BPIC 2013 VOLVO IT BELGIUM: ROOT-CAUSE ANALYSIS OF PING PONG BEHAVIOUR

EXECUTIVE SUMMARY

Business processes are likely to confront inefficiency in their performances. The widely popular and new discipline of process mining allows the use of event logs to identify efficiency opportunities. This paper discusses an approach to use the event log of Volvo IT Belgium provide as part of the Business Process Intelligence Challenge 2013 for the purpose of root-cause analysis of the ping pong behaviour through the application of a classification technique. The analysis identifies that there are some support teams which have higher influence on the occurrence of the ping pong between the support teams. We conclude that to there are three major support teams which cause the delay in the case completion time in

INTRODUCTION

With the rapid growth the world is experiencing, the businesses have realised that the best way to grow and compete in the current market is to manage the business processes with unique approaches [1]. The digitally expanding world has increased the significance of information systems in business processes [2]. They store a large amount of information generated daily, which records details of every activity performed – when was it executed and by whom. The availability of this invaluable raw event data can be used to extract process-related insights by the use of process mining – an amalgamation of process analysis and data mining techniques. Process mining is an approach that allows automatic extraction of processes from event logs so as to understand how the processes are implemented in reality, how they are actually executed. This extraction of knowledge from the data gives us room for improvement. Process mining is nothing but an initiative for business process improvement to identify the new ways to design current business processes. Traditionally, process mining focussed on process discovery, conformance checking and performance analysis [3]. However, lately there is wide interest amongst researchers for the use of data mining techniques to solve business process related challenges.

The Volvo IT Belgium's incident and problem management system called VINST contains data in the form of event logs. Event logs contain information about the activities that have been executed, the time they had occurred and the resources which carried out the activities. The organisation has put forward the issue of repeated back and forth sending of incidents between the support teams. They call this as the "ping pong behaviour" [4].

In this paper, we discuss a data mining technique to identify the causes for the ping pong behaviour within the system. We will use classification technique which will help us reach to the bottom of the issue and help in process optimisation. We transform the event log into a form so that it becomes possible to use classification method of decision trees to find the cause of delays. In particular, we capture the information in the event log and feed it into a decision tree learner which induces a decision tree. It identifies factors that are a possible reason of the delay.

The structure of the paper is as follows: Literature review presents the background on process mining, root-cause analysis and the related work on Volvo IT's VINST system. It is then followed by Background knowledge section that discusses the definition of the ping pong behaviour which is of utmost importance for this paper. Approach describes information on the data, the data mining technique used, and the tools used for the purpose of this analysis. Analysis details out the pre-processing of data, transforming the data to be classification ready and the evaluation methods for the models generated. Findings discusses the observations of the analysis and generates insights. Limitations puts forwards the limitations of this research and is followed by Conclusion, which is an overview of the study. Reflection demonstrates the learnings and observations during the journey of this research study and directs towards the future scope of this analysis with ideas of areas to be explored.

LITERATURE REVIEW

In this section, we talk about process mining, its importance and effectiveness to deal with organisational challenges, previous work related to Volvo's VINST system and how root-cause analysis (RCA) can help us solve the ping pong behaviour issue raised by the process owners of Volvo IT Belgium.

Process mining is the combination of process modelling and analysis and data mining. It is a widely popular discipline used for the extraction of insights available from event logs. Event logs are the data containing details of every activity performed. It contains information about case id, a task, a user, a timestamp to name a few. The main purpose of process mining is "to discover, monitor and improve real processes". It deals with extraction of process models from the event logs of the organisations so as to understand how the processes are implemented in reality. Traditionally, process mining focussed on process discovery, conformance checking and performance analysis [3]. Moreover, they further help in analysing deviations.

Business process mining has proven to be effective to manage organisational challenges. For example, the insights gained by Aaist and Weijters [5] from event logs led to an impressive impact on the business performance. Of the many other examples out there, a survey by Rojas [6] claimed that process mining has largely been accepted in the healthcare sector for process discovery and analyse social networks.

There has been substantial research done on the event log provided by Volvo IT Belgium in the 2013 BPI Challenge. Much of the academic literature revolves around the possible questions proposed by the process owners. For example, Arias and Rojas [6] discovers behaviour characteristics associated with products, resources and organisational lines. Most of the study involve the use of exploratory analysis to answer the specific questions. [7]

All these studies involve the use of process mining tools like Disco and ProM and manual analysis. However, there is absence of use of data mining techniques to tackle the issues of the organisation.

The past studies on the ping pong behaviour analyse the functions responsible for it, the support teams responsible, the products impacted by ping pong behaviour [7] [8] [9], impact of ping pong behaviour on case completion time, products with highest ping pong rate [7], linear and circular ping pongs [9]. However, none of them investigates the root causes of this behaviour. This work will be focused to identify the root causes of the ping pong behaviour between the support teams.

<u>Definition of Ping Pong behaviour and why is it an issue?</u>
The document provided by Volvo IT Belgium describes "ping pong behaviour" as repeated back and forth sending of incidents between support groups. It describes it as an "unwanted situation". Ideally, an incident is solved quickly with the involvement of minimum number of support teams [10]

This behaviour is a contributing factor to the inefficiency of the process. It may lead to waste of time and resources. It leads to increase in case completion time and customer unsatisfaction.

Professionals describe root-cause analysis (RCA) as a "means of retrospectively improving on ill-defined, inadequate processes and systems" [11]. To simplify, RCA aims to solve issues by identifying the underlying causes of the problems. The traditional RCA involved identifying contributing factors, ranking them in terms of likelihood of impact, classify into groups (correlated, unrelated, contributing factor or root cause) and the mapping them for better understanding.

