REVA UNIVERSITY
Bengaluru, India

A Project Report on

Key Driver Analysis for Customer Decommissioning using Machine Learning

Submitted in Partial Fulfilment for Award of Degree of

Master of Business Administration in Business Analytics

Submitted By

Aruna Kashinath

R19MBA54

Under the Guidance of

Mithun D J

Senior Manager – Data Science

RACE, REVA University

REVA Academy for Corporate Excellence - RACE

REVA University

Rukmini Knowledge Park, Kattigenahalli, Yelahanka, Bengaluru - 560 064

race.reva.edu.in

August, 2022

REVA UNIVERSITY
Bengaluru, India

Candidate's Declaration

I, **Aruna Kashinath** hereby declare that I have completed the project work towards the first year of Master of Business Administration in Business Analytics at, REVA University on the topic entitled **Key Driver Analysis for Customer Decommissioning using Machine Learning** under the supervision of **Mr. Mithun D. J.** This report embodies the original work done by me in partial fulfilment of the requirements for the award of the degree for the academic year **2022.**

Place: Bengaluru

Name of the Student: **Aruna Kashinath**

Aruna K

Date: 12th August 2022

Signature of Student

REVA UNIVERSITY
Bengaluru, India

Certificate

This is to Certify that the project work entitled **Key Driver Analysis for Customer Decommissioning using Machine Learning** carried out by **Aruna Kashinath** with SRN R19MBA54**,** is a bonafide student at REVA University, is submitting the first-year project report in fulfilment for the award of **Master of Business Administration** in Business Analytics during the academic year 2022. The Project report has been tested for plagiarism and has passed the plagiarism test with a similarity score of less than 15%. The project report has been approved as it satisfies the academic requirements in respect of the project work prescribed for the said degree.

Signature of the Guide
Name of the Guide:
Mr. Mithun D. J.

Signature of the Director
Name of the Director:
Dr. Shinu Abhi

External Viva
Names of the Examiners

1. Harsh Vardhan, Chief Digital Technology Architect, Capgemini
2. Pradeepta Mishra, Director – AT, L&T InfoTech
3. Rajib Bhattacharya, Director – Data & Analytics, Cargill

Place: Bengaluru
Date: 20th August 2022

REVA UNIVERSITY
Bengaluru, India

Acknowledgment

I am highly indebted to **Dr. Shinu Abhi**, Director, and Corporate Training for their guidance and constant supervision as well as for providing necessary information regarding the project and also for their support in completing the project.

I would like to thank my project guide **Mr. Mithun D. J.** for the valuable guidance provided to understand the concept and execute this project. It is my gratitude towards **Dr. Jay Bharateesh Simha** and all other mentors for their valuable guidance and suggestion in learning various data science aspects and for their support. I am thankful to my classmates for their aspiring guidance, invaluable constructive criticism, and friendly advice during the project work.

I would like to acknowledge the support provided by Hon'ble Chancellor, **Dr. P Shyama Raju**, Vice Chancellor, **Dr. M. Dhananjaya,** and Registrar, **Dr. N Ramesh**. It is sincere thanks to all members of the program office of RACE who were supportive of all requirements from the program office.

It is my sincere gratitude towards my parents, and my family for their kind co-operation and encouragement which helped me in the completion of this project.

Place: Bengaluru

Date: 12th August 2022

Similarity Index Report

This is to certify that this project report titled **Key Driver Analysis for Customer Decommissioning using Machine Learning** was scanned for similarity detection. Process and outcome are given below.

Software Used: **Turnitin**

Date of Report Generation: **12.08.2022**

Similarity Index in %: **15%**

Total word count: **10805**

Name of the Guide: **Mithun D J**

Place: Bengaluru

Name of the Student: **Aruna Kashinath**

Date: 12.08.2022

Aruna K

Signature of Student

Verified by: Dincy Dechamma

Signature

Dr. Shinu Abhi,

Director, Corporate Training

List of Abbreviations

Sl. No	Abbreviation	Long Form
1	ECS	Enterprise Cloud Services
2	MTTR	Mean Time to Resolve
3	SLA	Service Level Agreement
4	CDM	Client Delivery Manager
5	TSM	Technical Service Manager
6	SM	Server Management
7	AWS	Amazon Web Services
8	GCP	Google Cloud Platform
9	KNN	K-Nearest Neighbours
10	SOP	Standard Operating Procedure
11	WI	Work Instructions
12	AI	Artificial Intelligence
13	ML	Machine Learning
14	CHAID	Chi-Square automatic interaction detection
15	CART	Classification and Regression Trees
16	CRISP-DM	Cross-industry standard process for data mining
17	HANA	High-performance Analytic Appliance
18	EDA	Exploratory Data Analysis

List of Figures

List of Tables

Abstract

The customer decommissioning process at SAP is the last phase in the life cycle of a customer system and is the permanent removal of an IT service or any other configuration item, from the live customer environment. The purpose of decommissioning process in the SAP Enterprise Cloud Services delivery (ECS) team is to provide a structured and standardized way to decommission system/server and network and to facilitate transparency on the decommission process globally.

Decommissioning involves the deactivation of the customer system/tenant performed by the responsible cloud decommissioning personnel upon the customer contract end date or system end date and confirmation provided by the customer to decommission their system. Customer data is provided to them prior to decommissioning upon customer request and the data retention period of 30 days is applicable in case the customer requests data post-decommissioning their systems.

Various factors are contributing to the delays in the end-to-end decommissioning process resulting in a huge loss to SAP for retaining customer systems even after the contract has been terminated or the system end date has been reached.

The project aims at performing key driver analysis affecting delays using the Mean Time to Resolve (MTTR) of the tickets created for decommissioning requests. Further to identifying the factors affecting delays in the end-to-end decommissioning process, the business objective is to provide a roadmap for Cost savings beneficial by optimizing the process thereby reducing delays in the end-to-end process.

Data consisting of tickets created in the JIRA tool for customer decommissioning requests have been considered from Q1 of FY2021 to Q2 of FY2022 forming the basis of our study on factors that are most significant to higher MTTR.

The project exemplifies a machine learning classification that can be used to create a decision tree model for arriving at the key driver analysis of factors resulting in delays. The business impact of this project would result in cost optimization for SAP by performing customer decommissioning within the defined Service Level Agreement (SLA).

Keywords: MTTR analysis, Cost Optimization, Decision Tree, Predictive Modeling, Key driver analysis.

Contents

Chapter 1: Introduction

Every business aims at ways to contain costs in an organization. SAP ECS team should perform customer system/server decommissioning within the defined SLAs for the cost to be efficient. To achieve this, the ability to decrease costs and increase the bottom line by making processes more efficient is the key.

The customer decommissioning process has various scenarios which are followed in SAP ECS as listed below –

- Server Decommissioning
- System Decommissioning
- Network Decommissioning

The CMDB contains a vast amount of information related to customer systems such as contract end date, customer systems, servers, system end date, applications installed, license, system ID etc. (Krishna M V, 2016). Various roles are involved in performing customer decommissioning.

The Client Delivery Manager (CDM) is responsible to maintain customer relationship management. CDM is the backbone of the SAP ECS team interacts with the customer and is responsible for customer contract management.

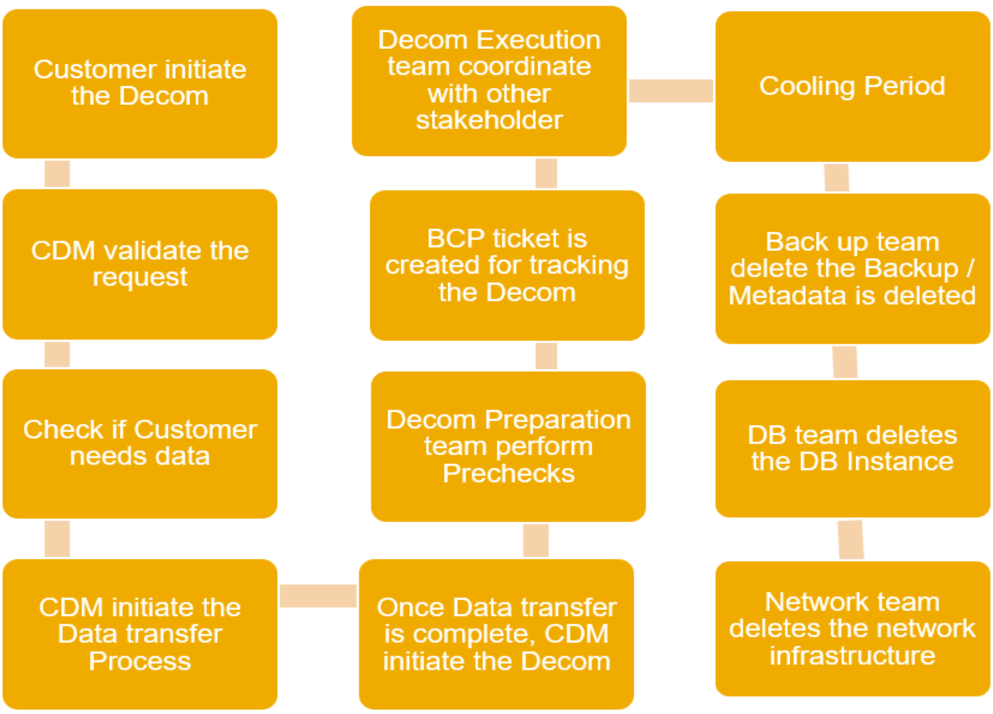


Figure 1.1 End-to-End Decommissioning Process

Customer decommissioning process is defined in Figure 1.1 which involves various roles being responsible to perform decommissioning of systems, servers, and network. Detailed explanation of process mentioned in Figure 1.1 is discussed below.

Process initiation: Initiation by approaching contract or equipment end date of one or multiple tiers, otherwise referred to as systems that has multiple servers connected within a customer landscape.

CDM is notified on an approaching end date <6 weeks prior system/contract end date. CDM will have to actively involve customer for discussions upon approaching contract end date. Upon customer confirmation to go ahead with a system/entire landscape decommission based on system/contract end date, decommissioning process commences.

CDM creates decommissioning requests in JIRA ticketing tool and various teams involved in the decommissioning process work on the ticket created for an individual customer request. JIRA ticket consists of an excel template that lists all the necessary fields needed for performing the decommissioning. CDM must capture all the details in the ticket.

- In case of system decommission request, various activities such as backup configuration removal, volume removal, and virtual machine deletion by Server Management, DB deletion and server decommission are executed along with monitoring removal.

- In case of server decommission request, server decommission is executed along with monitoring removal.

- In case of network decommission request, complete network segment decommissions will be performed.

Automated procedures perform decommissioning. Decommission teams are involved in the execution of procedures to perform system/server/network decommissioning.

The decommissioning team creates tickets for the database team for tenant deletion, the backup team for removing the backup configuration, and monitoring removal. They also create tickets for the server management team to decommission the server. Network tickets will be created for the network team to delete the network in case the customer is moving out of SAP ECS.

Various roles are involved in performing customer decommissioning

- Customer-facing roles – Client Delivery Manager (CDM) and Technical Service Manager (TSM)
- Decommissioning Team – Preparation and Execution team
- Support teams
- Server Management
- Database
- Backup and Network Support Team
- Public Cloud (AWS / Azure / GCP) Multi cloud Management team
- Capacity management team

System decommissioning or a Server decommissioning must undergo a cooling period (systems will be powered down) for 5 days in case the customer changes their mind to extend the contract or to avoid wrong decommissioning. Once the decommissioning process is completed, the JIRA ticket will be closed.

SAP ECS team has the following vendors, also called as sub-contractors working on the decommissioning requests -

- NTT
- Accenture
- HCL
- TCS
- Capgemini

Whenever a server is connected to a database or an application that is shared among multiple other servers, then that server must be isolated prior to decommissioning the server. This is an important process prior to decommissioning process and the customer has to confirm if the server can be isolated before decommissioning the same.

MTTR of the tickets have been considered a key metric to measure the process efficiency. Internal SLA of 30 days to 60 days has been defined based on the MTTR of previous tickets to be the optimal period for customer decommission. Beyond this period has been considered as delays in the end-to-end processing and requires root cause analysis to be performed to analyze the key drivers for the delays.

The value of the MTTR depends on a variety of factors (Nyarko-Boateng et al., 2020) such as the number of servers, vendors/sub-contractors working on the request, and teams working on different components such as database, backup, network team, etc.

In the project demonstrated below, the focal point is to reduce the cost incurred due to the delay in customer decommissioning as the customer system will be in the SAP landscape till it is decommissioned and that adds to the cost SAP has to bear due to delays in the decommissioning process. The project aims at creating decision trees for analyzing various key drivers influencing MTTR. Also, the KNN model has been built to predict the MTTR for the tickets that have the status being open so that we proactively identify the tickets that can be delayed due to key drivers analyzed in the project.

