Optimizing Customer Service Delivery for Lenovo Data Center Group

Carlos Gonzalez; Madhura Indurkar; Akshay Joshi; Archit Ramprasad
North Carolina State University

Project Report

Executive Summary

The Data Center Group (DCG) at Lenovo focuses on providing solutions for enterprise networks, storage, and high-quality servers. They provide High Volume server and networking solutions to the entire North America and 6 major countries in Latin America. In the last fiscal year (August 2016 through July 2017), the customer service department handled ~11K claims with only 233 escalations to L3. Lenovo would like to optimize customer service delivery to reduce costs and improve efficiency. The team took a data driven approach to develop the solution. Two key questions were answered: 'How to assign *optimal* service action for incoming service claims?' and 'How to assign next service action to hold down service costs?' Firstly, data preparation helped to standardize the raw information and create input datasets. EDA was performed to provide preliminary insights into the data while a Natural history model was developed to compute time estimates to better understand the service process. Based on these learnings, a logistic regression model was trained to select the *best* service type for an incoming service request. To answer the second question, a *Policy Iteration* methodology was implemented to find the optimal rules for service actions. Based on the analysis, the team recommends that initial service actions should be assigned based on the logistic regression model. It is also important that initial service action and age of the machine is looked at while assigning subsequent service actions to optimize costs and customer service delivery. The team believes that this data driven approach to customer service will lead to faster service resolution and lower costs. The recommended approach also ensures continuous improvement of customer service delivery. Finally, the team hedged the discussion with some future considerations to improve the current recommended methodology.

Introduction

The Data Center Group (DCG) at Lenovo focuses on providing solutions for enterprise networks, storage, and high-quality servers. The objective of this analysis was twofold, to understand the service process provided by the DCG and help optimize the customer service delivery model.

Lenovo has a streamlined process to handle customer service claims during the product lifecycle. When a customer faces a problem with a product purchased from Lenovo, they contact Lenovo's Level 1(L1) Technical Support Team and request for a solution to the problem. The technician who receives this call attempts to fix the problem over the phone using information he/she has about similar problems that the team has encountered in the past.

If the technician cannot resolve the problem or if it seems like a hardware issue, then it is forwarded to Lenovo's Level 2(L2) team for further inspection. Both the L1/L2 have the option of dispatching a Support Service Representative (SSR) to the customer's location to fix the problem. If none of the L1/L2 nor the SSR can fix the defect, then the problem will be escalated to Lenovo's Product Engineering team (Level 3) who will then attempt to fix the defect.

The service action process that L1/L2 teams employ to fix a problem are as follows:

1. Fix on Phone (FOP)
2. Customer Replaceable Unit (CRU)
3. Onsite (ONS)
4. Onsite and no part replaced (NPRA)

The report discussed and addressed concerns faced by Lenovo's Technical Support Team during the service process for High Volume products in North America and Latin America. High Volume products are standalone (Rack or Tower) 1 & 2 processor systems. They range from simple designs that are almost PC-like, up through close to a high-performance system.

The analysis attempted to answer two important questions which will help optimize the current service process.

1. How to assign *optimal* service action for incoming service claims?
2. How to assign next service action to hold down service costs?

These questions were answered by following a rigorous data driven approach. Firstly, the data was cleaned and prepared to standardize the information and create input datasets. Then, exploratory data analysis(EDA) was performed to identify underlying trends in the data and to gain better understanding of the data.

Based on the insights gained from EDA, a Logistic Regression model was used to select the best service type for an incoming service request on the machine characteristics and service history. Next, a natural history model was used to compute time estimates to better understand the service process. This information was then used to find the optimal rules for service actions using policy iterations methodology. Based on this analysis, the team provided high-level recommendations for Lenovo and hedged the discussions with current limitations and future considerations.

Methodology

The project was divided into three broad phases: Data Preparation and Analysis, Predictive Modelling and Stochastic Optimization. Output from each phase provided insights into the customer service operations and helped drive the final recommendations.

Data Preparation

Lenovo provided the team with 2 flat files containing High Volume customer service history for North and Latin American regions from August 2016 through July 2017. An additional file with dates of escalations for claims was given out based on clarifications. The raw data contained 14429 rows of information for 9883 machines (34 distinct machine types) from 9 countries (Refer Appendix 1 for data dictionary). Primary data cleaning was carried out using Excel. The team made a list of theoretical data issues and tested against the provided data. Non- standard and possibly erroneous information was removed. For e.g., Rows with non-null commodities when part count is zero were removed. In some cases, service type had to be recategorized from ONS to NPRA. Overall, 12.6 % of data was removed (1823 rows), bringing the row count to 12606 (Appendix 2). Next, the derived columns were created based on client input and heuristics. Reasonable assumptions were made for ease of modelling and implementation (Appendix 3).