RCA techniques have been used in various domains – be it risk management or process-based industry [12]. The use of data mining techniques in business processes is explored largely 16. In recent years, this area has gained a lot of attention. A major study proposed root-cause analysis for business process models [13]. Gheorghe performed RCA to improve "efficiency of maintenance" and reduce unexpected breakdowns of machines [14].

Previous RCA focussed work involve the use of decision trees to carry out decision point analysis with the use of Decision Miner in ProM [15] or regression analysis to identify the relationship between workload and performance [16] or the enrichment and transformation of event logs to get classification-ready log for RCA . Another paper used logical decision trees to discover the causes of process delays in event logs [17], which drives the basis of this analysis.

In this paper, we will perform root-cause analysis with the use of classification techniques (Decision tree) and process mining tools – Disco and ProM, the unexplored area. This work will be distinct in that as none of the existing work on BPIC 2013 Volvo event log uses data mining techniques to solve the issue of ping pong behaviour. The findings from this study will help to achieve precise accurate solutions to the underlying issue of ping pong behaviour and allow process optimization at Volvo IT Belgium.

BACKGROUND KNOWLEDGE

The term we will be widely using in this paper will be Ping Pong behaviour. The document provided by Volvo IT Belgium describes it as repeated back and forth sending of incidents between support groups. Ideally, an incident is solved quickly with the involvement of minimum number of support teams. It is an unwanted situation. It leads to exploitation of resources. It increases the case resolution time and causes delays and customer unsatisfaction.

APPROACH

This section describes the approach to the research topic. It discusses and explains the data and the data mining technique used. It further also talks about the tools used in this analysis and their purpose specific to this study.

ABOUT THE DATA

The data we use is the event log from the incident and problem management system called VINST of Volvo IT Belgium. There are three CSV files with incident or problem number as case ID. The available CSV files are - VINST cases incidents, VINST cases open problems and VINST cases closed problems. These datasets contain information of the resolved cases for the year 2012. This data was available as part of the Business Process Intelligence (BPI) Challenge 2013 and is available on its website. [10]

Table 1 About the data

Event Log	#Events
VINST cases incidents	65,533
VINST cases open problems	2,351
VINST cases closed problems	6,660
	74,544

In our study, we focus on the incidents log as it has the details of all the incidents. This would allow us to do a better analysis. It contains 65,533 events having 7,554 incident records. The attributes of the dataset are shown in the table [Table2] below.

Table 2 Data Attributes

Attribute	Description
Problem/Serial No	The unique ticket number for a problem or incident
Problem change Date + Time	Moment of status change
Problem Status	The status of the problem which can be queued, accepted, completed, or closed

Problem sub-status	The sub-status of an incident – assigned, awaiting assignment, cancelled, closed, in-progress, wait, unmatched
Involved ST function div	The IT organisation that provides the service. It is divided into functions
Involved Organisation line 3	The business area of the user reporting the incident to the service desk
Involved ST	The team responsible for resolving the incident
SR Latest impact	The impact of an incident to the customer: major, high, medium, low
Product	The identification of the product which originated the incident
Country	The country of the support team that takes ownership of the incident record
Owner Country	The country of the owner
Owner First Name	The person of the support team that is the owner of the reported incident

ABOUT THE DATA MINING TECHNIQUE

In this study, our focus is to go beyond the process mining tools like Disco and ProM and use pure data mining techniques for the purpose of finding the root causes of the ping pong behaviour experienced by Volvo IT Belgium. There are various techniques used for root-cause analysis. One past study uses logical decision trees to understand the causes of process delays in event logs [18]. Another study uses classification techniques using Weka to transform process-based event logs for the purpose of root cause analysis [18]. Therefore, for our analysis we use the classification technique and get decision trees to understand the root causes of the ping pong behaviour.

CLASSIFICATION TECHNIQUE – DECISION TREE

Classification is a type of supervised method, with Regression being the other one. Supervised method aims to discover the relationship between the input/independent attributes and the target/dependent attribute.

A Decision Tree is used for classification tasks. It is a graph with the resemblance of an upside-down tree. It is formed based on test conditions. They are used to classify an instance based on its attribute values into a predefined set of classes. It contains nodes known as leaves which have attribute test conditions, which are further connected to branches with the possible values of the attribute. All these nodes have one incoming branch. However, there is one node with no incoming branch and is known as the root node, which is the starting node [17].

We get our decision tree by using the Scikit-learn's DecisionTreeClassifier function in python.

TOOLS USED

Disco

Disco is the commercial tool created by Fluxicon for process mining analysis of datasets. It is a quick analysis tool capable of performing well for process discovery and statistical overviews. It is useful for filtering data as well. It is available for academic assistance. Disco was used for exporting the CSV file into a readable tabular form that can be used in both Microsoft Excel and Python.

<u>Microsoft Excel</u>

Microsoft Excel was used for pre-processing – to split the "Involved ST" column into "ST" with information of only about the support team that received the incident and another column named "Line" representing the level or the organisation line the support team belonged to. We used the if-else function in Microsoft Excel for this purpose. Further, the "Case ID" column was also organised into a more understandable format by remove the "#-" from it.

<u>Python programming language and Jupyer Notebook</u>

The programming language used to carry out the analysis was Python on a platform called Jupyter Notebook which is available as part of Project Jupyter [https://jupyter.org/]. It is advantageous as it allows to develop code and communicate about the steps and the results at the same time.

Specifically, we use the Scikit-learn which is a machine learning library for Python programming language. It has many classification, regression and clustering algorithms. It can be used along with 'numpy'. For the decision tree algorithm, we use the DecisionTreeClassifier. More details on the steps and analysis is provided in the following section.