Chapter 2: Literature Review

The delay prediction in this project entails using machine learning models, for instance decision tree and KNN to obtain the expected result. Machine learning algorithms receive and analyse input data to predict output values. Extensive study of some of the research papers have been reviewed under various topics related to decision trees and KNN and the different methodologies used to analyse the cost and process optimization techniques. Some papers have been discussed herein.

Organizations accumulate many legacy systems over time that are rarely accessed but contain huge amounts of data. These systems must be kept and maintained due to legal regulations, however holding this data comes at an enormous cost, both in terms of running expenses and data footprint making moving to new cloud systems really expensive.

Decommissioning can add significant value in scenarios such as mergers and acquisitions or moving to the cloud to name a few. Maintaining older systems can be very expensive because of infrastructure, hardware, and software support costs and internal personnel costs. Retiring legacy systems can make savings of the total cost of ownership.

The decommissioning process is a major challenge to plan and decide as the processes are resource exhaustive (Vuttipittayamongkol et al., 2021). The preparation of the transition to decommissioning and deleting is a key issue in order to minimize delays and undue costs. The different phases of decommissioning involve planning, preparation, execution, and completion of decommissioning activities (Radioactive Waste Management, 2018).

Decommissioning an IT infrastructure involves data centers, systems, servers, and networks to be decommissioned from a customer environment due to aging, broken, or reaching their end of life.

Vendor management has been pivotal in decommissioning process. One company cannot manage all decommissioning services through owned facilities. Organizations performing customer decommissioning or data center decommissioning must have a robust structure in place for auditing and managing multiple sub-contractors and partners. Data security is always a concern. Data Centers utilize on-site data destruction services for their sites to ensure data is removed prior to equipment being decommissioned. Hyperscale data centers are built with the infrastructure needed to meet the heavy expectations of businesses today. These data centers are built with thousands of servers that operate together through a high-speed network. In fact, Gartner predicts that by “2025, 80% of enterprises will have to shut down their traditional data center, versus the 10% today” (SIMS Lifecycle Services, 2020a).

It takes a huge amount of energy to manufacture, install, and run global data centers to meet huge data needs. With hyperscalers and wholesale data centers in existence, ensuring that data centers are as sustainable as possible is a concern of growing importance. When equipment reaches its end of life, it might involve the recovery, repair, reuse, and remarketing of parts and components within the data center environment or outside of it (SIMS Lifecycle Services, 2020b).

Regarding hard drive shredding or wiping, On-site services have been a popular choice for data center managers who want to have the benefit of hard drives being wiped and physically shredded on-site, with certificates of destruction in hand before equipment is transported off-site. Having a better understanding of customer equipment decommissioning services will help
provide assurance of security, compliance, and environmental responsibility. There is a long list of benefits associated with the responsible decommissioning of data center equipment. Some of these include ensuring data security, business continuity, compliance, and maximizing the revenue returned to the client (SIMS Lifecycle Services, 2020a).

Machine learning algorithms can be built on top of decommissioning requests to predict or identify key factors driving the dependent variables thereby reducing cost. The decision tree algorithm is part of the supervised learning algorithm, its main objective is to construct a training model that can be used to predict the class or value of the target variables through learning decision rules from the training data (Charbuty & Abdulazeez, 2021).

Zhang & Jiang in describes in their paper that a decision tree is a common, intuitive, fast classification method. There are many decision tree construction algorithms, ID3 algorithm proposed by Quinlan in 1986, is the first international, influential decision tree. It is considered a very simple decision tree algorithm. ID3 uses information gain as splitting criteria (Zhang & Jiang, 2012).

CART (Classification and Regression Trees) is proposed by (Breiman & Ihaka, 1984), it is the first algorithm to build a decision tree using continuous variables. When the CART-based decision tree algorithms deal with continuous variables, the division direction and division point are greedily optimized under the goal of maximizing the GINI gain, which mainly considers the degree of difference in data after division (Jiao et al., 2020).

Decision trees are built in this project to identify the key drivers affecting the delays in the decommissioning process. Predicting MTTR is the outcome of the decision tree indicating the delays. With this, the business objective of reducing the costs that are being incurred due to decommissioning delays can be obtained at an organizational level.

Chapter 3: Problem Statement

The decommissioning team is responsible for performing customer system/server decommissioning for the requests created in the form of JIRA tickets. An internal SLA of 30-60 days have been defined as the MTTR to perform the end-to-end decommissioning process. Any decommissioning request beyond 90 days is considered as delay.

There can be valid delays from the customer's end such as:

- The customer has requested not to go ahead with decommissioning during the process because the customer may want the system back and has extended the contract.
- The customer has not clarified for open questions from the CDM due to which the decommissioning has not commenced but the end date has been reached.
- The customer has not yet provided the clarity if they require the data to be backup prior to decommissioning.

Key Driver Analysis for MTTR – delay along with the delay period (in days) using appropriate Machine Learning Algorithms and to explore the key drivers leading to delay in the process. The purpose is to identify the factors affecting delays and in turn optimize the cost incurred by SAP for retaining the customer systems even after the contract has been ended.

Chapter 4: Objectives of the Study

The scope of this study is to identify the key drivers causing delays in the decommissioning process thereby increasing the MTTR. With this, it is imperative to focus on the important features causing delays in the end-to-end decommissioning process thereby optimizing the cost which SAP has been bearing due to the delays in decommissioning after the customer contract has been terminated.

SAP ECS team’s process management has detailed Standard Operating Procedure (SOP) and Work Instructions (WI) created for the decommissioning process. SOP captures the end-to-end decommissioning under various scenarios, task descriptions and roles performing the tasks have been outlined in the process. Various WI’s have been created that capture a detailed stepwise breakdown explaining the technical decommissioning process. However, there is scope for improving the process and there exist continuous process improvements being undertaken to make the process more efficient.

Using AI/ML approach to identify the key factors causing the delays can help optimize the process and provide a solution to operations. This can bring down the huge cost being incurred by SAP ECS for having the system in the SAP landscape even after the contract has been ended or the system end date has been reached.

A decision tree for a continuous variable which is MTTR in our project scope will be created by using CHAID (Chi-square automatic interaction detection) and CART (Classification and Regression Trees) techniques. MTTR is the independent variable, and the other variables will be the dependent variables for identifying the key drivers causing delays in the overall decommissioning process.

Machine learning algorithms are used in the project to solve the following objectives:

- *Key Driver Analysis that is leading to the delays in the process by predicting the MTTR using decision trees and KNN algorithms.*

- *Business objective further to the prediction – Cost savings beneficial by optimizing the process thereby reducing decommissioning delays.*

Chapter 5: Project Methodology

CRISP-DM framework has been used for this project. Cross-industry standard process for data mining, knowns as CRISP-DM is an open standard process model that describes common approaches used by data mining experts. It is a widely used analytics model.

CRISP-DM breaks the process of data mining into six phases: Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment. Figure 5.1 captures the phases. The sequences of phases are not strict and move back and forth between different phases as is always required. The arrows in the process diagram indicate the most important and frequent dependencies between phases. The outer circle in the diagram symbolizes the cyclic nature of data mining itself. A data mining process continues after a solution has been deployed. The lessons learned during the process can trigger new, often more focused business questions and subsequent data mining processes will benefit from the experiences of the previous one (Investopedia, 2021).

Figure 5.1 “CRISP-DM Framework” (Investopedia, 2021)

Chapter 6: Business Understanding

SAP ECS provides the transition of existing customer systems to the SAP HANA Enterprise Cloud. SAP ECS onboards the customers by provisioning the landscapes and once built, the systems are ready for operation and handed over to customers. Once the contract period ends, then the other process in the SAP ECS lifecycle is customer decommissioning process.

Customer decommissioning process has following scenarios in SAP ECS business:

- System decommissioning ("Sandbox no longer needed")
- Solution decommissioning ("Solution with multiple tiers is no longer needed")
- Landscape decommissioning ("End dates for entire landscape are reached/contract end date" of the customer).

Customer decommissioning can be the outcome of one of the following reasons:

- The customer contract has been ended.
- The customer wants to decommission only a system (also referred as tier) or a server or set of systems/servers.
- The customer has chosen early termination due to various reasons. Some of them are:
  - A customer moving out of one data center to another
  - The customer wants to move out of SAP
  - Customer is migrating to hyper scalar (either AWS or Azure or GCP)

Server: A server can be single or associated with the system –

- A server is a hardware running an operating system.
- A server can be virtual or physical.

- A server without an assignment to a certain system is just a ‘single server.’
- A server is not a system.

System: A system is a combination of many separate components –

- A running application on certain hardware.
- The application can provide different services and the server acts as a host of the application.
- The application saves the data in a database.
- The server is just one part of the system besides many others.
- Application, Services, Database, and Server are integrated into a huge environment
- All together is called "SYSTEM".

There are several trigger points for decommissioning in SAP ECS landscape. The customer decommissioning process has been categorized as Planned and Unplanned scenarios.

Planned Scenarios: Decommissioning needs to be triggered under the following circumstances. These are scenarios that will be communicated to the customer by the Client Delivery Manager (CDM) and have received customer confirmation to proceed with decommissioning.

1. Customer contract end date
2. Server replacement due to end of life
3. Server uplift/downgrade
4. Physical to Virtual server migration

Unplanned Scenarios: Following scenarios are said to be unplanned and it involves building a new server once the customer provides confirmation and the old server should be decommissioned.

1. Server is crashed
2. Server is not Online

3. Server issue

The end-to-end customer decommissioning process has been illustrated in the Figure 6.1. The process initiates with a JIRA ticket created by the Back-Office team which is a supporting team for CDM (Client Delivery Manager) organization.

JIRA ticket consists of the following important fields to be captured for decommissioning:

- Customer Contract End Date
- Is the Customer leaving SAP ECS?
- System/Server Decommission?
- Customer System Number
- Customer Server Number
- Network ID in case of a customer leaving SAP ECS
- Is customer data transfer required?
- Does server isolation is needed?
- Are there any customer data backup retention tickets?

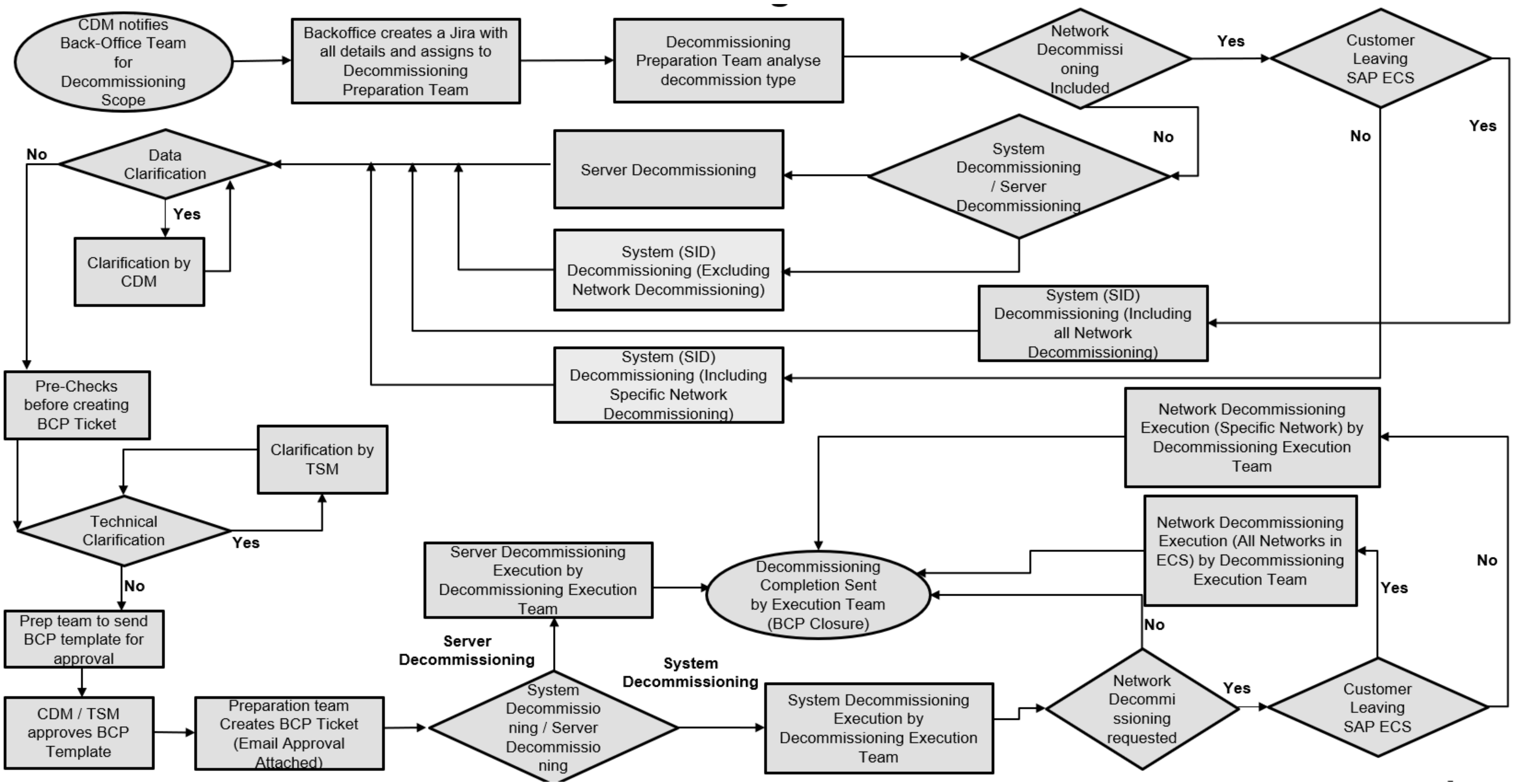


Figure 6.1 Decommissioning Process Flow

Based on the parameters in the JIRA ticket, specified decommissioning will be performed. Decommissioning automated procedures have been created to ease the decommissioning process. However, in case of procedure failure, the responsible team will be notified through tickets being created for the specified task and they perform the respective task. Post Processing has various steps to be completed before the customer has been notified that the decommissioning process has been completed.