These assumptions were based on input along with clarifications provided by the DCG team. Based on these assumptions, 6 new columns such as Machine Status, Problem number, Total cost of repair etc. created (Refer Appendix 4 for data dictionary of derived columns).

Once the final data was curated, multiple datasets were created for various levels of the data. For example, a dataset at machine number level was created to predict the service action in the next phase. The next section discusses descriptive statistics which were conducted on the mentioned cleaned machine level dataset.

Exploratory Data Analysis

The Exploratory Data Analysis provided the team with a better understanding of the data and identifying any underlying trends in the data. Some of the results obtained by answering some of the below questions were used in both building the predictive and Stochastic model.

The cleaned dataset was analyzed on Excel to answer the below questions:

1. What are the details of the different service actions performed and what are the probabilities of each of these actions following one another?
2. What are the fix rates of the different service types?
3. What are the fix rates by country and service types?
4. What are the common reasons for L3 escalations?
5. Which service types lead to L3 escalations most often?

The results of these questions are described in further detail in the Analysis Section.

Predictive Modelling

Using the insights gained from the EDA, a predictive model was built to predict the next service action for the machine. The dependent variable chosen was the status of the machine (Fixed or Broken). The covariates for the model were selected based on EDA and included variables describing machine characteristics and service history.

To choose the best service action for a machine, team has built a logistic regression model. This model will generate the probability that a machine will be fixed given the machine history based on the chosen service type. The probability with other decision variables such as budget of the customer will help Lenovo to assign service actions for the incoming customer requests. Since the model gives probability of machine getting fixed for all service actions (FOP, CRU and ONS), it will help Lenovo make a holistic decision.

Based on EDA performed on machine level data, it was further prepared for model building by removing redundant columns due to homoscedasticity (GEO_2012) and recategorized Machine_Type column to avoid overfitting. The data was divided into training and test datasets using which a logistic regression model was implemented in R. Further improvements in the model ensured that model was not overfitting data and the trained model was compared to the test dataset. The results of this model and certain modelling caveats are discussed in subsequent sections.

Natural History Stochastic Model

The Natural History Stochastic Model is the model that was developed based on historical transitions between the service actions for all the machines in the data.

Using the clean dataset, the team determined each of the service actions that a broken machine went through till the defect was fixed. The possible states considered in this model are FOP, ONS, CRU, L3 and Fixed state. The assumption made here is that both Fixed and L3 guarantee that the defect will be resolved when the machine reaches any of these states.
The data was split into subsets based on the different service actions and the respective transition probabilities were computed. Using the probabilities for each service action a 6x6 transition matrix was built which was used to answer the following questions:

1. What is the expected time a defect spends on a specific state?
2. What is the probability that the machine is fixed given that a specific service action was performed?

Service Action Policy

To find the Optimal Service Action Policy, the team modeled the Service action process as a Markov Decision Process (MDP). There were several approaches taken, defining different State Spaces such as Machine Age or number of attempts to repair. The process which used 'Machine Age' as states for the MDP is described below, with the process for 'number of attempts to repair' being analogous.

For state definition, the team has segmented the dataset according to age of Machine. The range of 'Machine Age' was found out to be 0-60 months. The range was divided into the following three segments - T0: (1-20), T1: (21-40) and T2: (41 - 60) arbitrarily.

States "Early", "Mid", and "Late" represent the broken machine with ages in ranges T0, T1 and T2 respectively. State "Fixed" represents the absorbing state of a broken machine being fixed. The Action Space of the MDP associated with the states of 'Early', 'Mid' and 'Late' is (FOP, CRU, ONS, L3) and the Action associated with state 'Fixed' is (DN).

Using the cleaned dataset, the team calculated the probability transition matrices associated with each of the service actions. (Appendix 17a for states being ‘Machine Age’ and Appendix 17b for states being ‘number of previous repair attempts’). Knowing the costs associated with each action, the team employed the Value Iteration method implemented using R to determine the optimal service action policies for each state (Refer Appendix 18b code).