ANALYSIS

DATA PRE-PROCESSING

This section discusses the tasks of data pre-processing, cleaning, transformation, and reduction. These are the initial and important steps of any data analysis project. The data might have incomplete information, contain missing values, outliers, or redundant values. Using this raw data can lead to wrong analysis and findings.

The data (VINST cases incidents.csv) was available in the form of a CSV file but was not properly organised in a tabular form. So, we used Disco to get a tabular form of the data accessible in both ProM and Microsoft Excel.

Table 3 Snippet of Data in microsoft Excel

Case ID	Status	Involved ST	Complete Tir	Variant	Variant inde›	Sub Status	Involved ST F	Involved Org	SR Latest Im	Product	Country	Owner Count	Owner First Name
1-36428576ᵹ	Accepted	V5 3rd	04:00.0	Variant 352	352	In Progress	A2_5	Org line A2	Medium	PROD582	fr	France	Anne Claire
1-36428576ᵹ	Queued	V30	04:00.0	Variant 352	352	Awaiting Assignment	A2_4	Org line A2	Medium	PROD582	fr	France	Anne Claire
1-36428576ᵹ	Accepted	V13 2nd 3rd	04:00.0	Variant 352	352	In Progress	A2_5	Org line A2	Medium	PROD582	fr	France	Anne Claire
1-36428576ᵹ	Completed	V13 2nd 3rd	04:00.0	Variant 352	352	Resolved	A2_5	Org line A2	Medium	PROD582	fr	France	Anne Claire
1-36428576ᵹ	Queued	V30	04:00.0	Variant 352	352	Awaiting Assignment	A2_4	Org line A2	Medium	PROD582	fr	France	Anne Claire
1-36428576ᵹ	Accepted	V30	04:00.0	Variant 352	352	In Progress	A2_4	Org line A2	Medium	PROD582	fr	France	Eric
1-36428576ᵹ	Queued	V5 3rd	04:00.0	Variant 352	352	Awaiting Assignment	A2_5	Org line A2	Medium	PROD582	fr	France	Eric
1-36428576ᵹ	Accepted	V5 3rd	04:00.0	Variant 352	352	In Progress	A2_5	Org line A2	Medium	PROD582	fr	France	Anne Claire

<u>Extraction of Organisation Line</u>

For our analysis (as explained later in this section), we needed a column providing information on the organisation line/level of the support team. For this we used the column with the name "Involved ST" which had combined information about the ST and the organisation line (For example, V50 2nd and 3rd) in the same field. With the use of Microsoft Excel, we spit this column into two – “ST” and “Line”, where the latter column had values of 1,2, or 3 depending on the level the support team belonged to. The summary is available in the table 4 below with the reasoning in the later part of the section. The following table is a snapshot of the data after splitting the original column into two new columns as mentioned above.

Table 4 Line explanation

If “Involved ST” involves __,	Then “Org Line” has the value __.
Nothing mentioned	“1”
“2nd”	“2”
“3rd”	“3”
“2nd 3rd”	“2”

Table 5 Data after pre-processing

Case ID	ST	Line	Complete Tir	Variant	Variant inde:	Status	Sub Status	Involved ST F	Involved Org	SR Latest Im	Product	Country	Owner Count	Owner First Name
364285768	V5	3	04:00.0	Variant 520	520	Accepted	In Progress	A2_5	Org line A2	Medium	PROD582	fr	France	Anne Claire
364285768	V30	1	04:00.0	Variant 520	520	Queued	Awaiting Ass	A2_4	Org line A2	Medium	PROD582	fr	France	Anne Claire
364285768	V13	2	04:00.0	Variant 520	520	Accepted	In Progress	A2_5	Org line A2	Medium	PROD582	fr	France	Anne Claire
364285768	V13	2	04:00.0	Variant 520	520	Completed	Resolved	A2_5	Org line A2	Medium	PROD582	fr	France	Anne Claire
364285768	V30	1	04:00.0	Variant 520	520	Queued	Awaiting Ass	A2_4	Org line A2	Medium	PROD582	fr	France	Anne Claire
364285768	V30	1	04:00.0	Variant 520	520	Accepted	In Progress	A2_4	Org line A2	Medium	PROD582	fr	France	Eric
364285768	V5	3	04:00.0	Variant 520	520	Queued	Awaiting Ass	A2_5	Org line A2	Medium	PROD582	fr	France	Eric
364285768	V5	3	04:00.0	Variant 520	520	Accepted	In Progress	A2_5	Org line A2	Medium	PROD582	fr	France	Anne Claire
364285768	V5	3	04:00.0	Variant 520	520	Accepted	Assigned	A2_5	Org line A2	Medium	PROD582	fr	France	Anne Claire

For remaining part of pre-processing like removing null values and changing datatypes of the columns to be able to get a decision tree, we used the Python programming language on Jupyter Notebook platform.

ATTRIBUTE SELECTION

All the attributes present in the original dataset were not significant for the purpose of our analysis. Hence, we removed them before moving further. We further assign the columns as feature and target attributes. Target attribute is the variable we want to analyse and the feature or the independent attributes are the variables contributing to determining the target variable. The data is then ready for classification analysis.

TRAINING AND TEST SET

After the attribute selection, the data was split into training and test data using the “train_test_split” function present in Scikit-learn model_selection package.

CLASSIFICATION – DECISION TREE BUILDING

The data was partitioned into training and test data to be able to fit the decision tree model in the training data. Then DecisonTreeClassifier from sklearn.tree was used for building our decision tree. Further tuning of parameters was done using the GridSearchCV function to get the optimal parameters and the best performing decision tree with less complexity and better accuracy.