Server isolation has been an important parameter that needs to be clarified before decommissioning process being triggered. If an application system consists of more than one server being connected or sharing the same database and if that server has been requested for decommissioning, then that server needs to be isolated from the system before commencing the decommissioning process. The customer should confirm if we are good to go ahead concerning isolating the server. Delays can be possible if this information is not captured on the ticket. If server isolation is not performed, then alerts might get generated if a server is associated with a system. Figure 6.2a and Figure 6.2b refer to server isolation. Figure 6.2a shows that server1 and 2 are connected to the same database tenant. TSM confirms if the server needs to be isolated. Once the TSM confirms that the server needs to be isolated, then the server will be isolated as shown in Figure 6.2b. This is a crucial step if the requirement is to decommission a single server so that the server needs to be isolated else there will be an inconsistency in the different tools used at SAP ECS.

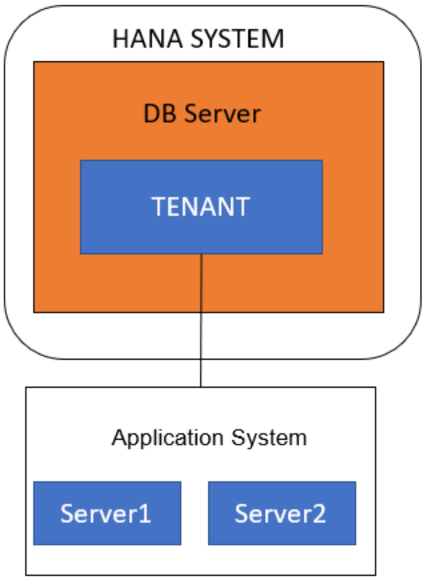


Figure 6.2a Before Server Isolation

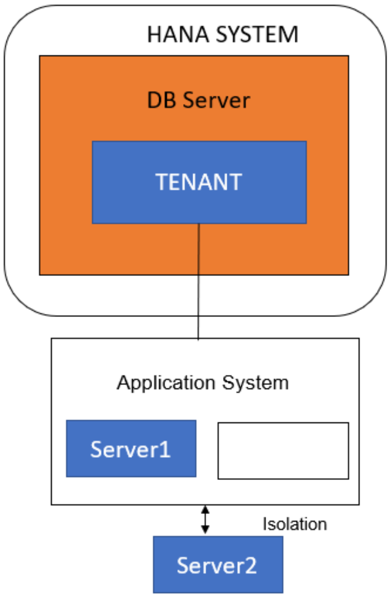


Figure 6.2b After Server Isolation

Decommission execution process consists of various tasks that will be performed for the successful completion of decommissioning. Figure 6.3 illustrates the various tasks under decommission execution. Hardware resizing refers to downsizing the hardware to 80% of its original size. This ensures the system is not completely decommissioned by giving the scope if the customer returns requesting for the system not be decommissioned as the contract might be extended. The cooling Period refers to the period where the system will be kept with a bare minimum configuration for 5 days to provide a small window in case the customer may not want to go ahead with decommissioning. During the post-cooling period, the other stages in the decommissioning lifecycle proceeds until the decommissioning has been completed. In case of procedure failure, the teams will be notified by tickets so that they can perform the necessary task to complete the process.

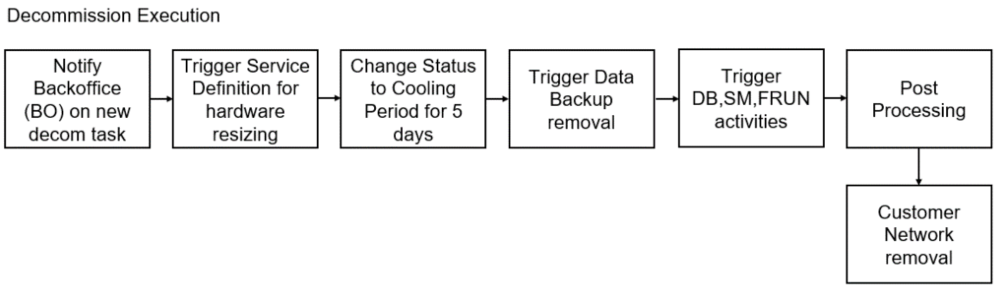


Figure 6.3 Decommission Execution Process

There have been significant delays in the end-to-end decommissioning process. The delays have been identified in some of the stages of processing such as deletion of backup and customer network removal causing huge delays in the process. Internal SLA has been aligned with the respective teams, however, there have been delays due to these two stages. The delays add up to costs SAP must bear retaining the customer system in the SAP landscape even after the contract has been terminated.

This project aims to provide algorithmic solutions to the team in identifying the key drivers for delays based on the features like several servers requested for decommissioning, type of server, if the customer leaving SAP, the number of days taken to isolate the server, days computed for every execution stage will be fed as the independent variable and MTTR being the dependent variable. Based on the predictions, delays in customer decommissioning can be reduced and recommendations can be provided to optimize the process thereby saving cost.

An attempt is being made to develop a classification system to identify the reason for delays by establishing the key driver analysis, based on a training set of data that includes cases with known outcomes.

Chapter 7: Data Understanding

7.1 Collecting the Initial data

The analysis in the following sections is based on the JIRA data set that provides technology products and services. The JIRA data is in monthly frequency and have collected data from January 2021 till July 2022.

Data collected from the JIRA tool are structured and in masked format. Data is masked so that the organizations and customers' confidentiality can be maintained.

Below is the list of features in the given dataset in the Table 7.1.

Features	**Description**
Key	Unique tickets created in JIRA
Summary	Ticket description
Status	Ticket status
Assignee	Last assignee worked on the ticket
Created	Ticket creation date
Resolved	Ticket resolution date
Updated	Ticket last updated date
Priority	Priority of the ticket
Sub-Contractor	Vendor worked on the ticket
Types of Decommission	Server, System, Network or customer leaving SAP
No. of Servers	Total number of servers to be decommissioned
Cost (Euros)	Cost for maintaining the server
MTTR (Days)	Mean Time to Resolve (Resolved – Created)
Isolation (Days)	Number of days taken to isolate a server
Downsizing (Days)	Number of days taken to downsize the server
Cooling Period (Days)	Number of days the server was kept on cooling period
Backup Removal (Days)	Number of days taken to delete data backup
Database Removal (Days)	Number of days taken to delete database
Volume's deletion (Days)	Number of days taken to delete server volume
FRUN (Days)	Number of days taken to delete FRUN
Network (Days)	Number of days taken to delete Network

Table 7.1 Data Dictionary

In Figures 7.1a, 7.1b, and 7.1c, a snapshot of the data considered for the study can be visualized. Data is extracted from JIRA tickets and outputted in an excel file. Various fields considered in the study have been mentioned.

Issue Type	Key	Summary	Status	Assignee	Resolution	Component/s
Decommissioning	HCBT-5468	Backoffice – Decommissioning Request – Stedin Holding N. V.	Resolved	Kammadanam, Suchitra (external - Service)	[+] Solved	EMEA
Decommissioning	HCBT-5454	Backoffice – Decommissioning Request - [TSUDAKOMA]	Verified & Closed	Vennapusa, Kaveri (external - Service)	[++] Verified	APJ
Decommissioning	HCBT-5436	Backoffice – Decommissioning Request – Blue Empreendimentos E Participacoes Ltda.	Resolved	Maradana, Prasanna kumar (external - Service)	*Unresolved*	LAC

Figure 7.1a Snapshot of Data

Created	Resolved	Updated	Priority	Sub-Contractor	Type of Decommission	No of Servers	No of Systems	No of Network
29-12-2021	01-04-2022	01-04-2022	Medium	NTT	Server	7	4	
29-12-2021	12-04-2022	12-04-2022	Medium	NTT	System	10	7	
28-12-2021	22-02-2022	22-02-2022	Medium	NTT	Server	29	26	

Figure 7.1b Snapshot of Data

Isolation (Days)	Downsizing (Days)	Cooling period (Days)	Backup removal (Days)	Db removal (Days)	Volume deletion (Days)	FRUN (Days)	Network (Days)
13.95	0.93	9.3	18.6	9.3	9.3	9.3	22.32
15.6	1.04	10.4	20.8	10.4	10.4	10.4	24.96
8.4	0.56	5.6	11.2	5.6	5.6	5.6	13.44

Figure 7.1c Snapshot of Data

7.2 Describing the data

The dataset contained 1,624 JIRA tickets created between January 2021 and July 2022. For confidentiality, customer wise details were not available for study and hence only ticketing data were considered.

Most fields in the dataset are continuous features by which certain level of correlation have been defined between levels. The most important numerical features are **MTTR** computed in **days**: the number of days taken to complete the decommissioning process. The **Cost** is a derived field. **Cost (in Euros)** has been derived considering 5000 Euros per month is needed to maintain the server by SAP. Cost is computed considering the total number of servers and the amount needed to maintain the server.

7.3 Target Setting

The main goal of the machine learning models should be able to predict the key drivers influencing the dependent variable MTTR. Decision trees will be built to identify the key drivers which can be analyzed, and recommendations based on the key drivers can be provided to the team to optimize the process thereby reducing costs to be incurred by SAP.

MTTR computed over time has been depicted in Figure 7.2. MTTR has been derived by considering the difference between the date the JIRA ticket has been created to the date the ticket has been resolved. The value considered is in days indicating the total number of days taken to complete the decommissioning. The Created month of the tickets is on X-axis and the total number of days to resolve is on Y-axis.

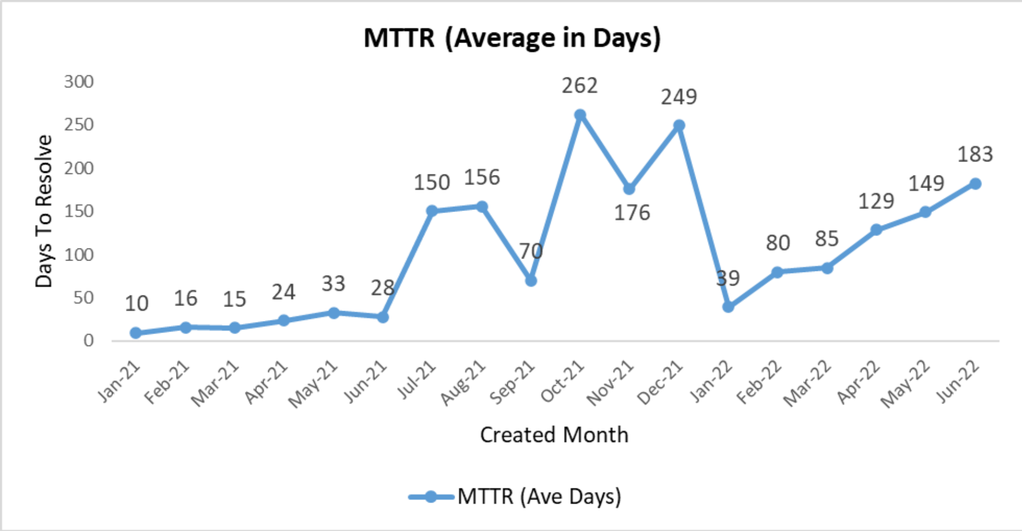


Figure 7.2 MTTR in Days

Figure 7.3 Boxplot for MTTR

The boxplot distribution for MTTR in days is shown in Figure 7.3. The median (middle quartile) marks the mid-point of the data and is shown by the line that divides the box into two parts. It indicates that more than half the scores are below 200, indicating the average of days taken to close the tickets is less than 200 days.

7.4 Exploratory Data Analysis (EDA)

There are three types of decommissioning in SAP ECS:

- **Application System Decommissioning:** A complete system identified by System Number and System ID shall be decommissioned. This includes evening that belongs to the system such as:
    - Application and Database
    - All assigned servers (hosts)
    - Repository changes (like lifecycle status) will be set to "Deleted"
    - Removal out of the Solution Manager monitoring
    - Removal out of the Backup configuration
    - Complete de-integration from SAP ECS environment

- **Server Decommissioning:** One or more servers (hosts) without a system relation shall be decommissioned. Some of the examples of server decommissioning are listed below:
    - A customer wants to downsize a system by deleting one application server
    - Server replacement due to a hardware failure
    - The server is redundant after migration.