To address the fact that it would not be always optimal to consider FOP as the first service action to be taken in our model, the process was modeled in diverse ways by changing the state space associated with the process. A second model was created considering the Number of service actions performed on a machine as the states. The State space of the model was defined as follows: (0): 0 Service actions performed; (1): 1 Service action performed; (2): 2 or more service actions performed on a machine.Consequently, the probability transition matrices were calculated from the data and optimal policies were determined for each Service Action. (Refer Appendix 17b).

Analysis & Results

Exploratory Data Analysis

Certain pointed questions were answered to help understand the data better and identifying underlying trends in the data. Multiple analytical questions (Refer Appendix 5) were answered, however only those which were used in building the predictive model and Stochastic model are discussed here. Please refer Appendix 6-10 for summary EDA results.

1. What are the details of the different service actions performed and what are the probabilities of each of these actions following one another?

   As mentioned previously there are 4 primary actions that were performed on a machine. They were FOP, ONS/NPRA, CRU. The most common service action performed is ONS (c.46% of machines), followed by NPRA (c.29% of machines. CRU and FOP contribute to 10% and 14% of cases overall.

   The probabilities of transitions between service action states are computed using the Natural History Model (Refer the transition Matrix in Appendix 7).
2. *2.* What are the fix rates of the different service types?

The service actions (FOP, ONS/NPRA and CRU) fixed c.92% of all machines that were sent to the Technical Service Center. ONS/NPRA contributed to fixing c.63% while FOP and CRU contributed to c.16% and c.12% (Appendix 6).

3. What are the fix rates by country and service types?
   Across both regions (Latin America/North America) ONS was the most effective at fixing cases (c.43% in NA and c.21% in LA). FOP contributed to fixing c.16% of cases across LA and NA, while CRU contributed to fixing c.13% of cases (Appendix 8).
4. What are the common reasons for L3 escalations?
   In total there were 233 L3 escalations that were found in the cleaned machine level dataset. The data tells us that of these L3 escalations (c.42% of machines) the commodity that is damaged/defective is unknown (NULL). The second most commonly defective commodity found in L3 escalations are System Boards (c.15% of machines). Followed by Raid Adapters and Processors (both c.6% of machines) (Appendix 9).
5. Which service types lead to L3 escalations most often?
   Of the total number of cases that have been escalated to L3 and a service action performed on them, ONS/NPRA are the most prevalent actions (c.82%) that have been escalated to L3. This shows that although it is more expensive to send a technician onsite it provides Lenovo with the benefit of identifying a service engineering problem more easily (Appendix 11).

Predictive Model

As discussed previously, a logistic regression model was implemented using R to predict the next service actions for the machines based on the machine characteristics and service history. Based on initial data analysis, 3 machine characteristics (Machine Age, Country and Machine type) and 2 service history characteristics (Number of escalations, Total service cost till date and Service type) were selected as co-variates. Machine status, an indicator column created by the team, was the dependent variable for the model.

Modeled using the robust GLM function in R (family: Binomial, Link: Logit), the logistic regression model showed promising results against the testing dataset. The model had a balanced accuracy of 99% (which suggests overfitting) and sensitivity of 98%. The lower negative predictive value of 70% can be attributed to the high prevalence of the data (97%). Refer Appendix 12 for the confusion matrix *(Appendix 12a)* and summary parameters *(Appendix 12b)*. As is evident from the preliminary results, logistic regression model shows promise in selecting the initial service action based on machine history.

There can be two approaches that can be taken to improve the predictability of the model. Lenovo can feed more information about the differentiating service history parameters to improve the negative predictive value of the model. Next steps on the modeling front can also be to create distinct model for specific service actions.

Natural History Stochastic Model

Firstly, the current state of the system was analyzed. Using the “clean” data from the data analytics section the team built descriptive tables to determine service action count *(Appendix 13)* and the top 5 paths taken to repair a problem *(Appendix 14)*.

It can be inferred from the top 5 paths taken that the first 3 paths taken are simple, one action paths. This is the ideal situation, as we want the problems to be solved in the shortest amount of time (least actions taken) and with the least cost.

Possible improvements can be reducing the count of multi-step paths, and reducing the costly ONS service action, if the repair can be done with cheaper service actions such as FOP (preferably as it is the cheapest) or CRU.

In the next analysis step, the 6x6 transition matrix with the states being the service actions (NPRA, FOP, ONS, CRU) as well as the absorbing states (L3, F) was divided into 4 submatrices to study the time the machines spent in the transient states (broken, and moving between different repair actions) and the probability of the machine ends up fixed or being escalated to level 3.