HYPERPARAMETER TUNING using Scikitlearn's GridSearchCV

The performance of the model highly depends on the input values of its parameters. It is a very time-consuming process to manually calculate the accuracy of each combination of the selected parameters. Hence, we use Scikitlearn's GridSearchCV tool for hyperparameter tuning to determine the optimal parameters of the model. We can include the parameters in the parameter grid. It then evaluates the model for all the possible combinations using the cross-validation method (grid search + k-fold cross validation). It then identifies the one with "best" parameters. Here we do a 10-fold cross validation.

In our study, we have focussed on the following three hyperparameters – criterion, max_depth and min_samples_leaf.

- Criterion - the function to measure the quality of split. There are two – "gini" for Gini impurity and "entropy" for information gain.
- Max_depth – It controls the maximum depth of the decision tree. We set our range from 1-20.
- Min_samples_leaf – The minimum number of samples required to be at a leaf node. We set the range of 1-25 with step of 5.

FEATURE IMPORTANCE

It is important to get more information and insights on the input variables or the features that hold higher stand in the decision-making process in the model. For this purpose, we carry out Feature Importance using feature_importances_ present in sklearn. In our analysis, we have generated outputs of feature names along with the values for better understanding.

PERFORMANCE EVALUATION

We evaluate the performance of the models obtained in terms of various statistics as explained below. The basis of evaluation lies on four ways:
True positives (TP) – When a case was positive and was predicted positive
False positives (FP) – When a case was negative but was predicted positive
True negatives (TN) – When a case was negative and was predicted negative
False negatives (FN) – When a case was negative but was predicted positive

Accuracy

Accuracy is the measure of predicting correctly classified instances from the total instances. We evaluate it by using the score() feature available. This helps us understand the performance of the model against both the training and the test data.

$$Accuracy = \frac{True\ positives + True\ negatives}{All\ samples}$$

Classification Report

We also use classification report to evaluate the performance of the classifier. In other words, it measures the quality of the predictions made by the model. It gives us a report of a few statistics – Precision, Recall, F1-score, and support calculated by using true positives, false positives, true negatives and false negatives.

Precision – It is the measure of the actual positive observations that were predicted correctly by the classifier.

$$Precision = \frac{True\ positives}{True\ positives + False\ positives}$$

Recall – It is the measure of all real positive observations that were predicted correctly by the classifier from actual positive observations.

$$Recall = \frac{True\ positives}{True\ positives + False\ negatives}$$

F1-score – It is the harmonic mean of precision and recall. Its values vary from 0(bad) to 1(good).

$$F1\ score = \frac{2 * (Precision * Recall)}{Precision + Recall}$$

Support – Number of instances in each class.

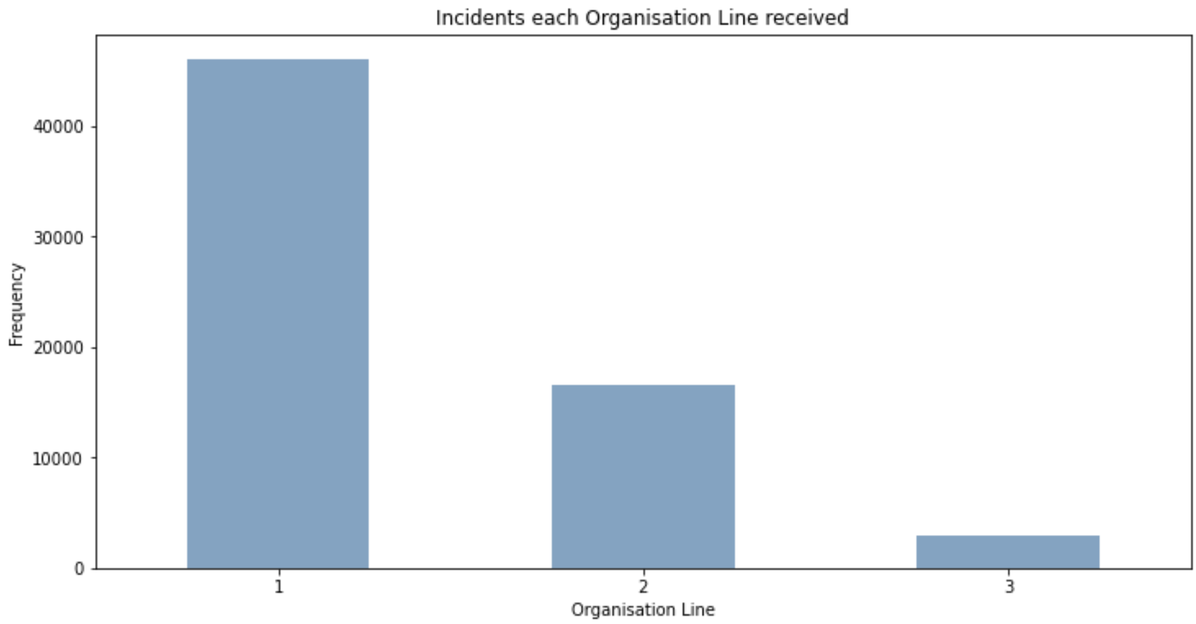
<u>Confusion Matrix</u>

We also use the methods of plotting the confusion matrix as a performance measurement technique. We perform this by using the plot_confusion_matrix() function available in scikit-learn's library. It gives us information on the correctly and incorrectly classified instances. It has true values as rows and predicted values as columns.

FINDINGS

To begin with, we got the number of cases resolved by each organisation line. Figure 6 shows that Line 1 had the maximum cases with more than 45000 cases. We observe that the difference in number of cases resolved by Line 2 and 3 in comparison to Line 1 is very huge. After this initial data exploration, we moved to building a decision tree.

Figure 6 Number of cases per line



We evaluated the performance of the models obtained – the default model and models obtained after tuning of parameters - which is summarised in the following table.