- **Network Segment Decommissioning:** When the customer is leaving SAP or terminating his contract with SAP then it is necessary to decommission the infrastructure (network) components. But it is also possible to raise a decommission request for a single network segment in case it is no longer needed.

Delay in the decommissioning process - The ideal runtime considered by the SAP ECS team based on previous data is that the average time to complete decommissioning is 30 days. In case of the total number of servers is huge, then the runtime can be averaged to 60 days. MTTR beyond this has been considered as a delay. The total number of tickets considered for the project is **1624** created between the period January 2021 to July 2022.

The percentage of tickets open vs closed is depicted in Figure 7.4. Out of **1624** tickets, **1366 (84%)** of the tickets are closed and **258 (16%)** of the tickets are still under various open statuses.

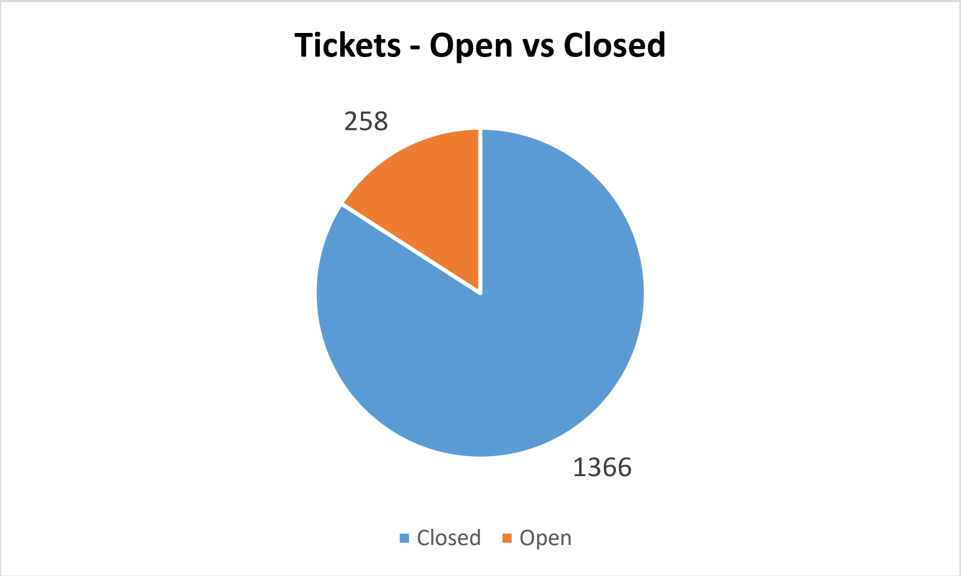


Figure 7.4 Percentage of tickets open vs closed

The ticket trend (Created vs Resolved) month over month Figure 7.5. On an average, over 90 tickets have been created for various decommissioning requests. It is unable to predict if there are any seasonal factors influencing the customers being terminated during the same period. Further study on seasonal variation shall be done as future analysis to identify if there are any patterns indicating the same.

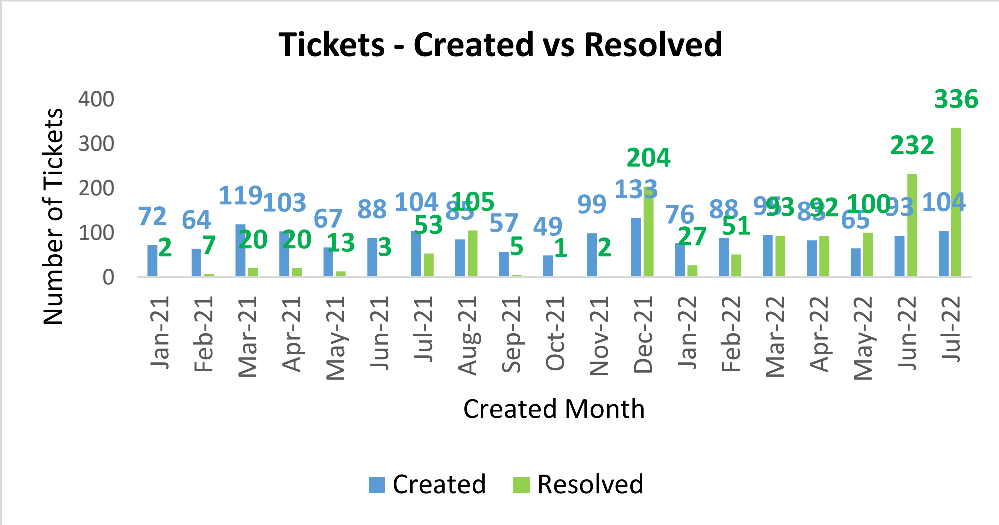


Figure 7.5 Trend – Created vs Resolved

One of the important analyses is whether the decommissioning requests are for systems/servers or customers leaving SAP and the total number of servers can be an important factor influencing MTTR because more the number of servers requested for decommissioning, the more the delays.

Types of Decommissioning requests: Figure 7.6 provides an insight into the types of decommissioning requests placed within SAP ECS. As per the data, only **7%** of customers are terminating their contract, hence leaving SAP ECS. Most decommissioning requests are server decommissioning contributing to 65% of the dataset. This indicates the number of servers plays a key role in a decommissioning request. System decommissioning on the other hand consists of servers to be decommissioned. The Client Delivery Manager (CDM)'s main responsibility is to convince the customers not to leave SAP. One of the developments in SAP ECS is migration of customers to hyper scalers (Cloud) such as AWS, Azure or GCP (Google Cloud Platform) which has been completed as a migration project and hence this is one of the key developments to retain the customers within SAP.

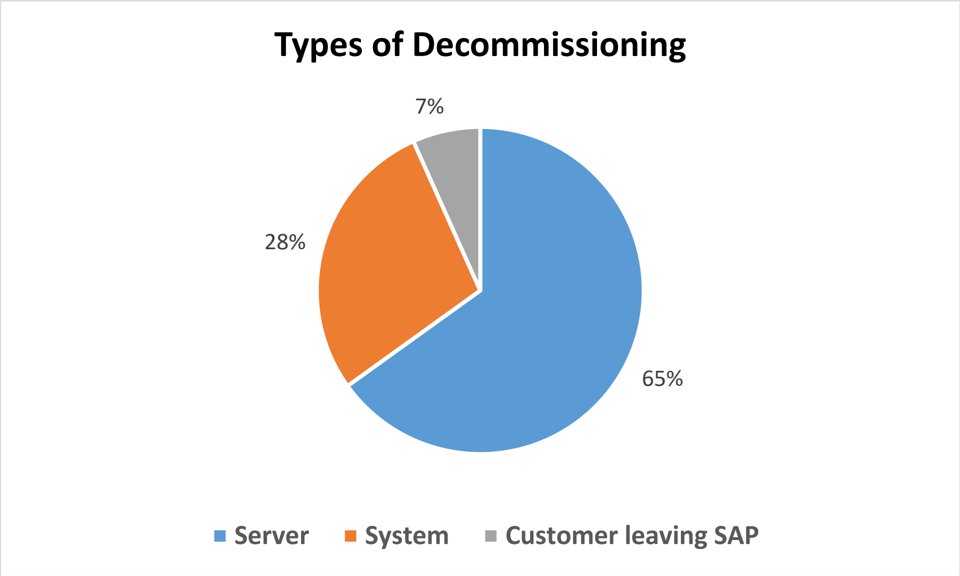


Figure 7.6 Types of Decommissioning

Tickets Open (in Days) since created: Open tickets have been categorized under different buckets as specified in Figure 7.7. 11% of the tickets are in the "Open" state between 60 – 90 days bucket and 34% of tickets beyond 90 days. The team need to focus on these tickets as they are way beyond 30 – 60 days runtime SLA and hence these tickets must be prioritized for completion.

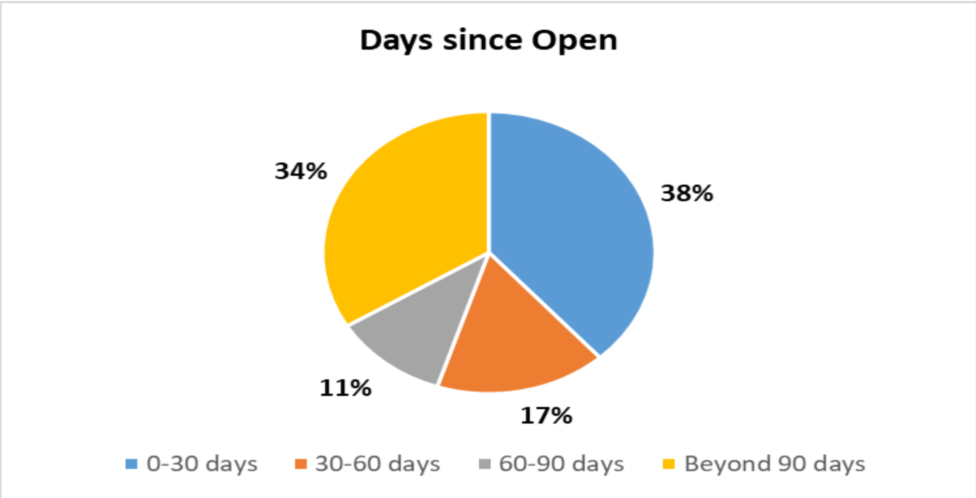


Figure 7.7 Tickets Open (in Days) since created

Statuses of Open tickets: Figure 7.8 shows the percentage distribution of tickets under various statuses.

Tickets under the following statuses are the tickets that have been considered for decommissioning execution. However, the team need to ensure there are no delays as the decommissioning process have already been commenced for 77% of “Open” tickets.

- Execution in process
- In Process by Decom Prep Team

Following statuses together contributes to 15% of open tickets need immediate attention as these tickets are awaiting responses from CDM (Client Delivery Manager), TSM (Technical Service Manager) as they need clarification prior to proceeding with decommissioning process. Further follow-up with CDM and TSM need to be done immediately so that the delays can be reduced.

- Pending for CDM Review
- Pending for TSM Review
- Execution on Hold

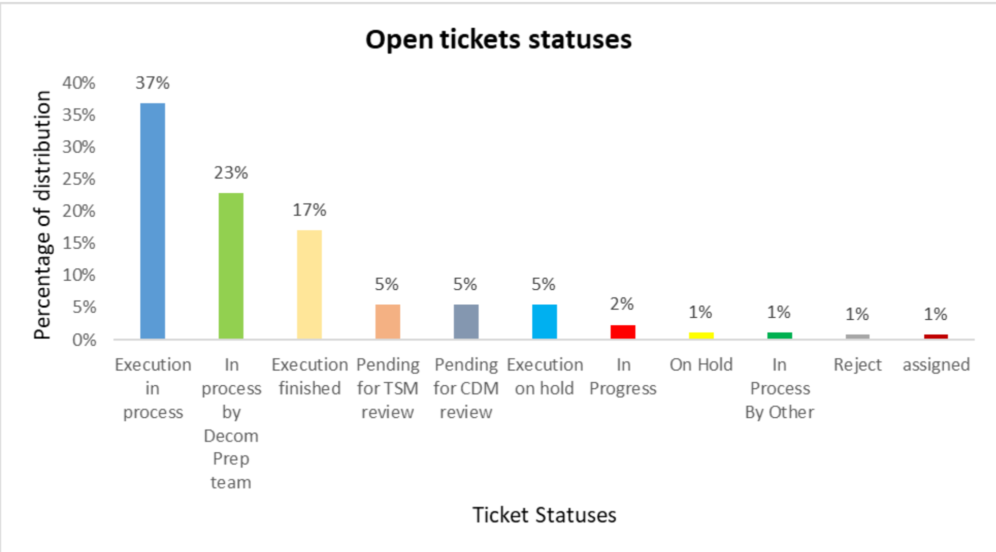


Figure 7.8 Statuses of Open tickets

Days since tickets last updated: Another important feature has been derived based on the date since the ticket last updated. This can be computed by considering the difference between days since last updated to today and all open tickets. Figure 7.9 shows the percentage distribution of tickets last updated indicating that **48%** tickets which are beyond 60 days have no updates. **5%** of the tickets have no updates between 45 – 60 days. A total of 53% of tickets need immediate attention from the team.