Operating with the transient part (upper left 4x4 section) of this transition matrix, the team was able to calculate the expected number of epochs until first absorption for each starting state (Appendix 15).

It is clearly visible from this result that FOP and ONS provide a quite similar and faster time until the machine exits the broken state and moves on to one of the two absorbing states (either escalated to L3 or Fixed), than the other two options (NPRA and CRU). This is not enough information, as ideally, we want all machines to be fixed, and not escalated to level 3, which is costly and time consuming. The team then proceeded to further operate with the result obtained to calculate the probability of absorption by each state (L3 and F) (Appendix 16).

In this study, this probability of absorption is also the steady state probability for the process. Once the process leaves the transient states (in this case the broken state with the respective service actions) it cannot return, ending in the absorbing states with the probabilities (Appendix 16).

Commenting on the results, once again FOP and ONS service actions provide comparable results while at the same time being better than NPRA or CRU services. These results provide insights into how the MDP was built and analyzed.

Service Action Policy

Using the cleaned dataset, the team calculated the probability transition matrices associated with each of the service actions. (Appendix 17a for states being Machine Age and Appendix 17b for states being # of previous repair attempts). Knowing the costs associated with each action, we employed the Value Iteration method implemented in R to determine the optimal service action policies for each state (Refer Appendix 18b code).

To address the fact that it would not be always optimal to consider FOP as the first service action to be taken in our model, we modeled the process in diverse ways by changing the state space associated with the process.

A second model was created considering the Number of service actions performed on a machine as the states. The State space of the model was defined as follows:
(0): 0 Service actions performed; (1): 1 Service action performed; (2): 2 service actions performed on the machine.

Consequently, the probability transition matrices were calculated from the data and optimal policies were determined for each Service Action. (Refer Appendix 17b).

Recommendations

The team has been able to identify recommendations and changes that can be made to the service process based on the data driven approach.

- The team suggests that initial action for incoming service claim should be made based on a logistic regression predictive model. This will ensure that *best* service action is chosen for expedient resolution of the problem.

- Once the service action is assigned, next service action can be made based on the MDP. The selected state space will ensure that machine characteristics (Age of machine) and previous service action are both taken into consideration while making the decision. The policy iteration methodology will ensure that cost to Lenovo is minimized.
- Following optimal rules are defined for service actions
    - If the machine is young/mid and incoming service action is FOP or CRU, the next service action should be FOP.
    - If the machine is old and incoming service action is CRU, a follow-up CRU action is recommended. The team would hedge its recommendation with the caveat that an old machine will probably need heuristic solution rather than a MDP recommendation.
    - At this point of time, team feels that if the initial service action is ONS, the next service action should be ONS as well, overriding the MDP recommendations. This decision is well founded, since the most common path after ONS, are Broken-ONS-ONS- fixed and Broken-ONS-ONS-ONS-Fixed.
    - The team also recommends not to use the MDP for L3 escalation, as the decision to escalate should be made based on monetary basis. A more intuitive MDP which takes time and customer good-will into considerations will be needed to make decision to escalate.
- The team also recommends exploratory data analysis and developing a *natural history model* to ensure continuous improvement.

Some recommendations have been summarized in the following section to ensure continuous improvement of the customer service in Lenovo:

- Improve the claim data entry process so that inherent issues with data fidelity are resolved. This will help make predictions with higher certainty.
- If some service history data was available, the predictive model could be improved by building a distinct model for each service type.
- Finer cost estimation will translate into better customized policies to ensure quicker turnover.
- If data regarding time spent during service action was known, policy recommendation could be fine-tuned to reflect Lenovo’s priorities.

- If Lenovo could do an estimate to calculate monetarily how much it is costing for a machine to stay in the broken state and a service action not fixing the problem, the stochastic process could add this penalty in the cost structure to better reflect the real situation.

Summary

Based on the High-volume data for the North America and Latin America regions provided, the team was able to pick out several findings and underlying trends from the data. Some of these key findings are as follows:

1. Sending a technician onsite (ONS) has the highest success rate of solving any customer complaint across both regions (c.43% and c.21% cases for NA, LA respectively)
2. The Engineering team of Lenovo was notified about escalations to L3 most often by sending a technician onsite (c.55% of cases)
3. Machines that were assigned to be fixed over the phone were relatively younger (c.22 months) than those assigned to CRU and ONS (c.25 months)

Using these key findings, the team was able to build a predictive model using logistic regression which had the capability to assign initial service actions in an optimal fashion to incoming service claims that the DCG received. The predictive model provided probabilities of fixing a machine using a service action with an accuracy of ~98% based on the machine characteristics (Age of the machine, Geographic Location, number of escalations, total parts replaced etc.).