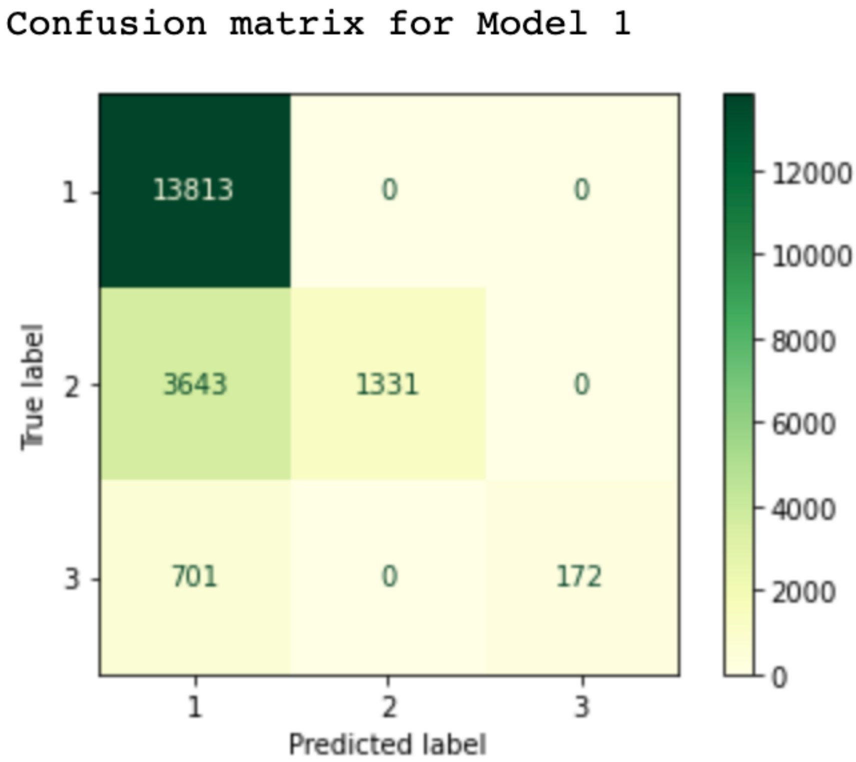
Table 7 Statistics for performance evaluation

	Default model	Model 1	Model 2
Accuracy			
- Train accuracy	100	77.97	76.93
- Test accuracy	99.75	77.90	76.83
Weighted avg of F1-score	1	0.73	0.71

We observe in the table above (table 7) that the default model has the highest accuracy in comparison to the other two models. However, we need to keep in mind that the default model generates a complex decision tree and hence, is not the best option to go further with. For model 1 and 2, we observe that model 2 has a higher accuracy of 77.90% as compared to the accuracy of model 2 of 76.83%. Hence, we can conclude that model two is our preferred model for our analysis.

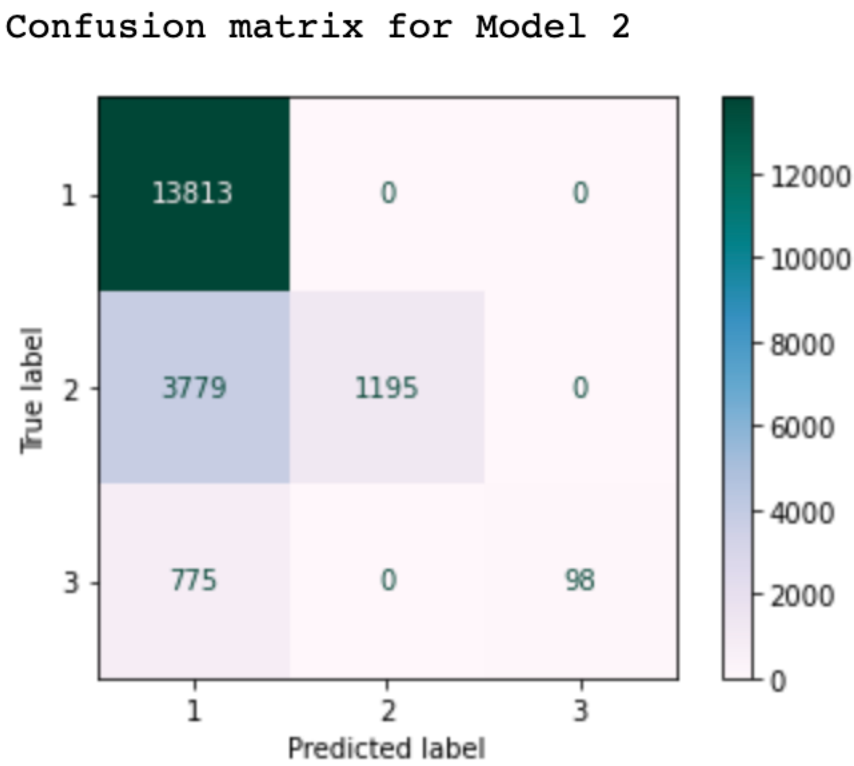
In terms of F1 score, again model 2 has a higher score of 0.73 than model 1 (0.71). The closer the value is to 1, the better it is. The F1 score implies the percentage of positive predictions that were correct. Hence, model 2 outweighs model 1 again. Therefore, we can say that model 1 made more correct positive predictions.

Figure 8 Confusion matrix for model 1



There are 13813 cases in organisation line 1 and the model was successful in identifying all (TP) correctly as 1. Similarly, there are 4974 cases in Line 2 and the model has correctly classified 1331 cases (TP) as 2, but 3643 were wrongly classified as 1. There are 701 cases in Line 3 and the model has predicted 172 cases correctly. One advantage of confusion matrix is that it helps us understand the type of error made by the classifier.