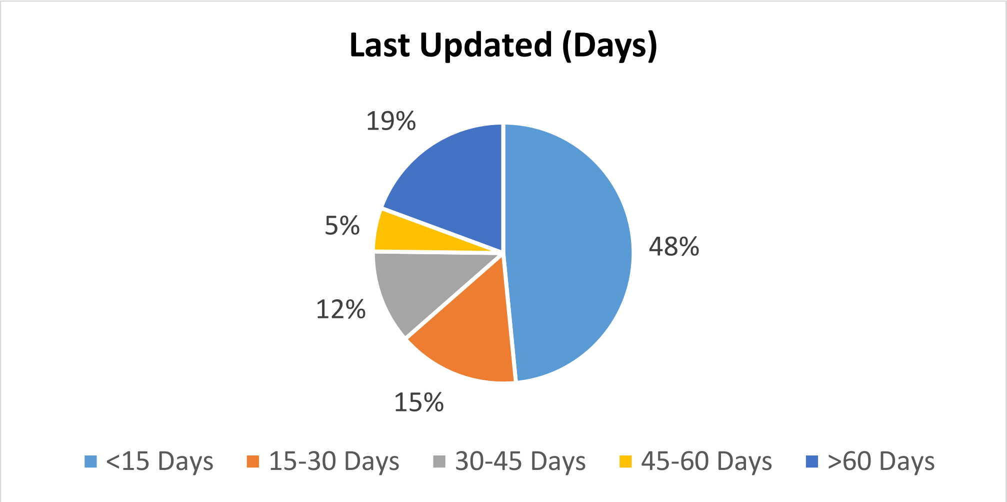


Figure 7.9 Days since tickets last updated

Sub-Contractors or Vendor analysis: Decommissioning requests are worked upon by different vendors of SAP. Figure 7.10 showcases the percentage distribution of vendors who worked on various decommissioning requests. In the study, the scope of vendor analysis is to identify if the type of vendor can be a key driver causing the delays in decommissioning. Further variables such as "Staff Competency", "Absenteeism rate", "Number of staff employed", and "staff experience" shall be captured in future studies that can help provide recommendations to vendors for optimizing the decommissioning process. Following vendors are responsible to work on decommissioning requests.

- HCL
- NTT
- TCS
- Accenture

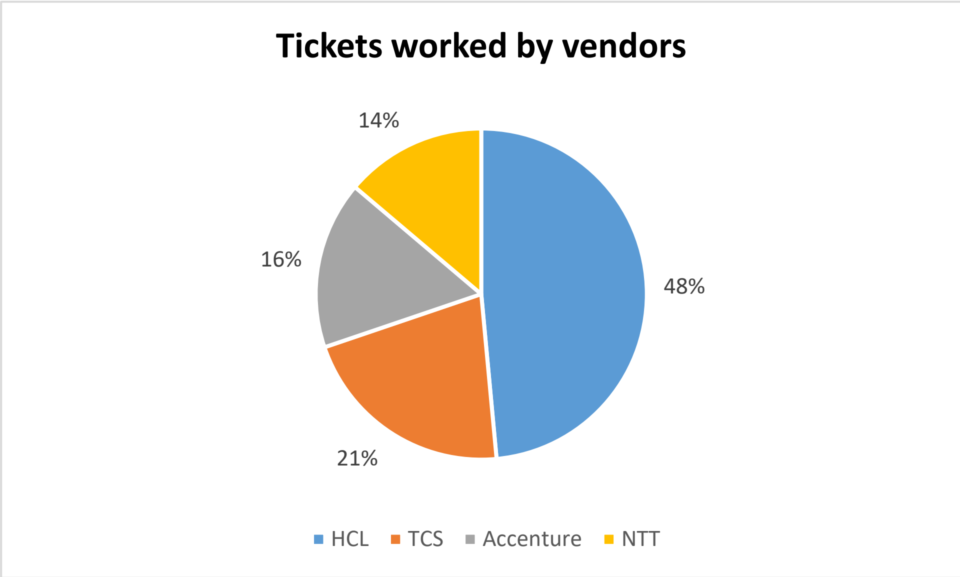


Figure 7.10 Distribution of tickets worked by vendors

Open tickets vendor wise: Figure 7.11 shows the percentage of tickets that are still open vendor wise. 37% of tickets are assigned to HCL and 31% tickets have been assigned to TCS.

Figure 7.12 shows the percentage of tickets combined under the following statuses for each vendor.

- Execution in Process
- Execution finished
- Pending for Review (CDM/TSM)
- On Hold

This provides a view to further analyze vendor management by gathering related features as part of future studies.

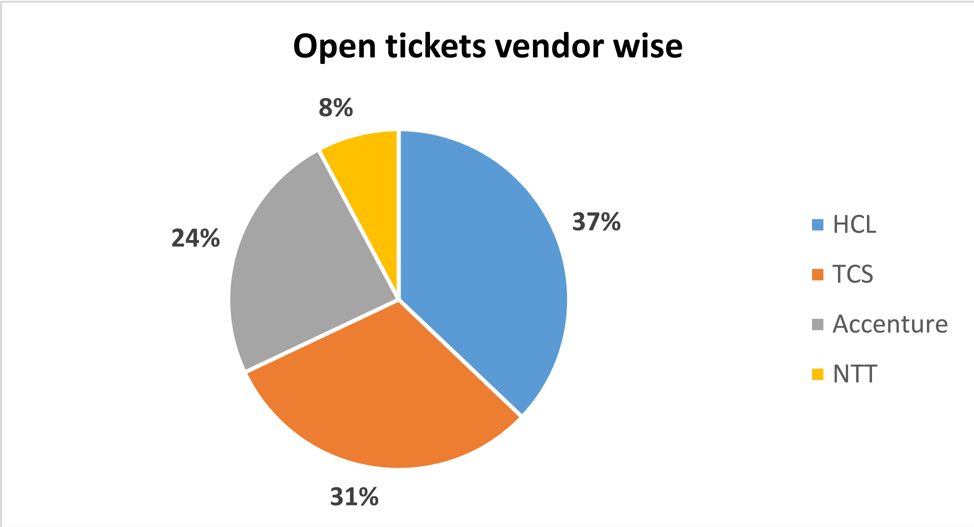
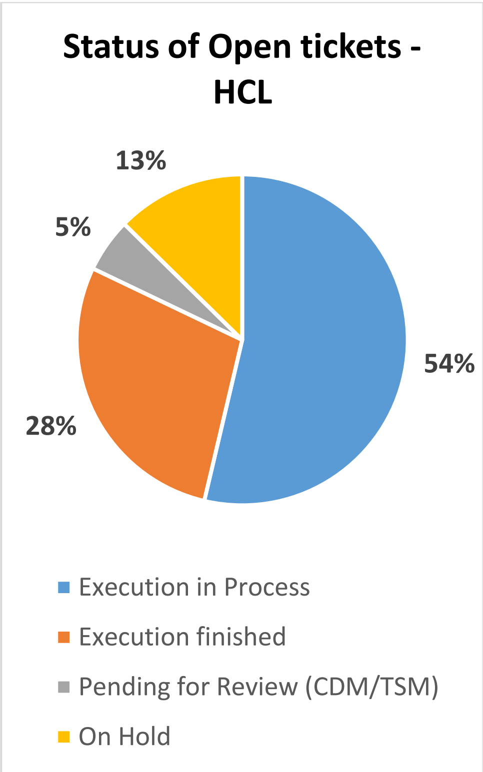
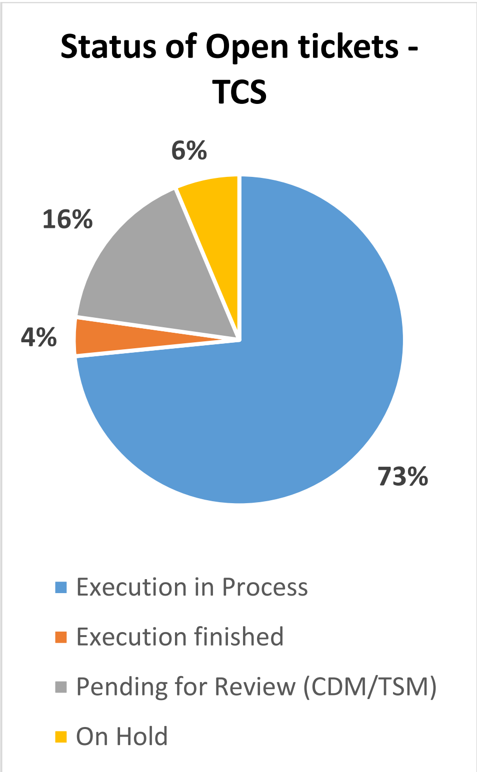


Figure 7.11 Tickets Open – Vendor wise





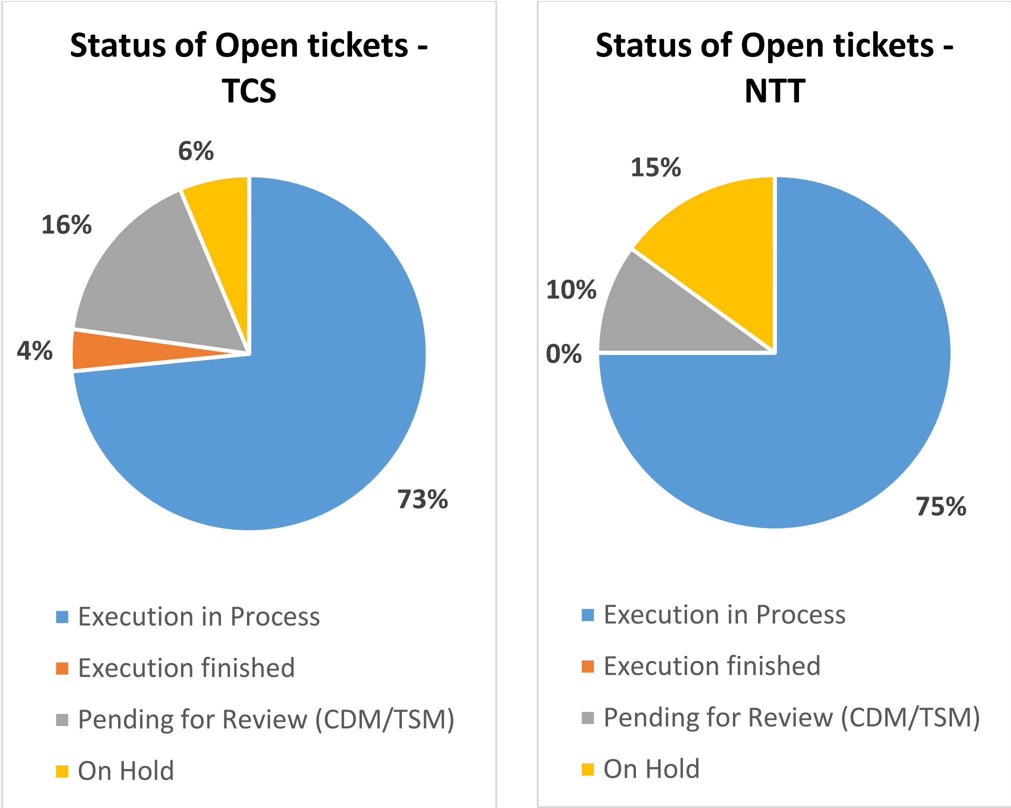


Figure 7.12 Status of Tickets (Open) – Vendor wise

Delays in decommissioning stages: Data with respect to the days spent has been captured for various stages in the decommissioning. Table 7.2 depicts the days on average has been spent on each stage in the decommissioning process. Various teams are involved in the decommissioning process. There has been an internal SLA agreed with the teams responsible for different stages. Deletion of the Network in case of a customer leaving SAP has been attributed to major delays. There has been an internal SLA of 7 days agreed with the infrastructure team for network decommissioning, however, on average, it takes 45 days to decommission the network adding to major delays in decommissioning. Also, deletion of data backup is another major factor contributing to delays in decommissioning. An SLA of 3 days has been agreed upon in the recent past with the team performing the backup deletion. Figure 7.13 depicts the delays in decommissioning stages.

Stages of Decommissioning	Days in Average	SLA in Days
Network (Days)	45	7
Backup removal (Days)	37	3
Isolation (Days)	28	1
Cooling period (Days)	19	5
Db removal (Days)	19	2
Volume deletion (Days)	19	2
Downsizing (Days)	2	1

Table 7.2 Delays in Decommissioning stages with internal agreed SLA

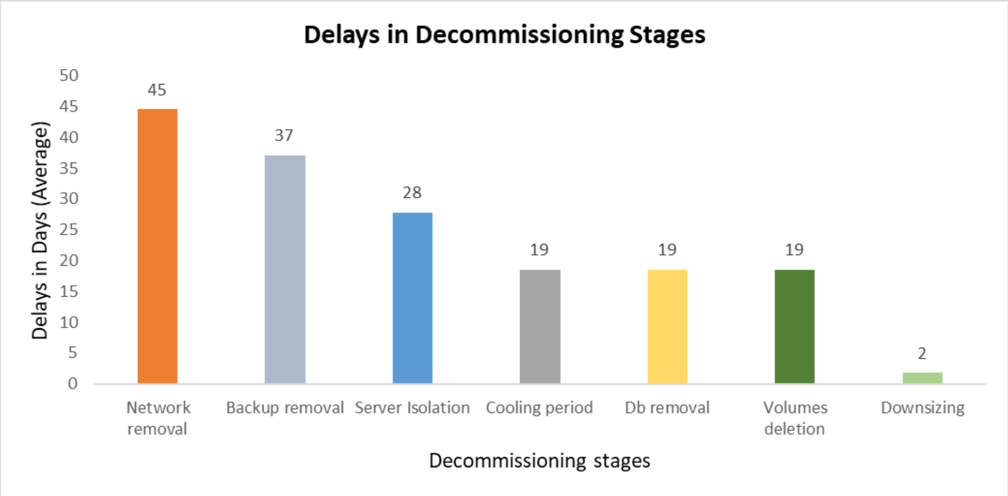


Figure 7.13 Delays in Decommissioning stages (Average in Days)

Cost Optimization: Customer Once the customer terminates the contract or if the system has reached its end date, then the decommissioning process should get completed at the earliest. The customer stops paying for the systems/servers once the end date has been reached. In case of delays in decommissioning process, the server maintenance cost per month is estimated to be 5000 euros per server, and SAP ECS has to bear the cost till the systems/servers have been decommissioned.