The next step in the optimization process was building a stochastic model by using age of the machine as states and the service actions as the action space. Once the initial service action was determined by the predictive model the Stochastic model helped in assigning subsequent service actions to optimize costs.

It must be remembered that there are certain limitations to the analyses which include the following:

1. The analyses consider escalations to L3 certainly fixes the problem however this need not necessarily be the case.
2. If a problem occurs for the same machine it is considered a separate problem.

Appendix

Appendix 1

Data Dictionary: Raw Data

Column Name	Description	Data Type
Claim_Nbr	Claim number	Text
Country_Code	2 Letter ASCII Code	Factor
Machine_Type	Model number of the machine	Text
Serial_Nbr	Serial number of the machine	Text
Commodity	Part replaced during service	Factor
Part_Count	Number of parts replaced	Number
Service_Date	Date of service of machine	Date
Ship_Date	Date the machine was shipped	Date
Month_in_Service	Age of the machine in months	Number
Box_Count	Number of claims for a machine and claim	Number
Service_Delivery_Type	Type of service carried out	Factor
STYLE	Style of the machine (High Volume)	Factor
GEO_2012	Region (LA/NA)	Factor
Has the case been escalated to L3?	Indicator flag (Yes/ No)	Factor
Date of escalation	Date of escalation of the machine	Date

Note: Box count column was removed after discussion with Dr. Orgut due to inherent data fidelity issues.

Appendix 2

Summary: Data Cleaning

Commodity	Action	Rows Affected
Commodity is null, part count is non-zero, service ONS, CRU	Remove Rows	185
Negative months in service	Remove Rows	34
Commodity is non-null, and service is FOP	Remove Rows	68
Commodity is null, part count is zero, service CRU	Remove Rows	1261
Commodity is null, part count is non-zero, service CRU	Remove Rows	73
Commodity is not null, part count is zero, service CRU	Remove Rows	3
Service date before ship date	Remove Rows	17
Machine belongs to two countries (S65984 and S69406)	Remove Rows	2
Remove Duplicates	Remove Rows	180
Commodity is null, part count is zero, service ONS	Recategorize as NPRA	79
Commodity is not null, part count is zero, service ONS	Recategorize as NPRA	1851
Claims with negative days to escalation (BNPV89P and 404DSZF)	Set days to escalation as zero	2
Total Rows Removed		1823

Note: These changes were incorporated post discussions with Dr. Orgut.

Appendix 3

Summary List of Assumptions

Data Preparation
Any repairs made on the same machine (serial number) within a 30-day window indicate the same problem. If no further repairs are made on the machine following 30 days, then we will assume that machine is "Fixed".
A "Problem" column was defined. Problem represents every instance a broken machine was fixed. Each problem may be made up of multiple claims

Predictive Model
Machine types were recategorized into 4 separate groups to avoid overfitting
Machine age was categorized into 4 levels using Quantiles in order to factorize the column
NPRA and ONS were combined but costs were computed separately
Dependent variable was machine status, independent variables were costs, parts replaced, no. of escalations, Geographic location, Age
Training set was assumed to be 70% of data and 30% was the test set

Natural History Model
Service actions FOP, ONS, NPRA, CRU and Fixed were the states
Assumption that if the problem is assigned to L3 it will be fixed
If a problem occurs after 30 days for the same machine it is considered a separate problem in itself

Appendix 4

Data Dictionary: Derived Columns

Column Name	Description	Data Type
Machine_Status	Status of the machine as of July 31, 2017 (Broken/Fixed)	Factor
Problem	Problem number of a machine	Factor
Total_Cost	Total cost for a service action	Number
History	The list of service actions for a specific *problem*	Text
Path_Cost	Cost incurred for each *problem*	Number
Country_Name	Friendly name of the country from country code	Text
Machine_Type_2	Revised Machine Type Classification	Factor

Note: The rationale behind creation of these variables can be found in Table 3.

Appendix 5

Exploratory Data Analysis Questions

Exploratory Data Analysis Questions
What are the common reasons for L3 escalations?
Which service actions lead to L3 escalations?
What are age statistics for machines undergoing NPRA?
Which Country has machine for NPRA?
Which machine types result in NPRA cases?
What is the most replaced commodity by region?
Which is the most replaced commodity and by which service type?
Which country has the most number of L3 escalations?
What are the common transitions between states?