Figure 9 Confusion matrix for model 2



The observations from the visualisation of the confusion matrix obtained from model 2 says that all cases for Line 1 were correctly predicted by the model, like for model 1. However, the number of correctly classified instances for line 2 and 3 is less than model 1. For line 2, the model was successful in identifying 1195 cases correctly and 98 for line 3.

Figures illustrate the decision trees built to predict the organisation line resolving the incident. We observe that the most relevant factor is the support team (ST) which takes up the case for resolution. In other words, we observe that some support teams have a major role to play in resolving cases.

The following is the decision tree with default parameter values. It is highly complex and hence, of no use for our analysis.

Figure 10 Default decision tree

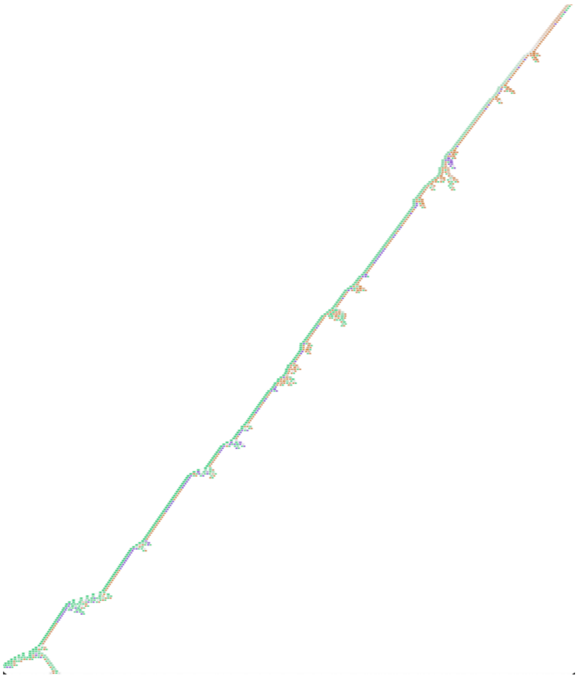


Figure 11 illustrates the decision tree (Model 1) obtained after using GridSearchCV. It has a max_depth of 15 with 5 being the minimum number of cases at each leaf.