Table 7.3 showcases the number of servers that are yet to be decommissioned and the cost associated with it. Hence it is highly crucial to decommission the systems at the earliest to save cost.

Year	No of Servers	Cost (Euros)
2021	**255**	**€ 12,75,000.00**
2022	**900**	**€ 45,00,000.00**
Grand Total	**1155**	**€ 57,75,000.00**

Table 7.3 Servers yet to be decommissioned and the cost associated with it

Descriptive statistics for the continuous variables considered further have been showcased in Table 7.4 that generalizes the data to the broader population.

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
No of Servers	1366	2.0	97.0	45.131	23.8780
Isolation (Days)	1366	.0000000000000000	66.3000000000000000	27.824890190336742	18.549574281688752
Downsizing (Days)	1366	.00	4.42	1.8550	1.23664
Cooling period (Days)	1366	.0000000000000000	44.2000000000000000	18.549926793557820	12.366382854459166
Backup removal (Days)	1366	.0000000000000000	88.4000000000000000	37.099853587115640	24.732765708918333
Db removal (Days)	1366	.0000000000000000	44.2000000000000000	18.549926793557820	12.366382854459166
Volume’s deletion (Days)	1366	.0000000000000000	44.2000000000000000	18.549926793557820	12.366382854459166
FRUN (Days)	1366	.0000000000000000	44.2000000000000000	18.549926793557820	12.366382854459166
Network (Days)	1366	.0000000000000000	106.0800000000000000	44.519824304538750	29.679318850702007
MTTR (Days)	1366	.0	442.0	185.499	123.6638
Valid N (listwise)	1366				

Table 7.4 Descriptive statistics for continuous variables

There is a crucial need to identify the delays and reduce the cost spent due to delays in decommissioning. The objective is to identify key drivers using machine learning algorithms for saving costs.

Chapter 8: Data Preparation

Data preparation is a step that requires the most time and effort. The Data Preparation stage aims to address different problems as mentioned below –

- Selecting columns and dropping duplicate rows
- Treating Missing Values
- Checking for Outliers
- Coherence Checking
- Data Transformation and Feature Engineering

JIRA tickets have been collected by downloading the data from the JIRA tool. Data is then masked and shared for analysis and study.

The time window considered for the study is January 2021 to July 2022. Data is missing the Resolution date for tickets that are still open. Out of 1634 tickets, 258 tickets are still open. Hence this data has been separated from analysis and only 1366 tickets that are closed are considered for our study. There are no missing data in any of the columns and hence no further imputing is required.

Based on the Exploratory Data Analysis (EDA) and the domain experts' suggestions, the following features have been removed for further analysis as shown in Table 8.1.

Features	**Comments for dropping the features**
Key	Unique tickets for decommissioning requests
Summary	Dropped as descriptive field
Assignee	Dropped
Component/s	Dropped as region is not a factor
Priority	Post factor feature, hence dropped
Number of Systems	Dropped as total number of servers has been retained
Number of Network	Insignificant as customer leaving SAP is marginal

Table 8.1 Reasons for dropping features

Data Preparation Steps:

- Dataset is split into two categories training and test set.
- The training and test dataset contain MTTR, a continuous variable providing the days taken to resolve the ticket and complete decommissioning.
- The test set is used as the deployment data for the re-validation of the model.

8.1 Selecting data and dropping duplicate rows

Out of 24 columns, only 14 have been selected candidates for feature engineering part. It was based on several characteristics of each column. In short, irrelevant columns that do not represent valuable information to the problem were dropped.

Categorical columns with too many levels are also irrelevant. In fact, their use will undermine the model's performance as there will be fewer examples available for each level of training, which will reduce the ability of each level to have any impact on the model's outcome (Quiry et al., 2009).

It was checked and found that there were no **duplicate rows** on the rows level in the dataset.

8.2 Treating Missing Values

Missing values refer to data points that are not present in the dataset. They could be missing at random, for example, because the data was entered incorrectly, but their absence could also be indicative of something so that their absence is not random (Kang, 2013).

There are no missing data in any of the columns and hence no further imputing is required.

8.3 Checking for Outliers

Outliers are values that differ significantly from the rest of the data. It is possible that they represent errors that were accidentally inserted into the dataset, or they may simply represent extreme values that are demonstrating variance (W et al., 2018). Domain knowledge can be used to differentiate between the two.

Decommissioning requests that are pending for a very long time are rare and usually involve special contracts that should be considered separately.

8.4 Feature Extraction

The last step of Data Preparation stage is feature engineering. In machine learning, models learn by searching for patterns in their input data. That input is what practitioners call the features: an informative representation of the data in numerical form (Zeng et al., 2008). During the feature engineering phase, data is transformed in order to create predictive features for the model to interpret.

The Data Understanding component outlined the details of all the attributes considered in this dataset. Following are some of the issues with respect to attributes:

- It is possible that the details of attributes are enormous and dispensable for management.
- There have been scenarios in which categorical data is saved numerically.
- Vendor information is not sufficient but can be gathered.

By using machine learning to identify the key drivers, a subgroup of attributes has been chosen to create a new set of attributes in order to promote discoveries as well as to boost the interpretation of the results.

New features have been created to train the classifier. A delay in decommissioning process may be the result of multiple factors. There is a direct correlation between the number of servers requesting for decommissioning and the delays (MTTR) due to the volume. Lesser number of servers can be decommissioned sooner. Feature extraction and selection become highly crucial in increasing the accuracy of the model.

New features extracted for learning

New features have been created to increase the efficiency and performance of machine learning models. Table 8.2 details out the features created in this study.

Features	Explanation
Cost (in Euros)	Cost calculated in Euros
MTTR (in Days)	Time taken to resolve the ticket (Resolved – Created)
Tickets Open (in Days)	Number of tickets still in Open status
Days since open buckets	Days since the tickets Open
Tickets closed buckets	Tickets closed in buckets
Open since last update	Last update in days for open tickets
Last update bucket for open tickets (in Days)	Last update in days bucket for open tickets

Table 8.2 List of Features Created

Chapter 9: Modeling

In this chapter, an overview of machine learning procedures is provided and explained the process of predictive modeling.

Data discussed in the previous section has been fed into multiple models to get the key drivers influencing the MTTR. Based on the problem statement and the data availability, the Decision Tree algorithm, a supervised learning technique has been considered which is used to predict the key drivers and their relationship with MTTR based on the training set. Programs learn from the given data and classify it or groups.

Before implementing the machine learning models, data sets are arbitrarily divided into training and testing sets. 70% of data is used as training sets and 30% of data is used as testing sets. The training set is used to train the decision tree model with certain features of tickets, such as the number of servers, cost, isolation, downsizing, cooling period, backup removal, database removal, volume deletion, FRUN, and network deletion. A Testing set is used to evaluate the performance of classifiers.

9.1 Evaluation metrics

The metrics that have been used in the model are detailed below:

Entropy: It is the measure of randomness in the data. It provides the impurity present in the dataset. It varies between 0 to 1.

Information Gain: It is the measure of Entropy. It calculates information a feature can provide us about a class. Based on the value of information gain, the node is split to build a decision tree.

Gini Index: It is the measure of purity or impurity used in a decision tree in CART. An attribute with a lower Gini Index has to be preferred.

Chi-Square: It is used to find out the statistical significance between the sub-nodes and parent nodes. The bigger the value of chi-square implies a bigger statistical significance of differences between sub-nodes and parent nodes.

Variance: Variance reduction is used for continuous target variables. This algorithm uses the standard formula of variance to choose the best split.

9.2 Machine Learning Algorithms for Supervised learning

The following algorithms were applied for this dataset:

- Decision Trees
- K-Nearest Neighbors

9.2.1 Decision Tree classification

The decision tree methodology is used for establishing classification based on multiple covariates or even for developing predictions for a target variable. (Song & Lu, 2015) Decision trees are one of the effective methods for data mining. The tree structure provides an easy way to interpret the results.

Common usage of decision tree models includes the following:

Variable selection: The number of independent variables considered for the model includes continuous variables which form the tasks in the end-to-end decommissioning process. MTTR is the dependent variable and all the independent variables considered are various factors contributing to delays in MTTR.

The relative importance of variables: Once the independent variables have been identified, the key factors driving the delays can be ascertained by the model. Variable importance is computed based on the reduction of the model accuracy when a variable is removed.

Handling of missing values: Omitting missing values is an incorrect way of handling missing values. A decision tree can classify missing values as a separate category or replace the missing ones with predicted values.

Prediction: Using the tree model derived from historical data, it's easy to predict the result for future records.

Data manipulation: In the case of highly skewed continuous data as in our study, decision tree models help in binning the continuous variables into a manageable number of ranges in an optimal way.

Decision Tree model output:

Figure 9.1 shows the decision tree built by considering **MTTR** as a dependent variable and the following as independent variables to determine the key drivers influencing MTTR.

Independent variables: Sub-Contractor, No of Servers, Isolation (Days), Downsizing (Days), Cooling period (Days), Backup removal (Days), Db removal (Days), Volume deletion (Days), FRUN (Days), Network (Days).

CHAID (Chi-squared Automatic Interaction Detector) has been used as a growing method and the key drivers are - Isolation (Days), Sub-Contractor (Vendor).

The MTTR based on two input parameters – Isolation (Days) and Vendor analysis has been evaluated in Figure 9.1. If the isolation days are between 2.7 – 6.7 days, then the predicted MTTR is 66 days. If the isolation is less than 2.7 days, then the predicted MTTR is 9 days.

Also, the sub-contractor or vendor can perform decommissioning with MTTR being 101 days and the isolation being between 12.9 to 19.3 days.

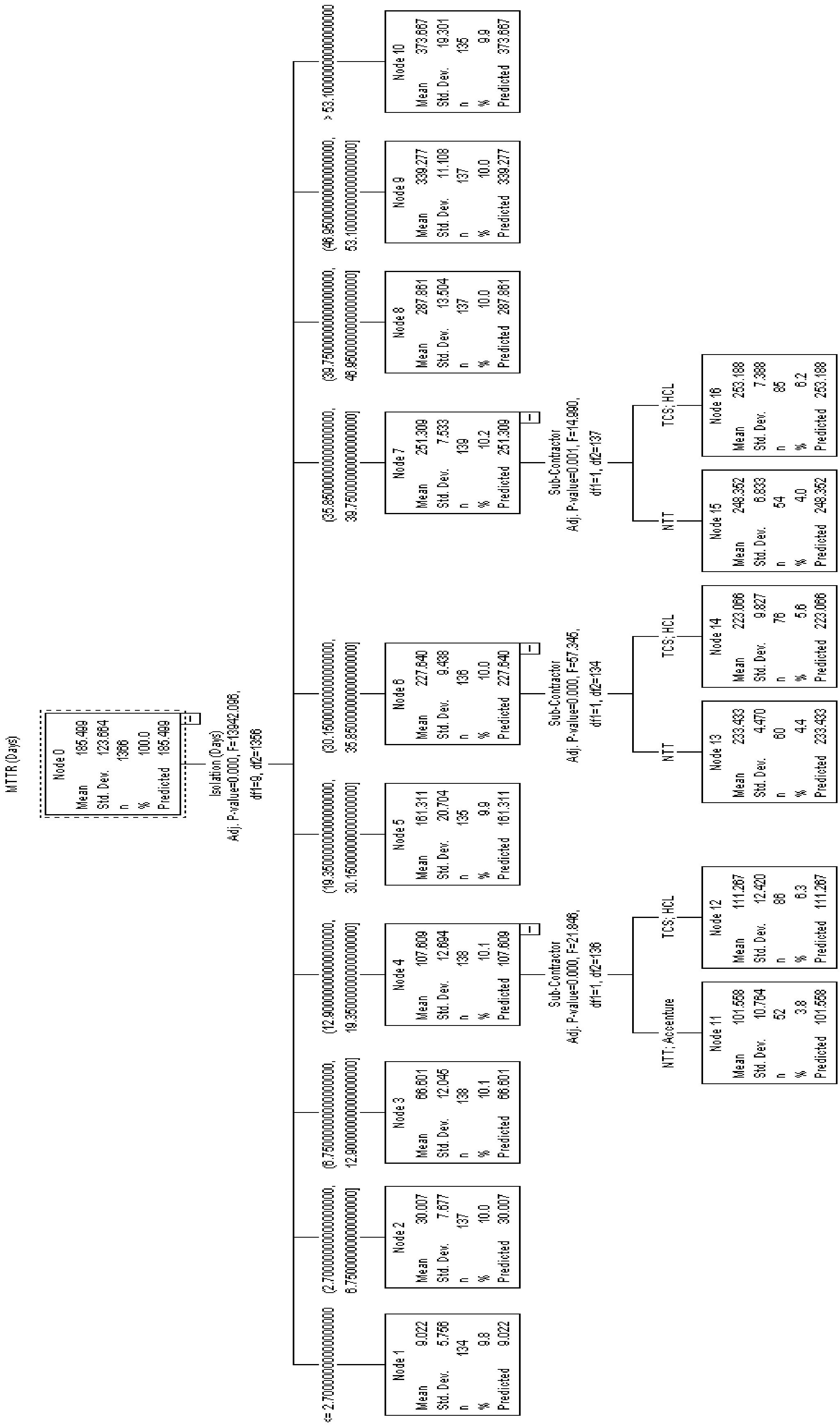


Figure 9.1 Model Summary – Classification Tree Output

Another decision tree evaluates the MTTR based on one input parameter – The cooling period (Days). If the cooling period is between 1 – 8 days, then the predicted MTTR is 31 days. If the cooling period is >8 days, then the predicted MTTR is 129 days. SLA for the cooling period is 5 days, however, there are scenarios that the team has missed reviving from the cooling period due to priorities and hence this adds to delays in overall decommissioning.