Appendix 6

Fix rate by Service Type

Service Action	**Machine Status***	**Cases**	**% Cases**
FOP	0	93	1%
ONS	0	579	6%
CRU	0	74	1%
FOP	1	1510	16%
ONS	1	5877	63%
CRU	1	1138	12%

Appendix 7

Natural History Model Transition Matrix

	NPRA	**FOP**	**ONS**	**CRU**	**L3**	**Fixed**
NPRA	0.08	0.01	0.16	0.00	0.02	0.73
CRU	0.02	0.02	0.03	0.05	0.01	0.88
FOP	0.02	0.08	0.03	0.00	0.01	0.87
ONS	0.10	0.01	0.16	0.00	0.02	0.71
L3	0	0	0	0	1	0
Fixed	0	0	0	0	0	1

Appendix 8

Fix rate by Service Action and Country

Service Action	Geography	Machine Status*	Cases	Fix rate
CRU	LA	1	162	2%
CRU	LA	0	8	0%
CRU	NA	1	976	11%
CRU	NA	0	66	1%
FOP	LA	1	444	5%
FOP	LA	0	31	0%
FOP	NA	1	1066	11%
FOP	NA	0	62	1%
ONS	LA	1	1934	21%
ONS	LA	0	136	1%
ONS	NA	1	3943	43%
ONS	NA	0	443	5%

Note: Machine Status - 1 - Fixed, 0- Broken

Appendix 9

Common reasons for L3 escalations

Commodity	L3 Cases	% of L3 Cases
NULL	98	42%
SYSTEM BOARDS	36	15%
RAID ADAPTERS	15	6%
PROCESSORS	15	6%
DASD BACKPLANES	7	3%
MEMORY MODULES	7	3%
RAID MEMORY	7	3%
HARDFILES SAS	5	2%
HARDFILES SSD	5	2%
NETWORK ADAPTERS	5	2%
POWER SUPPLIES	5	2%
CABLES	5	2%
HARDFILES SATA	3	1%
MECHANICAL ASSEMBLIES	3	1%
VOLTAGE REGULATORS / PADDLE CARD	2	1%
POWER SUPPLIES	2	1%
OTHER BACKPLANES	2	1%
RAID SUPERCAP	2	1%
SYSTEM CARDS	2	1%
RISERS	2	1%

Appendix 11

Type of service leading to L3 escalations

Service Action	Cases	% of Cases
ONS	127	55%
NPRA	86	37%
FOP	12	5%
CRU	8	3%

Appendix 12

Logistic Model Parameters

Appendix 12a

Confusion Matrix

	Predicted	
Actual	**0**	**1**
0	56	4
1	23	2520

Appendix 12b

Model Statistics

Statistics	
Accuracy	98%
Balanced Accuracy	96%
Sensitivity	97%
Specificity	93%
Positive Prediction	99%
Negative Prediction	70%

Appendix 13

Service Action Count

Service Action	Number of Claims	Assignment Probability
CRU	1323	10%
ONS	5482	45%
FOP	1812	14%
NPRA	3685	29%
L3	233	2%
TOTAL	**12835**	**100%**

Appendix 14

Top 5 paths taken to repair a problem

Rank	Path	Number of Problems
1st	Broken-ONS-Fix	5193
2nd	Broken-FOP-Fix	1396
3rd	Broken-CRU-Fix	1094
4th	Broken-ONS-ONS-Fix	1022
5th	Broken-ONS-ONS-ONS-Fix	307

Appendix 15

Computations of time parameters

Appendix 15a

Expected time spent in every state

$$E = (I - Q)^{-1} = \left(\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} - \begin{bmatrix} 0.083 & 0.012 & 0.155 & 0.002 \\ 0.015 & 0.018 & 0.025 & 0.054 \\ 0.022 & 0.077 & 0.026 & 0.003 \\ 0.103 & 0.009 & 0.156 & 0.003 \end{bmatrix}\right)^{-1} = \begin{bmatrix} 1.096 & 0.027 & 0.176 & 0.004 \\ 0.024 & 1.022 & 0.039 & 0.056 \\ 0.027 & 0.082 & 1.034 & 0.008 \\ 0.118 & 0.025 & 0.18 & 1.005 \end{bmatrix}$$

Appendix 15b

Summary table for time until first absorption

Starting service action	**Epochs until first absorption**
CRU	1.328
NPRA	1.303
ONS	1.151
FOP	1.141