Figure 11 Model 1

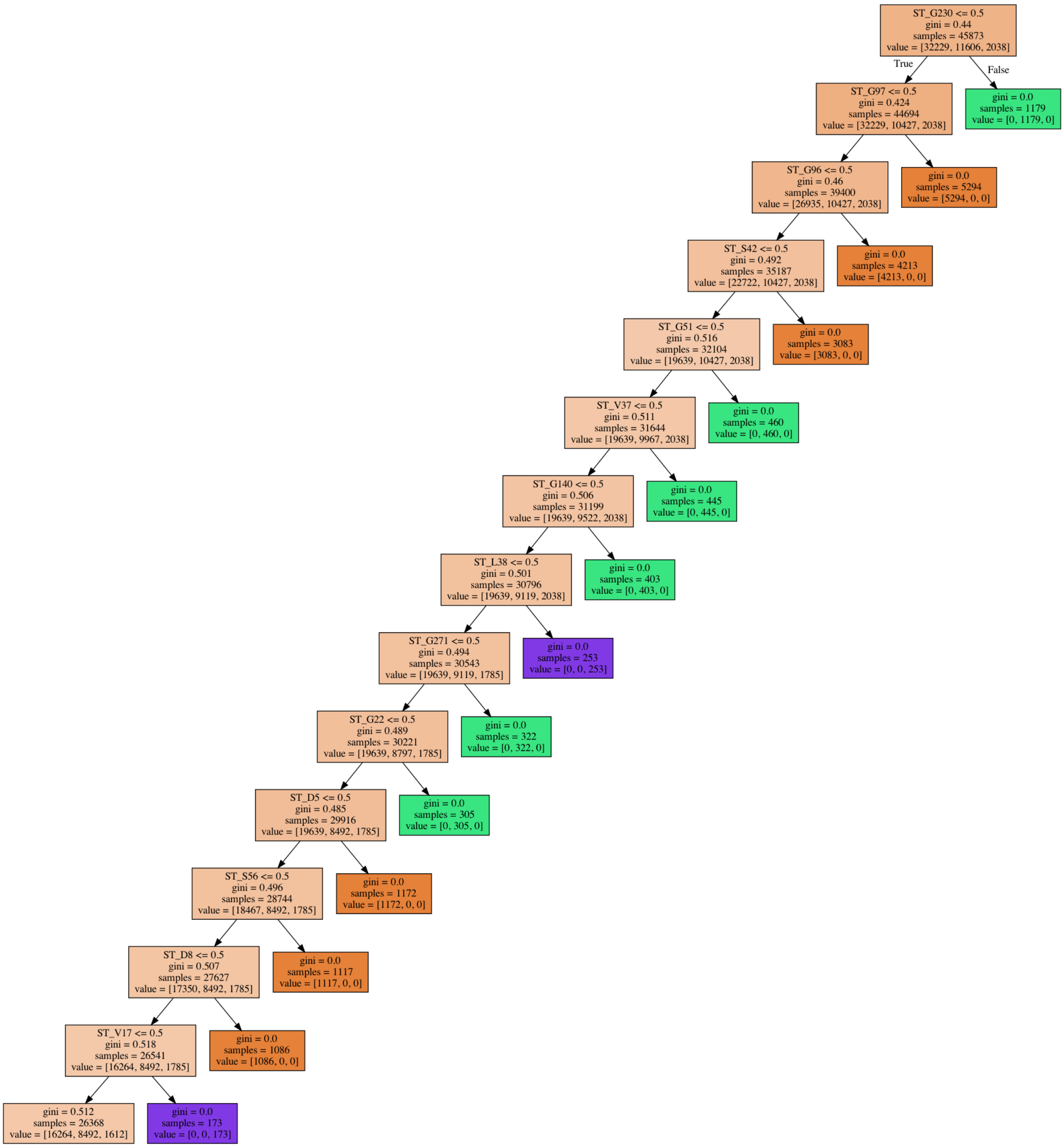
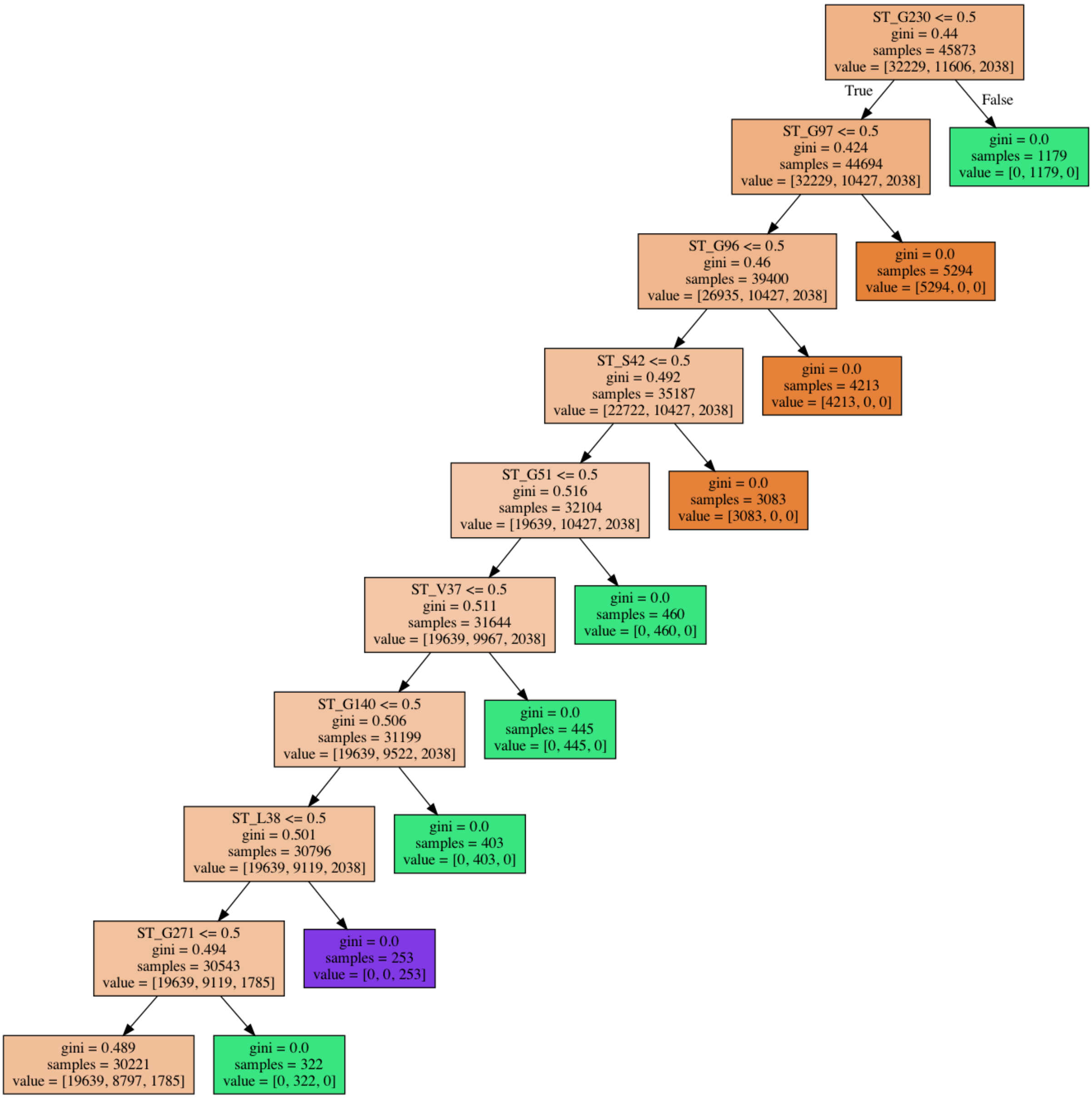


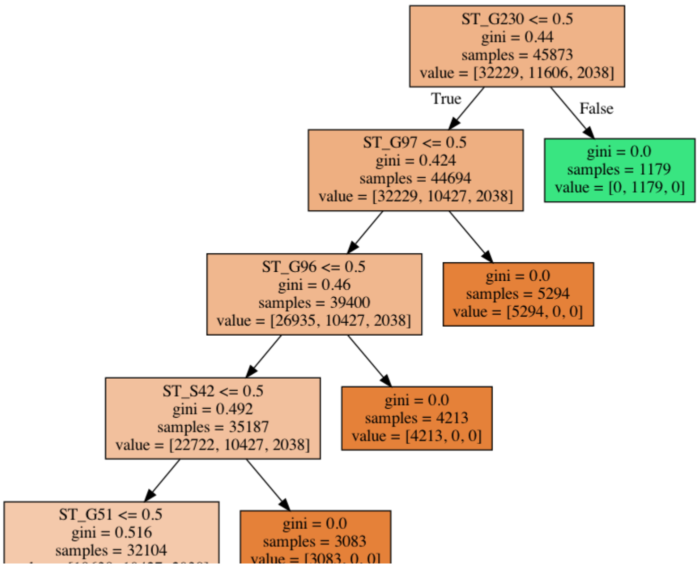
Figure 12 shows the decision tree obtained when max_depth was 9 and min_samples_leaf as 5. The root node indicates that there are 45873 instances with 32229 belonging to line 1. The first decision is based on whether the case is taken up by ST_G230 or not.