The decision tree built by considering **MTTR** as a dependent variable and the following as independent variables to determine the key drivers influencing MTTR are Cooling Period (Days), Sub- Contractor and No. of Servers.

Independent variables: Sub-Contractor, No of Servers, Isolation (Days), Downsizing (Days), Cooling period (Days), Backup removal (Days), Db removal (Days), Volume deletion (Days), FRUN (Days), Network (Days).

CHAID (Chi-squared Automatic Interaction Detector) has been used as a growing method and the key drivers are – Cooling Period (Days).

9.2.2 k-Nearest Neighbour (KNN) Algorithm

In this algorithm, we are dividing the dataset into training and test data. Selecting the value of K, determining which distance function should be used, choosing a sample from test data (as a new sample) and computing the distance to its n training samples, sorting distances gained and taking k-nearest data samples, and finally, assigning the test class to the sample on the majority vote of its k neighbours (Nabipour et al., 2020). Table 9.1 provides the summary in KNN. A sample of 697 tickets was considered for training and 321 tickets for testing.

Case Processing Summary			
		N	Percent
Sample	Training	697	68.5%
	Testing	321	31.5%
	Valid	1018	100.0%
	Excluded	617	
	Total	1635	

Table 9.1 KNN model summary

Figure 9.2 provides the predicted value (MTTR in Days) for the data considered in the model. An error of 12% deviation is estimated from the original value. Figure 9.3 provides Normalized importance for variables in the KNN algorithm.

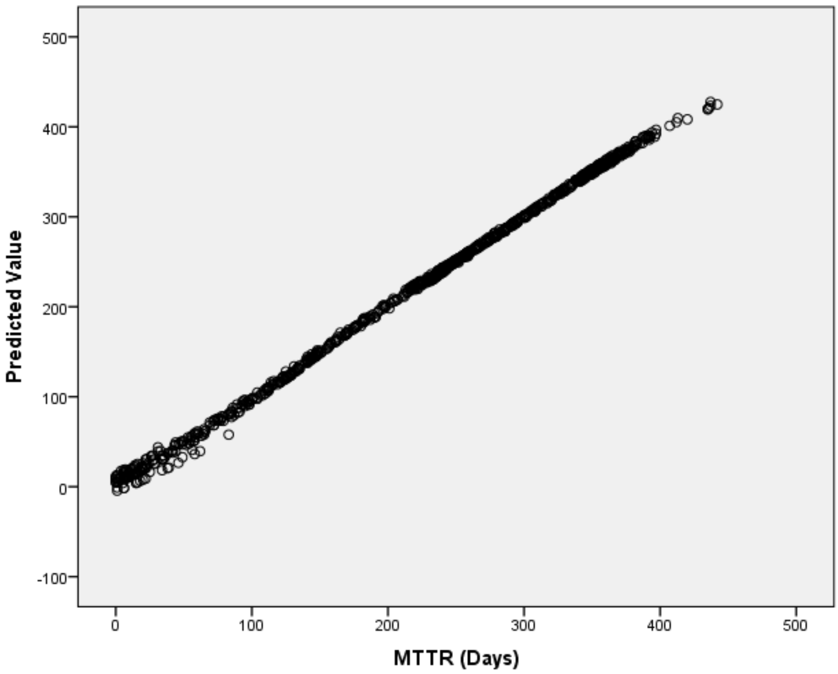


Figure 9.2 MTTR Predicted value using KNN algorithm

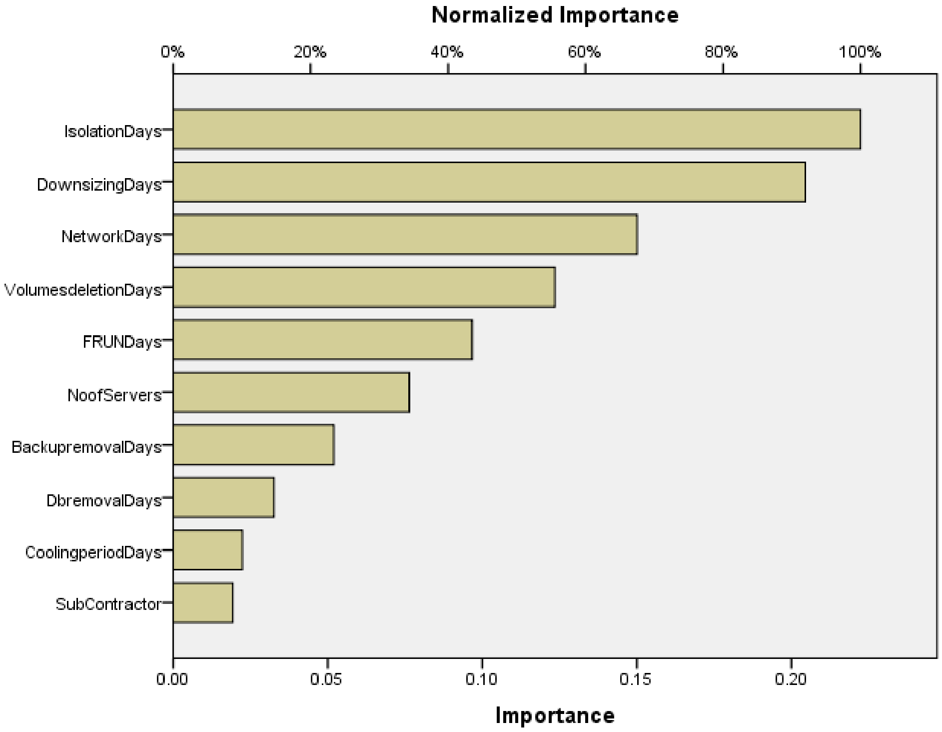


Figure 9.3 Normalized Importance using KNN algorithm

In Figure 9.3, if the scale of features is very different then normalization is required. This is because the distance calculation done in KNN uses feature values. When one feature values are bigger than the other, then the feature will dominate the distance, hence the outcome of the KNN.

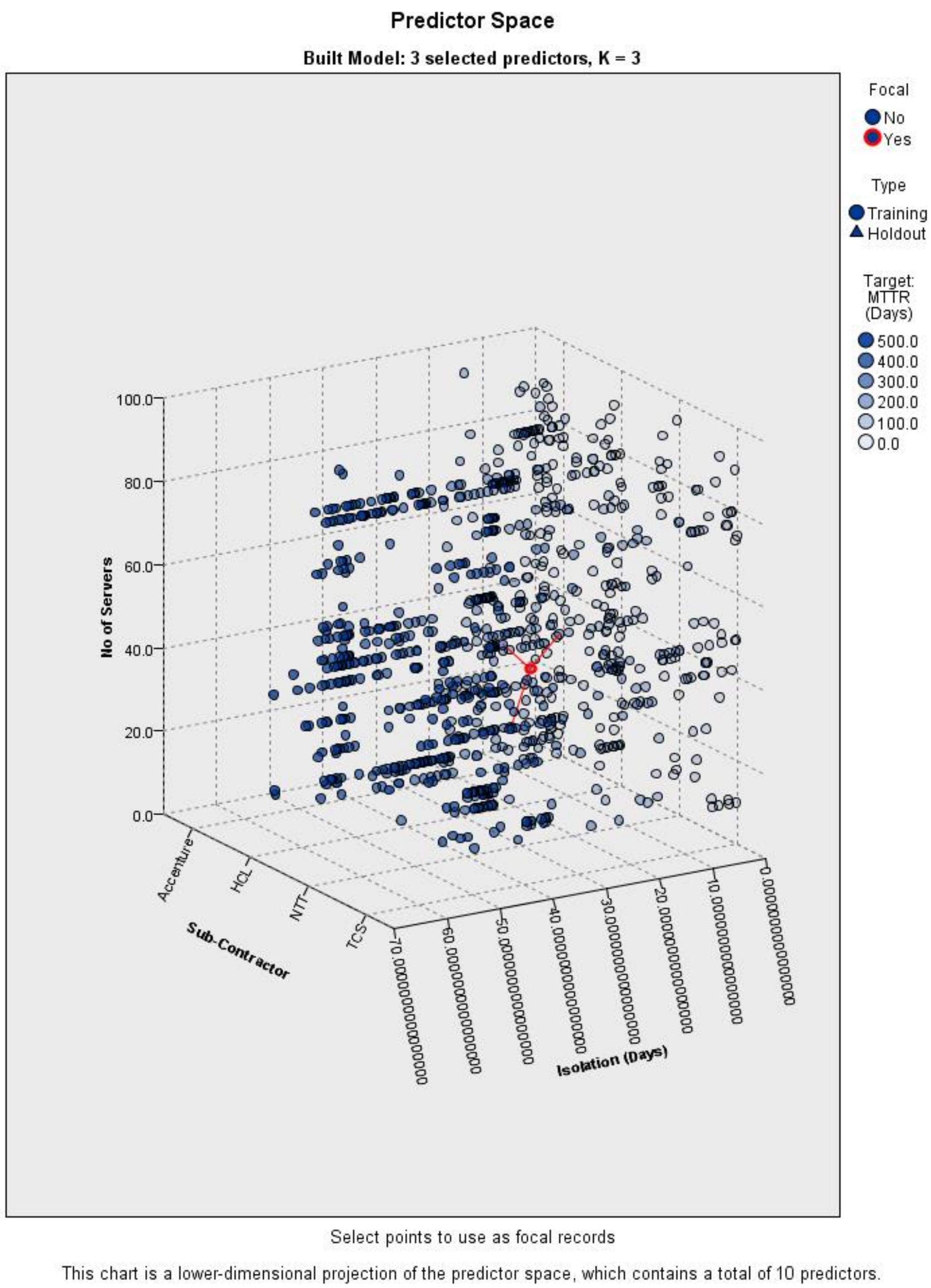


Figure 9.4 KNN model consisting of several servers, sub-contractor, and isolation (days)

The MTTR (days) by considering 3 independent variables - number of servers, sub-contractor (vendor), and isolation (days) has been showcased in Figure 9.4. This algorithm provides the nearest neighbours for a variable. KNN is a non-parametric algorithm as it does not assume anything about the training data and hence makes it useful for non-linear data as in our study.

Chapter 9: Model Evaluation

Decision tree classification algorithms consist of several types that are used to generate decision trees such as Classification and Regression Tree CART, Chi-squares Automatic Interaction Detector (CHAID), Iterative Dichotomies 3 (ID3), Multivariate Adaptive Regression Splines (MARS), Generalized, Unbiased, Interaction Detection and Estimation (GUIDE), Conditional Inference Trees (CTREE), Classification Rule with Unbiased Interaction Selection and Estimation (CRUISE), Quick, Unbiased and Efficient Statistical Tree (QUEST) (Charbuty & Abdulazeez, 2021).

In the study, a Decision Tree with CHAID has been used, with MTTR being the independent variable. The study has been done to identify the key drivers influencing MTTR in turn causing delays in the end-to-end decommissioning process.

The decision tree was built by considering the MTTR based on two input parameters – Isolation (Days) and Vendor analysis. If the isolation days are between 2.7 – 6.7 days, then the predicted MTTR is 66 days. If the isolation is less than 2.7 days, then the predicted MTTR is 9 days.
Also, the sub-contractor or vendor can perform decommissioning with MTTR being 101 days and the isolation being between 12.9 to 19.3 days.