Appendix 16

Absorption Probabilities

Note: An issue will be resolved given that a specific service action was chosen

Appendix 16a

Mathematical computation for probabilities

$$A = E * R = \begin{bmatrix} 1.096 & 0.027 & 0.176 & 0.004 \\ 0.024 & 1.022 & 0.039 & 0.056 \\ 0.027 & 0.082 & 1.034 & 0.008 \\ 0.118 & 0.025 & 0.18 & 1.005 \end{bmatrix} * \begin{bmatrix} 0.023 & 0.725 \\ 0.006 & 0.882 \\ 0.007 & 0.865 \\ 0.022 & 0.707 \end{bmatrix} = \begin{bmatrix} 0.027 & 0.973 \\ 0.008 & 0.992 \\ 0.009 & 0.991 \\ 0.026 & 0.974 \end{bmatrix}$$

Appendix 16b

Summary table of Probabilities

Starting Service Action	Probability of Fixed	Probability of escalation
FOP	99.2%	0.8%
ONS	99.1%	0.9%
CRU	97.4%	2.6%
NPRA	97.3%	2.7%

Appendix 17

MDP Iterations

Appendix 17a

State Space (Based on Machine Age)

Refer to the below table prior to calculating the transition matrices for the MDP

"EARLY": 0-20 months old

"MID": 21-40 months old

"LATE": 41-60 months old

FOP-EARLY	FOP-EARLY-FIX	FOP-MID	FOP-MID-FIXED	FOP-LATE	FOP-LATE-FIXED
891	787	795	681	126	110

CRU-EARLY	CRU-EARLY-FIX	CRU-MID	CRU-MID-FIXED	CRU-LATE	CRU-LATE-FIXED
517	448	660	595	146	129

ONS-EARLY	ONS-EARLY-FIX	ONS-MID	ONS-MID-FIXED	ONS-LATE	ONS-LATE-FIXED
3943	2851	4555	3304	969	705

Corresponding Probability Matrices

$$P_{FOP} = \begin{array}{c} \begin{bmatrix} 0.1167 & 0 & 0 & 0.8833 \\ 0 & 0.1434 & 0 & 0.8566 \\ 0 & 0 & 0.1270 & 0.8730 \\ 0 & 0 & 0 & 1 \end{bmatrix} \end{array} \begin{array}{c} EARLY \\ MID \\ LATE \\ FIXED \end{array}$$

$$P_{CRU} = \begin{array}{c} \begin{bmatrix} 0.1335 & 0 & 0 & 0.8665 \\ 0 & 0.0985 & 0 & 0.9015 \\ 0 & 0 & 0.1164 & 0.8836 \\ 0 & 0 & 0 & 1 \end{bmatrix} \end{array} \begin{array}{c} EARLY \\ MID \\ LATE \\ FIXED \end{array}$$

$$P_{ONS} = \begin{array}{c} \begin{bmatrix} 0.2769 & 0 & 0 & 0.7231 \\ 0 & 0.2746 & 0 & 0.7254 \\ 0 & 0 & 0.2724 & 0.7276 \\ 0 & 0 & 0 & 1 \end{bmatrix} \end{array} \begin{array}{c} EARLY \\ MID \\ LATE \\ FIXED \end{array}$$

Appendix 17b

States Space based on number of actions performed

Refer table prior to calculating the transition matrices for the MDP.

Number of actions: 0, 1, 2+

FOP-0	FOP-0-FIX	FOP-1	FOP-1-FIXED	FOP-2+	FOP-2+-FIXED
1550	1396	169	136	34	28

CRU-0	CRU-0-FIX	CRU-1	CRU-1-FIXED	CRU-2+	CRU-2+-FIXED
1220	1094	81	68	16	6

ONS-0	ONS-0-FIX	ONS-1	ONS-1-FIXED	ONS-2+	ONS-2+-FIXED
6840	5193	1677	1101	572	335

Corresponding Probability matrices

$$P_{FOP} = \begin{array}{c} \begin{bmatrix} 0.0994 & 0 & 0 & 0.9006 \\ 0 & 0.1953 & 0 & 0.8047 \\ 0 & 0 & 0.1765 & 0.8235 \\ 0 & 0 & 0 & 1 \end{bmatrix} \end{array} \begin{array}{c} 0 \\ 1 \\ 2+ \\ FIXED \end{array}$$

$$P_{CRU} = \begin{bmatrix} 0.1033 & 0 & 0 & 0.8967 \\ 0 & 0.1605 & 0 & 0.8395 \\ 0 & 0 & 0.625 & 0.375 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{array}{c} 0 \\ 1 \\ 2+ \\ FIXED \end{array}$$

$$P_{ONS} = \begin{bmatrix} 0.2408 & 0 & 0 & 0.7592 \\ 0 & 0.3435 & 0 & 0.6565 \\ 0 & 0 & 0.4143 & 0.5857 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{array}{c} 0 \\ 1 \\ 2+ \\ FIXED \end{array}$$

