
```

K-Means

에러 로그를 정상 로그로 판단
(모델의 학습 상태가 나쁨)

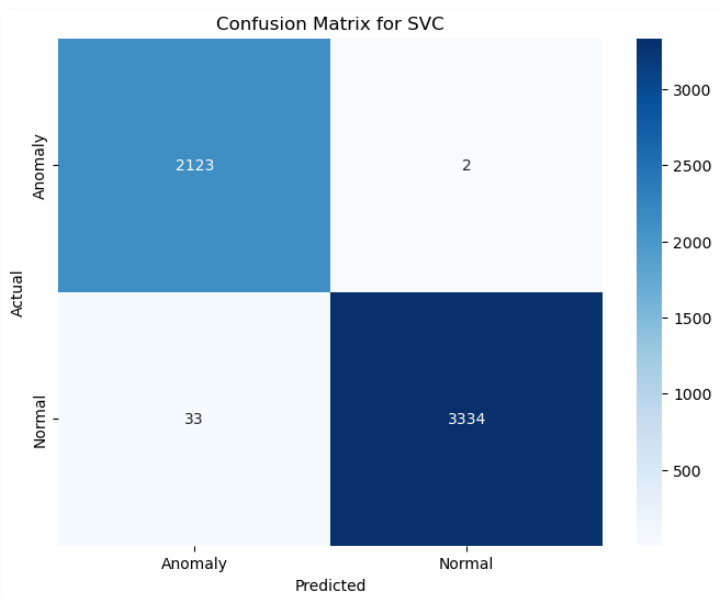
[Log Detection with K-Means]

```
[Line 1] INFO: [ NTPDate ] Initial NTP sync finished. Status: 0
>> Normal Log
[Line 2] Nov 01, 2024 8:23:11 AM alien.api.DispatchSSLClient checkIdleConnection
>> Error Log
[Line 3] WARNING: Exception filling the values for resource_status / null
>> Normal Log
[Line 4] WARNING: Exception moving 87184252 to cgroup 1923
>> Normal Log
[Line 5] SEVERE: Problem in construction of ComputingElement:
>> Normal Log
[Line 6] SEVERE: Exception checking the identity: /home01/gsd24a01/test/
>> Normal Log
[Line 7] INFO: Exception while waiting for release of CPU pinning lock
>> Normal Log
[Line 8] INFO: Error creating Agent Startup file: /scratch/gsd24a01/ALICE/alien-log/2024-11-02/jobagen
t-4231-00039.sh
>> Normal Log
[Line 9] INFO: We have jobs that might be optimised (TTL)
>> Normal Log
[Line 10] INFO: Printing client information:
>> Normal Log
[Line 11] INFO: Running JCentral with host cert:
>> Normal Log
[Line 12] WARNING: HTTP port is defined but not as a number for
>> Normal Log
[Line 13] WARNING: Exception accepting a client
>> Normal Log
[Line 14] WARNING: Cannot setSoTimeout for the initial communication
>> Normal Log
[Line 15] SEVERE: No TTL optimizations could be applied. Errors while parsing the entries.
>> Normal Log
[Line 16] SEVERE: Host certificate is not valid any more, please renew it and restart the service
>> Normal Log
[Line 17] INFO: Queue is not open! Check queueinfo
>> Normal Log
[Line 18] INFO: Did not shutdown gracefully
>> Normal Log
[Line 19] INFO: Can't instantiate JobOptimizer!
>> Normal Log
[Line 20] INFO: Can not instantiate JobOptimizer!
>> Normal Log
```


5. 결론

❖ 모델 학습 결과 + 새로운 데이터에 대한 에러 탐지 결과 ⇒ [Best] SVM

Classification Report:					
	precision	recall	f1-score	support	
0	0.98	1.00	0.99	2125	
1	1.00	0.99	0.99	3367	
accuracy			0.99	5492	
macro avg	0.99	0.99	0.99	5492	
weighted avg	0.99	0.99	0.99	5492	



```
[Log Detection with SVC]
[Line 1] INFO: [ NTPDate ] Initial NTP sync finished. Status: 0
>> Normal Log
[Line 2] Nov 01, 2024 8:23:11 AM alien.api.DispatchSSLClient checkIdleConnection
>> Normal Log
[Line 3] WARNING: Exception filling the values for resource_status / null
>> Error Log
[Line 4] WARNING: Exception moving 87184252 to cgroup 1923
>> Error Log
[Line 5] SEVERE: Problem in construction of ComputingElement:
>> Error Log
[Line 6] SEVERE: Exception checking the identity: /home01/gsd24a01/test/
>> Error Log
[Line 7] INFO: Exception while waiting for release of CPU pinning lock
>> Error Log
[Line 8] INFO: Error creating Agent Startup file: /scratch/gsd24a01/ALICE/alien-log/2024-11-02/jobagen
t-4231-00039.sh
>> Error Log
[Line 9] INFO: We have jobs that might be optimised (TTL)
>> Normal Log
[Line 10] INFO: Printing client information:
>> Normal Log
[Line 11] INFO: Running JCentral with host cert:
>> Normal Log
[Line 12] WARNING: HTTP port is defined but not as a number for
>> Error Log
[Line 13] WARNING: Exception accepting a client
>> Error Log
[Line 14] WARNING: Cannot setSoTimeout for the initial communication
>> Error Log
[Line 15] SEVERE: No TTL optimizations could be applied. Errors while parsing the entries.
>> Error Log
[Line 16] SEVERE: Host certificate is not valid any more, please renew it and restart the service
>> Error Log
[Line 17] INFO: Queue is not open! Check queueinfo
>> Error Log
[Line 18] INFO: Did not shutdown gracefully
>> Error Log
[Line 19] INFO: Can't instantiate JobOptimizer!
>> Normal Log
[Line 20] INFO: Can not instantiate JobOptimizer!
>> Error Log
```