Figure 12 Model 2



The following is the closer look of the model 1. This analysis result suggests that the support team working on the case seems to have an influence on the occurrence of ping pong behaviour. We observe that there are three support teams – G230, G97 and G96 play a major cause for the ping pong behaviour. The root node indicates that there are 45873 instances with 32229 belonging to line 1. The first decision is based on whether the case is taken up by ST_G230 or not. We then see that there are 1179 cases which are directly taken up by the line 2, which is improper.

Table 13 Zoom-in of model 1



LIMITATION

This research is an approach to use data mining technique in business processes for the purpose of root-cause analysis of an event log specifically for the log available for Volvo IT Belgium. It uses the classification technique to identify the features responsible for the causes of the ping pong behaviour and the possible delays in the case completion time. However, the research poses some limits. It does not consider the types of ping pong behaviour that can exist, that is, it can either be linear or circular ping pong OR direct or indirect ping pongs. Linear/direct PP means support teams not working on the same case more than once after it is being moved to other support team, for example, A->B->C. On the other hand, circular PP means support teams not working on the same case more than once after it is being moved to other support team, for example, A->B->C->A. This study does not consider the Product variable. It is another limitation as it can help us gain insight onto whether the type of product has any influence on the ping pong behaviour between support teams.

CONCLUSION

The huge amount of data available to us from companies in the form of event logs can be widely used for making better business decisions and for process improvement. The IT systems present in these business processes records all the activities and has records of a lot of information from time, type of activity and resource to name a few. The huge amount of cases received by the incident management system leads the focus on resource and process optimisation and aims to reduce delays in case duration and resolution time along with the ping pong behaviour between support teams. Existing algorithms in data mining like classification methods can be used to dig deep into finding the factors that influence the case duration teams and hence, cause delays. In our study, we have used the classification technique to identify the causes of the ping pong behaviour observed in Volvo IT Belgium from the event log of the VINST system.

REFLECTION

The root-cause analysis for the ping pong behaviour is a topic of interest and can be further explored. I have realised that there are a number of factors or variables that need to be taken into consideration apart from the support teams involved to get a better understanding of the causes. The future scope of the research is that one can consider enriching the event log by deriving the data for the purpose of root cause analysis. One can explore the values or the data that can be derived using the current data that is available rather than simply using the available information in the logs. Another future scope of the research area is the consideration of different types of ping pongs. Ping pongs can be categorised into two types – linear and circular. Linear meaning when cases move from one line to another but do not return back to the same line again (For example, L1->L2->L3), where else circular ping pong means when cases return back to the same line after visiting other line (For example L1->L2->L3->L2). The same concept can be used for support teams as well.

Bibliography

[1] J. Gantz and D. Reinsel, *Extracting values from chaos,* 2011.

[2] B. Jokonowo, J. Claes, R. Sarno and S. Rochimah, “Process Mining in Supply Chains: A Systematic Literature Review,” *International Journal of Electrical and Computer Engineering,* vol. 8, no. 6, pp. 4626-4636, December 2018.

[3] W. Van der Aalst, Process Mining: Discovery, Conformance and Enhancement of Business Processes, Springer, 2011.

[4] W. Van der Aaist, A. Weijters, H. A. Reijers, B. Dongen, M. Song, H. Verbeek and A. Alva de Medeiros, “Business Process Mining: An industrial application,” *Information Systems,* vol. 32, pp. 713-732, 1 January 2007.

[5] E. Rojas, J. Munoz-Gama, M. Sepulveda and D. Capurro, “Process mining in healthcare: A literature review,” *Journal of Biomedical Informatics,* vol. 61, pp. 224-236, 2016.

[6] M. Arias and E. Rojas, *BPIC 2013: Volvo Incident and Problem Management Behaviour Analysis,* 2013.

[7] P. V. d. Spiegel, L. Dieltjens and L. Blevi, *BPI Challenge 2013 - Applied process mining techniques for incident and problem management,* 2013.

[8] A. D. Bautista, S. M. K. Akbar, A. Alvarez, T. Metzger and M. L. Reaves, *Process Mining in Information Technology Incident Management: A Case Study at Volvo Belgium,* 2013.

[9] E. Dudok and P. v. d. Brand, *BPIC’13: Mining an Incident Management Process,* 2013.

[10] Volvo Information Technology, *VINST - User Guide.*

[11] P. F. Wilson, L. D. Dell and G. F. Anderson, “Root Cause Analysis: A Tool for Total Quality Management,” *Journal For Healthcare Quality,* vol. 18, no. 1, p. 40, 1996.

[12] M. Horev, Root Cause Analysis in Process-Based Industries Journal For Healthcare Quality, Trafford Publishing, 2008.

[13] M. Heravizadeh, J. Mendling and M. Rosemann, *Root Cause Analysis in Business Processes,* Queensland University of Technology, 2008.

[14] D. Gheorghe, “Applying Data Mining for root-cause analysis of machine faults in complex continuous processes,” 2010.

[15] R. A and W. Van der Aaist, *Decision Mining in ProM.*

[16] J. Nakatumba and W. M. P. Van der Aaist, “Analyzing Resource Behavior Using Process Mining,” in *International Conference on Business Process Management*, 2009.

[17] S. Suraidi, C. Ouyang, W. van der Aalst and A. H. M. Hofstede, *Root Cause Analysis with Enriched Process Logs,* 2013.

[18] D. Ferriera and E. Vasileyev, Using logical decision trees to discover the cause of process delays from event logs.

[19] O. Maimon and L. Rokach, Data Mining With Decision Trees: Theory And Applications, 2014.

[20] S. K. L. M. v. Broucke, J. Vanthienen and B. Baesens, *Third International Business Process Intelligence Challenge (BPIC'13): Volvo IT Belgium,* 2013.