Another decision tree was built considering the MTTR based on one input parameter – The cooling period (Days). If the cooling period is between 1 – 8 days, then the predicted MTTR is 31 days. If the cooling period is >8 days, then the predicted MTTR is 129 days. SLA for the cooling period is 5 days, however, there are scenarios that the team has missed reviving from the cooling period due to priorities and hence this adds to delays in overall decommissioning.

In summary, we were able to identify the following key drivers influencing MTTR using the decision tree model -

- Isolation Days

- Sub-Contractor or Vendor
- Cooling Period (in days)

Recommendations have been provided in the "Analysis and Results" section below as an outcome of the machine learning models built during the study.

To improve the algorithm probability, it is highly imperative to gather data on vendor analysis, seasonality and other features that can contribute to the delays.

Chapter 10: Deployment

The model shall be used to forecast delays for the future data with the improvement or recommendations shall be provided to the relevant teams to optimize the decommissioning process thereby reducing the overall cost being incurred by SAP for not decommissioning the systems on time. Power BI dashboard has been created capturing all the relevant KPIs as shown in Figure 10.1.

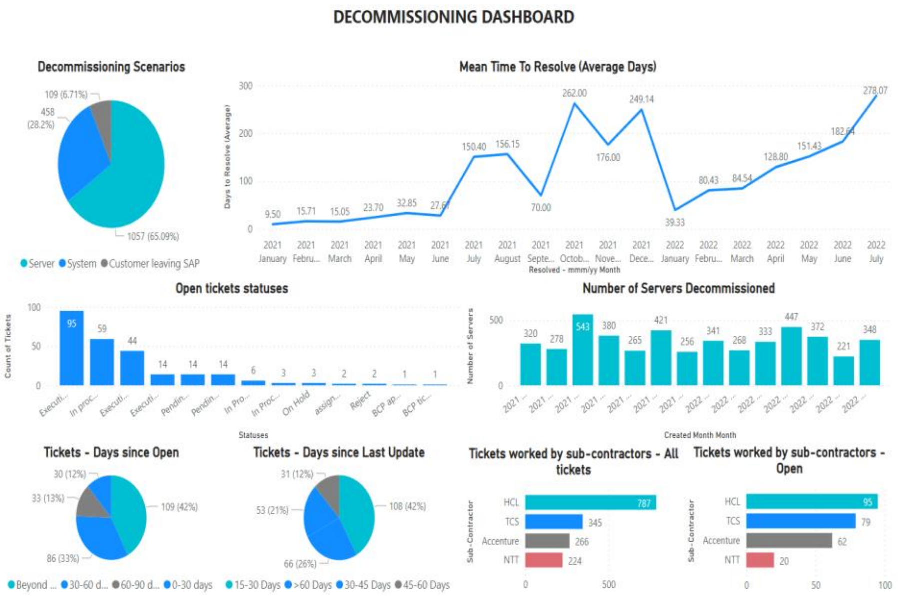


Figure 10.1 Power BI Dashboard for Decommissioning KPIs

Chapter 11: Analysis and Results

In the study presented in this project, the dependent variable MTTR is a continuous variable, and the independent variables are also continuous variables. To handle continuous variables, the decision tree created a threshold and then split the list into those whose attribute value is above the threshold and those that are less than or equal to it.

The main goal of the algorithm was to find an attribute that will be used for the first split. In the case of variables considered for the decision tree, isolation (Days) was used for the first split. In the case of the continuous type of variable, the 'standard deviation reduction' metric was used.

Decision Tree	
Accuracy	**90.30%**
Mean Absolute Error	**9.67**
Min Absolute Error	**1%**
Max Absolute Error	**23%**

Table 11.1 Decision Tree

KNN	
Accuracy	**96.70%**
Mean Absolute Error	**3.27**
Min Absolute Error	**1%**
Max Absolute Error	**27%**

Table 11.2 K Nearest Neighbours

The model accuracy for the decision tree has been showcased in Table 11.1 and Table 11.2 indicates the model accuracy for the KNN model.

The accuracy obtained in the decision tree is **90.3%** with Mean Absolute Error (MAE) being 9.67. In this project, the continuous variable decision tree outcome can be predicted based on multiple variables rather than a single variable. The accuracy achieved with KNN is **96.7**% with MAE being 3.27%.

Key drivers influencing the MTTR as shown in the decision tree are –

- Isolation (Days)
- Cooling Period
- Sub-Contractor (Vendor)

Other factors such as Backup removal and Network deletion have also indicated contributing to delays based on the analysis as there is a dependency on the

teams to complete the process. Internal SLA has been defined with the teams performing the tasks, however, there need to be recommendations and process optimization to ensure the delays will be reduced.
If the isolation days are between 2.7 – 6.7 days, then the predicted MTTR is 66 days. If the isolation is less than 2.7 days, then the predicted MTTR is 9 days. Also, the sub-contractor or vendor can perform decommissioning with MTTR being 101 days and the isolation being between 12.9 to 19.3 days.

Another decision tree was built considering the MTTR based on one input parameter – The cooling period (Days). If the cooling period is between 1 – 8 days, then the predicted MTTR is 31 days. If the cooling period is >8 days, then the predicted MTTR is 129 days. SLA for the cooling period is 5 days, however, there are scenarios that the team has missed reviving from the cooling period due to priorities and hence this adds to delays in overall decommissioning.

The following recommendations shall be provided based on the output of the decision tree –

- Server Isolation needs an internal SLA of 2 days to be defined with the respective stakeholder. Sub-contractor or vendor should follow up to ensure the server isolation is done within the SLA.
- A cooling Period of 5 days being the internal SLA must be adhered to by the teams in the decommissioning process.
- Sub-contractor or vendor analysis must be done to identify the delays and provide the necessary training and knowledge transfer needed for the teams.
- The process must be optimized wherever necessary based on the output from the machine learning model.

Chapter 12: Conclusions and Recommendations for future work

The main objective of the project was to identify Key Drivers influencing MTTR in the customer decommissioning process delays. Machine learning has been applied by building the decision tree models for continuous variables.

The outcome of the model is the detection of key drivers causing huge delays in the end-to-end decommissioning process. Those key drivers influencing MTTR, once handled well within the process, in turn, reduce the cost being incurred by SAP for not decommissioning on time. This is the main intention of the study and with the recommendations provided in the previous section, the objective of the project will be met.

Recommendations for further work: This project does not cover vendor analysis. In the future study, the scope of vendor analysis is to identify if the type of vendor can be a key driver causing the delays in decommissioning. Further variables such as “Staff Competency”, “Absenteeism rate”, “Number of staff employed”, and “staff experience” shall be captured in future studies that can help provide recommendations to vendors for optimizing the decommissioning process.

Bibliography

Breiman, L., & Ihaka, R. (1984). *Nonlinear Discriminant Analysis Via Scaling and Ace NONLINEAR DISCRIMINANT ANALYSIS VIIA SCALING AND ACE*.

Charbuty, B., & Abdulazeez, A. (2021). Classification Based on Decision Tree Algorithm for Machine Learning. *Journal of Applied Science and Technology Trends*, *2*(01), 20–28. https://doi.org/10.38094/jastt20165

Investopedia. (2021).

Jiao, S. R., Song, J., & Liu, B. (2020). A Review of Decision Tree Classification Algorithms for Continuous Variables. *Journal of Physics: Conference Series*, *1651*(1). https://doi.org/10.1088/1742-6596/1651/1/012083

Kang, H. (2013). The prevention and handling of the missing data. *Korean Journal of Anesthesiology*, *64*(5), 402–406. https://doi.org/10.4097/kjae.2013.64.5.402

Krishna M V. (2016). *asset-data-linkage-using-artificial-intelligence-method-of-machine-learning-201*.

Nabipour, M., Nayyeri, P., Jabani, H., S, S., & Mosavi, A. (2020). Predicting Stock Market Trends Using Machine Learning and Deep Learning Algorithms Via Continuous and Binary Data; a Comparative Analysis. *IEEE Access*, *8*, 150199–150212. https://doi.org/10.1109/ACCESS.2020.3015966

Nyarko-Boateng, O., Adekoya, A. F., & Weyori, B. A. (2020). Using machine learning techniques to predict the cost of repairing hard failures in underground fiber optics networks. *Journal of Big Data*, *7*(1). https://doi.org/10.1186/s40537-020-00343-4

Quiry, P., Vernimmen, P., Fur, Y., Dallocchio, M., & Salvi, A. (2009). *Corporate Finance: Theory and Practice*.

Radioactive Waste Management, N. (2018). *Preparing for Decommissioning During Operation and After Final Shutdown Radioactive Waste Management and Decommissioning 2018*.

SIMS Lifecycle Services. (2020a). *SOLUTIONS FOR GLOBAL DATA CENTER DECOMMISSIONING*.

SIMS Lifecycle Services. (2020b). *SUSTAINABLE DATA CENTER DECOMMISSIONING*.

Song, Y. Y., & Lu, Y. (2015). Decision tree methods: applications for classification and prediction. *Shanghai Archives of Psychiatry*, *27*(2), 130–135. https://doi.org/10.11919/j.issn.1002-0829.215044

Vuttipittayamongkol, P., Tung, A., & Elyan, E. (2021). A Data-Driven Decision Support Tool for Offshore Oil and Gas Decommissioning. *IEEE Access*, *9*, 137063–137082. https://doi.org/10.1109/ACCESS.2021.3117891

W, B. A., A, K. K., & B, A. D. (2018). Issues with data and analyses: Errors, underlying themes, and potential solutions. *Proceedings of the National Academy of Sciences*, *115*(11), 2563–2570. https://doi.org/10.1073/pnas.1708279115

Zeng, S., Melville, P., Lang, C. A., Boier-Martin, I., & Murphy, C. (2008). Using Predictive Analysis to Improve Invoice-to-Cash Collection. *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1043–1050. https://doi.org/10.1145/1401890.1402014

Zhang, X., & Jiang, S. (2012). A splitting criteria based on similarity in decision tree learning. *Journal of Software*, *7*(8), 1775–1782. https://doi.org/10.4304/jsw.7.8.1775-1782

Appendix

Plagiarism Report

Report_Key Driver Analysis for MTTR - Aruna K

by Aruna Kashinath

Submission date: 12-Aug-2022 05:35PM (UTC+0530)
Submission ID: 1881735313
File name: Report_Key_Driver_Analysis_for_MTTR_-_Aruna_K.docx (3.3M)
Word count: 10805
Character count: 58780

Report_Key Driver Analysis for MTTR - Aruna K

ORIGINALITY REPORT

15%	13%	5%	8%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	hdl.handle.net Internet Source	2%
2	www.simslifecycle.com Internet Source	1%
3	www.caedes.net Internet Source	1%
4	www.coursehero.com Internet Source	1%
5	www.packtpub.com Internet Source	1%
6	S R Jiao, J Song, B Liu. "A Review of Decision Tree Classification Algorithms for Continuous Variables", Journal of Physics: Conference Series, 2020 Publication	1%
7	srs-brochure.s3.amazonaws.com Internet Source	1%
8	docplayer.net Internet Source	1%

9	nmbu.brage.unit.no Internet Source	1 %
10	Submitted to Sogang University Student Paper	1 %
11	www.ncbi.nlm.nih.gov Internet Source	<1 %
12	Submitted to Georgia Institute of Technology Main Campus Student Paper	<1 %
13	towardsdatascience.com Internet Source	<1 %
14	Submitted to Universiti Teknologi Petronas Student Paper	<1 %
15	www.bec.nu.ac.th Internet Source	<1 %
16	stats.stackexchange.com Internet Source	<1 %
17	journalofbigdata.springeropen.com Internet Source	<1 %
18	moneymatch.com Internet Source	<1 %
19	Submitted to University of Glasgow Student Paper	<1 %

20	"Proceedings of the 13th International Conference on Soft Computing and Pattern Recognition (SoCPaR 2021)", Springer Science and Business Media LLC, 2022 Publication	<1 %
21	open.uct.ac.za Internet Source	<1 %
22	Submitted to Oklahoma State University Student Paper	<1 %
23	docshare.tips Internet Source	<1 %
24	"Explore Business, Technology Opportunities and Challenges After the Covid-19 Pandemic", Springer Science and Business Media LLC, 2023 Publication	<1 %
25	Harsh Baheti, Anuradha Thakare, Yash Bhople, Sudarshan Darekar, Om Dodmani. "Machine Learning Algorithm for Detection And Classification of Tomato Plant Leaf Disease", 2022 IEEE 7th International conference for Convergence in Technology (I2CT), 2022 Publication	<1 %
26	Submitted to Regional Centre for Biotechnology Student Paper	<1 %

		<1%
39	Submitted to University of Liverpool Student Paper	<1%
40	www.rrce.org Internet Source	<1%
41	zenodo.org Internet Source	<1%
42	Submitted to Southern New Hampshire University - Continuing Education Student Paper	<1%
43	Usir.Salford.Ac.Uk Internet Source	<1%
44	trap.ncirl.ie Internet Source	<1%

Exclude quotes	On	Exclude matches	< 10 words
Exclude bibliography	On		