Appendix 18

Appendix 18a

R Codes for training and testing the Logistic Regression model

```
#Setup Libraries
library(readxl)
library(caret)
RawData <- read_excel("**/ModelInformation.xlsx", na = "null")

#Coerce columns as factors

RawData$Machine_Status<-as.factor(RawData$Machine_Status)#Derived
Column
RawData$Machine_Type<-as.factor(RawData$Machine_Type)
RawData$Machine_Type_2<-as.factor(RawData$Machine_Type_2) #Derived
Column
RawData$Country_Code<-as.factor(RawData$Country_Code)
RawData$Service_Delivery_Type<-
as.factor(RawData$Service_Delivery_Type)

#Creation of training and testing dataset

set.seed(123) #Seed to sample the data
train_sample <- sample(seq_len(nrow(RawData)), size = floor(0.7 *
nrow(RawData)))
Train_RawData<-RawData[train_sample, ]
Test_RawData <- RawData[-train_sample, ]

#Logistic Regression Model

#Model Training
Model<-glm(Machine_Status~Service_Delivery_Type +
No_of_Escalations+Total_Cost+Country_Code+Age_2017+Machine_Type_2,fami
ly=binomial(link='logit'),data = Train_RawData)
summary(Model)

#Model Testing
Prediction<-cbind.data.frame(Test_RawData,predict(Model, newdata =
Test_RawData , type = "response"))
colnames(Prediction)[12]<-"Predict_Probability"
plot(Prediction$Machine_Status,Prediction$Predict_Probability)
confusionMatrix(round(Prediction$Predict_Probability,0),Prediction$Mac
hine_Status,positive = "1")
```

Appendix 18b

R Codes for running MDP Policy Iterations

```
library(MDPtoolbox)

#Define Probability Transition Matrices for Action Space

FOP<- matrix(rbind(c(0.1167,0,0,0.8833),c(0,0.1434,0,0.8566),
c(0,0,0.127,0),c(0,0,0,1)),nrow = 4,ncol = 4)

CRU<-matrix(rbind(c(0.1335,0,0,0.8665),c(0,0.0985,0,0.9015),
c(0,0,0.1164,0.8836),c(0,0,0,1)),nrow = 4,ncol = 4)

ONS<-matrix(rbind(c(0.2769,0,0,0.7231),c(0,0.2746,0,0.7254),
c(0,0,0.2724,0.7276),c(0,0,0,1)),nrow = 4,ncol = 4)

L3 <-matrix(rbind(c(0,0,0,1),c(0,0,0,1),c(0,0,0,1),c(0,0,0,1)),nrow =
4,ncol = 4)


#Corresponding Rewards Matrices

a<--0.5
b<--1.475
c<--2.028
d<--10

RFOP<-matrix(rbind(c(a,0,0,a),c(0,a,0,a),c(0,0,a,a),c(0,0,0,a)),nrow =
4,ncol = 4)

RCRU<-matrix(rbind(c(b,0,0,b),c(0,b,0,b),c(0,0,b,b),c(0,0,0,b)),nrow =
4,ncol = 4)

RONS<-matrix(rbind(c(c,0,0,c),c(0,c,0,c),c(0,0,c,c),c(0,0,c)),nrow =
4,ncol = 4)

RL3<-matrix(rbind(c(0,0,0,d),c(0,0,0,d),c(0,0,0,d),c(0,0,0,d)),nrow =
4,ncol = 4)



#Set Up MDP Array

#Probability Transition Arrays
```

```
S<-list()
S[[1]]<-FOP
S[[2]]<-CRU
S[[3]]<-ONS
S[[4]]<-L3

#Reward Arrays

K<-list()
K[[1]]<-RFOP
K[[2]]<-RCRU
K[[3]]<-RONS
K[[4]]<-RL3

#Initial Set Up
Policy<-c(1,1,1,1)
V0<-c(0,0,0,0)
epsilon<-0.02
mdp_check(S,K)

#Policy Iterations
mdp_policy_iteration_modified(S,K,0.9)
```